Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-21-668

Reviewer A

This is an interesting paper describing the methods for designing and testing a CDSS the predict diagnoses and recommend treatment for childhood asthma in China. I think there is much value is adapting such an approach for the specific context of China and I read the paper with great interest.

I was left wanting more clarity and details about the methods and not enough understanding of key issues such as the workflow integration of the CDSS. Without more detail, the reader is left without enough information to reproduce the methods or vet the rigor of the methods. Major suggestions for improvement are as follows:

1. Why was gradient boosting decided on? What is the rationale for choosing this model over random forest or a more traditional logistic regression model? Often with clinical prediction models, different models/approaches are compared and the best performing model is selected, but that does not seem to be the approach the authors took - why?

Reply 1: Thanks for your comments.

We chose extreme gradient boosting (XGBoost) algorithms to model asthma diagnoses and treatment because it is an advanced implementation of a gradient-boosting decision-tree algorithm and has been used in a few studies to predict asthma hospital visit. Although there is no study comparing the prediction accuracy of XGBoost algorithm and gradient boosting decision tree algorithm in predicting the diagnoses and treatment of asthma, a retrospective analysis compared the differences of accuracy in predicting hospitalization risk among different machine learning algorithms. The area under curves for each model were: logistic regression 0.82 (95% CI: 0.81-0.82), random forests 0.82 (95% CI: 0.81-0.83), and gradient boosting machines 0.85 (95% CL 0.84- 0.86), which showed that gradient boosting machines model was the most successful at predicting need for hospitalization at the time of triage in pediatric patients presenting with asthma exacerbation. XGBoost is advantageous because of its high speed and performance, making it dominant in applied machine learning for structured data and also because it offers regularized gradient boosting and feature importance scores using a trained predictive model, which can be used for feature selection. Above all, we chose XGBoost algorithms.

Changes in the text: We have added the reason for choosing XGBoost algorithms in *[Discussion]* section paragraph two.

2. For the control subjects with asthma like symptoms, was their diagnoses of asthma ruled out? A true negative is important to the accuracy of the predictions.
Reply 2: Thanks very much for your comments.

In our study, the definition of asthma diagnosis and treatment were from the recommendations derived by a panel of advisors including pediatricians from China after consideration of international guidelines[]. Each clinical diagnoses and treatment will be determined by an adjudication panel comprising three consultant paediatric clinicians (median 10 years of specialist practice). Two members will review each subject independently, with a third member acting as tie-breaker in the event of non-agreement. The panel will arrive at diagnoses and prescriptions after assessment of all available clinical data. There will be three outcomes: "YES", "NO" or "UNSURE". The outcome of "UNSURE" indicates that the case is not entirely met due to lack of information and these cases will be excluded from the endpoint.

Changes in the text: We have revised and clarified the descriptions of the clinical diagnoses and in *[Statistical analysis]* section *[Analysis of Accuracy]* part. And we also revised Figure 1. Study schema of Phase I.

3. The measure for "applicability" is unclear as stated on page 4. "Applicability" is not a common term used in the CDSS space, so I was eager to get clarity on what this meant. Does applicability refer to effectiveness of changing prescribing behavior?
Reply 3: Thanks for your comments. Sorry for the confusion made by the wording. We delete the word "applicability". The aim of Phase II is to evaluate the accuracy of treatment prediction of the CAMCDS. The accuracy of treatment prediction of CDSS model will be based on the comparison between ROC curve and the results of pediatrician's prescriptions.

Changes in the text: We checked the manuscript and changed the description in *[Title]*, *[Abstract]*, *[Introduction] and [Method]* section. And added more descriptions of study design of Phase II in *[CAMCDS model]* section paragraph five.

4. There needs to be greater description of the CDSS design and implementation. Will it be integrated in EHR workflows, external to the EHR? Will there be an alert that interrupts providers? Will patient data be automatically entered into the CDSS or will manual entry be required? If these decisions will be made based on a user-centered design process with stakeholder engagement, then these methods need to be articulated.
As is, there is not enough detail for me to know what the end result/intervention/product will be so it is hard to interpret the impact of this work.
Reply 4: Thanks for your comments. All of the data will be collected by trained study nurses via case report forms and will be manually entered into the CDSS. For aggregation, every site will get an analysis script to run on their data that produces an aggregated output that will be collected at Children's Hospital of Shanghai for the

final analysis. M.M. Na Dong (Department of Respiration, Children's Hospital of Shanghai) will be responsible for data collecting and gathering. While M.M. Beirong Wu (Department of Respiration, Children's Hospital of Shanghai) will be responsible for data analysis and interpretation.

Changes in the text: We have added the above information in *[Data collection]* section paragraph *two.*

5. I also noted that this paper would benefit from editorial review for grammatical and semantic errors.
Reply 5: Thanks very much for your comments. We have asked MedSci, a professional language editing service, to improve the grammar and readability. Please see if the revised version met the English presentation standard.

Once again, we thank you for the time you put in reviewing our paper and look forward to meeting your expectations.