Reviewer’s report

Title: Importance of Medical Data Preparation in Predictive Modeling and Risk Factor Discovery for the Frailty Syndrome

Version: 1 Date: 15 Apr 2018

Reviewer: Bonnie Westra

Reviewer's report:

The authors address a topic of importance and apply data mining techniques with an emphasis on data preparation. I am including my comments as an uploaded document as it is difficult to maintain formatting to understand the comments. While the authors addressed previous reviewer comments, there remains multiple issues of concern.

The paper continues to be longer than one would expect for a journal. There continues to be considerable redundancy in sections that could be combined and sections that could be reduced or removed as noted in my comments submitted.

The focus of the paper is on predicting the frailty syndrome, yet a definition is lacking until p. 12 and then the future work identifies frailty as a disease - clear, consistent use of this term is needed and should be introduced very early in the paper.

The background section has been reduced, but it continues to lack integration based on the main arguments for why a study is included - the "so what" is missing to justify the need for the current study based on previous research. This section should be about key points supported by the literature to show how you are building on previous studies and the gap that this investigation will address or to support what the investigators will be doing.

The most significant problem is still not addressed based on a previous reviewer suggestion a traditional format of intro, methods, results, conclusion/discussion) - these headings were used - however, results included methods, findings (sometimes limited) and interpretation. Often the results starts with additional comments.

PCA - it is not clear how this is useful for the rest of the paper. After describing this section - it doesn't seem to inform further analysis

Please see uploaded comments for more details with additional comments

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

No
**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:

Needs some language corrections before being published

**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