Author's response to reviews

Title: Prediction of Adverse Cardiac Events in Emergency Department Patients with Chest Pain Using Machine Learning for Variable Selection

Authors:

Nan Liu (liu.nan@sgh.com.sg)
Zhi Xiong Koh (koh.zhi.xiong@sgh.com.sg)
Junyang Goh (kengoh83@gmail.com)
Zhiping Lin (ezplin@ntu.edu.sg)
Benjamin Haaland (benjamin.haaland@duke-nus.edu.sg)
Boon Ping Ting (tingbp@gmail.com)
Marcus Eng Hock Ong (marcus.ong.e.h@sgh.com.sg)

Version: 3
Date: 2 August 2014

Author's response to reviews: see over
2 August 2014

Dear Editor,
BMC Medical Informatics and Decision Making

Research Article MS: 1298913294135322 (Revision)
Prediction of Adverse Cardiac Events in Emergency Department Patients with Chest Pain Using Intelligent Variable Selection and Heart Rate Variability

We would like to thank the editor and the reviewers for their constructive comments and suggestions. We have revised the manuscript accordingly. Specific point-by-point replies to the comments are given below. We hope that the revised paper is now acceptable for publication.

Thank you.

Corresponding Author:
A/Prof Marcus Eng Hock Ong
Department of Emergency Medicine
Singapore General Hospital
Outram Road, Singapore 169608
Tel: (65) 6321 3590
Fax: (65) 6321 4873
E-mail: marcus.ong.e.h@sgh.com.sg
Editor's Commentary

I think these reviewers suffice (I have assessed the manuscript by myself too)

Author reply: We thank the editor for giving the opportunity to review our manuscript for further consideration. We have addressed all the reviewer comments, which are shown in both this reply letter and the revised manuscript.

Please provide a detailed line-by-line response to all of reviewers’ questions. Also, in addition to the reviewers’ suggestions, please provide in the manuscript the performance obtained by fitting a simple main-effects linear model with any variable selection method (e.g. bi-directional stepwise). This should help in understanding if a simple re-fitted scoring model is as good as a more complex non-linear model.

Author reply: We have addressed all the reviewer comments. The detailed line-by-line response to all of reviewers’ questions can be found as follows in this letter.

We also conducted a testing with linear logistic regression with stepwise variable selection. Brief results have been reported in the revised manuscript (Line 291-296). Since this regression model is for the purpose of simple comparison while TIMI and MEWS are two established scoring systems, we don’t included the results by logistic regression in Figure 4 and Table 6. We hope this is acceptable by the editor.

Additional Editorial Requirement:
1. We recommend that you copyedit the paper to improve the style of written English.

Author reply: We have invited a native English speaker to improve the quality of this manuscript. Furthermore, we have tried to polish the writing by ourselves.

Reviewers #1 Comments

The paper describes a method to perform variable reduction and its implementation in predicting adverse cardiac events in emergency department patients through a previously published scoring system. The paper is interesting to read and the analysis is overall well conducted. However, some points have to be addressed before being suitable for publication.

Author reply: We thank the reviewer for the kind words. We have addressed the comments of the reviewer as follows.

Minor essential revisions:
In the description of the variable selection methods it is not stated how many trees were used in random forest (from line 163).

**Author reply:** We thank the reviewer for pointing this out. In this study, we used 500 trees in the random forest method. We have mentioned this in the revised manuscript. (Line 175)

In the description of the variable selection methods, it is used a 1:1 ratio between MACE and not MACE. Why was this ratio and not others (e.g. 1:2, 1:5) used? Furthermore, during the bootstrap procedure the systematic inclusion of all MACE patients could introduce a bias of selection. Why were not statistical techniques to deal with unbalanced data (e.g. SMOTE, https://www.jair.org/media/953/live-953-2037-jair.pdf) used? (from line 163)

**Author reply:** We thank the reviewer for the suggestion. In creating the subset, we chose to use 1:1 ratio so as to build a balanced dataset. When applying variable selection / classification methods on an imbalanced dataset, we will need extra efforts to handle the bias generated from the data as majority class of the data tends to dominate model building.

Regarding the reviewer’s suggestion on the use of bootstrap procedure, we may have a different opinion. According to the SMOTE algorithm, synthetic samples are generated from training samples, assuming that these samples have similar distribution that is found in real training data. That is to say, these synthetic samples may or may not approximate to some of real testing samples in terms of similarity. Looking forward, we plan to recruit more patients so that we will have sufficient patients with the outcome MACE in the dataset. By then, we will be able to use a subset of MACE group for variable selection.

In the revised manuscript, we have included this into the limitation section of the discussion. (Line 379-381)

In the description of the variable selection methods it is not explained why the authors decided to select 8 variables. At the moment this results an arbitrary decision that needs to be justified (from line 163).

**Author reply:** We thank the reviewer for pointing this out. In this study, we empirically decide to select 1/3 of the variables in this study as we feel that 8 variables would be able to produce a trade-off between accuracy and efficiency. However, as suggested by the reviewer, there is a need to determine the optimal number of variables. We have mentioned this as future work in the last paragraph of Discussion. (Line 387-391)

Which was the statistical test (e.g. Delong’s) used to pair-wise compare AUCs? (from line 213)
**Author reply:** In this study we used bootstrap method to conduct the pair-wise comparison. We have added this information into the Method section. (Line 242-243)

In the results, authors report cut-offs for analysed model. How were these cut-offs calculated? (from line 257)

**Author reply:** We determined this “optimal” cutoff score by looking for a point that is nearest to the upper-left corner of the ROC curve. We have included this information into the revised manuscript. (Line 281-282)

Overall, English could be improved.

**Author reply:** We thank the reviewer for the suggestion. We have invited a native English speaker to improve the quality of this manuscript. Furthermore, we have tried to polish the writing by ourselves.

---

**Reviewers #2 Comments**

Overview: Authors of this article tried to use random forest approach to select meaningful variables for prediction of adverse cardiac events, and then fed the selected variables to an intelligent scoring system to derive risk scores. It is an interesting research topic, and it offers a pragmatic approach to aid clinicians in emergency department to do triage of patients with chest pain.

**Author reply:** We thank the reviewer for the kind words.

1. Title
The focus and the objective of the study were to discover a set of rational and meaningful variables that can be used to predict risk of adverse cardiac events. Heart rate variability is a predictor of adverse cardiac events, and it can be expressed by a set of parameters. Intelligent variable selection means a computerized way to choose variables. It seems that it is not proper to put ‘intelligent variable selection’ and ‘heart rate variability’ in parallel in the title as a proposed way to predict adverse cardiac events. I would like to suggest the authors to change the title to another one which is closely related to the research focus as presented in the article.

**Author reply:** We thank the reviewer for the suggestion. We have changed the title to “Prediction of Adverse Cardiac Events in Emergency Department Patients with Chest Pain Using Machine Learning for Variable Selection” to focus on the contribution of our work. Term “heart rate variability” is removed from the title as it is only a type of variable. Term “machine learning” is added in because it best describes the nature of the technique.
2. Abstract
In the Methods part, data and analytic methods need to be better described. In Conclusions part, the conclusion from the study needs to be more related to the objective of the study.

Author reply: We have revised the abstract according to reviewer's suggestion by adding in descriptions of data and analytic methods as well as improving the writing of conclusion. (Line 45-51; Line 65-69)

3. Background
a) It's good that the authors mentioned current risk scoring systems such as TIMI, GRACE, and MEWS and their previous intelligent scoring system in this section. However, it would be better if the authors can introduce the advantages and disadvantages of current risk scoring systems.

Author reply: We thank the reviewer for this suggestion. We have added some descriptions on the advantages and disadvantages of current risk scoring systems into the background section. (Line 93-94; Line 97-100)

3. Background
b) The aim of the study is to discover meaningful variables for prediction of adverse cardiac events. It would be better if what other researchers have done in the area of prediction variable selection can be discussed here.

Author reply: We have added in the background section some descriptions on variable selection for prediction. (Line 113-116)

4. Method
a) Study design: 1) Vital signs usually specifically include body temperature, pulse, blood pressure, and respiratory rate. It seems that ‘vital signs’ used in this article consist of more variables than the above. Is ‘clinical signs’ more appropriate? 2) Glasgow coma scale is also a risk scoring mechanism. Does it belong to vital signs? 3) Clear data description should be placed here.

Author reply: We thank the reviewer for the valuable comments.

1) We have changed the description of “vital signs” to “clinical signs” in the revised manuscript where appropriate.

2) Since we have used term “clinical signs”, GCS could be included as one of the parameters. Moreover, GCS alone is not able to provide much predictive power.
3) Detailed data description has been included and summarized in Table 1 and Table 4.

4. Method
b) Predictive variable selection: 1) What variables have been previously used by the authors or peer researchers for cardiac risk prediction need to be discussed. 2) The authors mentioned that they used the random forest (RF) approach to do variable selection. A brief introduction of RF and the reason for the authors to use it should be provided. 3) In the variable selection algorithm as presented in Figure 1, why didn’t the authors derive statistical significance for each variable first, and then use the RF method to discover most significant variables in a smaller variable pool which contains only variables with significance? The authors choose 8 most significant variables for each selector, why? A clear description of the algorithm needs to be provided here. 4) Since there are demographic and medical history variables in the dataset, why didn’t the authors put those variables into the variable pool for selection?

Author reply: We thank the reviewer for the valuable comments.

1) We have mentioned in the revised manuscript that all 15 HRV parameters and 8 clinical signs were used in the previously risk prediction method. (Line 219-220)

2) We chose random forest approach for variable selection because it has been proved to be effective and several studies show evidence on this. In the revised manuscript we added 2 new references to support our use of RF. We also mentioned that there are other approaches worthy trying in future works. We have included a short introduction to RF in the Method section. (Line 175-179)

3) Machine learning-based variable selection examines interactions among various variables and selects a subset of them. There are possibilities that non-statistically significant variables are good predictors. Therefore, if we derive statistical significance for each variable first, and then use the RF method to discover most significant variables in a smaller variable pool, we will have high chance to miss out some discriminatory but non-significant variables. Regarding the reason of choosing 8 variables, we have replied the same question raised by reviewer #1.

4) Demographic and medical history data are potential predictive variables. The reason that they are not included in the prediction is because the adopted scoring system does not work well on binary variables. The scoring method is based on geometric distance so that continuous variables are more suitable. In the revised manuscript, we have integrated reviewer’s suggestion into future work. (Line 386-387)

4. Method
c) Risk score prediction: My understanding of risk score prediction using the selected variables is to validate the performance of the selected variables. From the article, it seems that the variable selection was just for the intelligent scoring system which the authors developed before. But how the machine learning algorithm and support vector machine work in the intelligent scoring system needs to be clearly presented.

Author reply: We thank the reviewer pointing this out. We have clarified in the revised manuscript that variable selection (main contribution of this paper) and risk score prediction are two separate steps. The intelligent scoring system simply serves as the platform where different sets of variables are the inputs. The reference to the detailed description of the ML method is given. (Line 203-204; Line 216)

4. Method
d) Statistical analysis: How the leave-one-out-validation strategy was used to evaluate performance, and what performance dataset was used to do ROC analysis need to be clearly described here.

Author reply: We thank the reviewer for pointing this out. We have included a detailed description on how performance is evaluated. (Line 233-238)

5. Results
a) It will be better if description and discussion about collected data are placed in Methods section.

Author reply: We thank the reviewer for this suggestion. On this point, we may have a different opinion. In studies on emergency medicine, we usually report baseline characteristics (Table 2) in the first paragraph of Result section. Therefore, we feel it is more proper to place the description of the data in Result section instead of Method section.

5. Results
b) It will be better if description and discussion about previously used prediction variables are placed in Methods section as well.

Author reply: We thank the reviewer for this suggestion. We also have a different opinion on this point. Following our reply to the previous question, we usually organize univariate analysis (Table 4) next to baseline characteristics. Therefore, we feel that it may be proper to keep the current layout.

5. Results
c) The authors mentioned that ML score with variable selection is correlated to ML score without variable selection (AUC diff. 0.076; p value 0.28). Does this mean that the system
prediction performance has no statistically significant difference after the variable selection? If so, is there any need for variable selection?

**Author reply:** We thank the reviewer for the comments. In the revised manuscript, we have explained in the Discussion section that “only systolic BP instead of all 8 clinical signs is required for risk prediction, which dramatically simplifies the scoring system and saves a lot of time in measurement”. Even though variable selection may not statistically improve the prediction performance, it is able to simplify the scoring system. (Line 316-318; Line 361-363)

6. Discussion
a) The authors mentioned that in a machine learning method, variable selection is performance-driven where prediction performance dominates the selection process while statistical significance and correlation play a comparably minor role. Does this correct? The final objective of this study is to aid clinicians in the ED to do triage of patients with chest pain. If clinicians think the selected variables for risk prediction are closely related and without clinical meaning, and they do not accept the selected variable set, then what’s the meaning of this study?

**Author reply:** We thank the reviewer for the comments. In the revised manuscript, we have changed our claims that both prediction performance and statistical significance are important (Line 350-352). The aim of the risk score prediction is to show evidence on how severe the patient’s condition is. It provides clinicians a second opinion but does not attempt to make decisions for them.

6. Discussion
b) Since the authors mentioned that age and some other variables are also predictive, and those variables are included in the dataset, why not put all available variables into the variable pool for selection?

**Author reply:** We thank the reviewer for the suggestion. We mentioned in the previous reply that we are not able to include some demographic and medical history data as they are binary variables. Age is a continuous variable; however, we notice that in this study age is not discriminatory between MACE and no MACE groups (seen from Table 2). Therefore, in the revised manuscript, we have removed our claims that age is potentially a predictive variable.

Overall, it is an interesting research topic being explored in the article, and the authors do have done much work in this study. But deeper analyses and more details need to be provided, and the article needs to be reorganized.

**Author reply:** We thank the reviewer for this encouraging comment.