Peer Review File

Article information: https://dx.doi.org/10.21037/tcr-23-2304

Reviewer A

1) *While the initial study conducted to develop the PLCO model using the NLST sample is cited, there have been several studies conducted since this study that has applied the PLCO model in real-world settings including in diverse populations. The lack of inclusion of these studies and comparing these results with the findings from the present study seems incomplete.*

Response: We added the three articles as advised in (page 4, lines 9-101, 103-109) (page 5, lines 125-128), (page 11, lines 273-275).

"Specifically, the United States has a disproportionate impact on African American individuals, leading to higher occurrence, diagnosis at advanced stages and lower survival rates (2). The National Lung Screening Trial (NLST) has shown that the low-dose computed tomography (LDCT) as a screening tool for LC has significantly improved detection and survival rates (2-5)." (page 4, lines 9-101)

"the study cohort consisted of 91% White and 4.5% African American participants (2). In addition, by current screening criteria, only 50% of those who will develop LC are currently eligible for LDCT monitoring (5). The United States Preventive Services Task Force (USPSTF) offers recommendations for identifying individuals at risk of lung cancer, advocating for LDCT screening. Their criteria for LC screening appear to disproportionately favor whites and the male population (2, 6,7). Neglecting to consider racial and sex differences in LC risk can lead to inadequate screening for minorities, such as African Americans, and females. (2, 6, 7)". (page 4, lines 9-103-109)

"A retrospective study in 2022 showed that PLCO2 has a greater sensitivity in predicting LC among African American population, women, and men compared to USPSTF. However, this study have been conducted in populations already diagnosed with LC (7). Further research is needed to enhance sensitivity in identifying individuals at a higher risk of LC for effective LC screening". (page 5, lines 125-128)

"Pasquinelli et al. also affirms the significance of employing more broader prediction models in racially diverse populations to address disparities in LC screening and outcomes (2)". (page 11, lines 273-275).

2) *Additionally, it is currently unclear as written if the SNH model has been validated using a sample other than the BMC safety-net patient population.*

Response: The SNH model was validated only on the BMC population. We added some information to clarify.

"We used this database to apply both PLCO and SNH model. The SNH model has been exclusively implemented within the specific demographic of the BMC safety-net patient sample, while the PLCO model has been used throughout the literature." (page 6, lines 144-146)

3) *While acknowledged in the discussion, the small number of lung cancer cases is a limitation, and the results and conclusions seem to overstate what can be drawn from this current analysis.*

Response: We added additional information and an explanation on our expectations of using a small sample size.

"Moreover, our small sample size is consistent with existing literature, suggesting that a significant portion of screened individuals showed positive outcomes, with only a minimal percentage of the study population being diagnosed with LC (6)". (page 11, lines 281-283).

Reviewer B

1) *The authors appear to present results for their SNH model in the development data whereas for the PLCOm2012 model, the results are for external validation data.*

Response: The review is correct that the PLCOm2012 is a model that was developed from external data. In the current work, we applied the SNH and the PLCOm2012 in the same study population and compared their performances.

2) *It has been suggested that for prediction modelling one should have perhaps around 15 outcomes, lung cancers, per predictor evaluated in the model or overfitting will be a problem. In the authors model the sample size was small with only 38 outcomes. Overfitting is a real concern.*

Response: The review is correct; the sample size is small. We added additional information and an explanation on our expectations of using a small sample size.

"Moreover, our small sample size is consistent with existing literature, suggesting that a significant portion of screened individuals showed positive outcomes, with only a minimal percentage of the study population being diagnosed with LC(6)". (page 11, lines 281-283).

3) No internal or external validation was presented.

Response: Thank you for the very valid comment. We presented our internal validation by performing a regression analysis to assess the difference between each model, where we controlled for confounders. This can be found in the following sections.

a) Method section
"Univariate linear and logistic regression was conducted to evaluate the association between the differences among the SNH and PLCO models and the individual covariates. Multivariate regression was applied to control for confounders (age, sex, race, BMI, education, emphysema, COPD, personal history of lung cancer, family history of lung cancer, smoking status, and pack-year)." (page 7, lines 177- 180)

b) Result section
"The LC group showed positive coefficients for age, sex, race, BMI, and emphysema; however, only emphysema was statistically significant ($P< 0.001$) (Table 3). After adjusting for all variables, emphysema remained statistically significant ($P < 0.001$) (Table 4). On the other hand, the presence of COPD, personal history of cancer, family history of LC, and greater smoking pack-years was associated with a reduction in difference between the two models (Table 3). Thus, both models demonstrated the same precision regarding LC prediction, when the individual had a high pack-year history. However, there was no statistically significant difference among these variables.
Among the non-LC cohort, the alignment between SNH and PLCO risk scores per patient was optimized with the use of several patient variables (Table 3), with statistical significance specifically observed in the patient characteristics of age ($P < 0.001$), personal history of cancer ($P < 0.001$), and pack year ($P< 0.001$). When controlling for all the variables, age, personal history of cancer, and pack year maintained their statistical significance. Given the negative coefficient values derived from univariate regression analysis of these three variables, these results indicate that both models performed similarly among patients of a younger age, without a personal history of cancer, and with less smoking pack years. Whereas the difference between models increased with high BMI ($P< 0.001$), other education ($P = 0.03$), and family history of LC ($P = 0.006$) (Table 3). After controlling for all the variables, BMI and family history of LC continued to show statistical significance with respect to an associated difference between the models (Table 4). Thus, the SNH model was more accurate to predict non-LC when the patient had low BMI and no family history of LC". (page 9, lines 218-234).

4) The approach to analysis is contorted and deficient.

Response: Thank you so much for your valid comment. Currently, several studies compare the precision of the PLCOm2012 with the USPSTF guidelines among African Americans and a consistent theme is shown; there is a high false positive rate during LDCT screening. Therefore, the goal of our study was to evaluate other types of precision models to properly represent a more diverse population. Since the SNH model was previously published, we believed this model should be compared to the more established PLCOm2012 model for LC screening. (PLCOm2012). This model was derived from a predominantly Caucasian population and its effectiveness in a safety net hospital (SNH) population is unknown.
Our study showed the comparison of these two models by comparing their risk classification score and their sensitivity, positive and negative predictive value, and specificity. This is shown in the statistical analysis section.

"The primary objective of the study was to compare the performance of the PLCO and the SNH models. This was accomplished by evaluating the models' score within each group (LC and non-LC) using descriptive statistics. Student t-test was run on the mean score to determine whether the models differed in each group. Following score submission, risk was then classified as either low, moderate, and high. The risk classification was described by using counts (total number of scores in each risk) and their percentage within each model for both groups.
Per risk classification, data points for sensitivity, positive and negative predictive value, and specificity were gathered to determine which model was best tailored to the population under review. Sensitivity was defined as the probability of being in the moderate or high-risk group among patients with LC. Similarly, specificity was interpreted as the probability of being in the low-risk group among patients without LC. Positive predictive value defined the probability that a patient who has moderate or high-risk classification actually has LC, while negative predictive value was interpreted as the probability that a subject with low risk actually does not have LC". (page 7, lines 165-176)

5) *That race showed no relationship in the SNH model may be a consequence of small sample and smaller yet subsample categories and does not match observed data and validated models.*

Response: The review is correct. The sample size might have an impact on this result. Race did not influence the SNH model to predict LC in comparison to the PLCO model. (page 9, lines 218-223)