Reviewer’s report

Title: Statistical approaches to identify subgroups in meta-analysis of individual participant data: a simulation study

Version: 0 Date: 20 Nov 2018

Reviewer: David Fisher

Reviewer’s report:

This is a well written article on an important area of research which merits continued investigation.

The central work done -- the simulation study -- has not been done to this level of detail before. I fully support the authors' aims. However, I have a few reservations.

- My greatest reservation is that ecological bias was not modelled at any point. In fact, if I understand the data generating mechanism correctly, ecological bias *cannot* exist in the simulated datasets. Hence, the conclusions are arguably only valid in scenarios where ecological bias is known not to exist; which I would argue is of limited use in practice. This is not to say that the conclusions are not useful. They confirm what may already have been suspected: (i) that one- and two-stage models are asymptotically equivalent, but that one-stage models have the edge in certain scenarios such as with sparse data, low study numbers and binary outcomes (two-stage models typically assume a normal distribution for the study effects, which is clearly less likely to hold in such scenarios); and (ii) that PS-MA and meta-regression are generally inadequate (unless under very specific estimand definitions) and cannot be recommended. I do agree with the Conclusion, that centered one-stage IPD-MA is preferable; but given that the effects of ecological bias are only assumed and not actually tested, I find the statement "naive and centered one-stage IPD-MAs performed equally well" to be misleading.

[ N.B. my suggestion for how to simulate ecological bias is as follows: firstly, generate treatment and covariate as usual; then when generating the outcome via a GLM, the linear predictor should include, instead of bx*treatment*smoking [eq 1], these two terms: bx_within*trt*meandiff and bx_across*trt*meansmk where "meansmk" is the prevalence of smoking per trial (i.e. the trial mean of the 0/1 smoking variable), and meandiff = smoking - meansmk, i.e. the trial-centered covariate, c.f. the centered one-stage IPD-MA. In my experience, the coefficient bx_across needs to be fairly large relative to bx_within to generate a meaningful amount of ecological bias. ]

- The estimand of interest is not defined -- implicitly it is the treatment-covariate interaction effect, but this is not explicitly stated. See e.g. https://arxiv.org/abs/1712.03198 for more discussion about best practice as regards simulation studies. Furthermore, although (what I assume to be) bias statistics are presented within the supplementary materials, this is not made clear, and nor is it discussed anywhere in the article body. Notwithstanding my comments above
r.e. ecological bias, it would be interesting to know whether e.g. PS-MA or meta-regression are biased even if ecological bias is absent. (At a glance, it appears that they are).

- Although it is true that two-stage models may cause problems e.g. likelihood issues with small trials/sparse data, one-stage models are not problem-free. In particular, time-to-event models with random-effects and/or study-level error specifications often struggle to achieve convergence. Furthermore, it is less obvious how to present the results of one-stage models e.g. in terms of forest plots.

On a related note, it seems that the authors classify both PS-MA and MA-IT as "two-stage" methods. I do understand why, as both clearly consist of two analysis stages. However, when discussing one-stage vs two-stage approaches, this immediately becomes disingenuous, as one-stage IPD-MA and two-stage PS-MA are simply not comparable. This only serves to deepen the "one-stage good, two-stage bad" narrative which in my opinion is far more nuanced.

Also: In practice, PS-MA is often performed using fixed-effects models in each subgroup. The authors here only consider random-effects models. I would expect both bias and FPR to be worse under fixed-effects. Did the authors consider this variation?

- Finally, I'm not sure I agree with the final two sentences of Section 4 (Discussion): "Although our findings were based on binary, continuous and survival endpoints, it is highly likely that they will be applicable to other types of outcome ... this supports the generalisability of our recommendations." Is this not a circular argument? Or have I misunderstood?

In conclusion: The abstract states that "Clear guidance on the analyses [N.B. I would use "methods" here] is lacking ... our aim is to ... provide recommendations over which should be preferred."

Given my comments above, I do not feel that these aims have quite been achieved.

**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.

No
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 organisation 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