scientific reports

Check for updates

OPEN

Development and validation a nomogram to predict long-term mortality risks of PRISm and mild-to-moderate COPD based on NHANES 2007–2012

Xiangqing Hou[1,2] & Pixin Ran[1,2]✉

Chronic Obstructive Pulmonary Disease (COPD) can be prevented in the pre-clinical and early stages. However, very limited prediction models of COPD focus on Preserved Ratio Impaired spirometry (PRISm) and early stages. To fill this gap, this study aimed to develop and validate a nomogram to predict long-term mortality risks of PRISm and early COPD. We obtained data of participants in the US National Health and Nutrition Examination Surveys 2007–2012 and the available mortality follow-up data from the date of survey participation to Dec 31, 2019. The study population (n = 1043) was randomly divided into training and validation datasets at a ratio of 7:3. The cox proportional hazards model was applied to select significant prognostic risk factors of COPD in the training dataset. Besides, the predictive power and clinical usage value were assessed by the area under time dependent receiver operating characteristic curve (time-dependent AUROC), calibration curves and decision curve analysis (DCA). Moreover, directed acyclic graph (DAG) was utilized to plot causal associations between risk factors and mortality. We developed an accurate and easy to use nomogram using six predictors (age, passive smoking, alkaline phosphotase, gamma glutamyl transferase, lactate dehydrogenase, potassium). The nomogram had satisfactory predictive performance, as the time-dependent AUROC with 95% confidence interval (CI) at 7.5 years was 0.78 (0.69–0.84) and 0.80 (0.67, 0.87) in the training and validation datasets, respectively. The calibration curves and DCA also showed that the nomogram had good clinical usage value. Compared with the low-risk groups, the Hazard Ratio in the high-risk group was 2.25 (95% CI 1.29–3.94) in the validation datasets, respectively. DAG shown that there had directly associations of passive smoking and lactate dehydrogenase with all-cause mortality. The nomogram has the potential to identify high-risk populations in the pre-clinical and early stages of COPD.

Keywords PRISm, COPD, Nomogram, Mortality, NHANES, DAGs

Abbreviations

COPD	Chronic obstructive pulmonary disease
PRISm	Preserved ratio impaired spirometry
the time-dependent AUROC	The area under time dependent receiver operating characteristic curve
NHANES	The US National Health and Nutrition Examination Surveys
ALP	Alkaline phosphatase
GGT	Gamma glutamyl transferase
LDH	Lactate dehydrogenase
FENO	Fractional exhaled nitric oxide
NB	Net benefit

[1]Guangzhou National Laboratory, No. 9 XingDaoHuanBei Road, Guangzhou International Bio Island, Guangzhou 510005, Guangdong Province, China. [2]State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou Medical University, 151 Yanjiang Xi Rd., Guangzhou 510120, Guangdong Province, China. ✉email: pxran@gzhmu.edu.cn

LLNs	Lower limits of normal
FEV1	Forced expiratory volume in the first 1.0 s
FVC	Forced vital capacity

Chronic obstructive pulmonary disease (COPD) has become the third leading cause of death in the world according to the 2019 Global Burden of Disease Data Report[1]. COPD is a prevalent chronic, progressive, inflammatory airway disease with a lengthy course, and it can be prevented in its early stages. Therefore, the early identification of COPD patients at high risk of death is crucial for precision health management. Moreover, COPD is a heterogeneous disease, thus the establishment of an effective death risk predictive model enables risk stratification and provides personalized treatment strategies. Pre-clinical COPD patients refer to individuals in whom spirometry cannot detect airflow limitation, but who are likely to develop overt airflow obstruction and eventually progress to COPD[2]. Preserved Ratio Impaired Spirometry (PRISm) is defined as a condition where the ratio of forced expiratory volume in the first second to forced vital capacity (FEV1/FVC) is greater than or equal to 0.7, but the FEV1 is less than 80% of the predicted value. This condition is recognized as a form of pulmonary dysfunction that does not meet the spirometric criteria for chronic obstructive pulmonary disease (COPD). PRISm has been significantly associated with an increased risk of worse dyspnea, reduced walking distance, emphysema, decreased total lung capacity, and an elevated segmental bronchial wall area percentage. Additionally, it is linked to heightened respiratory symptoms, systemic inflammation, and increased mortality[3]. However, not all PRISm patients ultimately develop COPD. For PRISm or even mild to moderate COPD, there is an unresolved balance when considering whether to implement early intervention, given the cost-effectiveness. Therefore, early identification and prevention of PRISm and mild to moderate COPD through predicting long-term mortality risks are of considerable significance in the management of COPD.

Despite numerous prognostic models for chronic obstructive pulmonary disease (COPD) having been proposed, few have focused on preserved ratio impaired spirometry (PRISm) and mild to moderate COPD. The primary reasons for this gap include: (1) Early diagnosis of COPD remains clinically challenging due to significant individual variability in clinical manifestations; (2) PRISm, while associated with pulmonary dysfunction, does not meet the spirometric criteria for COPD, leading to the exclusion of such patients in prognostic model development; and (3) PRISm is widely regarded as a precursor state of COPD, necessitating long-term follow-up data to observe mortality outcomes. To address these limitations, this study aimed to integrate PRISm and mild to moderate COPD cohorts into a unified "early-stage COPD" category for prognostic model development, thereby preventing the exclusion of PRISm individuals. The study population was derived from the NHANES database, which provides spirometry data and approximately 10 years of follow-up data on all-cause mortality. The proposed prognostic model serves as a critical supplement to existing frameworks, offering broader applicability to early disease stages and enhancing opportunities for timely intervention. Moreover, the existing prognostic models for outcome prediction in COPD still have significant room for improvement and require additional assessment for clinical utility and cost-effectiveness[4]. Currently, there is a lack of an effective early warning model for COPD mortality risk in clinical practice. Possible reasons include severe methodological limitations in previous relevant studies, a lack of external validation, and insufficient representativeness of the sample. To address these issues, we followed a step-by-step guide[5] to develop and evaluate a clinical nomogram model for predicting long-term mortality risks of PRISm and mild to moderate COPD. A nomogram is a straightforward and easily interpretable predictive tool, designed to be accessible not only to specialized researchers but also to patients in clinical practice[6]. The outcome risk can be calculated and indicated by accumulation of risk scores corresponding to multiple predictors of diseases. The nomogram excels in visualizing information for patients and facilitating immediate decision-making at the bedside, offering a clear advantage over traditional, complex mathematical formulas[7–9]. When developing a clinical risk predictive model, both clinical predictive performance and simplicity must be considered simultaneously, in accordance with Occam's razor criteria[10].

Many studies[4,11] have emphasized that combining multiple prognostic predictors of COPD to establish death risk models would provide a more precise method of evaluation. The National Health and Nutrition Examination Survey (NHANES) has been examining a nationally representative sample of the US population since 1999, continuously collecting health-related data. Recently, data collected from several NCHS population surveys have been linked with death certificate records from the National Death Index, providing mortality follow-up data up to December 31, 2019. In this study, to build up a more comprehensive prognostic model, we obtained demographics, spirometry (pre and post-bronchodilator), laboratory and questionnaire data of patients administered by highly trained medical personnel from the NHANES program (2007–2012). Demographics, such as age, gender, BMI, etc., have been reported to be associated with a poor prognosis in COPD[12], however, the predictive power for long-term mortality of the demographics is still unclear. Several studies[13,14] also suggested that blood biomarkers are easily accessible and could be reliable strong predictors of mortality in COPD. Nevertheless, the findings of various researches are inconsistent, perhaps due to the heterogeneity of the study populations. Therefore, whether blood biomarkers can accurately predict mortality in COPD still requires more evidence through comprehensive studies. Although evidence suggests that radiological markers, such as the CT emphysema index, can be effective prognostic predictors of COPD patients[15] because they are significantly linked with increased mortality. However, they are not easily accessible and often have a higher cost-effective than routine blood biomarkers.

This study aimed to establish an accurate nomogram for predicting long-term personalized mortality risks of PRISm and mild to moderate COPD by integrating easily accessible demographics and blood biomarkers.

Methods and materials

Study population and data description

NHANES (https://www.cdc.gov/nchs/nhanes/index.htm) is a national representative, large-scale, ethically approved epidemiological survey program of the US population. It aims to determine the prevalence of major diseases and risk factors by continuedly collecting demographic, socioeconomic, dietary and health-related data, including medical, dental, and physiological measurements, as well as laboratory tests. We downloaded three waves of cross-sectional survey data from NHANES (2007–2012) based on the availability of spirometry data and linked this data with the National Death Index (NDI). Detailed information about the NDI has been reported elsewhere[16]. The study population was limited to 18 years and above because mortality follow-up data is available for this age group. This study included only Caucasian, African-American and Mexican–American participants to calculate LLNs (Lower Limits of Normal) by inputting age, height, gender and ethnicity. The LLNs were computed for various spirometry parameters using NHANES III equations[17]. The exclusion criteria were as follows: (1) aged less than 18 years; (2) spirometry examination was "missing" or its quality was not rated 'A' or 'B'; (3) race was "missing" or categorized as "Other Hispanic" or "Other Race—Including Multi-Racial"; (4) refusal to report smoking history or secondhand smoking history. Ultimately, a total of 1043 participants (595 were males and 448 were females) aged 18 to 79 years met the inclusion criteria and were included in the formal analysis.

Data on demographics, including age, gender, BMI, and race, were obtained to provide baseline individual information. Lung function parameters, specifically Forced Expiratory Volume in the first 1.0 s (FEV1) and Forced Vital Capacity (FVC), were obtained from pre- and post-bronchodilator spirometry examinations to define preserved ratio impaired spirometry (PRISm) and mild to moderate COPD. A standard biochemistry profile, covering 24 serum biomarkers, was measured to reflect lipid metabolism, respiratory and metabolic systems, as well as other disorders. Additionally, complete blood count measurements provided 20 biomarkers to describe overall health and indicate a wide range of conditions, such as anemia. The measurement of fractional exhaled nitric oxide (FENO) was used to depict respiratory health and was included in the analysis of this study. All laboratory test results underwent standard processing, proper storage, and shipping before being analyzed in the NHANES Laboratory. A detailed description of the laboratory methods, quality assurance, and quality control can be found on the NHANES website (https://www.cdc.gov/nchs/nhanes/index.htm).

Main variable definition

Participants with a post-bronchodilator FEV_1/FVC ratio of less than 0.7 were considered to have COPD. Mild to moderate COPD was represented by Global Initiative for Chronic Obstructive Lung Disease (GOLD) Stage 1 and 2, according to the GOLD criteria[18]. GOLD Stage 1 is classified as mild COPD with an FEV1 ≥ 80% of the predicted value; GOLD 2 stage also considered to moderate COPD if the FEV_1 value is between 50 and 79%.

PRISm is characterized by a decline in the one-second ratio with normal pulmonary function. In this study, the definition of PRISm[19] was based on a pre-bronchodilator FEV_1/FVC ratio of less than 0.7 and FEV1% less than 80% to expand the study samples[20]. We calculated the predicted values of spirometry parameters using NHANES III equations in this study.

'Smoking history' was defined as a 'Yes' answer to the question '*Have you smoked at least 100 cigarettes in your entire life*', and 'passive smoking history' was defined as a 'Yes' answer to the question '*Does anyone smoke inside home?*'.

Statistical analysis

The sample size estimate of this study was completed according to the method provided by Richard[21]

$$n = \frac{P}{(S-1)\,ln\left(1-\frac{R^2_{cs}}{S}\right)}$$

Our nomogram risk model for predicting time-to-event outcomes was developed using six predictors (P), additionally, the anticipated Cox-Snell R squared statistic (R^2_{cs}) and the targeted expected shrinkage factors (S) were set at 0.1 and 0.9, respectively[22]. Consequently, the required sample size of this study was determined to be at least 509, which would meet the necessary criteria.

For missing value, continuous variables were imputed using the median, and categorical variables were filled with the mode. The description was conducted before and after data filling (Supplementary Table 1). We used a completely randomized sampling method to divide the data into a training set and a validation set in a ratio of 7:3. The baseline demographics and laboratory test results were described and compared between the training and validation datasets to assess balance and comparability. Continuous variables were described as mean and standard deviation if the variable met a normal distribution; otherwise, the variables were depicted as median and interquartile range. Categorical variables were reported as frequency and proportion. The comparison between training and validation datasets was performed using Chi-square tests and Kruskal–Wallis tests. The prediction model was constructed based on the training set, and the prediction effect of the model was verified on the validation set. Based on the final prediction model, the time-dependent ROC, calibration curve, and DCA curve were used to comprehensively evaluate the prediction accuracy, calibration, and clinical net benefit of the model. The nomogram was formulated to calculate the personalized risk of death and identify high-risk population. Based on the final prediction model, the time-dependent ROC, calibration curve, and DCA curve were used to comprehensively evaluate the prediction accuracy, calibration, and clinical net benefit of the model. Additionally, X-tile software was used to determine the optimal cut-off value of the nomogram risk score, allowing patients to be divided into low and high-risk groups. Implementation of Spirometry Equations

was completed using the R package '*rspiro*'. Directed acyclic graph was constructed based on causal discovery with latent confounders of the 6 predictors using PC algorithm.

All data cleaning and statistical analysis were completed using R-software and SAS 9.4 software (Copyright 2002–2012 by SAS Institute Inc., Cary, NC, USA). The figures were drawn using R-studio 4.4.1 (Copyright 2009–2019 RStudio, Inc.) and X-tile software (Robert L Camp, MD., Ph.D. Copyright Yale University 2003-05). A *P* value with 2-sides less than 0.5 was considered statistically significant.

Role of funders
The funding sources listed in the manuscript did not have a role in the study design, analysis, interpretation, or writing of the manuscript. The decision to submit the manuscript for publication was made solely by the authors listed.

Results
Characteristics of participants
Finally, this study included a total of 1043 patients with PRISm (n = 638) and mild (n = 266) to moderate (n = 139) COPD in the formal analysis. The median age with interquartile range of the participants was 55 (42, 65) years. More than half of the participants (58.36%) were Non-Hispanic White. The median follow-up time was 9.8 years (IQR 8.2, 11.2), ending on the research's conclusion date (Dec 31, 2019). The all-cause mortality rate of the sample during the follow-up period was 16.59% (173/1043). The top three leading causes of death were malignant neoplasms (n = 54), heart diseases (n = 36), and chronic lower respiratory diseases (n = 10).

All samples were divided into training (n = 730) and validation (n = 313) datasets to establish a mortality prediction model and further validate the model's performance. The comparison between the training and validation datasets revealed that the demographics and laboratory test results were balanced and comparable ($P > 0.05$), with the exception of red blood cell count and passive smoking history (Table 1).

Predictors screening depending on the training set
Based on the training datasets, a series of univariate Cox regression models was developed to identify 24 significant ($P < 0.05$) candidate predictors of mortality for PRISm and mild to moderate COPD. Ultimately, a multiple Cox regression model was formulated to select the most important predictors for predicting mortality risks, including these 24 variables (Table 2). The results indicated that passive smoking history, age, lactate dehydrogenase (LDH), alkaline phosphatase (ALP), gamma glutamyl transferase (GGT) and potassium remained associated with COPD mortality after controlling for other potential confounders. To assess the stability of the model and ensure that the performance estimates are not inflated, The Cox Lasso regression model was additionally employed to evaluate the variable selection approach in this study. The results (Supplementary Fig. 1) demonstrate that the final model includes age, passive smoking, alkaline phosphatase, gamma-glutamyl transferase, lactate dehydrogenase, potassium, and 11 other confounders. This finding aligns with the results obtained from the stepwise selection approach described in our manuscript. However, in accordance with Occam's Razor[10], which emphasizes the superiority of models with fewer variables for achieving optimal results, we aimed to develop a streamlined yet highly predictive model. Consequently, we ultimately selected 6 significant factors that exhibit excellent discriminatory ability, as evidenced by an area under the ROC curve (AUROC) exceeding 0.7 for each. Besides, we conducted a multicollinearity test and found that the variance inflation factors (VIFs) for the six predictors—age, passive smoking, alkaline phosphatase, gamma-glutamyl transferase, lactate dehydrogenase, and potassium—were close to 1. This indicates that multicollinearity was not a significant concern in the formal analysis. Additionally, we performed Spearman correlation analysis among the six predictors and found no strong associations between the variables, as all Spearman's correlation coefficients were below 0.60. These results further confirm the absence of significant multicollinearity in our analysis (Supplementary Tables 2, 3). Besides, The Kaplan–Meier curve analysis suggested that passive smoking history was associated with an increased risk of death ($P = 0.028$, Supplementary Fig. 2). Additionally, the cubic spline curves revealed a distinct positive exposure–response relationship between age and the mortality risk of COPD. Our results also demonstrated that the risk of mortality in COPD significantly increased as the levels of ALP exceeded 80 U/L, and the association between GGT and death risk showed an increasing trend, particularly at higher levels of GGT. Furthermore, we identified an S-shaped association between LDH levels and all-cause mortality in participants with PRISm and mild to moderate COPD. Moreover, a significant positive association existed between potassium levels and the risk of death when potassium measurements were more than 4 mmol/L (Supplementary Fig. 3).

Development and validation of the nomogram
In this part, an accurate nomogram score system was developed using passive smoking history, age, lactate dehydrogenase (LDH), alkaline phosphatase (ALP), gamma glutamyl transferase (GGT) and potassium (Fig. 1). The calculated formula of the nomogram is as follows:

$$Nomogram = Passivesmokinghistory \begin{Bmatrix} Yes, 15.15 \\ No, 0 \end{Bmatrix} + (Age \times 1.2 - 1.8) + (ALP \times 0.125) + (GGT \times 0.099) + (LDH \times 0.1836 - 9.18) + (Potassium \times 15.32 - 38.3)$$

The nomogram demonstrated good predictive performance in forecasting long-term mortality risks for PRISm and mild to moderate COPD, as the time-dependent AUROC reached 0.767 (95% CI, 0.693–0.841), 0.780 (95% CI, 0.693–0.837), and 0.757 (95% CI, 0.723–0.811) at 5, 7.5, and 10 years, respectively, in the training dataset. Moreover, the corresponding time-dependent AUROC were 0.762 (95%CI, 0.667–0.857), 0.804 (95%CI, 0.667–0.869) and 0.776 (95%CI, 0.739–0.843) in the validation dataset (Fig. 2). Moreover, the integrated AUC (iAUC)

Label	Total (n = 1043)	Training (n = 730)	Validation (n = 313)	*P* value
Gender				0.310
Male	595(57.1)	409(56.0)	186(59.4)	
Female	448(42.9)	321(44.0)	127(40.6)	
Race				0.502
Mexican American	121(11.6)	90(12.3)	31(9.9)	
Non-Hispanic White	617(59.2)	426(58.4)	191(61.0)	
Non-Hispanic Black	305(29.2)	214(29.3)	91(29.1)	
Smoking history				0.154
Yes	666(63.8)	456(62.5)	210(67.1)	
No	377(36.2)	274(37.5)	103(32.9)	
Passive smoking history				0.027
Yes	295(28.3)	192(26.4)	103(33.1)	
No	748(71.7)	536(73.6)	208(66.9)	
Participants				0.616
Prism	638(61.2)	448(61.4)	190(60.7)	
GOLD1	266(25.5)	181(24.8)	85(27.2)	
GOLD2	139(13.3)	101(13.8)	38(12.1)	
Age, years	55.0(42.0,65.0)	55.0(43.0,65.0)	54.0(40.0,63.0)	0.396
BMI	29.5(25.0,34.8)	29.7(25.1,35.0)	28.9(24.5,33.9)	0.140
Albumin (g/dL)	4.2(4.0,4.4)	4.2(4.0,4.4)	4.2(4.0,4.4)	0.181
ALT (U/L)	21.0(17.0,28.0)	21.0(17.0,28.0)	21.0(17.0,28.0)	0.664
AST (U/L)	23.0(20.0,28.0)	23.0(20.0,28.0)	23.0(20.0,27.0)	0.575
Alkaline phosphotase (U/L)	68.0(57.0,82.0)	68.0(58.0,83.0)	68.0(56.0,81.0)	0.196
Blood urea nitrogen (mg/dL)	12.0(10.0,15.0)	12.0(10.0,15.0)	12.0(10.0,15.0)	0.175
Total calcium (mg/dL)	9.4(9.2,9.6)	9.4(9.2,9.6)	9.4(9.2,9.7)	0.753
Cholesterol (mg/dL)	193.0(168.0,220.0)	193.0(169.0,221.0)	193.0(167.0,220.0)	0.618
Bicarbonate (mmol/L)	25.0(24.0,27.0)	25.0(24.0,27.0)	25.0(24.0,27.0)	0.739
Creatinine (mg/dL)	0.9(0.8,1.0)	0.9(0.8,1.0)	0.9(0.8,1.0)	0.619
GGT (U/L)	22.0(16.0,33.0)	22.0(16.0,33.0)	22.0(16.0,33.0)	0.808
Glucose (mg/dL)	94.0(87.0,106.0)	94.0(87.0,106.0)	94.0(84.0,106.0)	0.260
Iron (ug/dL)	78.0(61.0,99.0)	78.0(61.0,100.0)	78.0(59.0,97.0)	0.475
LDH (U/L)	132.0(119.0,146.0)	132.0(119.0,147.0)	132.0(120.0,145.0)	0.526
Phosphorus (mg/dL)	3.7(3.4,4.0)	3.7(3.4,4.0)	3.7(3.3,4.1)	0.883
Total bilirubin (mg/dL)	0.7(0.6,0.8)	0.7(0.6,0.8)	0.7(0.6,0.9)	0.144
Total protein (g/dL)	7.1(6.8,7.4)	7.1(6.8,7.4)	7.1(6.9,7.4)	0.833
Triglycerides (mg/dL)	127.0(86.0,194.0)	127.0(86.0,195.0)	127.0(88.0,190.0)	0.824
Uric acid (mg/dL)	5.6(4.8,6.5)	5.6(4.8,6.5)	5.6(4.8,6.5)	0.897
Sodium(mmol/L)	139.0(138.0,141.0)	139.0(138.0,141.0)	139.0(138.0,141.0)	0.929
Potassium (mmol/L)	4.0(3.8,4.2)	4.0(3.8,4.2)	4.0(3.8,4.2)	0.491
Chloride (mmol/L)	104.0(102.0,106.0)	104.0(102.0,106.0)	104.0(102.0,106.0)	0.833
Osmolality (mOsm/kg)	278.0(276.0,281.0)	278.0(276.0,282.0)	278.0(276.0,281.0)	0.629
Globulin (g/dL)	2.9(2.6,3.2)	2.9(2.6,3.2)	2.9(2.6,3.2)	0.481
White blood cell count (1000 cells/uL)	7.2(6.0,8.6)	7.2(6.0,8.5)	7.2(5.9,8.8)	0.208
Lymphocyte percent (%)	29.3(24.7,34.2)	29.3(24.9,34.2)	29.3(24.7,33.7)	0.559
Monocyte percent (%)	7.6(6.3,9.0)	7.6(6.2,9.0)	7.6(6.4,9.0)	0.769
Segmented neutrophils percent (%)	59.1(53.6,64.7)	59.1(53.8,64.5)	59.1(52.9,64.8)	0.634
Eosinophils percent (%)	2.6(1.7,3.7)	2.6(1.7,3.7)	2.6(1.7,3.7)	0.746
Basophils percent (%)	0.6(0.4,0.9)	0.6(0.4,0.9)	0.6(0.4,0.8)	0.165
Lymphocyte number (1000 cells/uL)	2.1(1.7,2.6)	2.1(1.7,2.6)	2.1(1.7,2.6)	0.584
Monocyte number (1000 cells/uL)	0.5(0.4,0.6)	0.5(0.4,0.6)	0.5(0.4,0.7)	0.136
Segmented neutrophils num (1000 cell/uL)	4.2(3.3,5.3)	4.2(3.3,5.3)	4.2(3.3,5.5)	0.258
Eosinophils number (1000 cells/uL)	0.2(0.1,0.3)	0.2(0.1,0.3)	0.2(0.1,0.3)	0.333
Basophils number (1000 cells/uL)	0.0(0.0,0.1)	0.0(0.0,0.1)	0.0(0.0,0.1)	0.501
Red blood cell count (million cells/uL)	4.7(4.4,5.0)	4.7(4.3,5.0)	4.7(4.4,5.1)	0.024
Hemoglobin (g/dL)	14.4(13.4,15.3)	14.4(13.4,15.3)	14.4(13.5,15.4)	0.217
Hematocrit (%)	41.9(39.2,44.6)	41.9(39.0,44.6)	41.9(39.7,44.9)	0.163
Continued				

Label	Total (n = 1043)	Training (n = 730)	Validation (n = 313)	*P* value
Mean cell volume (fL)	89.9(86.6,93.1)	89.9(86.8,93.2)	89.9(86.4,92.8)	0.160
Mean cell hemoglobin (pg)	30.9(29.5,32.2)	30.9(29.6,32.2)	30.9(29.5,32.0)	0.371
Mean cell hemoglobin concentration (g/dL)	34.3(33.6,34.9)	34.3(33.6,34.9)	34.3(33.6,34.9)	0.708
Red cell distribution width (%)	12.9(12.3,13.5)	12.9(12.4,13.5)	12.9(12.3,13.5)	0.451
Platelet count (1000 cells/uL)	240.0(205.0,286.0)	240.0(203.0,285.0)	240.0(206.0,286.0)	0.728
Mean platelet volume (fL)	8.0(7.4,8.6)	8.0(7.4,8.6)	8.0(7.4,8.4)	0.360
FENO[¶]	12.5(8.0,20.0)	12.5(8.5,20.0)	12.5(8.0,20.0)	0.316

Table 1. The baseline characteristics of population in the training and validation datasets. ¶: Mean of two reproducible FENO measurements, in parts per billion. *ALT* alanine transaminase, *AST* aspartate transaminase, *GGT* Gamma-glutamyl transferase, *LDH* Lactate dehydrogenase.

over the entire follow-up period were 0.806 and 0.795 in the training and validation datasets, respectively (Supplementary Fig. 4). The time-dependent AUROC based on the various time points suggested that the nomogram had good prediction stability during follow up (Supplementary Fig. 5).

The calibration curves (Fig. 3) confirmed that the observed and predicted probabilities of 5-, 7.5-, and 10-year mortality according to the nomogram were well-aligned in both the training and validation datasets. Furthermore, the decision curve analysis indicated that the nomogram offered a clear net benefit in clinical usage over treating all or none (Fig. 4). More specifically, the nomogram achieved maximum clinical benefit compared to Model 1 (age, passive smoking history) and Model 2 (ALP, GGT, LDH, and potassium), suggesting that our nomogram is more useful than models based on individual clinical laboratory indicators and demographics.

Performance of the nomogram in stratifying risk of patients

To identify the high death risk population among PRISm and mild to moderate COPD patients, we further divided all study samples into high and low risk groups based on the selected cut-offs of the nomogram (Table 3). The cut-off value (121.45) was determined by X-tile analysis (Supplementary Fig. 6). The numbers and proportions of high- and low-risk groups were 133 (18.22%) and 597 (81.78%) in the training dataset, and 55 (17.57%) and 258 (82.43%) in the validation dataset, respectively. The hazard ratio (HR) in the high-risk group was 6.20 (95% CI 4.30–8.95) compared to the low-risk group in the training dataset. Additionally, the HR in the high-risk group was 2.25 (95% CI 1.29–3.94) compared to the low-risk group in the validation dataset. The Kaplan–Meier curve also suggested that there is a significant difference between high- and low-risk groups (Fig. 5). The C-index of the nomogram achieved 0.771 (95% CI, 0.727–0.816) in the training dataset and 0.737 (95% CI, 0.681–0.793) in the validation dataset, suggesting that the predicted results closely matched the actual results. Furthermore, the high predictive accuracy of the nomogram was also observed in the training (AUC = 0.81) and validation datasets (AUC = 0.76). The difference in mortality between high- and low-risk groups was significant, determined to be 37.57% in the training and 17.61% in the validation datasets.

Directed acyclic graph of all-cause mortality in COPD reveals biological signatures of mortality risk

To further reveal the direct or indirect causal relationship between 6 predictors and COPD mortality, we applied PC algorithm to draw causal discovery with latent confounders for the 6 predictors (Supplementary Fig. 7). Figure 6 shown that passive smoking directly related to the death risk of COPD patients. Besides, GGT perhaps an important confounder which influence the associations of ALP to LDH and increased the risk of mortality. The death risk of COPD patients ultimately attributed to potassium provide the key biological signatures and potential biological explanation.

Discussion

PRISm and mild to moderate COPD are pivotal stages for early screening of patients at high risk of mortality and for preventing poor prognoses. Although many mortality risk models have been proposed previously, there is limited research focused on predicting long-term mortality risks among PRISm and mild to moderate COPD patients using simple tests. In this study, we developed and validated an accurate and simple nomogram to predict long-term mortality risks of PRISm and mild to moderate COPD based on NHANES (2007–2012) database. The nomogram consists of 6 predictors (passive smoking history, age, LDH, ALP, GGT, and potassium) and is very convenient for clinical use. Additionally, it has satisfactory predictive power and is acceptable in terms of cost-effectiveness. All predictors can be obtained from routine epidemiological surveys and laboratory tests; therefore, our nomogram has the potential for widespread application in predicting the mortality of COPD patients, especially in resource-limited areas. An AUROC of 0.78 indicated that the model has moderate discrimination ability with an overall satisfactory prediction performance. An AUC greater than 0.7 was commonly recognized as a clinically meaning separation[23]. The calibration curves demonstrated a strong predictive consistency between the observed and predicted risks. Furthermore, the decision curve analysis (DCA) highlighted the clinical utility and superiority of the nomograms in terms of their practical impact. To the best of our knowledge, this study represents the first effort to develop a nomogram with superior performance for predicting survival probability in individuals with Preserved Ratio Impaired Spirometry (PRISm) and mild to moderate COPD, utilizing easily obtainable laboratory indicators and demographic characteristics. We believe this research is not

Parameter	Crude HR (95%CI)	*P* value	Model1 HR (95%CI)	*P* value
Female (ref = Male)	0.888(0.614,1.285)	0.529		
Race				
Non-Hispanic White (ref = 'Mexican American')	2.488(1.148,5.392)	0.021	1.795(0.786,4.097)	0.165
Non-Hispanic Black (ref = 'Mexican American')	2.184(0.961,4.963)	0.062	2.328(0.923,5.875)	0.074
Smoking history (ref = 'No')	0.506(0.331,0.775)	0.002	0.659(0.404,1.074)	0.094
Passive smoking history (ref = 'No')	0.654(0.447,0.957)	0.029	0.550(0.351,0.860)	0.009
Age, years	1.065(1.048,1.083)	0.001	1.062(1.041,1.083)	< 0.001
BMI	0.992(0.969,1.016)	0.508		
Albumin (g/dL)	0.549(0.306,0.986)	0.045	1.497(0.742,3.018)	0.260
ALT (U/L)	1.002(0.993,1.012)	0.672		
AST (U/L)	1.003(0.990,1.016)	0.634		
Alkaline phosphotase (U/L)	1.013(1.007,1.019)	0.001	1.007(1.001,1.013)	0.016
Blood urea nitrogen (mg/dL)	1.078(1.053,1.102)	0.001	1.014(0.974,1.055)	0.500
Total calcium (mg/dL)	1.181(0.711,1.962)	0.520		
Cholesterol (mg/dL)	0.998(0.994,1.003)	0.414		
Bicarbonate (mmol/L)	0.968(0.894,1.049)	0.427		
Creatinine (mg/dL)	1.608(1.351,1.913)	0.001	1.066(0.761,1.494)	0.710
GGT (U/L)	1.006(1.003,1.010)	0.001	1.005(1.001,1.009)	0.027
Glucose (mg/dL)	1.003(1.000,1.007)	0.060		
Iron (ug/dL)	1.002(0.997,1.008)	0.441		
LDH (U/L)	1.011(1.006,1.015)	0.001	1.010(1.002,1.017)	0.008
Phosphorus (mg/dL)	1.230(0.881,1.719)	0.224		
Total bilirubin (mg/dL)	1.180(0.650,2.141)	0.586		
Total protein (g/dL)	1.018(0.681,1.522)	0.932		
Triglycerides (mg/dL)	0.999(0.998,1.001)	0.419		
Uric acid (mg/dL)	1.173(1.034,1.330)	0.013	1.060(0.918,1.223)	0.427
Sodium(mmol/L)	0.919(0.845,1.000)	0.050		
Potassium (mmol/L)	3.019(1.896,4.805)	0.001	2.219(1.304,3.776)	0.003
Chloride (mmol/L)	0.941(0.884,1.002)	0.058		
Osmolality (mOsm/kg)	1.031(0.993,1.071)	0.115		
Globulin (g/dL)	1.349(0.907,2.006)	0.139		
White blood cell count (1000 cells/uL)	1.090(1.004,1.183)	0.039	0.970(0.598,1.574)	0.903
Lymphocyte percent (%)	0.955(0.932,0.980)	0.001	0.961(0.876,1.054)	0.401
Monocyte percent (%)	1.070(0.991,1.155)	0.085		
Segmented neutrophils percent (%)	1.030(1.008,1.053)	0.007	0.980(0.891,1.078)	0.684
Eosinophils percent (%)	1.006(0.920,1.101)	0.888		
Basophils percent (%)	0.829(0.587,1.170)	0.286		
Lymphocyte number (1000 cells/uL)	0.818(0.619,1.081)	0.158		
Monocyte number (1000 cells/uL)	3.835(1.512,9.724)	0.005	0.964(0.227,4.089)	0.961
Segmented neutrophils num (1000 cell/uL)	1.146(1.040,1.262)	0.006	1.158(0.584,2.296)	0.674
Eosinophils number (1000 cells/uL)	1.980(0.725,5.411)	0.183		
Basophils number (1000 cells/uL)	1.398(0.112,17.438)	0.795		
Red blood cell count (million cells/uL)	0.418(0.293,0.596)	0.001	0.056(0.001,4.182)	0.191
Hemoglobin (g/dL)	0.870(0.779,0.972)	0.014	5.240(0.388,70.856)	0.213
Hematocrit (%)	0.954(0.916,0.994)	0.026	0.736(0.241,2.252)	0.591
Mean cell volume (fL)	1.089(1.052,1.127)	0.001	1.288(0.765,2.166)	0.341
Mean cell hemoglobin (pg)	1.092(1.038,1.148)	0.001	0.366(0.109,1.228)	0.104
Mean cell hemoglobin concentration (g/dL)	0.903(0.759,1.076)	0.254		
Red cell distribution width (%)	1.259(1.136,1.396)	0.001	0.972(0.821,1.150)	0.738
Platelet count (1000 cells/uL)	0.999(0.997,1.002)	0.687		
Continued				

Parameter	Crude HR (95%CI)	*P* value	Model1 HR (95%CI)	*P* value
Mean platelet volume (fL)	0.900(0.741,1.093)	0.286		
FENO¶	0.991(0.977,1.004)	0.181		
Participants				
Gold 1 (ref = 'PRISm')	1.083(0.692,1.695)	0.726	0.761(0.462,1.253)	0.284
Gold 2 (ref = 'PRISm')	2.113(1.343,3.324)	0.001	1.380(0.835,2.279)	0.209

Table 2. The univariate and multivariable Cox regression model based on the training datasets. ¶: Mean of two reproducible FENO measurements, in parts per billion.

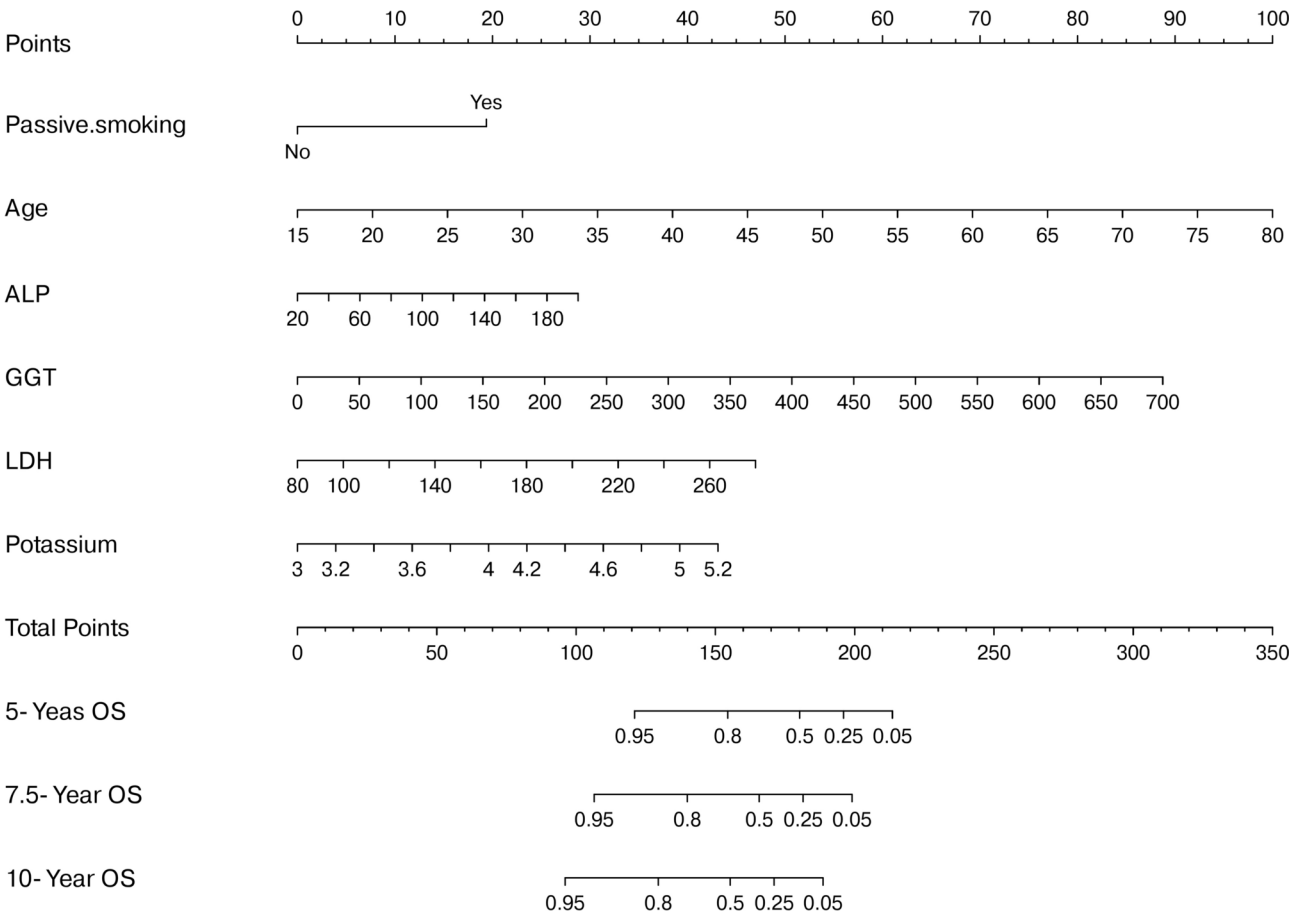

Fig. 1. Nomogram to predict the mortality risk of PRISm and mild and moderate COPD. Shown are clinical predictors and blood biomarkers as well as their corresponding nomogram points. The total points indicate the addition of the scores for each variable, then a vertical line can be drawn to determine the risk of mortality according to the total points.

only methodologically robust but also makes a significant contribution to the existing literature. The findings have the potential to greatly advance precise prevention and personalized management strategies for early-stage COPD.

Ideally, both p-values and the minimal clinically important difference (MCID) should be considered when determining the effectiveness of a given intervention[24]. However, as there is no widely accepted MCID for assessing mortality in COPD patients during the development and validation of prognostic models, this study references Marloes et al.'s paper[25], which summarizes healthcare professionals' preferred efficacy endpoints and MCID values for evaluating new COPD medicines. The authors recommend a 23% mortality reduction as the preferred efficacy endpoint. In this study, participants were stratified into high- and low-risk groups based on the selected best cutoff value. The observed difference in mortality between these groups was 37.57% in the training datasets, exceeding the referenced MCID threshold of 23%. Consequently, the effect sizes (HR = 6.20) derived from the nomogram score further support the conclusion that the high-risk group had a significantly greater mortality risk than the low-risk group.

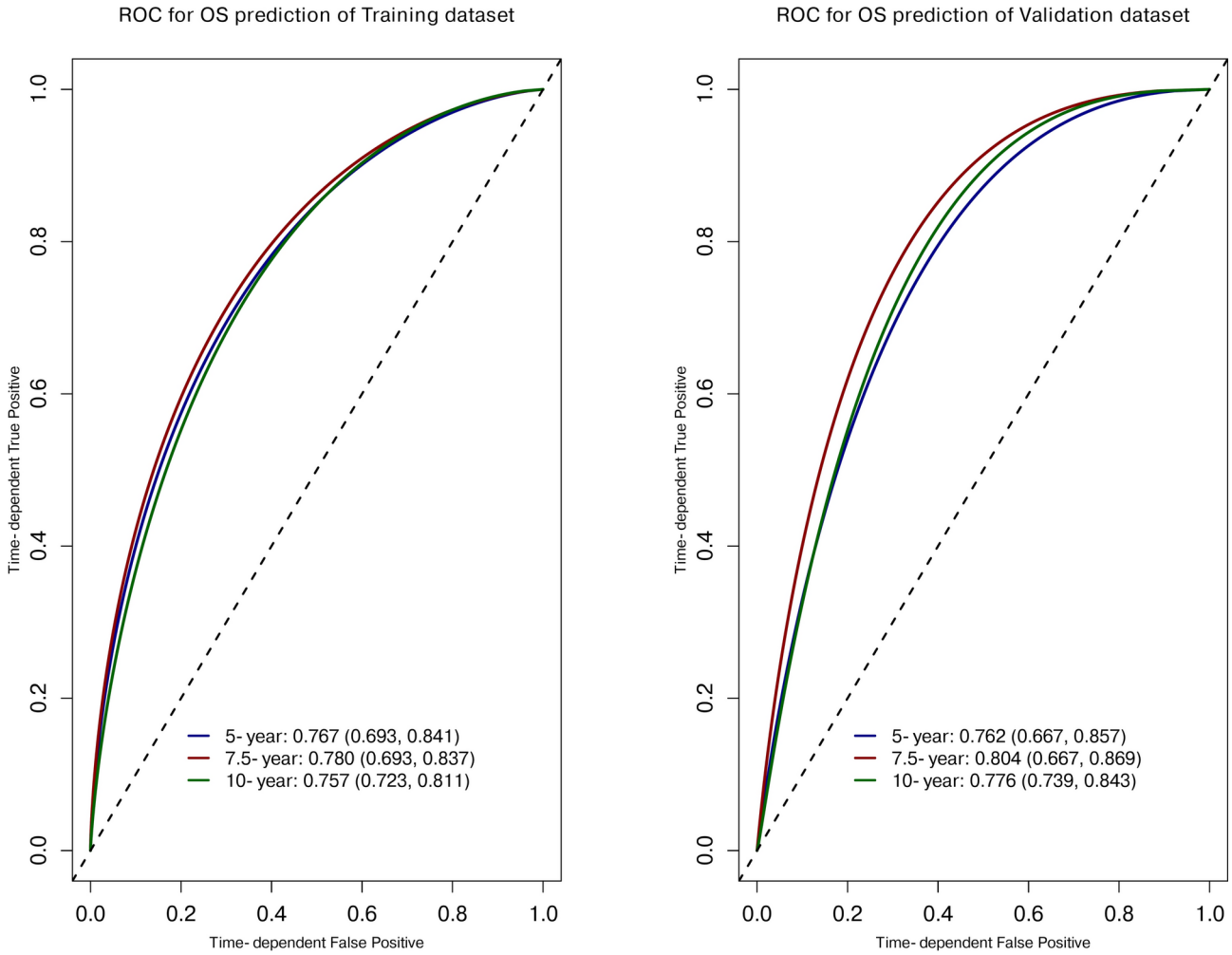

Fig. 2. Time-dependent ROC curves of the nomogram in the training and validation datasets. The labels show the AUC and its 95% confidence interval in the nomogram at 5 years (blue line), at 7.5 years (red line) and at 10 years (green line).

According to a previous systematic review[4], the most commonly used predictors in prognostic models for outcome prediction in COPD patients are age, FEV_1, sex, BMI and smoking. Similarly, our study also emphasized that age and passive smoking are significant risk factors associated with mortality in early-stage COPD. In addition to these common predictors, our nomogram includes several serum biomarkers to construct a death risk prediction model. Recently, blood biomarkers have attracted much interest among scientists due to their easy access and assay reproducibility. Studies on COPD serum biomarkers have explored various potential prognostic biomarkers, such as fibrinogen, C-reactive protein, and white cell count[14], yet their reproducibility still faces significant challenges due to a low rate of external validation. The review also reported that the summary C-index estimates ranged from 0.611 to 0.769 across the 12 prognostic models included in the meta-analysis. However, the C-index of our nomogram is 0.771, which is notably higher than the highest C-index among the 12 prognostic models. Moreover, BARC index[26], including age, airflow obstruction, body mass index, smoking, exacerbations and comorbidities has been proposed performed better than existing tools in predicting 1-year mortality among COPD patients. However, eighteen variables were included in the BARC which is obvious less simple than our nomogram based on the close AUC (0.78) of BARC and the nomogram. In clinical utility, our nomogram is best suited for predicting long-term mortality by integrating six predictors specific to early-stage COPD. In contrast, the BARC index demonstrates greater utility for short-term mortality risk prediction. Additionally, the VAPORED mortality risk score[27], developed specifically for patients with COPD, has shown significantly superior performance compared to the ADO, BODE, and updated BODE indices in predicting all-cause mortality. However, a key limitation of the VAPORED score and other existing models is their exclusive focus on COPD patients, thereby neglecting individuals with preserved ratio impaired spirometry (PRISm). Moreover, this nomogram offers simplicity by eliminating the need to collect time-consuming, costly, and labor-intensive data such as exercise capacity measurements and chest computed tomography (CT) imaging results. Compared with previous studies, this study has many noteworthy findings, and our nomogram model is more useful in clinical practice.

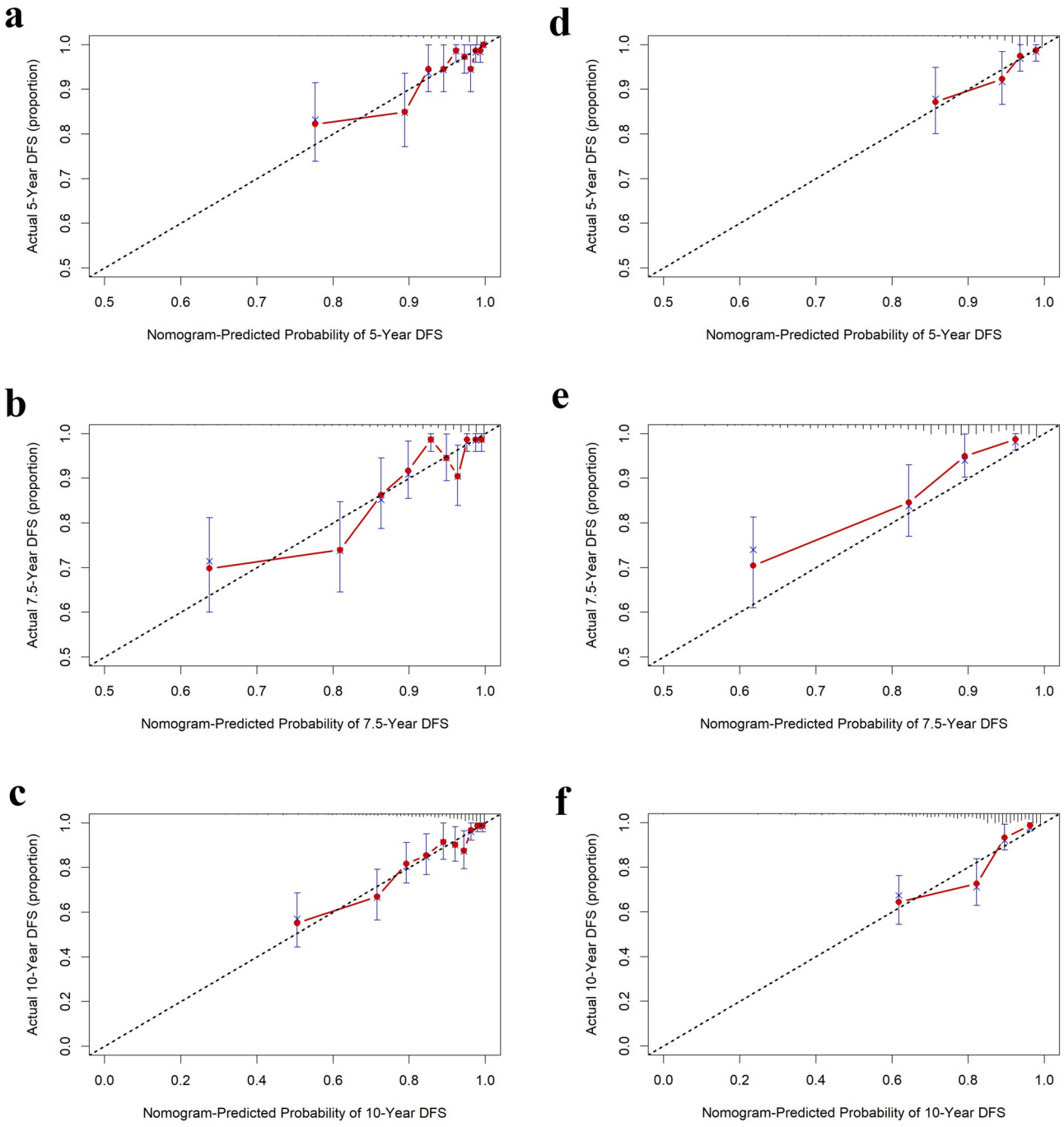

Fig. 3. Calibration curves of the nomogram in the training and validation datasets. Shown are the nomogram predicted probability (PP) for PRISm and early COPD mortality on the x-axis and the observed probabilities (OP) on the y-axis. Calibration curves of the prognostic nomogram for 5-year (**a**) disease free survival (DFS), 7.5-year DFS (**b**) and 10-year DFS (**c**) in the training set; calibration curves for 5-year DFS (**d**), 7.5-year DFS (**e**), and 10-year DFS (**f**) in the validation set.

This study suggests that lactate dehydrogenase (LDH), alkaline phosphatase (ALP), gamma glutamyl transferase (GGT) are significant predictors of mortality in COPD. LDH is an important enzyme of the anaerobic metabolic pathway which has been discussed frequently associated with cancer metabolism and inflammation[28]. A cross-sectional study has suggested that serum LDH is negatively and linearly correlated with pulmonary function indices[29], possibly due to hypoxia, inflammation, and cell damage, making it a potential biomarker for the progression of COPD. However, whether there is a causal association between LDH and poor prognostics of COPD remains unclear. A recent study[30] based on NHANES (2007–2012) found an obvious U-shaped relationship between LDH and all-cause mortality in patients with COPD. In contrast, we revealed the associations between LDH and the risk of pre or early COPD are tended to 'S' shape. This outcome difference

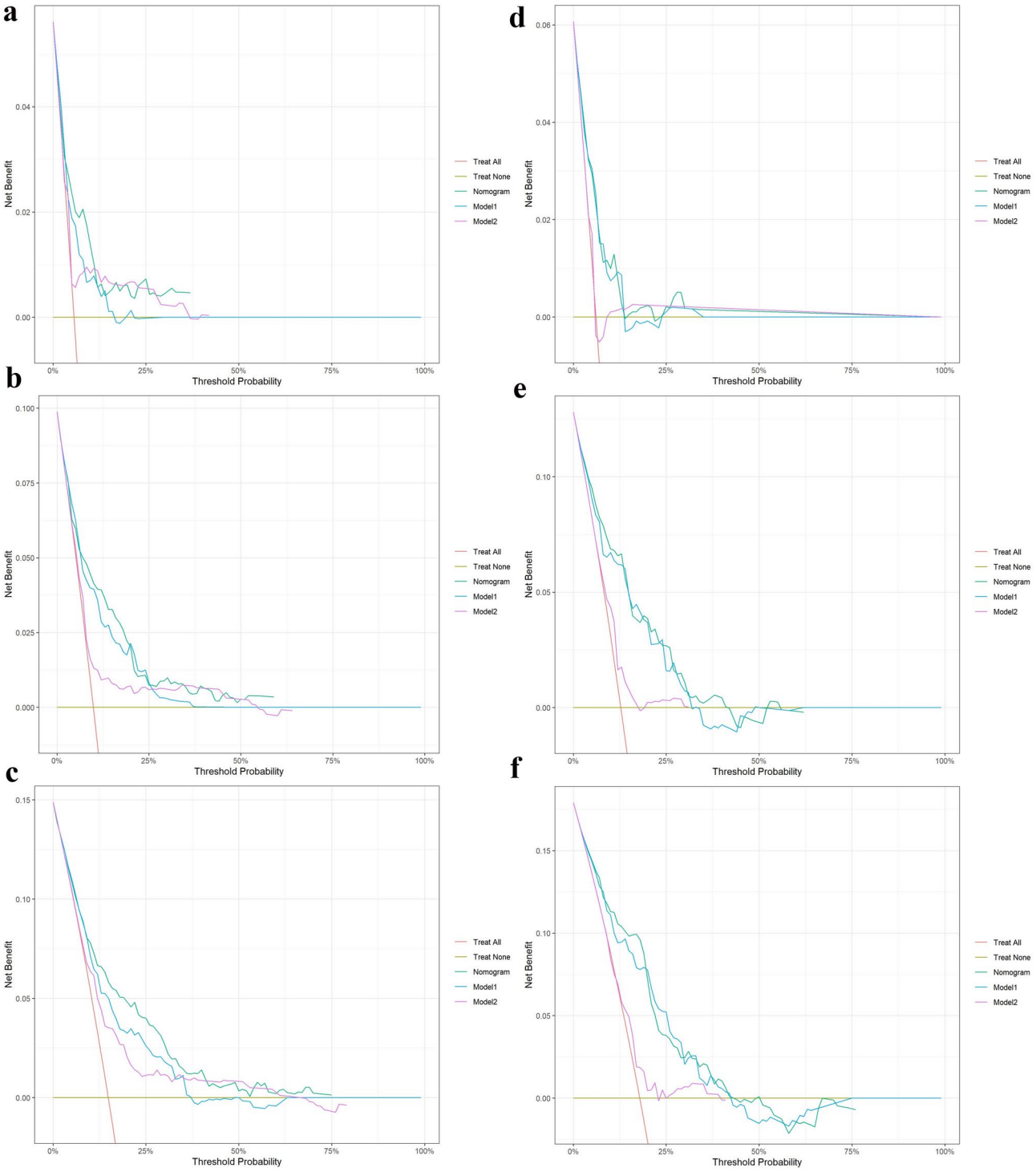

Fig. 4. DCA curves for comparisons of the nomogram, model1 and model2 in the training (**a**), validation (**b**) datasets. Shown are the threshold probability (TP) on the x-axis and the net benefit on the y-axis. Assuming the TP is predetermined, the models which close to the top regions can obtain the maximum net benefit compared with treating all or treating none or other interventions. The nomogram consists of age, passive smoking history, ALP, GGT, LDH, and potassium. The model 1 consists of age and passive smoking history, while the model 2 includes ALP, GGT, LDH, and potassium.

is likely due to differences in the population represented by the sample. More than 60% of participants in our research were PRISm which is significant from the previous study population. Circulating liver function markers, including GGT and ALP, have been proven to be associated with COPD exacerbations and mortality by several prospective studies and animal models[31]. Additionally, previous research[32] suggested that higher levels

	Training cohort (N = 730)				Validation cohort (N = 313)			
	N (%)	Death (%) £	HR (95%CI)	*P*	N (%)	Death (%)	HR (95%CI)	*P*
Risk group								
Low-risk (< 121.45)	597(81.78)	54(9.05)	Ref	Ref	258(82.43)	39(15.12)	Ref	Ref
High-risk (≥ 121.45)	133(18.22)	62(46.62)	6.20(4.30,8.95)	< 0.001	55(17.57)	18(32.73)	2.25(1.29,3.94)	0.004
Risk score performance AUC (95% CI)	0.81(0.76,0.85)				0.76(0.70,0.82)			
C-index(95%CI)	0.771(0.727,0.816)				0.737(0.681,0.793)			
Difference in mortality**ξ**	37.57%				17.61%			

Table 3. Mortality by risk group in participants of the nomogram score in the training and validation cohorts. **£**: Mortality of each risk group. **ξ**: Difference in mortality between high- and low-risk groups (%), calculated by [mortality in high-risk group]—[mortality in low-risk group].

of GGT and ALP were nonlinearly associated with elevated COPD risk. Similarly, we also observed significant positive associations of GGT and ALP with the risk of COPD mortality. The links between liver function markers and COPD can be explained by the "lung-liver axis"[33]. Serum GGT is recognized as a sensitive biomarker of oxidative stress which is the key pathogenesis of COPD[34]. ALP levels are an important indicator for skeletal muscle dysfunction which is common complication of COPD patients[35]. This study not only confirmed the clinical relevance of LDH, ALP, and GGT with COPD mortality but also provided novel evidence regarding the roles of these three serum biomarkers for early prediction of mortality in PRISm and mild to moderate COPD participants. However, this study only included the baselines blood biomarkers test results in the analysis, therefore, further study should utilize repeated measurements of these biomarkers to improve reliabilities of the relationship. Previous studies suggested that age, sex and BMI are potential confounding factors for LDH, ALP and GGT. Therefore, this study also well considered these potential confounders when establishing and validating the nomogram model.

Previous studies have reported that most patients with COPD die from non-respiratory disorders, and nearly one-fifth of patients with COPD deaths are attributed to cardiovascular disease (CVD)[36]. Similarly, in this study, the proportion of deaths attributed to heart diseases (I00-I09, I11, I13, I20-I51) was approximately 20% (36 out of 173). Potassium plays a crucial role in maintaining the normal functions of muscles and the heart, as well as preserving a regular heart electrical rhythm. Population cohort studies have reported that serum potassium levels are associated with mortality in chronic heart failure patients[37] and CVD. Therefore, potassium may reflect, and even predict, long-term mortality in COPD patients by monitoring cardiovascular disease risk. Our findings suggest that potassium levels > 4 mmol/L are significantly associated with an elevated risk of mortality in PRISm and early-stage COPD. This could provide a potential intervention strategy to reduce death risks in COPD by maintaining serum potassium homeostasis.

Interestingly, this study found that compared to participants with Preserved Ratio Impaired Spirometry (PRISm), COPD patients at GOLD stage 1 had a reduced risk of death, whereas those at GOLD stage 2 showed an increased risk of death, though the associations did not reach statistical significance after adjusting for confounders. Consequently, we combined PRISm and mild-to-moderate COPD (GOLD 1 and 2) to predict long-term mortality risk in this study. Notably, smoking history was associated with a reduced risk of death in this cohort. While smoking is widely recognized as a major risk factor for developing COPD in the general population (with smokers or passive smokers facing higher COPD incidence than non-smokers), its role in mortality among individuals with PRISm or early-stage COPD remains unclear. This study revealed that smoking history was linked to a lower mortality risk in early-stage COPD patients. As a result, non-smoking COPD patients may require closer monitoring when assessing long-term mortality risk, as their outcomes could differ unexpectedly from those with a smoking history.

It has been reported that the prevalence of COPD and PRISm ranged from 13.1–14.3% and 9.6–10.2% in the United States from NHANES 2007–2012, and the calculated overall all-cause mortality was 8.15% with a median follow-up of 9.4 years[38]. We hypothesize a country with a population of 100 million, where the number of pre-clinical and early COPD patients approximately 23 million. A benchmark, based on a previous study, suggests that pre-clinical and early COPD patients should receive clinical interventions if the predicted mortality exceeds 10%. In this case, our nomogram can yield net benefit $(NB)_{5y} = 0.015$, $NB_{7.5y} = 0.04$ and $NB_{10y} = 0.07$, respectively. This means that, compared to a strategy of treating none, our nomogram can identify an additional 0.3, 0.9, and 1.6 million people at higher risk of death at 5 years, 7.5 years, and 10 years in advance among PRISm and early COPD patients[39,40]. Furthermore, the results also indicate that the nomogram is superior to models using single serum biomarkers (LDH, ALP, GGT, and potassium) and clinical models (age and passive smoking) when the risk thresholds range from 10 to 12%. Therefore, our nomogram can obtain a maximum benefit at a low cost when predicting long-term mortality risks of pre-clinical and early COPD as a large-scale population screening tool.

This study has several limitations. First, all participants were from the US population, and races were limited to 'Mexican American', 'Non-Hispanic White', and 'Non-Hispanic Black'; therefore, our nomogram should be verified with other populations, such as those in Asia. Second, we only included NHANES data from 2007 to 2012 in the final analysis due to the availability of spirometry-Pre and Post-Bronchodilator test results. This may introduce potential time trend bias in the study samples. However, this study primarily focuses on predicting long-term mortality risks for PRISm and mild to moderate COPD, so the median follow-up time of nearly

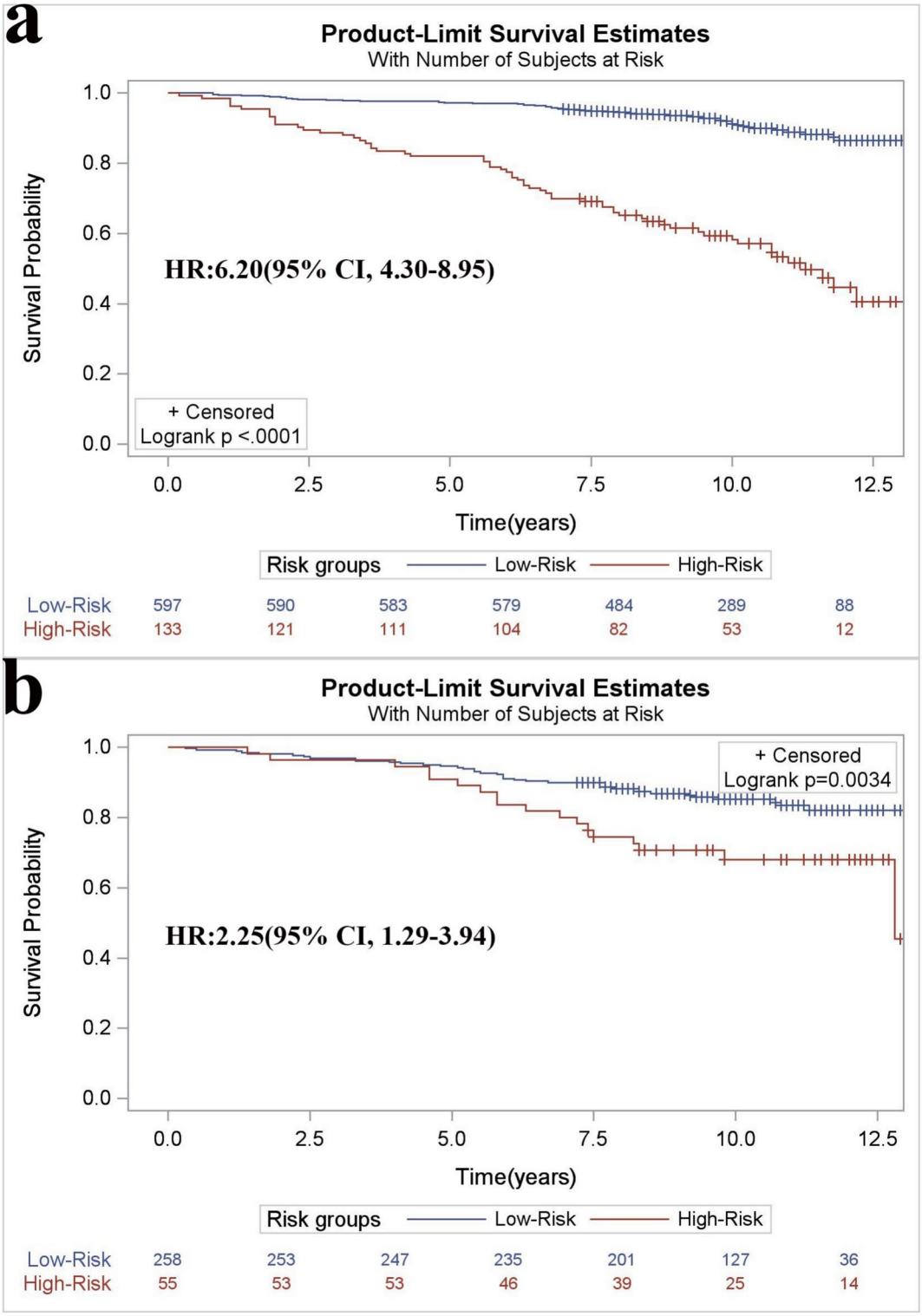

Fig. 5. Survival curves stratified by the nomogram scoring system in the training (**a**) and validation (**b**) groups. The high-risk group represents participants with nomogram score was large than 121.45, whereas the low-risk group indicates those who had nomogram score was less than 121.45.

10 years is beneficial for observing the occurrence of the outcome. Third, since GOLD standards recommend the use of post-bronchodilator data to define and classify COPD patients, this led to a relatively small sample size compared to previous studies[38]. Nevertheless, the sample size estimate analysis confirmed that this study's sample has reliable statistical power.

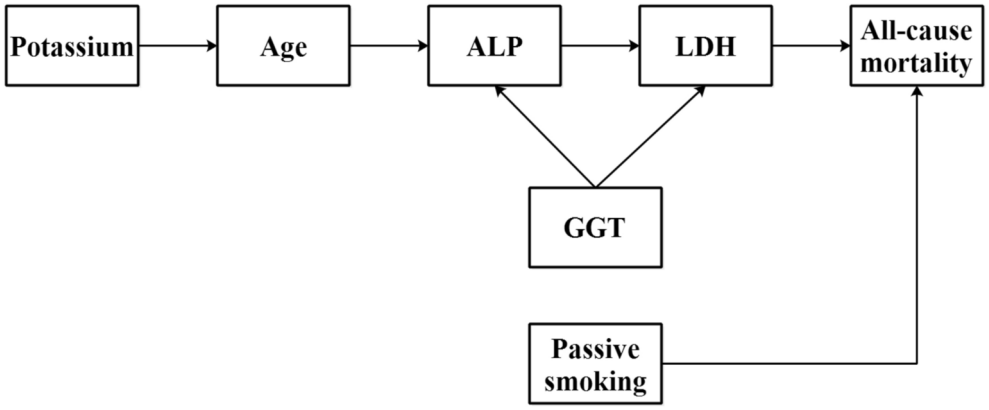

Fig. 6. Directed acyclic graph between six risk factors and all-cause mortality.

In conclusion, we developed and validated an easy-to-use and overall accurate nomogram for predicting long-term mortality risks in PRISm and mild to moderate COPD. This study provided new evidence about early identification high risk population among pre-clinical and early COPD patients.

Data availability

Data used in the preparation of this article were obtained from the US National Health and Nutrition Examination Surveys (NHANES, https://www.cdc.gov/nchs/nhanes/index.htm).

Received: 7 January 2025; Accepted: 13 March 2025
Published online: 08 May 2025

References

1. Diseases, G. B. D. & Injuries, C. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**(10258), 1204–1222 (2020).
2. Han, M. K. et al. From 0 GOLD to Pre-COPD. *Am. J. Respir. Crit. Care Med.* https://doi.org/10.1164/rccm.202008-3328PP (2020).
3. Wan, E. S. et al. Association between preserved ratio impaired spirometry and clinical outcomes in US adults. *Jama J. Am. Med. Assoc.* **326**(22), 2287–2298 (2021).
4. Bellou, V., Belbasis, L., Konstantinidis, A. K., Tzoulaki, I. & Evangelou, E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *Bmj Brit. Med. J.* https://doi.org/10.1136/bmj.l5358 (2019).
5. Efthimiou, O. et al. Developing clinical prediction models: a step-by-step guide. *Bmj Brit. Med. J.* https://doi.org/10.1136/bmj-2023-078276 (2024).
6. Hou, X. et al. Development and validation of a prognostic nomogram for HIV/AIDS patients who underwent antiretroviral therapy: Data from a China population-based cohort. *EBioMedicine* **48**, 414–424 (2019).
7. Cai, X. T. et al. Derivation and validation of a prediction model for predicting the 5-year incidence of type 2 diabetes in non-obese adults: A population-based cohort study. *Diabetes Metab. Syndr. Obes.* **14**, 2087–2101 (2021).
8. Cai, X. et al. A prediction model based on noninvasive indicators to predict the 8-year incidence of type 2 diabetes in patients with nonalcoholic fatty liver disease: A population-based retrospective cohort study. *Biomed. Res. Int.* **2021**, 5527460 (2021).
9. Cai, X. et al. Establishment and validation of a nomogram that predicts the risk of type 2 diabetes in obese patients with non-alcoholic fatty liver disease: a longitudinal observational study. *Am. J. Transl. Res.* **14**(7), 4505–4514 (2022).
10. Van Den Berg, H. A. Occam's razor: from Ockham's via moderna to modern data science. *Sci. Prog.* **101**(3), 261–272 (2018).
11. Celli, B. R. Predictors of mortality in COPD. *Respir. Med.* **104**(6), 773–779 (2010).
12. van Dijk, W. D. et al. Multidimensional prognostic indices for use in COPD patient care. A systematic review. *Resp. Res.* https://doi.org/10.1186/1465-9921-12-151 (2011).
13. Mendy, A., Forno, E., Niyonsenga, T. & Gasana, J. Blood biomarkers as predictors of long-term mortality in COPD. *Clin. Respir. J.* **12**(5), 1891–1899 (2018).
14. Fermont, J. M. et al. Biomarkers and clinical outcomes in COPD: a systematic review and meta-analysis. *Thorax* **74**(5), 439–446 (2019).
15. Shaw, J. G. et al. Biomarkers of progression of chronic obstructive pulmonary disease (COPD). *J. Thorac. Dis.* **6**(11), 1532–1547 (2014).
16. Hyattsvile M. *Office of Analysis and Epidemiology. Public-Use Linked Mortality National Center for Health Statistics*; https://www.cdc.gov/nchs/data-linkage/mortality-public.htm. (2015).
17. Bowerman, C. et al. A race-neutral approach to the interpretation of lung function measurements. *Am. J. Resp. Crit. Care Med.* **207**(6), 768–774 (2023).
18. (GOLD) GIfCOLD. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: 2024 report. (2024).
19. Wan, E. S. et al. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Resp. Res.* https://doi.org/10.1186/s12931-014-0089-y (2014).
20. Zheng, Y. Y. et al. Associations of dietary inflammation index and composite dietary antioxidant index with preserved ratio impaired spirometry in US adults and the mediating roles of triglyceride-glucose index: NHANES 2007–2012. *Redox Biol.* https://doi.org/10.1016/j.redox.2024.103334 (2024).
21. Riley, R. D. et al. Calculating the sample size required for developing a clinical prediction model. *Bmj Brit. Med. J.* https://doi.org/10.1136/bmj.m441 (2020).
22. Guo, C. N. et al. Machine-learning-based plasma metabolomic profiles for predicting long-term complications of cirrhosis. *Hepatology* https://doi.org/10.1097/HEP.0000000000000879 (2024).

23. Greiner, M., Pfeiffer, D. & Smith, R. D. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prevent. Vet. Med.* **45**(1–2), 23–41 (2000).
24. Horita, Nobuyuki et al. Minimal clinically important difference (MCID) of effect sizes other than mean difference. *J. Clin. Question* **1**(3), 116–127 (2024).
25. Dankers, Marloes et al. Healthcare professionals' preferred efficacy endpoints and minimal clinically important differences in the assessment of new medicines for chronic obstructive pulmonary disease. *Front. Pharmacol.* **10**, 1519 (2020).
26. Bloom, C. I., Ricciardi, F., Smeeth, L., Stone, P. & Quint, J. K. Predicting COPD 1-year mortality using prognostic predictors routinely measured in primary care. *BMC Med.* **17**(1), 73 (2019).
27. Lovelace, T. C. et al. Development and validation of a mortality risk prediction model for chronic obstructive pulmonary disease: a cross-sectional study using probabilistic graphical modelling. *EClinicalMedicine* **75**, 102786 (2024).
28. Li, X. et al. Lactate metabolism in human health and disease. *Sig. Transd. Target. Ther.* **7**, 305 (2022).
29. Hu, S. et al. Serum lactate dehydrogenase is associated with impaired lung function: NHANES 2011–2012. *PloS ONE* **18**(2), e0281203 (2023).
30. Huang, L. H. et al. U-shaped relationship between serum lactate dehydrogenase with all-cause mortality in patients with chronic obstructive pulmonary disease. *Int. J. Chronic Obstr.* **18**, 305–316 (2023).
31. Young, R. P., Hopkins, R. J. & Marsland, B. The gut-liver-lung axis modulation of the innate immune response and its possible role in chronic obstructive pulmonary disease. *Am. J. Resp. Cell Mol.* **54**(2), 161–169 (2016).
32. Du, W. C. et al. Circulating liver function markers and the risk of COPD in the UK Biobank. *Front. Endocrinol.* https://doi.org/10.3389/fendo.2023.1121900 (2023).
33. Hilliard, K. L. et al. The lung-liver axis: A requirement for maximal innate immunity and hepatoprotection during pneumonia. *Am. J. Resp. Cell Mol.* **53**(3), 378–390 (2015).
34. Domej, W., Oettl, K. & Renner, W. Oxidative stress and free radicals in COPD implications and relevance for treatment. *Int. J. Chronic Obstr.* **9**, 1207–1224 (2014).
35. Sciacqua, A. et al. Alkaline phosphatase affects renal function in never-treated hypertensive patients: effect modification by age. *Sci. Rep. UK* https://doi.org/10.1038/s41598-020-66911-z (2020).
36. Sin, D. D. et al. Mortality in COPD: role of comorbidities. *Eur. Respir. J.* **28**, 1245–1257 (2006).
37. Aldahl, M. et al. Associations of serum potassium levels with mortality in chronic heart failure patients. *Eur. Heart J.* **38**(38), 2890–2896 (2017).
38. Cadham, C. J. et al. The prevalence and mortality risks of PRISm and COPD in the United States from NHANES 2007–2012. *Resp. Res.* https://doi.org/10.1186/s12931-024-02841-y (2024).
39. Calster, B. V. et al. Reporting and interpreting decision curve analysis: A guide for investigators. *Eur. Urol.* **74**(6), 796–804 (2018).
40. Hou, X. & Wong, G. Nomogram for early prediction of Parkinson's disease based on microRNA profiles and clinical variables. *J. Parkinsons Dis.* **13**(4), 473–484 (2023).

Acknowledgements

We would like to acknowledge the US National Health and Nutrition Examination Surveys (NHANES) as the source of all data in the study.

Author contributions

Xiangqing Hou and Pixin Ran designed the study, accessed the raw data, completed the data management and data cleaning as well as interpretation. Xiangqing Hou performed the statistical analysis. Xiangqing Hou and Pixin Ran drafted and edited the manuscript. All authors contributed to the critical revision of the manuscript and approved the final version. Xiangqing Hou and Pixin Ran had full access to the raw data in this study and had final responsibility for the decision to submit it for publication.

Funding

This work was supported by the Foundation of Guangzhou National Laboratory (SRPG22-016 and SRPG22-018) and the Young Scientists Program of Guangzhou Laboratory, Grant No (QNPG24-15), as well as the Non-communicable Chronic Diseases-National Science and Technology Major Project (2023ZD0516600), and National Natural Science Foundation of China (82270043).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-94399-y.

Correspondence and requests for materials should be addressed to P.R.