Proceedings of Anticancer Research, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/PAR
ISSN Online: 2208-3553
ISSN Print: 2208-3545

Association between Glutathione S-transferase Gene Polymorphisms and Lung Squamous Cell Carcinoma in Patients with Chronic Obstructive Pulmonary Disease: A Case-control Study

Mengdan Zhao, Jiqing Hao*

Department of Oncology, The First Affiliated Hospital of Anhui Medical University, Hefei 230032, Anhui, China

**Author to whom correspondence should be addressed.*

Abstract: This case-control study evaluated the frequency of Glutathione S-transferase Mu 1 (GSTM1) deletion and Glutathione S-transferase Alpha 1 (GSTA1) mutation in chronic obstructive pulmonary disease (COPD) patients, whether they had concomitant lung squamous cell carcinoma (LSCC) or not, to assess their connection with cancer susceptibility. By means of multivariate logistic regression analysis, the GSTM1 null genotype serves as a significant standalone risk factor for LSCC, in addition to variables like age, smoking history, emphysema, body mass index (BMI), albumin level, and neutrophil-to-lymphocyte ratio (NLR). A predictive model incorporating these factors demonstrated superior discriminative ability compared to the established COPD Lung Cancer Screening Score (COPD-LUCSS).

Keywords: Case-control study; Chronic obstructive pulmonary disease; Gene polymorphism; Glutathione S-transferase; Lung squamous cell carcinoma

Online publication: December 8, 2025

1. Introduction

Lung cancer (LC) persists as a global health priority due to persistently elevated incidence and mortality rates. Despite some improvements in the diagnostic and treatment methods, its five-year survival rate is not high, and it is about 22% [1]. LC originates from either epithelial cells of the bronchi or alveoli, and according to the World Health Organization (WHO), it is classified into two distinct subtypes: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC is divided into several different subtypes, with the lung adenocarcinoma and the lung squamous cell carcinoma (LSCC) being the most prevalent [2]. Chronic obstructive pulmonary disease (COPD) additionally represents a significant predictor for LC, particularly squamous cell carcinoma (SCC) [3–6].

COPD is a highly heterogeneous chronic pulmonary disorder marked by persistent airflow limitation,

caused by chronic airways and/or alveolar abnormalities. This key pathophysiological feature leads to ongoing respiratory manifestations like progressive dyspnea, cough, and sputum production [7]. With the acceleration of global population aging, worsening environmental pollution, and the widespread prevalence of risk factors such as smoking, the incidence and mortality of COPD have risen significantly [8], making it a critical public health issue worldwide. Studies have reported that the annual incidence of lung cancer among COPD patients ranges from 0.8% to 1.7% [9,10], underscoring the urgency and importance of early and accurate identification of high-risk individuals susceptible to LC progression within this population.

Elucidating the underlying mechanisms driving the progression from COPD to LC may provide a theoretical basis for early intervention. This process involves multiple complex biological mechanisms, including genetic susceptibility, epigenetic alterations, oxidative stress, epithelial-mesenchymal transition, and chronic inflammation [11]. Among these biological elements, glutathione S-transferases (GSTs) exert a vital function. As a family of multifunctional enzymes extensively expressed in the human body, GSTs participate in cellular detoxification processes, regulation of oxidative stress responses, and modulation of inflammatory reactions [12]. Among the key roles, GSTs promote conjugation of the reduced glutathione (GSH) with electrophilic xenobiotics and convert them to water-soluble metabolites, which may be excreted via urine or bile to avoid oxidative and poisonous damage in the cells.

COPD is linked to persistent bronchial inflammation and oxidative damage. Prolonged inflammatory responses lead to genotoxic stress and oxidative damage, which play key roles in LC pathogenesis. By clearing reactive oxygen species (ROS) and metabolizing carcinogens, GSTs can mitigate such damage. However, impaired GST function may not only accelerate the progression of COPD but also lay a pathological foundation for LC development [13].

The GST superfamily includes multiple isoenzymes. Among them, Glutathione S-transferase Mu 1 (*GSTM1*) is key for metabolizing and detoxifying hydrophobic carcinogens, such as epoxides from polycyclic aromatic hydrocarbons [14]. The *GSTM1* gene exhibits a heritable homozygous deletion polymorphism. This null genotype results in a complete loss of enzymatic activity, thereby significantly increasing an individual's susceptibility to malignant tumors [15]. In lung tissue, *GSTM1* null genotype individuals have increased levels of polycyclic aromatic hydrocarbon-deoxyguanosine monophosphate (dGMP) adducts, which induce gene mutations [16] and increase the risk of LC. Conversely, Glutathione S-transferase Alpha 1 (*GSTA1*) assembles into complexes with c-Jun N-terminal kinase (JNK) to affect apoptosis [17]. *GSTA1* expression is strictly modulated by single-nucleotide polymorphisms (SNPs) within its promoter region, where rs3957356 and rs3957357 serve as key variants. These variants alter *GSTA1* transcriptional activity (via interactions with transcription factors and synergy with other SNPs) and are linked to a higher risk of smoking-related LC [18].

This study analyzes the distribution of these gene variants in COPD patients with and without concurrent LSCC. The study wants to know whether these genetic changes will affect the development of the disease and whether they can be used as a marker to predict the risk of LSCC. This work hopes to provide new ideas and reliable scientific foundation for identifying and preventing LSCC in COPD patients.

2. Methods

2.1. Study design and patients

The study conducted a case-control study that enrolled 73 patients with COPD only, all admitted to the Department of Respiratory Medicine at the First Affiliated Hospital of Anhui Medical University between December 2023 and June 2025. COPD is determined by a comprehensive evaluation of potential risk factors (including smoking),

characteristic clinical manifestations, and observable physical examination results. To make a definite diagnosis, there must be irreversible airflow obstruction, which is confirmed by the fact that the ratio of forced expiratory volume in 1 second (FEV1)/forced vital capacity (FVC) ratio is < 70% after using bronchodilators. During the same period, the study recruited 72 people diagnosed with LSCC from the respiratory medicine and oncology department of the same hospital. These patients not only met the aforementioned diagnostic criteria for COPD but also received pathological confirmation of LSCC via a transthoracic lung biopsy or surgical resection. This research's protocol was endorsed by the Ethics Committee of Anhui Medical University's First Affiliated Hospital (Approval Number: 2022457).

The study gathered data on the case and control groups from patient medical records, including age, gender, smoking history, alcohol consumption history, residential area, presence of hemoptysis, presence of chest pain, presence of emphysema, body mass index(BMI), serum albumin level, white blood cell (WBC) count, Neutrophil count, Lymphocyte count, Monocyte count, Eosinophil count, Basophil count, Platelet count, neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), FEV1/FVC ratio, Global Initiative for Chronic Obstructive Lung Disease (GOLD) Grade and tumor stage distribution (restricted to the COPD with LSCC group). According to the COPD Lung Cancer Screening Score (COPD-LUCSS) criteria, the variables age, smoking status, and BMI were dichotomized as follows: age was categorized using a cutoff of 60 years into > 60 and ≤ 60; smoking history was classified based on pack-years with a threshold of 60, where > 60 pack-years was defined as smoking history (1 pack-year = smoking one pack per day for one year), and ≤ 60 pack-years as no smoking history; BMI was divided into two groups with BMI of 25 or above and below. BMI was calculated by dividing weight (kg) by the square (kg/m) of height (m). In addition, NLR and PLR were calculated by these formulas: NLR was equal to the ratio of neutrophil count to lymphocyte count, and PLR was the ratio of platelet count to lymphocyte count.

2.2. Molecular analysis

After an overnight fast, 2 mL of venous blood was obtained from all subjects. Following anticoagulation treatment, 500 μL of whole blood was transferred to a 1.5 mL small centrifuge tube. Samples were stored at -80 ℃ for subsequent genomic analysis. Peripheral blood underwent genomic DNA extraction via a magnetic bead-based method. The complete deletion of the *GSTM1* gene was ascertained via polymerase chain reaction (PCR). In order to verify the success of the experiment, the study amplified a short segment of the *GSTA1* gene by PCR. Then, the study did bidirectional Sanger sequencing on these PCR products to verify the existence of homozygous mutations in the *GSTA1* gene.

When the study did the PCR experiment, the study used 25.0 μL of mixed liquid. This includes: 12.5 μL of 2× Taq Plus Master Mix II (which contains Taq DNA polymerase, dNTPs, and reaction buffer), 0.5 μL of forward primer, 0.5 μL of reverse primer, and 2.0 μL of genomic DNA template. The remaining volume was adjusted to 25.0 μL with nuclease-free water (ddH_2O).The amplification protocol comprised an initial denaturation at 95 ℃ for 5 minutes, followed by 30 cycles of denaturation at 95 ℃ for 30 seconds, annealing at 60 ℃ for 30 seconds, and extension at 72°C for 1 minute, with a final extension at 72 ℃ for 10 minutes. The study did Sanger sequencing of the *GSTA1* gene. The mixed solution for sequencing is 5.0 μL, including 2.0 μL of purified *GSTA1* PCR product, 2.0 μL of sequencing primer (forward or reverse, depending on which direction you want to sequence) and 1.0 μL of sequencing enzyme reagent. The first step of the PCR experiment is to denature the DNA at 96 ℃, which lasts for 2 minutes. Followed by 32 cycles, each cycle includes denaturation at 96 ℃ for 10 seconds, primer binding at 50 ℃ for 5 seconds, and extension at 60 ℃, which takes 2 minutes.

The appearance of the *GSTM1* gene was confirmed by detecting a 215 bp amplification product. An internal control targeting the *GSTA1* gene, which produces a 528 bp fragment, was included in all samples. The following primers were used: for *GSTM1*, forward 5′-GAA CTC CCT GAA AAG CTA AAG C-3′ and reverse 5′-GTT GGG CTC AAA TAT ACG GTG G-3′; for *GSTA1*, forward 5′-GCC CTT CAC AGA CAT CCT CTC-3′ and reverse 5′-CAC CAA GAC GGC ACA ATA TGA G-3′. After adding 5 μL of ethidium bromide to 2% agarose gel for electrophoresis, the study saw DNA fragments under an ultraviolet lamp and compared them with a DNA ladder reference (**Figure 1**). Genotyping was performed using SnapGene software. Polymorphism analysis was independently conducted by two individuals in a blinded manner. In order to check whether the genotyping process is accurate, the study randomly selected about 10% samples for retesting. The analysis results are completely consistent, and the results of all repeated tests are correct (**Figure 2**).

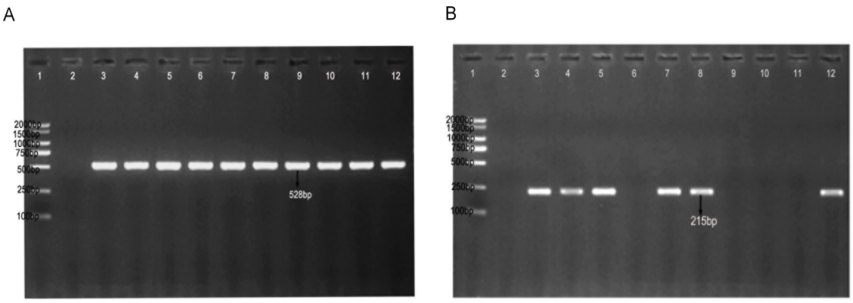

Figure 1. Agarose gel electrophoresis analysis of PCR amplification products for *GSTA1* (528 bp) and *GSTM1* (215 bp) genes. Amplified products of blood group *GSTA1* and *GSTM1*: (A) The *GSTA1* amplification product is 528 bp in length. (B) The *GSTA1* amplification product is 215 bp in length. Lane 1 is the DNA Marker, Lane 2 is the blank control, Lanes 3–7 are the control group, and Lanes 8–12 are the case group.

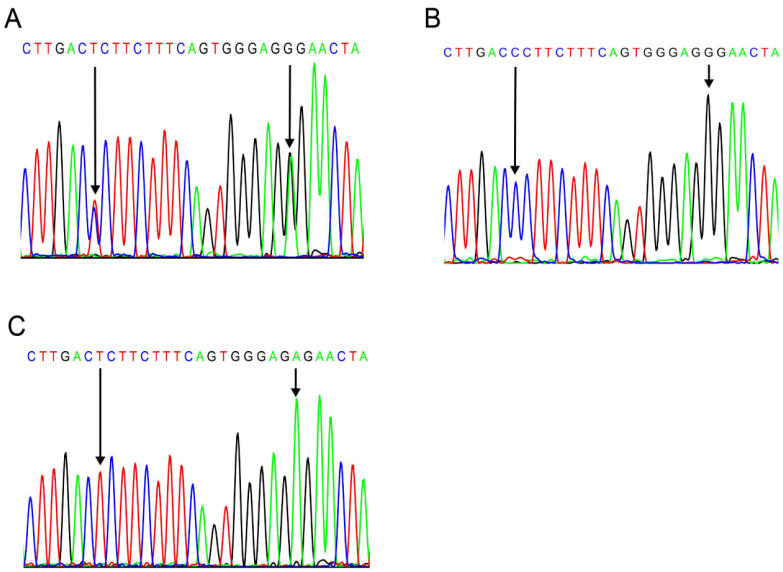

Figure 2. Sanger sequencing chromatograms showing the genotype variations at rs3957356 and rs3957357 loci in the promoter region of the *GSTA1* gene.Sanger sequencing chromatograms of rs3957356 and rs3957357 loci in the *GSTA1* gene. (A) The double peak marked by the arrow represents the heterozygous genotype AB. (B) The single peak marked by the arrow represents the homozygous mutant genotype BB. (C) The single peak marked by the arrow represents the wild-type homozygous genotype AA.

2.3. Data analysis

The study uses Social Sciences (SPSS) software (version 27.0) and R (version 4.4.2) for statistical calculation, and also uses the toolkits of stats, epitools, ggplot2, MASS, and genetics. For continuous data, the study first checks whether they conform to a normal distribution. The study use "mean ± standard deviation" to represent the data that conforms to normal distribution, and use an independent samples t-test to compare the differences between different groups. Non-normally distributed data are presented as medians with interquartile ranges (IQR) and evaluated using the Mann-Whitney U test. The study uses quantity (percentage) to express the classified data, and the study uses the chi-square test to test it. In order to see if the genes of the control group are representative, the study used the genetics package of R software to check the frequencies of *GSTA1* genotypes to see if they conform to the Hardy-Weinberg equilibrium (HWE).

To discern the discrepancy in the prevalence of genetic anomalies and the cumulative impact they have, the study first used a method called univariate logistic regression to compare the COPD with LSCC group and the COPD-only group. Crucial factors identified via univariate analyses were incorporated into a logistic regression model via backward elimination for variable selection. The variance inflation factor (VIF) was calculated to assess and mitigate potential multicollinearity issues that could adversely affect model stability [19].

A nomogram was developed to visualize the contribution of each predictor in the final model. The predictive ability of this model has been checked by us in many ways. The study looked at its receiver operating characteristic (ROC) curve to judge whether it was accurate or not, used calibration plots to see whether it was stable or not, and analyzed whether it was useful when seeing a doctor by decision curve analysis (DCA). Model stability and reliability were confirmed via 500 bootstrap resamples. Comparative analysis against the existing COPD-LUCSS model examined discriminative capacity (ROC curve areas) and clinical utility (DCA curves).

3. Results

3.1. Characteristics of the study population

Table 1 presents univariate analysis outcomes for the case and control groups. Group comparisons showed significant differences ($P < 0.05$) across twelve variables. These characteristics are: age ($P = 0.021$), gender ($P = 0.022$), smoking ($P = 0.004$), hemoptysis ($P = 0.014$), emphysema ($P = 0.034$), BMI ($P = 0.007$), albumin level ($P < 0.001$), white blood cell count ($P = 0.018$), neutrophil count ($P = 0.003$), monocyte count ($P = 0.037$), NLR ($P < 0.001$), and PLR ($P = 0.032$). Within the COPD with LSCC group, lung cancer staging was predominantly stage IV (32 cases, 44.4%) and stage III (29 cases, 40.3%), followed by stage II (7 cases, 9.7%) and stage I (4 cases, 5.6%). Regarding GOLD grades, 50% of the cases in the COPD with LSCC group were classified as GOLD 2, whereas 49.3% of the COPD-only group were GOLD 3, suggesting an overall worse GOLD grade distribution in the latter group. This inconsistency may be because people with COPD usually only consult doctors when their illness is serious or when their illness has developed to a more serious GOLD stage.

Table 1. Univariate analysis comparing characteristics of cases and controls

Variables	Cases (LSCC + COPD) ($n = 72$)	Controls (COPD) ($n = 73$)	*P*-value
Age (years old)			0.021*
≤ 60	8 (11.1)	19 (26.0)	
> 60	64 (88.9)	54 (74.0)	

Table 1 (Continued)

Variables	Cases (LSCC + COPD) (*n* = 72)	Controls (COPD) (*n* = 73)	*P*-value
Gender			0.022*
Male	68 (94.4)	60 (82.2)	
Female	4 (5.6)	13 (17.8)	
Smoking			0.004*
≤ 60	47 (65.3)	30 (41.1)	
> 60	25 (34.7)	43 (58.9)	
Drinking			0.340
Yes	25 (34.7)	20 (27.4)	
No	47 (65.3)	53 (72.6)	
Residence			0.895
Rural	52 (72.2)	52 (71.2)	
Urban	20 (27.8)	21 (28.8)	
Weight loss			0.059
Yes	15 (20.8)	7 (9.6)	
No	57 (79.2)	66 (90.4)	
Hemoptysis			0.014*
Yes	22 (30.6)	10 (13.7)	
No	50 (69.4)	63 (86.3)	
Chest pain			0.060
Yes	4 (5.6)	11 (15.1)	
No	68 (94.4)	62 (84.9)	
Emphysema			0.034*
Yes	48 (66.7)	32 (43.8)	
No	24 (33.3)	41 (56.2)	
BMI (kg/m^2)			0.007*
< 25	62 (86.1)	49 (67.1)	
≥ 25	10 (13.9)	24 (32.9)	
Albumin (g/L)	37.9 ± 4.3	34.6 ± 4.6	< 0.001*
White blood cell count (*10^9/L)	6.0 (4.5, 7.3)	6.6 (4.9, 8.3)	0.018*
Neutrophil count (*10^9/L)	4.0 (2.8, 4.8)	4.5 (3.4, 6.2)	0.003*
Lymphocyte count (*10^9/L)	1.3 (1.0, 1.7)	1.1 (0.9, 1.5)	0.093
Monocyte count (*10^9/L)	0.4 (0.3, 0.6)	0.5 (0.4, 0.6)	0.037*
Eosinophil count (*10^9/L)	0.1 (0.1, 0.2)	0.1 (0.1, 0.2)	0.102
Basophil count (*10^9/L)	0.03 (0.02, 0.04)	0.02 (0.02, 0.04)	0.281
Blood platelet count (*10^9/L)	196.0 (157.8, 252.3)	197.0 (163.0, 272.0)	0.614
NLR	2.8 (1.9, 3.6)	4.2 (2.7, 6.7)	< 0.001*
PLR	147.1 (109.4, 201.1)	183.0 (126.0, 254.7)	0.032*
FEV1/FVC	60.8 (53.7, 66.9)	58.3 (49.60,66.00)	0.087

Table 1 (Continued)

Variables	Cases (LSCC + COPD) (*n* = 72)	Controls (COPD) (*n* = 73)	*P*-value
GOLD grade			0.005
Grade 1	4 (5.5)	3 (4.1)	
Grade 2	36 (50.0)	21 (28.8)	
Grade 3	30 (41.7)	36 (49.3)	
Grade 4	2 (2.8)	13 (17.8)	
TNM Stage			
Stage I	4 (5.6)		
Stage II	7 (9.7)		
Stage III	29 (40.3)		
Stage IV	32 (44.4)		

The table above presents the clinical data and laboratory parameters for both groups. An asterisk (*) indicates statistical significance.

3.2. Analysis of genes GSTA1 and GSTM1

In control subjects, the study compared the genotype distribution of the *GSTA1* variant with the theoretical value according to the requirements of HWE. The result ($P > 0.05$) indicates that the control population is genetically representative and suitable for genetic association studies, as detailed in **Table 2**.

Table 2. Hardy-Weinberg equilibrium test

Genotype	Observed	Expected	*P*-value
AB	23	23.24	1.000
BB	47	46.88	
AA	3	2.89	

The study studied the relationship between *GSTM1* gene changes and LSCC risk, and found that *GSTM1* deletion type accounted for 62.5% (45/72) in patients with COPD and LSCC at the same time, and 43.8% (32/73) in patients with COPD only. This difference is statistically significant (OR = 2.135, 95% CI: 1.105–4.186, $P = 0.025$). See **Table 3** for details. These results indicate that the lack of the *GSTM1* gene will promote the development of LSCC in COPD patients. However, the *GSTA1* genotype frequencies showed no significant variation between the cohorts (OR = 1.257, 95% CI: 0.629–2.530, $P = 0.518$).

Table 3. Separate analysis of *GSTM1*、*GSTA1* between cases and controls

Genotype	LSCC+COPD, n(%)	COPD, n(%)	OR	95% CI	*P*-value
GSTM1			2.135	1.105-4.186	0.025*
Null	45(62.5)	32(43.8)			
Present	27(37.5)	41(56.2)			
GSTA1			1.257	0.629-2.530	0.518
BB	50(69.4)	47(64.4)			
AB/AA	22(30.6)	26(35.6)			

OR: Odds Ratio; CI: Confidence Interval. OR measures the strength of association between an exposure and an outcome. The 95% CI indicates the precision of the estimate. An asterisk (*) indicates statistical significance.

3.3. Risk factors and prediction model

Following univariate assessment and genetic studies, key variables were chosen for multivariate logistic regression modeling for identifying risk elements of the progression of COPD to LSCC. These factors include: age, gender, smoking history > 60 pack-years, hemoptysis, emphysema, BMI reaching or exceeding 25 kg/m², albumin level, white blood cell count, neutrophil count, monocyte count, NLR, PLR and *GSTM1* genotypes. The final model shows that there are seven independent factors that can predict whether COPD patients will get LSCC: age > 60 years old, smoking history > 60 pack-years, presence of emphysema, BMI ≥ 25 kg/m², albumin level, NLR, and *GSTM1* genotype. A predictive model was constructed based on these factors. **Table 4** exhibits the odds ratios (ORs) and their associated 95% confidence intervals (CIs) for the risk factors.

Table 4. Univariate and multivariate logistic analyses of COPD group and COPD complicated with LSCC group

Variables	Univariate Analysis, OR (95% CI)	*P*-value	Multivariate Analysis, OR (95% CI)	*P*-value
Age (years)	2.815 (1.177–7.296)	0.025*	4.028 (1.300–12.477)	0.016*
Gender	0.271 (0.073–0.814)	0.029*		
Smoking	2.695 (1.386–5.34)	0.004*	3.481 (1.432–8.463)	0.006*
Hemoptysis	2.772 (1.230–6.622)	0.017*		
Emphysema	2.562 (1.316–5.081)	0.006*	2.362 (1.019–5.475)	0.045*
BMI (kg/m^2)	0.329 (0.139–0.735)	0.009*	0.130 (0.042–0.409)	0.000*
Albumin (g/L)	1.186 (1.095–1.295)	0.000*	1.166 (1.050–1.295)	0.004*
White blood cell count (*10^9/L)	0.805 (0.685–0.933)	0.006*		
Neutrophil count (*10^9/L)	0.735 (0.603–0.874)	0.001*		
Monocyte count (*10^9/L)	0.173 (0.031–0.848)	0.036*		
NLR	0.798 (0.686–0.902)	0.001*	0.847 (0.731–0.980)	0.026*
PLR	0.997 (0.993–0.999)	0.027*		
GSTM1	0.468 (0.239–0.905)	0.025*	0.278 (0.114–0.679)	0.005*

An asterisk (*) indicates statistical significance.

3.4. Development of the predictive nomogram

Nomogram is a convenient chart, which can predict the possibility of a specific result by considering many different factors. As shown in **Table 4**, the study constructed a nomogram using the final variables identified from the multivariate logistic regression analysis, including: 'Age', 'Smoking', 'Emphysema', 'BMI', 'Albumin', 'NLR', and '*GSTM1*' genotype, to predict the probability of LSCC development in COPD patients. **Figure 3** displays the nomogram that was produced.

Based on the established nomogram model, the study developed an online interactive prediction tool for assessing LSCC risk in COPD patients using the Shiny framework. This tool, accessible to clinicians at https://lcnomo.shinyapps.io/dynnomapp/, allows real-time risk estimation by inputting patient-specific parameters including age, smoking history, emphysema status, BMI, albumin level, NLR, and *GSTM1* genotype. The tool provides immediate predictive outcomes to support clinical decision-making.

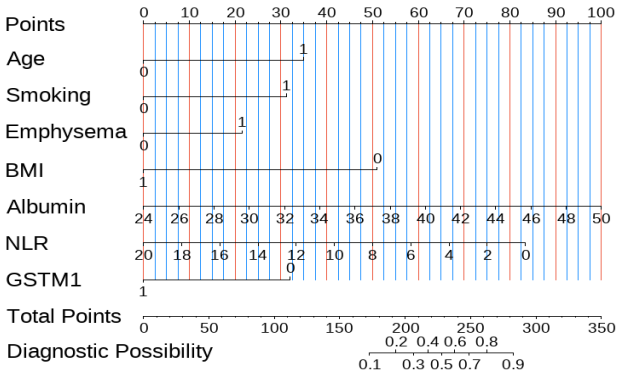

Figure 3. Nomogram based on the predictive model of backward stepwise multivariate logistic regression analysis. The predictive nomogram, built from variables identified through multivariate logistic regression, predicts LSCC risk in COPD patients. To use it, draw a vertical line from each predictor variable's axis to the top "points" axis to determine the score. Sum these scores and locate the total on the bottom "probability" axis to find the risk probability.

3.5. Predictive ability evaluation

Figure 4 illustrates that the model that is developed in this paper has a high level of performance: the model has an AUC of 0.854 (95% CI: 0.792–0.915), which indicates a great level of discrimination; the decision curve analysis (DCA) represented a high level of net benefits in a set of risk levels, and the calibration curve registered a high level of consistency between the anticipated and actual probabilities, which supports its high clinical utility. An AUC of 0.828 (95% CI: 0.763 –0.894) was obtained with internal validation using 500 bootstrap resamples.

The nomogram model produced a considerably higher AUC (0.854 vs. 0.744; Z = 3.124, $P = 0.001$) than the current COPD-LUCSS score. Furthermore, DCA indicated greater clinical net benefit for the nomogram across relevant threshold probabilities (**Figure 4**).

4. Discussi

This research examined LSCC vulnerability in COPD patients and developed a predictive tool for identifying at-risk individuals. Following an assessment of the associations between genetic susceptibility, clinical characteristics (such as demographic, living environment, and living habits), laboratory parameters, and the onset of LSCC in the COPD population, the study incorporated 13 clinical indicators, 10 laboratory measures, and 2 genes into a comprehensive analysis. Univariate logistic analysis identified 13 variables associated with the disease, while multivariate logistic regression further refined these into seven independent predictors, including age, smoking history, emphysema, BMI, albumin, NLR, and *GSTM1* genotype, consistent with the 10 EPV principle [20]. The model showed impressive discriminative power with an AUC of 0.854, maintained solid calibration, and demonstrated clear clinical value. To help clinicians put this into practice, the study created an easy-to-access online nomogram that enables early risk stratification for COPD patients.

Our findings indicate that *GSTM1* null genotype carriers demonstrate greater LSCC susceptibility relative to those with the wild-type variant(OR = 2.135, 95% CI: 1.105–4.186, $P = 0.025$). This suggests that *GSTM1* may be a valuable biomarker of the risk of LSCC among COPD patients. Individuals carrying the *GSTM1* null genotype may exhibit reduced detoxification capacity of GSTs, leading to higher concentrations of carcinogens in lung tissue and consequently an increased risk of developing LSCC. This finding is consistent with several

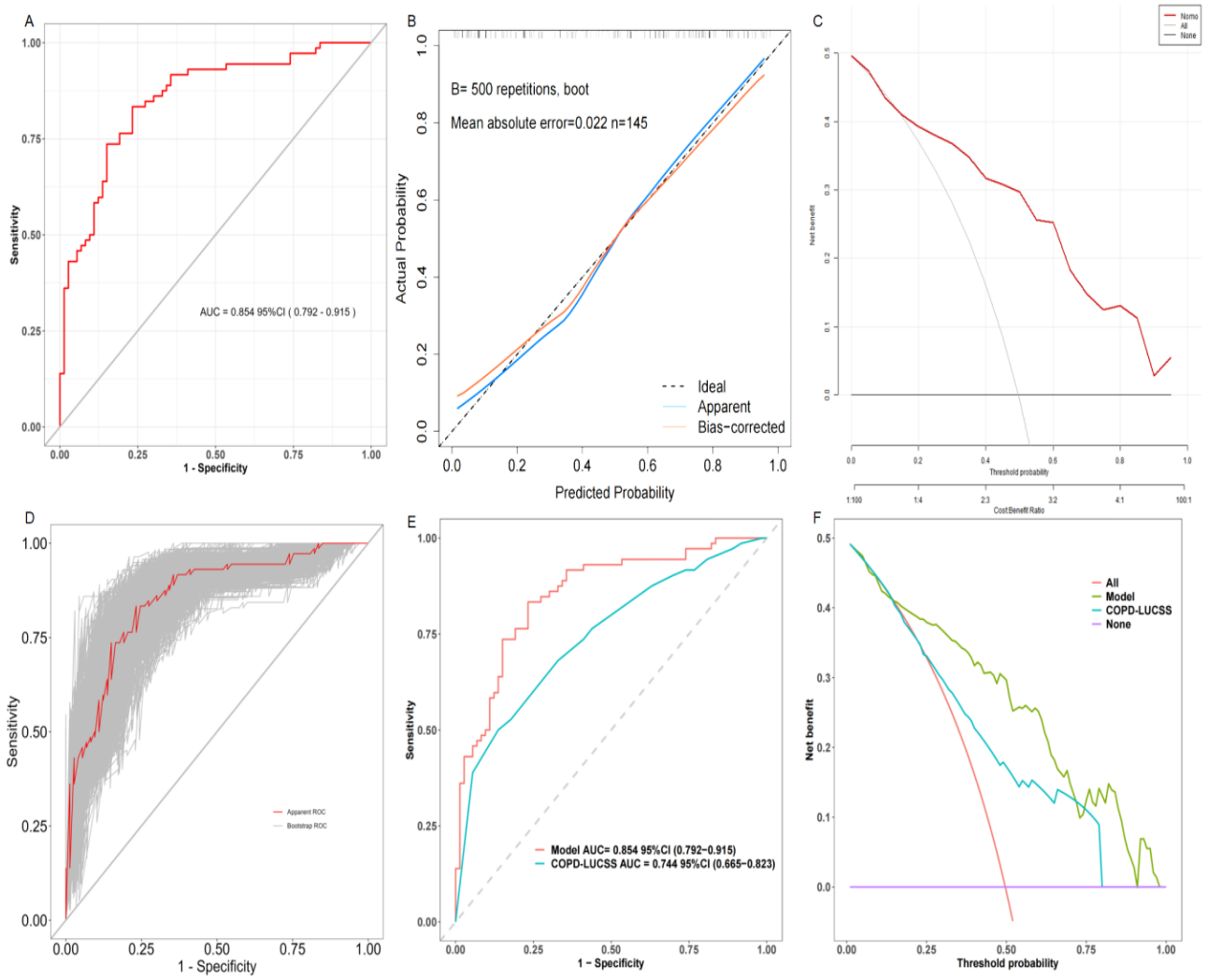

Figure 4. Evaluation of the Predictive Model's Performance and Comparison with the Established COPD-LUCSS Score. Comprehensive evaluation, validation, and comparative analysis of the predictive model's performance. (A) Receiver Operating Characteristic (ROC) curve. (B) Calibration plot comparing predicted and observed probabilities. (C) Decision Curve Analysis (DCA) assessing clinical utility. (D) Bootstrap-validated ROC curve (500 replications), with the shaded area representing the bootstrap AUC distribution and the solid line indicating apparent performance. (E) Comparison of ROC curves between the nomogram model and the COPD-LUCSS score. (F) Comparison of Clinical Decision Curve Analysis (DCA) between the nomogram model and the COPD-LUCSS score.

previous studies [13,21,22]. For example, Young et al. conducted a case-control analysis involving 669 COPD patients, 454 lung cancer patients, and 488 individuals with normal lung function. Their analysis showed a significantly increased frequency of the *GSTM1* null allele in smokers diagnosed exclusively with COPD (OR = 1.30, 95% CI 1.02–1.66, $P = 0.031$) as well as in those with lung cancer. Notably, the connection was even more pronounced in participants suffering from both COPD and lung cancer (OR = 1.50, 95% CI 1.06–2.12, $P = 0.018$) [23]. Moreover, a North Indian study with 540 participants, half being lung cancer patients and the other half controls, uncovered a link between the lack of the GSTM1 gene and an increased likelihood of developing lung cancer (OR = 1.65,

95% CI: 1.16–2.3, $P = 0.005$) [24]. Conversely, Lee et al. [25] and López-Cima et al. [26] reported no substantial link between *GSTM1* genetic variation and lung cancer susceptibility.

No notable link was detected between the *GSTA1* genetic variant and the likelihood of LSCC occurrence among COPD patients, in contrast to the *GSTM1* gene (OR = 1.257, 95% CI: 0.629–2.530, $P = 0.518$). Similarly, a case cohort study in Denmark detected no association between lung cancer susceptibility and *GSTA1* gene polymorphism [27]. However, the enzyme encoded by *GSTA1* plays a pivotal part in neutralizing carcinogens, environmental toxins, and harmful oxidative stress byproducts, and SNPs in its promoter region can regulate gene expression by altering promoter activity [24]. Furthermore, studies have shown that *GSTA1* promotes invasion and adhesion of lung cancer cells and contributes to metastasis through the induction of epithelial-mesenchymal transition (EMT) [28], as well as facilitating tumor cell growth [29]. In the present study, no significant interactive effect between *GSTM1* and *GSTA1* polymorphisms on LSCC risk in COPD patients was observed (OR = 2.651, 95% CI: 0.944–7.848, $P = 0.069$). This lack of significance may be attributed to the relatively small sample size and variations in LSCC clinical staging.

Regarding clinical indicators, advanced age, low BMI, greater smoking history, and the presence emerged as independent LSCC risk indicators in COPD individuals. Advanced age is established as a lung cancer susceptibility factor. A large-scale retrospective study [30] found that each decade of age correlated with an 18% heightened likelihood of lung cancer development. Consistently, our study found that age > 60 years was a solitary risk predictor for LSCC in those with COPD. BMI, which measures body fat proportionate to height and weight, has demonstrated a negative link to lung cancer risk across numerous investigations [31,32]. A comprehensive review of ten forward-looking cohort studies revealed that individuals with a body mass index below 18.5 kg/m² faced a higher likelihood of developing lung cancer, while each 5 kg/m² rise in BMI corresponded to a 21% lower risk of the disease [33]. Similarly, our findings revealed a negative correlation between BMI and LSCC probability among COPD patients (OR = 0.130, 95% CI: 0.042–0.409). Smoking contributes to both COPD and lung cancer, which significantly contributes to the onset of each condition. Our research findings indicated that, when accounting for other variables, individuals with a smoking history exceeding 60 pack-years were 2.48 times more likely to develop LSCC, which underscores the significant impact of nicotine's cancer-causing processes. Emphysema, a type of smoking-related lung injury, is acknowledged as an independent predictor of lung cancer [34]. Our findings (OR = 2.362, 95% CI: 1.019–5.475) corroborate Wang et al.'s findings [35], indicating that emphysema-predominant COPD increases LSCC risk (OR = 1.73, 95% CI: 1.03–2.89), even after adjusting for age, sex, BMI, and smoking background.

With respect to laboratory indicators, serum albumin, a hepatically synthesized protein, exerts a key function in the transport and regulation of nutrients, hormones, and pharmaceuticals. Reduced serum albumin concentrations are linked to systemic inflammatory activation and an elevated risk of malnutrition. Albumin safeguards tissues against inflammatory harm. In the Chinese population, albumin levels and lung cancer incidence were shown to be inversely correlated in a large prospective study [36]. However, our results indicated that albumin level acted as a protective factor against LSCC in COPD individuals (OR = 1.166, 95% CI: 1.050–1.295, $P = 0.004$). COPD-related airway inflammation, oxidative stress, and lung damage could influence the connection between albumin's association with oncogenesis probability. In this specific population, albumin may reflect not only nutritional status but also involvement in shared pathological processes such as inflammation-driven carcinogenesis and metabolic reprogramming, potentially explaining the positive association observed here. Meanwhile, a cohort study by Sprague et al. [37] found no correlation between albumin and lung cancer.

An example of a typical systemic inflammatory marker is the NLR, with NLR > 5 often indicating sustained inflammatory activity [38]. In a large retrospective investigation, Ma et al. discovered that NLR is a standalone factor for LSCC in COPD patients [39]. Interestingly, our results suggested that NLR was a protective factor against LSCC in the COPD individuals (OR = 0.847, 95% CI: 0.731–0.980, P = 0.026). This may be explained by the more severe COPD in the control group (49.3% classified as GOLD 3), which was associated with stronger chronic inflammation and higher NLR levels (median: 4.24). In contrast, the COPD with LSCC group had milder COPD (50% classified as GOLD 2) and lower NLR (median: 2.80). Thus, elevated NLR in this context may reflect aggravated COPD rather than cancer risk, resulting in an OR below 1. NLR has also been recognized as an independent prognostic predictor in numerous cancers, notably in lung cancer, where elevated NLR is linked to poorer outcomes in NSCLC patients [40,41].

The risk indicators for LSCC in COPD patients identified in our study (age > 60 years old, BMI < 25 kg/m², smoking > 60 pack-years, and presence of emphysema) are consistent with those included in the COPD-LUCSS score [42]. Furthermore, by incorporating *GSTM1* and *GSTA1* genetic analysis, *GSTM1* was confirmed as an independent genetic risk factor. An integrated genetic-clinical predictive model was subsequently developed using seven common clinical variables. This model demonstrated superior performance in predicting LSCC risk in COPD patients and is accompanied by an online nomogram to facilitate rapid clinical assessment.

5. Limitations of the study

The participants were exclusively of Han Chinese ethnicity, necessitating external validation in other populations; a rather limited number of subjects and patchy clinical data, such as COPD duration and personal history of lung carcinoma, were incomplete; the single-center design may introduce selection bias. Moreover, the model has not yet been applied in clinical practice and requires further validation in large prospective multicenter cohorts.

6. Conclusion

This research shows that the *GSTM1* null genotype significantly increases lung squamous cell carcinoma among individuals with COPD. On the flip side, there was no high score that was associated with the *GSTA1* mutation and the risk of LSCC in this population. The LSCC risk prediction model and online tool developed in this study offer a new approach for early prediction and personalized intervention of LSCC in COPD patients. Subsequent work should concentrate on improving and broadly testing the model.

Funding

Anhui Province's Clinical Medical Research Transformation Special Initiative (Project No.: 202204295107020016)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Sung H, Ferlay J, Siegel R, et al., 2021, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin, 71(3): 209–249.

[2] Nicholson A, Tsao M, Beasley M, et al., 2022, The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. J Thorac Oncol, 17(3): 362–387.

[3] Torres J, Bastarrika G, Wisnivesky J, et al, 2007, Assessing the Relationship Between Lung Cancer Risk and Emphysema Detected on Low-Dose CT of the Chest. Chest, 132(6): 1932–1938.

[4] Torres J, Marín J, Casanova C, et al., 2011, Lung Cancer in Patients With Chronic Obstructive Pulmonary Disease: Incidence and Predicting Factors. Am J Respir Crit Care Med, 184(8): 913–919.

[5] Mannino D, Aguayo S, Petty T, et al., 2003, Low Lung Function and Incident Lung Cancer in the United States: Data From the First National Health and Nutrition Examination Survey Follow-up. Arch Intern Med, 163(12): 1475–1480.

[6] Young R, Hopkins R, 2012, Diagnosing COPD and Targeted Lung Cancer Screening. Eur Respir J, 40(4): 1063–1064.

[7] Celli B, Fabbri L, Criner G, et al., 2022, Definition and Nomenclature of Chronic Obstructive Pulmonary Disease: Time for Its Revision. Am J Respir Crit Care Med, 206(11): 1317–1325.

[8] Adeloye D, Song P, Zhu Y, et al., 2022, Global, Regional, and National Prevalence of, and Risk Factors for, Chronic Obstructive Pulmonary Disease (COPD) in 2019: A Systematic Review and Modelling Analysis. Lancet Respir Med, 10(5): 447–458.

[9] Young R, Hopkins R, Christmas T, et al., 2009, COPD Prevalence Is Increased in Lung Cancer, Independent of Age, Sex and Smoking History. Eur Respir J, 34(2): 380–386.

[10] Wilson D, Weissfeld J, Balkan A, et al., 2008, Association of Radiographic Emphysema and Airflow Obstruction With Lung Cancer. Am J Respir Crit Care Med, 178(7): 738–744.

[11] Qi C, Sun S, Xiong X, 2022, From COPD to Lung Cancer: Mechanisms Linking, Diagnosis, Treatment, and Prognosis. Int J Chron Obstruct Pulmon Dis, 17: 2603–2621.

[12] Singh R, Reindl K, 2021, Glutathione S-Transferases in Cancer. Antioxidants, 10(5): 701.

[13] Pathak A, Husain N, Kant S, et al., 2021, Independent and Interactive Effect of CYPs and GSTs Genetic Variants and Tobacco Smoking on the Risk of Non-Small Cell Lung Carcinoma. Arch Med Res, 52(7): 719–730.

[14] Alexandrov K, 2002, CYP1A1 and GSTM1 Genotypes Affect Benzo[a]pyrene DNA Adducts in Smokers' Lung: Comparison With Aromatic/Hydrophobic Adduct Formation. Carcinogenesis, 23(12): 1969–1977.

[15] Norppa H, 2004, Cytogenetic Biomarkers and Genetic Polymorphisms. Toxicol Lett, 149(1–3): 309–334.

[16] Ramzy M, Solliman M, Abdel-Hafiz H, et al., 2011, Genetic Polymorphism of GSTM1 and GSTP1 in Lung Cancer in Egypt. Public Health, 3(1): 41–51.

[17] Romero L, Andrews K, Ng L, et al., 2006, Human GSTA1-1 Reduces c-Jun N-Terminal Kinase Signalling and Apoptosis in Caco-2 Cells. Biochem J, 400(1): 135–141.

[18] Mlakar V, Curtis P, Armengol M, et al., 2021, The Analysis of GSTA1 Promoter Genetic and Functional Diversity of Human Populations. Sci Rep, 11(1): 5038.

[19] Li S, Li M, Wu J, et al., 2024, Developing and Validating a Clinlabomics-Based Machine-Learning Model for Early Detection of Retinal Detachment in Patients With High Myopia. J Transl Med, 22(1): 405.

[20] Wijma K, Wijma B, Zar M, 1998, Psychometric Aspects of the W-DEQ: A New Questionnaire for the Measurement of Fear of Childbirth. J Psychosom Obstet Gynecol, 19(2): 84–97.

[21] Ford J, 2000, Glutathione S-Transferase M1 Polymorphism and Lung Cancer Risk in African-Americans. Carcinogenesis, 21(11): 1971–1975.

[22] Dey T, Dutta P, Kalita J, et al., 2015, Glutathione S-Transferase Gene Polymorphism and Lung Cancer in Indian Population: A Meta-Analysis of Case-Control Studies. Curr Sci, 109(3): 536–540.

[23] Young R, Hopkins R, Hay J, et al., 2011, GSTM1 Null Genotype in COPD and Lung Cancer: Evidence of a Modifier or Confounding Effect? Appl Clin Genet, 137–144.

[24] Sharma N, Singh A, Singh N, et al., 2015, Genetic Polymorphisms in GSTM1, GSTT1 and GSTP1 Genes and Risk of Lung Cancer in a North Indian Population. Cancer Epidemiol, 39(6): 947–955.

[25] Lee K, Kang D, Clapper M, et al., 2008, CYP1A1, GSTM1, and GSTT1 Polymorphisms, Smoking, and Lung Cancer Risk in a Pooled Analysis Among Asian Populations. Cancer Epidemiol Biomarkers Prev, 17(5): 1120–1126.

[26] López-Cima M, Álvarez-Avellón S, Pascual T, et al., 2012, Genetic Polymorphisms in CYP1A1, GSTM1, GSTP1 and GSTT1 Metabolic Genes and Risk of Lung Cancer in Asturias. BMC Cancer, 12: 433.

[27] Sørensen M, Autrup H, Tjønneland A, et al., 2005, Genetic Polymorphisms in CYP1B1, GSTA1, NQO1 and NAT2 and the Risk of Lung Cancer. Cancer Lett, 221(2): 185–190.

[28] Wang W, Liu F, Wang C, et al., 2017, Glutathione S-Transferase A1 Mediates Nicotine-Induced Lung Cancer Cell Metastasis by Promoting Epithelial-Mesenchymal Transition. Exp Ther Med, 14(2): 1783–1788.

[29] Liu H, Yang Z, Zang L, et al., 2018, Downregulation of Glutathione S-Transferase A1 Suppressed Tumor Growth and Induced Cell Apoptosis in A549 Cell Line. Oncol Lett, 16(1): 467–474.

[30] Sandelin M, Mindus S, Thuresson M, et al., 2018, Factors Associated With Lung Cancer in COPD Patients. Int J Chron Obstruct Pulmon Dis, 13: 1833–1839.

[31] Chen J, Ke K, Liu Z, et al., 2023, Body Mass Index and Cancer Risk: An Umbrella Review of Meta-Analyses of Observational Studies. Nutr Cancer, 75(4): 1051–1064.

[32] Yu D, Zheng W, Johansson M, et al., 2018, Overall and Central Obesity and Risk of Lung Cancer: A Pooled Analysis. JNCI J Natl Cancer Inst, 110(8): 831–842.

[33] Kawai S, Lin Y, Tsuge H, et al., 2024, Body Mass Index and Lung Cancer Risk: Pooled Analysis of 10 Prospective Cohort Studies in Japan. Cancer Sci, 115(4): 1346–1359.

[34] Yong P, Sigel K, de-Torres J, et al., 2019, The Effect of Radiographic Emphysema in Assessing Lung Cancer Risk. Thorax, 74(9): 858–864.

[35] Wang W, Xie M, Dou S, et al., 2018, The Link Between Chronic Obstructive Pulmonary Disease Phenotypes and Histological Subtypes of Lung Cancer: A Case–Control Study. Int J Chron Obstruct Pulmon Dis, 13: 1167–1175.

[36] Yang Z, Zheng Y, Wu Z, et al., 2021, Association Between Pre-Diagnostic Serum Albumin and Cancer Risk: Results From a Prospective Population-Based Study. Cancer Med, 10(12): 4054–4065.

[37] Sprague B, Trentham-Dietz A, Klein B, et al., 2008, Physical Activity, White Blood Cell Count, and Lung Cancer Risk in a Prospective Cohort Study. Cancer Epidemiol Biomarkers Prev, 17(10): 2714–2722.

[38] Zahorec R, 2001, Ratio of Neutrophil to Lymphocyte Counts—Rapid and Simple Parameter of Systemic Inflammation and Stress in Critically Ill. Bratisl Med J, 102(1): 5–14.

[39] Ma A, Wang G, Du Y, et al., 2022, The Clinical Relevance of Neutrophil-to-Lymphocyte Ratio and Platelet-to-Lymphocyte Ratio in Chronic Obstructive Pulmonary Disease With Lung Cancer. Front Oncol, 12: 902955.

[40] Bronte V, Brandau S, Chen S, et al., 2016, Recommendations for Myeloid-Derived Suppressor Cell Nomenclature and Characterization Standards. Nat Commun, 7(1): 12150.

[41] Sarraf K, Belcher E, Raevsky E, et al., 2009, Neutrophil/Lymphocyte Ratio and Its Association With Survival After Complete Resection in Non–Small Cell Lung Cancer. J Thorac Cardiovasc Surg, 137(2): 425–428.

[42] de-Torres J, Wilson D, Sanchez-Salcedo P, et al., 2015, Lung Cancer in Patients With Chronic Obstructive Pulmonary

Disease: Development and Validation of the COPD Lung Cancer Screening Score. Am J Respir Crit Care Med, 191(3): 285–291.