Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-23-931

Reviewer A

Comment 1: Please elaborate how AECOPD was evaluated. Is it coming from diagnosis from doctors or self-report from patients? If this information is not available, please discuss potential limitations.

Reply 1: I agree that we should explain it more clearly. AECOPD is characterized by a sudden worsening of symptoms, accelerated deterioration in respiratory function, reduced quality of life, and increased healthcare costs. In our study, the diagnosis of AECOPD from doctors mainly depends on the clinical course of acute onset, that is, the deterioration of the respiratory system exceeds the daily variation, and the main symptoms is the aggravation of dyspnea, often accompanied by wheezing, chest tightness, cough, and increased sputum volume, et al. Thus, we evaluated AECOPD mainly by worsening symptoms of COPD.

Changes in the text: We have modified our text as advised (see Page 7, line 114-118).

Comment 2: Please include specific numbers, directionality when interpreting different parts of results: "significant differences were found in infrequent and frequent groups...", Please do not just include p-values.

Reply 2: Thank you for pointing it out. We have added specific numbers, directionality when interpreting different parts of results.

Changes in the text: We have modified our text as advised (see Page 9-10, line 173,178-185,192-194).

Comment 3: The temporality between blood indicators and AECOPD was not very clear. It seems these variables were measured cross-sectionally. This should be discussed in the limitation part. Also, please elaborate how these indicators help predict AECOPD within one year.

Reply 3: These blood indicators results were the first measurements when they were admitted into the hospital, as the baseline characteristics. The results of multivariate logistic regression analysis indicated that neutrophils $\leq 5.350\times10^9$/L, monocytes $\leq 0.255\times10^9$/L, eosinophils $>0.275\times10^9$/L, DBil ≤3.415 umol/L, GGT ≤22.500 U/L and GLR >3.675 were independent predictors for frequent exacerbator phenotype of AECOPD. We have discussed this limitation and elaborated how these indicators help predict AECOPD within one year in the discussion part.

Changes in the text: We have modified our text as advised (see Page 12-13, line 225-229,243-246).

Comment 4: The paper mentioned previous publication regarding biomarkers associated with AECOPD. Please discuss why the blood indicators identified in this manuscript are more efficient (than biomarkers identified in previous publications) in prognosis of AECOPD.

Reply 4: Thank you for your comment. Compared with biomarkers identified in previous publications, most of them have to be measured in the laboratory, such as Sphingosine-1-phosphate, microalbuminuria, TNF-α, IL-10 and IL-8, et al, we predicted the prognosis of AECOPD based on more rapid, inexpensive, and easily obtained routine markers in this manuscript. Furthermore, we discussed the frequent exacerbator phenotype of AECOPD, which is a significant phenotype of COPD characterized by experiencing at least two exacerbations per year and associated with the number of exacerbations，and it's also different from previous studies.

Changes in the text: We have modified our text as advised (see Page 12-13, line 235-242).

Comment 5: I think there should be more strengths and limitations discussed. For strengths, please include discussion regarding why these indicators are more convenient or superior than the ones identified in previous studies. For limitations, please discuss any possible misclassification of AECOPD and future validation through other data source. Without replication through other studies, I don't think you can conclude this a

reliable and effective tool to identify AECOPD.

Reply 5: I agree with you that more strengths and limitations should be discussed. Compared with other previous publication regarding biomarkers associated with AECOPD， our study has a strength in that it establishes a novel prediction model using rapid, inexpensive, and commonly available blood indicators and a large sample size. Furthermore, we discussed the frequent exacerbator phenotype of AECOPD, which is a significant phenotype of COPD characterized by experiencing at least two exacerbations per year and associated with the number of exacerbations. Therefore, we can draw up the treatment plan of stable period of COPD according to the evaluation results. In our study, we divided patients into the infrequent exacerbation group if they experienced less than two exacerbations within one year, while this group may be admitted to another hospital when experience another acute exacerbation. It remains to be determined whether the predictive performance we were able to detect in our cohort can likewise be observed in other groups, so we will validate our predictive model in future through data source from other medical centers. We have discussed other strengths and limitations in the discussion part.

Changes in the text: We have modified our text as advised (see Page 12-14, line 234-268).

Reviewer B

Comment 1: In this paper, the authors investigated the predictors of frequent exacerbations in COPD patients. This retrospective study was based on a large cohort of COPD patients (more than 2000), among which 18% had 2 or more exacerbations within 1 year. The authors identified a number of predictors of frequent exacerbations, mainly based on blood test data, and created a nomogram to predict frequent COPD exacerbations.

This article has a relatively simple and clear design, well structured, however I have a number of major concerns:

- patient characterization is not very complete (no pulmonary function, no clinical signs,

drug therapy, etc.), meanwhile, this is not a very typical COPD patient population, since 53% of patients never smoked;

- it is not clear why, in addition to blood test parameters and "general characteristics", other available parameters were not included in the model - lung function? Saturation? Dyspnea? Sputum characteristics? possibly, image data? etc.?

Reply 1: Thank you for pointing it out. We should add it as a limitation of this paper. We totally agree pulmonary function, clinical signs and drug therapy are essential characteristics for COPD patients, and those are part of the reason why the AUC of the prediction model is only 0.681. As mentioned in the methods part, all the data are extracted from the big data platform. The difficulties in the construction this platform lie in how to integrate multisource heterogeneous data. Take pulmonary function as an example, data were stored in a third-party supplier and were failed to be plugged into this platform due to extremely high cost at this moment. We expect to negotiate this out soon. As for drug therapy, they are recorded by physicians and are not easily collected from the big data platform. Referring to smoking results, it was extracted from patients' complaints in text form by natural language processing. People who didn't mention the smoking status may be categorized into never-smoker group.

Thank you for your question. Similar to the above question. The data of our study came from the big data platform our hospital constructed in 2022. This is the first research paper based on the platform to our knowledge. The granularity of the data needs to be continuously optimized upon research needs. We will include the image data in the near future, either by extract the key words of the image report by nature language processing or the image itself using deep learning image recognition. After gathering all the relevant data, we believe the precision of the prediction model will be highly improved and other interesting topics could be investigated.

Changes in the text: We have modified our text as advised (see Page 12-14, line 234-268).