Reviewer’s report

Title: Patterns and correlates of physical activity in adult Norwegians: a forecasted evolution up to 2025 based on Machine Learning approach

Version: 0 Date: 15 Mar 2018

Reviewer: Steven Cohen

Reviewer's report:

This is an interesting and well-presented manuscript using innovative statistical methods to address a major public health issue facing many countries today. That said, there are a number of elements that could be fixed or improved upon in a revision. I also recommend that the manuscript be proofread by a native English speaker to improve the numerous, but extremely minor grammatical and syntactic issues throughout the manuscript. Specific elements that could be fixed are listed by section here:

Abstract

1. In the introduction of the abstract, it would be helpful to have a sentence motivating/justifying the sex-stratified analysis.

2. The authors could probably delete "using a machine learning approach" in the conclusion subsection.

Background

3. It would be helpful to have a sense of what percentage of Norwegians (or other relevant geographic area) meet the current WHO requirements.

4. (minor) Paragraph 2: Should be "work" not "works" in the second line.

5. I would suggest rewriting some of the second paragraph of the background section, in general. For example, the message of the two parts of the second sentence (beginning with "National surveys…") are not linked easily. I would recommend splitting this sentence into two smaller sentences. Also, the last sentence is difficult to interpret. Some overall numbers would be helpful here (see earlier comment).

6. Third paragraph: Rewrite first sentence. It is not grammatically correct.
7. This sentence should be edited: "As a matter of fact, thanks to the huge quantity of data that are now recordable by survey and technologies, Machine Learning approach could help to discover disease risk factors related to lifestyle or the environment helping people to understand and then modify their wrong behaviours." In addition to some minor grammatical issues, the sentence seems to be a bit biased, colloquial, and subjectively written. Consider substituting for words such as "huge" and "wrong".

8. Next sentence: Delete "actually".

9. Machine Learning should be defined clearly somewhere in the beginning of this paragraph.

10. Relatedly, please provide some additional references for the use of Machine Learning in previous studies. Currently, it appears as if one is referenced (ref 8).

11. The approach of using those specific measures (frequency, duration, and intensity) should be better justified previously in the Background section.

12. Last paragraph: Delete or reword "thanks to this approach".

13. Last paragraph: Consider specifying what "deeply investigated" means.

14. (minor) "Physical activities" should be abbreviated as PA (see last sentence).

Methods

15. For the statement "Answer options such as 'I don't know', 'I'm not sure', 'I cannot answer', in the independent features were deleted from the dataset." What percent were deleted? Were they consistent across cases? In other words, were respondents who answered in this way to one question more, less, or similarly likely to respond this way to other questions? This information would be extremely important to have to assess bias.

16. (minor) Consider rewording "First of all" in the Statistical Analysis paragraph.

17. The methods described under the subheading "Experiments" should have appropriate references.

18. What software was used to conduct this extensive, computationally-intensive analysis?
19. The explanation of the "feature" and "label" values found in the last paragraph of the methods section is difficult to follow. Some additional explanation/more basic terminology would be helpful for interpretability.

20. More information on the predictive models would be helpful in the last paragraph.

Results

21. In general, there is so much information contained in this section that it is hard to follow how the results flow logically from the study aims. Consider using subheadings to better bridge the aims with the results.

22. (minor) The first paragraph is one standalone sentence.

23. I recommend condensing some of the text in the descriptive results (first two paragraphs).

24. (minor) Add the phrase "statistically significant" in the first sentence of the results section.

25. Second paragraph: Consider rewording "males reduce". This sounds causal and should not be.

26. The results in the third paragraph should be stated more clearly and succinctly. I recommend listing the significant interactions and providing a brief interpretation of one of them.

27. For the PA Frequency, Duration, and Intensity paragraphs that follow and reference Table 3, the presentation of the specific results can be confusing and hard to see where the findings are listed in the table. Numbers representing the associations observed should be cited in the text to support the explanation.

28. Although I recommend condensing and clarifying much of the results section to better focus on the key results (and relevant interpretations of those results), I recommend discussion the predictive models presented at the very end of the results section in more detail.

29. Relatedly, are the confidence bounds or any other measure of uncertainty on these estimates for Table 4 (predictive model)?

30. (minor) Delete the use of "like" in the presentation of the results.
Discussion

31. Third paragraph: What are potential explanations for the increase in PA levels (slightly contradicts some of the arguments in the Background section). Were there events/historical changes that could account for these trends?

32. The fourth paragraph that describes the machine learning approach and the findings contained therein would benefit from a discussion of some public health approaches and theories that might explain or corroborate the specific findings of social determinants and PA.

33. I would also recommend providing a bit more detail about Machine Learning and its contributions to other public health issues in Norway or elsewhere. Some limitations are listed in the limitations paragraph, but since this is a fairly new approach, it would be helpful to have more detail.

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 recommend additional statistical review

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

Not suitable for publication unless extensively edited
**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/](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.