Reviewer’s report

**Title:** Weight loss as a predictor of cancer and serious disease in primary care: an ISAC approved CPRD protocol for a retrospective cohort study using routinely collected primary care data.

**Version:** 1  **Date:** 05 Jul 2017

**Reviewer:** Katherine Morley

**Reviewer's report:**

In light of the author responses to my reviewer comments, and the fact that DAPR is a new journal, I sought advice from the journal editors regarding the expectations for this type of article. Specifically, while I completely understand that the methods of an approved ISAC protocol cannot be changed, I sought clarification about whether it is appropriate for reviewers to request more information/clarification about aspects of the protocol, or whether the published article should simply be verbatim publication of the approved protocol. I include the relevant portion of the editors’ response below to provide context for my subsequent reviewer feedback:

"As the protocol has already received ethical approval and funding, the authors/investigators are unlikely to be able to make major changes to their protocol design so review should be focused on reporting, with a focus on clarification rather than changes to the intended research. However, even if changes cannot be made to methods it's sometimes useful for the authors to discuss some of the recommendations from reviewers in the Discussion section in relation to their study."

Based on this I have reiterated a few of my previous points below in which I requested further information about aspects of the protocol, but did not suggest any alterations.

[1] It appears that they propose the development of an EHR algorithm for weight loss - what steps do the authors propose to take in order to validate the algorithm developed?

**Author Response -** We will develop this algorithm using a small subset of clinical records, and then cross-validate is using the remaining dataset.

**Reviewer Response -** It would be helpful to include this information in the article (e.g., under Aim 1 in the Data/Statistical Analysis section). It would also be helpful to have more information about the methods you plan to use in developing the algorithm. I may have completely missed it, but it wasn't clear to me whether you're proposing rule-based methods or some type of machine learning approach or something entirely different.
[2] If I have understood correctly, the authors also appear to be proposing the development of another algorithm for what they have termed "Serious Disease", but it was not clear what conditions might be included in the endpoint, or how this algorithm might be validated.

**Author Response** - This serious disease codelists are being developed by literature review to identify likely candidate conditions that are known to cause weight loss.

**Reviewer Response** - As in my query above, are you able to provide more information on the methods that will be used to create the algorithm once codelists are defined, and how it will be validated?

[3] The cohort definition was somewhat unclear given that different time windows are proposed for different variables. I appreciate that extracting relevant information from EHR is complex and there will be rationales for the authors decisions, but these are currently hard to follow. A clear statement of inclusion/exclusion, cohort entry, outcome/exit definitions is needed, as well as the time windows used for each set of variables - a diagram may be helpful here?

**Author Response** - This is wording used in the approved ISAC protocol.

**Reviewer Response** - I appreciate that you cannot change these definitions, but I think further explanation of this would be helpful for the reader. If you do not feel that it is appropriate to expand on this in the text, then a diagram may be the best way forward in terms of explaining the sample definitions for the descriptive and cohort analyses without actually having to alter the original text.

**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

**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 financial 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.