Reviewer’s report

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

**Version:** 0 **Date:** 08 Nov 2017

**Reviewer:** Dustin French

**Reviewer's report:**

I agree with your idea of using EHR info for prediction analytics.

1. Streamline Intro and Lit Review to reflect your contribution in geriatrics. I would not get into drug surveillance and other disease predictors outside of your area of emphasis.

2. There are many predictors in this model. From a stats point of view this presents problems of Multicollinearity and degrees of freedom problems

3. Many of these variable are part of AGS guidelines for fall prevention. They are also RAP triggers in the MDS for LTC. You may want to use these clinical tools/guidelines as justification for variables.

4. The number of figures and info is overwhelming. Again can you condense what you have? Think about what you are adding to the literature. There have been other papers prediction 30day readmissions, falls in the elderly, LTC,etc. I did not see these here but you should think about adding them in.

**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.

Unable to assess

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

Unable to assess
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:

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