Reviewer’s report

Title: Online Randomized Controlled Experiments at Scale: Lessons and Extensions to Medicine

Version: 1 Date: 24 Oct 2019

Reviewer: Sarah Lensen

Reviewer's report:

Thank you for the opportunity to review this revised manuscript, which I believe has been much improved. I think interest among Trial readers could be increased further by reducing discussion regarding the detailed experiences of these companies in using A/B testing, and increasing examples (or even theoretical applications) of A/B testing to healthcare. Many of the experiences of these companies do not appear to be relevant to the goal of this paper - which is (presumably) to get medical researchers thinking about using these platforms for testing interventions in healthcare. For example, the description of the evolution towards a 'test everything model' and the section on choice of metric are less interesting. The authors suggest a relationship between core outcome sets and metric selection, which are rather different concepts, and this forced relationship highlights the lack of relevance of the metric testing for this paper.

The authors have added some examples of where A/B testing has been used in healthcare, as summarised by Horwitz et al. For me - this is the most interesting part, and I think this section should be expanded. Providing detailed examples of where A/B testing has been used is key to help readers understand why this is interesting, what sorts of our trials could be run within these frameworks (including real-life examples and even theoretical trials), and how we could improve health impact by embracing this model.

Minor points:

1. Pg 10, line 32 "body temperature or listening to the patient" is listening to the patient really a medical intervention or (optional) diagnostic technique? I think this sentence reads oddly and would suggest removing "listening to the patient"

2. Pg 13, line 39 "For some of the tested interventions, the system revealed that they were routine, but ineffective" do you mean to say the interventions were already routinely used but found to be ineffective by the testing? Presumably the system is not revealing they are routine - this is already known. The system is revealing they were ineffective, suggest re-word e.g. "The system also tested interventions which were in routine use, and revealed many to be ineffective..."

3. Pg 14, line 31 "There is a push towards having more trials that are simple and compare usual care modifications are already implemented somewhere or would be implemented anyway without ethical approval" - suggest some grammatical changes are made to this sentence
4. Box 2 "Methodological issues that are possible in online experiments to-date, difficult in traditional medical RCTs, but potentially relevant in future large-scale medical RCTs" - do you mean to say issues that are possible to OVERCOME with online experiments?

5. It is mentioned that many designs or promising alterations were not found to be beneficial in A/B testing. Traditional RCTs are expensive and we can only choose a few of the most promising looking interventions to test. It could therefore be mentioned that the application of A/B testing would help us to test more interventions, and this is important because promising interventions sometimes fail to have an impact, and other overlooked interventions (that we can't afford to test with standard RCTs) may have unanticipated effects.

**Level of interest**

Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

**Quality of written English**

Please indicate the quality of language in the manuscript:

Acceptable

**Quality of figures**

All images and figures within the manuscript should be genuine i.e. without evidence of manipulation. No specific feature within an image may be enhanced, obscured, moved, removed, or introduced. If you have concerns about the veracity of the figures you should choose the first option below.

**Statistical review**

Is it essential that this manuscript is seen by an expert statistician? If so, please give your reasons in your report.
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

Were you mentored through this peer review?

No