Reviewer's report

Title: Opinion versus practice regarding the use of rehabilitation services in home care: An investigation using machine learning algorithms

Version: 2 Date: 23 March 2015

Reviewer: Gregor Stiglic

Reviewer's report:

This paper presents an interesting approach to identify items from the interRAI HC instrument that are most predictive of whether a client will receive rehabilitation services in actual clinical practice, and to compare them with items selected by domain experts.

Major Compulsory Revisions:

1. Please clarify the meaning of X_1,….,X_6 on page 12. Is this only an example containing 6 features?

2. I like the proposed approach that combines LASSO and Random Forest's VIM to account for possible interactions very much. However, it is not clear how the threshold for keeping only the features ranked above 20 in both rankings was selected? Why not Top-10 or Top-50? Since the proposed approach primarily relies on LASSO (and uses RF only as an additional ranking) it could benefit from using a statistically robust way of selecting a group of relevant features from LASSO based on bootstrapping. A generally accepted method to do this is called Stability Selection [1] and also has a very user-friendly implementation in R [2]. At least point this out in the paper as a solution to obtain more statistically sound threshold value.


Minor Essential Revisions:

1. Definition of an impaired person is repeated in lines 275-279 (it already appears in the same form in lines 198 - 202). There seems to be no need to have this definition in the paper twice.

2. Explain the abbreviations or add more description when first using the terms interRAI and InfoRehab.

3. Revise or remove the sentence in line 317 as it does not add any information in the current form.
Overall, methodology and results presented in the paper are very interesting, especially given the fact that the list based on predictive performance significantly differs from the one that was prepared by the domain experts.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.