Open Access Full Text Article

ORIGINAL RESEARCH

Longitudinal Analysis of Risk Factors for Pulmonary Function Decline in Chronic Lung Diseases Over Five Years

Lu Li*, Jiaqi Meng*, Jiquan Chen

Department of Pulmonary and Critical Care Medicine, Third Affiliated Hospital of Naval Medical University, Shanghai, 200438, People's Republic of China

*These authors contributed equally to this work

Correspondence: Jiquan Chen, Department of Pulmonary and Critical Care Medicine, Third Affiliated Hospital of Naval Medical University, Shanghai, 200438, People's Republic of China, Tel +86-13065135302, Email chenjq9932@163.com

Objective: Chronic lung diseases (CLDs) are a major global health concern, characterized by a progressive decline in pulmonary function that severely impacts quality of life. It is essential to identify and predict the primary risk factors for CLDs. This study aims to establish a predictive model to assist healthcare providers in the early identification of high-risk patients and timely interventions and treatment options.

Methods: This study utilized questionnaire data from the China Health and Retirement Longitudinal Study (CHARLS) collected in 2011, 2013, and 2015. A latent class growth model (LCGM) was established using CLDs as the baseline sample. This model stratified the patients based on the extent of the decline in Δpeak expiratory flow (ΔPEF), which served as the target variable. Independent variables included age, gender, smoking status, body mass index, education level, and comorbidities. A random forest model was developed using Python, and the importance of the feature was visualized through the SHAP method. The predictive performance of the model was evaluated using receiver operating characteristic (ROC) curve analysis, calibration curve analysis, and decision curve analysis.

Results: After screening, a total of 553 patients with CLDs were included in the study. The random forest model pinpointed grip strength, age, education level, gender, and asthma as the top five risk factors for pulmonary function decline. Specifically, the model demonstrated robust predictive performance with an area under the ROC curve (AUC) value of 0.77, affirming its accuracy and clinical applicability. Both calibration and decision curves further substantiated the reliability of the model in identifying patients at increased risk for pulmonary function decline.

Conclusion: The predictive model developed in this study serves as a valuable tool for clinicians to target early interventions and optimize treatment strategies to enhance the quality of care and patient outcomes in the management of CLDs.

Keywords: chronic lung diseases, pulmonary function decline, latent class growth modeling, random forest model, health services, machine learning in healthcare

Introduction

Chronic lung diseases (CLDs), characterized mainly by chronic inflammation, structural changes, and functional impairment in the lungs, encompass various disease types such as chronic obstructive pulmonary disease (COPD), bronchial asthma, chronic bronchitis, and interstitial pneumonia. Currently, it is the third leading cause of death worldwide.[1] Chronic obstructive pulmonary disease (COPD) is a predominant contributor, accounting for 330 out of every 400 deaths attributed to CLDs.[1] This alarming issue is especially severe in the Western Pacific region.[2] In this context, the precise identification and prediction of factors leading to the decline in pulmonary function among patients with CLDs hold immense clinical and public health significance. The China Health and Retirement Longitudinal Study (CHARLS), a nationwide, large-scale longitudinal survey, provides valuable data resources to examine the health status of the middle-aged and older population in China, including various conditions such as CLDs and asthma.[3]

Received: 16 July 2024
Accepted: 27 November 2024
Published: 5 December 2024

COPD is a highly heterogeneous group of diseases characterized by symptoms such as dyspnea and coughing.[4,5] Although smoking is widely acknowledged as a significant risk factor,[2,4,5] The development and progression of COPD are influenced by a complex interplay of factors, including genetic predispositions, environmental pollutants, occupational hazards, chronic inflammation, and dysregulation of the lung microbiome.[2,4–6] Moreover, COPD is frequently associated with comorbid conditions, such as coronary artery disease, cognitive impairments, and psychological disorders.[7] This evidence points to the necessity of viewing C OPD as a systemic condition and highlights the need for comprehensive management strategies.

Peak expiratory flow (PEF) is recognized as a critical indicator for assessing lung function in patients with COPD and has demonstrated superior predictive value for mortality compared to other pulmonary function tests.[8] Therefore, identifying the key factors influencing alterations in PEF is crucial for predicting the disease progression trajectory in COPD patients. Nonetheless, given the heterogeneity and complexity of COPD, a single factor is insufficient to comprehensively reflect the disease progression. It is essential to employ a multidimensional and dynamic methodology to accurately evaluate the risks and progression patterns associated with COPD. In recent years, machine learning in medicine has brought a new paradigm. The random forest (RF) model is recognized in chronic disease research for its predictive capabilities and data processing.[9] The latent class growth model (LCGM) can identify different developmental trajectories within patient populations, thereby providing a scientific basis for creating personalized treatment strategies.[10] The integration of advanced statistical and machine learning methods with the CHARLS database offers promising avenues for in-depth analysis of the dynamic characteristics and influencing factors of pulmonary function changes in patients with COPD.

This study utilized patient data from the CHARLS database to identify different trajectories of PEF changes in patients with CLDs using an LCGM. Concurrently, the RF model was utilized to explore the pivotal risk factors that caused pulmonary function decline in patients with COPD over a five-year period. This comprehensive predictive model was constructed by integrating multidimensional data on patients' demographic characteristics, lifestyle choices, comorbid conditions, and physical indices. The research aimed to identify COPD pulmonary function change patterns, determine critical risk factors, and establish a predictive model for early identification of high-risk patients. The outcomes gained from this study were expected to underpin the prevention strategies, early interventions, and tailored treatments for COPD, thereby enhancing patient quality of life, lessening disease burden, and informing more effective public health strategies.

Methods

Data Source and Study Population

Data for the study were collected from the CHARLS database for the years 2011, 2013, and 2015, involving 6930 participants who underwent three consecutive PEF measurements. The baseline data were acquired from 620 participants diagnosed with CLDs during their initial participation in the 2011 CHARLS survey. During the data filtering process, patients with low compliance, inadequate data reliability, or conditions such as psychiatric disorders, stroke, or memory impairments were excluded. Data completeness was ensured by excluding records with missing information on body mass index (BMI), activities of daily living (ADL), the 10-item Center for Epidemiologic Studies Depression Scale (CESD-10), blood pressure, grip strength, and sleep duration. Ultimately, data from 553 patients were included in the study. A detailed participant flow diagram is provided in Figure 1.

The CHARLS is a nationwide representative research project designed to support scientific investigations into the health and socioeconomic conditions of the middle-aged and older population in China, specifically those aged 45 and above.[3] The baseline survey of this project commenced in 2011, involving 17,708 participants from 28 provinces and 150 counties. Detailed information on the geographic distribution of the participants, included subjects, and those who tested positive are available in Figure 2.

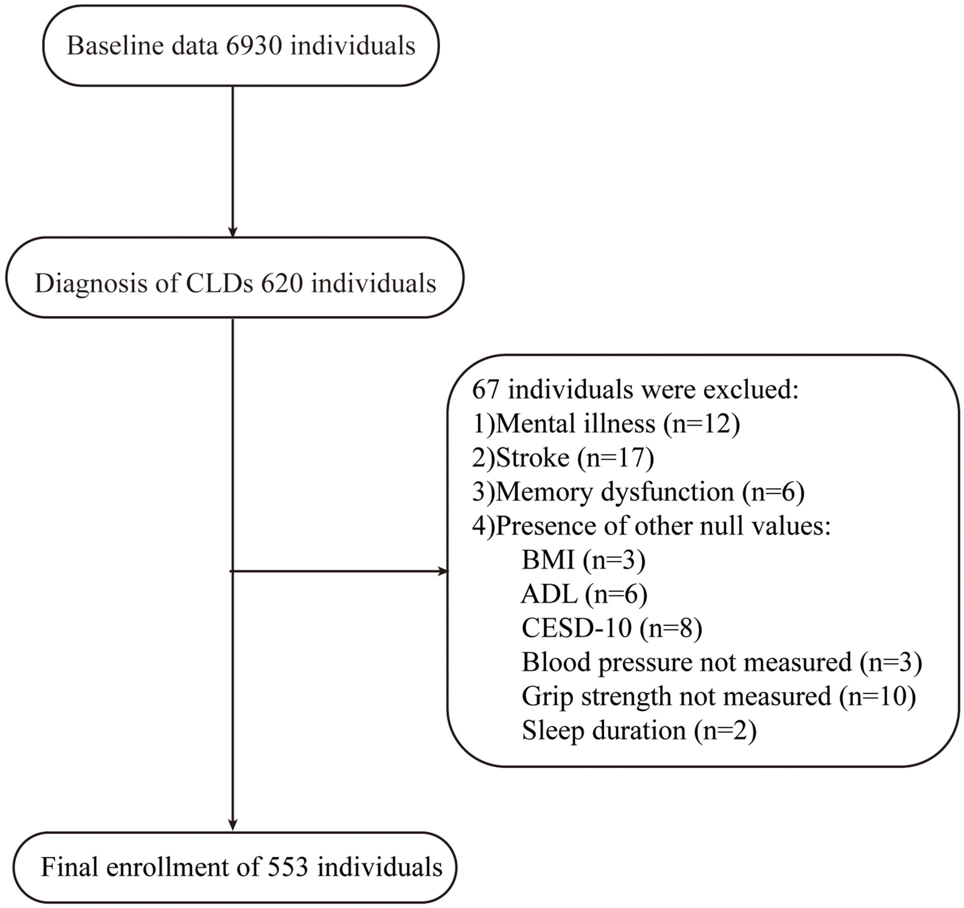

Figure 1 Flowchart of participant enrollment for the study. Selection and filtration process from a baseline dataset of 6930 individuals, leading to the final inclusion of 553 participants after accounting for various exclusion criteria related to health conditions and data completeness.
Abbreviations: CLDs, Chronic lung diseases; BMI, Body mass index; ADL, Activities of daily living; CESD-10, 10-item Center for Epidemiologic Studies Depression Scale.

Research Measures and Analytical Variables

Assessment of Lung Function

PEF is defined as the highest flow achieved during an expiratory effort from maximum lung inflation.[8] It is an important indicator for evaluating lung function, and its measurement results are supported by accuracy and effectiveness in many aspects.[8] In CHARLS, PEF assessments were conducted using a peak flow meter (Everpure TM, Shanghai, China). The flow meter is designed based on the principles of fluid mechanics, and its internal structure can accurately measure gas flow. It has also undergone a strict calibration process before being put into use to ensure the accuracy and reliability of the measurement results. Participants were required to perform the operation according to the standard specification and were instructed to repeat the measurement three times with a 30-s interval between each measurement. The highest value of the three readings was then selected for inclusion in the research data. The highest value can more effectively reflect the lung function level of the participants in the best breathing state and avoid affecting the accuracy of the final result due to possible errors in a single measurement.

Assessment of Depressive Symptoms

The baseline measurements of depressive symptoms were conducted using the Chinese version of the CESD-10.[11] This scale comprises ten items designed to evaluate the emotional and behavioral states of participants over the preceding week. Responses were scored on a four-point scale, with total scores ranging from 0 to 30. A higher score on this scale was indicative of more severe depressive symptoms.

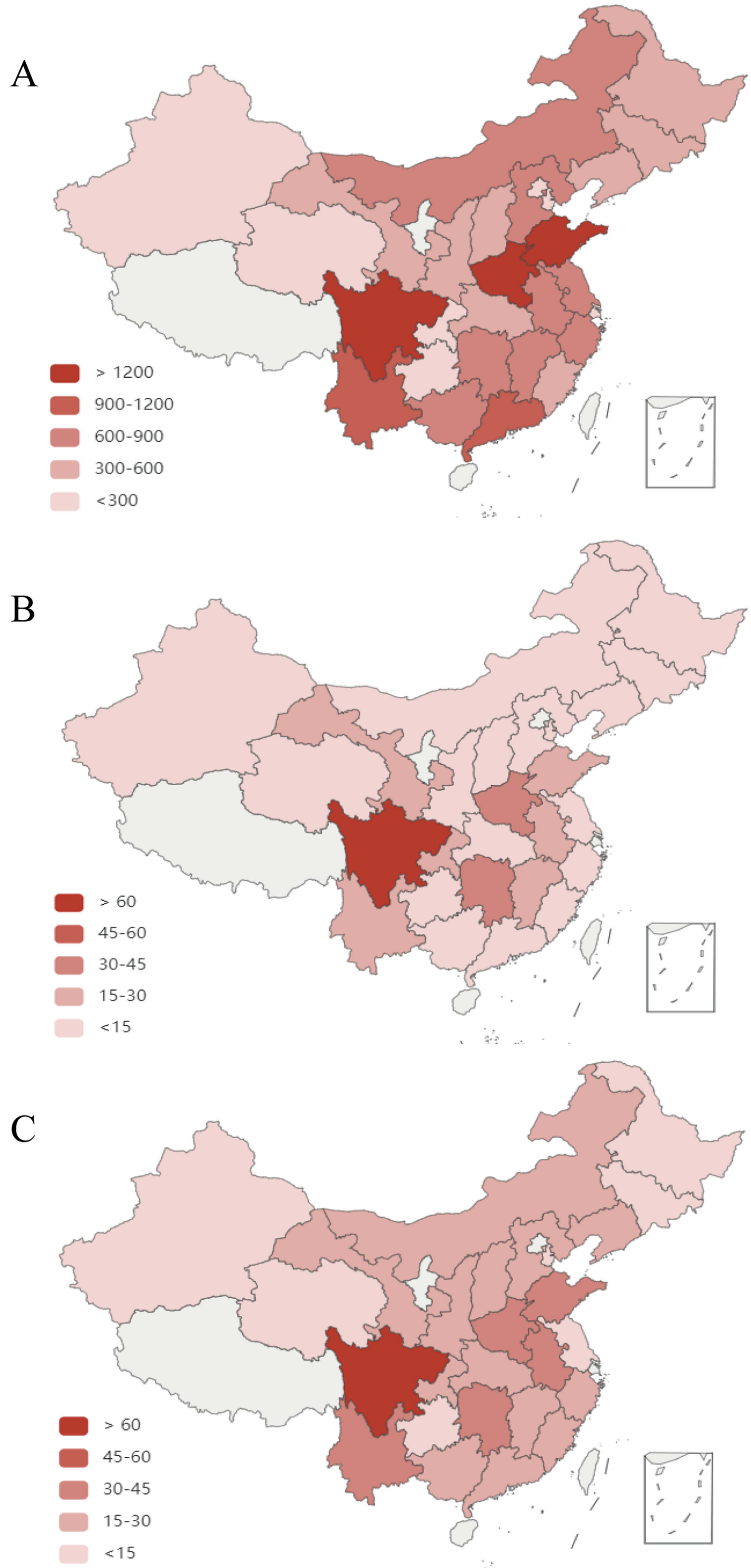

Figure 2 Geographic representation of participants. (**A**) Overall distribution of all 17,708 respondents across 28 provinces and 150 counties in the China Health and Retirement Longitudinal Study. (**B**) Distribution of 553 patients with CLDs specifically included in this study. (**C**) Regional representation of the 395 patients classified in the PEF decrease group in the study.
Abbreviation: CLDs, Chronic lung diseases.

Assessment of Daily and Instrumental Activities

Participants were asked to rate their difficulty in performing ADL and instrumental activities of daily living (IADL). The responses were categorized into four levels: (1) no difficulty, (2) some difficulty but can complete tasks independently, (3) difficulty requiring assistance, and (4) inability to complete tasks. The ADL index, which includes essential self-care tasks such as dressing, bathing, eating, sleeping, toileting, and continence, ranges from 0 to 6, with higher scores indicating more severe disability.[12] Similarly, the IADL index covers more complex activities such as household tasks, cooking, shopping, financial management, and medication management, scored from 0 to 5, where higher scores reflect greater disability.[13]

Cognitive Function Measurement

Cognitive function was evaluated based on two dimensions, mental integrity and episodic memory, according to the CHARLS data.[14] Mental integrity was measured with a subset of items from the Telephone Interview for Cognitive Status (TICS), which included date recollection, day awareness, seasonal recognition, picture replication, and the ability to perform serial subtractions of seven (conducted five times). Scores ranged from 0 to 11, with higher scores indicating better cognitive status. Episodic memory was assessed through immediate and delayed recall of a list of ten words, with each correctly recalled word scoring one point, for a total possible score of 0 to 10. The total cognitive function score was the sum of the mental integrity and episodic memory scores, ranging from 0 to 21, with higher scores indicating stronger cognitive abilities.[15]

Evaluation of Other Baseline Risk Factors

Additional baseline risk factors were collected, which included demographic information such as age and gender, physiological measurements such as BMI, and socio-economic indicators including marital status, residence area (urban or rural), educational attainment, health insurance, and total household income. Detailed health-related behaviors such as comorbidities, smoking, and alcohol consumption over the previous year were documented along with physical health metrics, including pain, sleep duration, social activities, disability, physical exercise, grip strength, and self-assessed health status. Marital status was categorized as "married" and "other". Educational attainment was classified as "no normal education", "primary school", "middle school", and "high school and above". Smoking and drinking status over the past year were classified as "current smoker/drinker" and "not non-smoker/non-drinker". Sleep duration was defined as nighttime sleep measured in hours. Health insurance, pain, social activities, disability, and physical exercise were categorized as either "yes" or "no". Comorbidities were diagnosed with chronic diseases, including "hypertension, diabetes, cancer, heart disease, dyslipidemia, and asthma". Grip strength was measured using the maximum value from the dominant hand. Self-assessed health status was scored into "very good (5 points), good (4 points), average (3 points), poor (2 points), and very poor (1 point)".

Statistical Analysis

Data analysis was conducted using SPSS version 27.0. Continuous variables with normal distribution were reported as mean ± standard deviation, and comparisons between two groups were performed using the *t*-test. For continuous variables that did not meet the normality assumption, data were reported as median (interquartile range) and compared using non-parametric tests. Categorical data were expressed as numbers and percentages (n, %), with comparisons of proportions made using the chi-square test. Statistical significance was defined as a p-value of less than 0.05.

LCGM was developed with the MPLUS 8.3 software. The number of classes in the model was incrementally increased from "1", and a comprehensive evaluation of the results was performed to select the optimal model. The primary fit indices used were the Akaike information criterion (AIC), Bayesian information criterion (BIC), and sample size-adjusted BIC (saBIC), where lower values indicated a superior model fit. Entropy values, ranging from 0 to 1, were also assessed, with an entropy value of 0.8 reflecting a classification accuracy exceeding 90%. The model comparisons were conducted using the bootstrapped likelihood ratio test (BLRT) and the Lo-Mendell-Rubin (LMR) likelihood ratio test. A p-value less than 0.05 signified that model K provided a significantly better fit than model K-1.[16]

The RF model was developed using Python, with patient data classified into training and testing sets in a 7:3 ratio through the sklearn package. Model optimization was achieved via grid search, tuning critical parameters such as the number of decision trees, the maximum depth of the trees, the maximum number of features considered for splitting, the minimum number of samples required for node splitting, and the minimum number of samples required at leaf nodes. To prevent overfitting, all training sessions were subjected to 10-fold cross-validation. The SHAP method was employed to visualize the importance of features in the model. The performance of the RF model was assessed using receiver operating characteristic (ROC) curves, calibration curves, and decision curve analysis.[9]

Results

Classification of PEF Trajectories Using LCGM

The LCGM was applied to categorize changes in PEF values over three waves: the first wave in 2011, the second wave in 2013, and the third wave in 2015. This analysis revealed five latent classes, with more classes correlating with decreased AIC, BIC, and saBIC. The LMR test results indicated non-significance ($p > 0.05$) for models with three or five classes. However, the LMR test reached significance with two or four latent classes ($p < 0.05$), accompanied by comparatively lower AIC, BIC, and saBIC values. Notably, the two-class model demonstrated the highest entropy value, suggesting the highest classification accuracy for this model. Therefore, the LCGM model with two classes was selected, as shown in Table 1. Based on this model, two subgroups within the CLD population were delineated for PEF change patterns: a PEF increase group and a PEF decrease group, as depicted in Figure 3.

Characteristics Influencing Group Classification Based on PEF Trajectories

Based on the trajectories of PEF, 553 patients with CLDs were divided into a PEF increase group (158 individuals) and a PEF decrease group (395 individuals). Notable differences in baseline characteristics were identified between the two groups. The PEF decrease group had a higher average age, a greater proportion of females, and lower educational attainment and household income. In terms of physical condition, this group exhibited lower BMI, weaker grip strength, and a lower percentage of alcohol consumption. Their health status was comparatively poorer, with higher incidences of hypertension and asthma. Higher scores on the CESD-10, ADL, and IADL, and lower cognitive ability scores were recorded for the PEF decrease group (Table 2).

It is important to note that no significant differences were found between the groups in terms of smoking status, pain status, sleep duration, social activities, disability status, and exercise habits ($p > 0.05$).

Optimization and Results of the RF Model

An RF model was developed using Python, incorporating all variables from the training set. The model was optimized through grid search. All training processes underwent 10-fold cross-validation to reduce overfitting. The optimal model was achieved when the number of decision trees was set at 91, the maximum depth at 11, and the minimum number of

Table 1 Model Fitting for LCGMs in Patients with CLDs Based on PEF Change Patterns

Model	K	AIC	BIC	saBIC	Entropy	LMR	BLRT	Class Probability (%)
1	5	20633.456	20655.033	20639.160	–	–	–	–
2	8	19954.088	19988.611	19963.216	0.871	0.000	0.000	71.4/28.6
3	11	19768.726	19816.195	19781.276	0.848	0.104	0.000	32/9.6/58.4
4	14	19692.885	19753.300	19708.858	0.800	0.006	0.000	32/5.2/17.8/45
5	17	19677.338	19750.699	19696.734	0.766	0.476	0.000	4/17.9/31.1/39.4/7.6

Notes: Model number (1–5) indicates different potential category model settings. K is the number of categories in the model. AIC, BIC, and saBIC measure the model's goodness of fit, with smaller values being better. Entropy (0–1) indicates classification accuracy (closer to 1 is better). LMR compares model fit; p < 0.05 means the current model fits better than the one with one less category. BLRT is for model comparison. Class Probability (%) is the probability percentage of each category.

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; saBIC, sample size-adjusted BIC; LMR, Lo-Mendell-Rubin; BLRT, bootstrap likelihood ratio test; CLD, chronic lung disease; LCGM, latent class growth model; PEF, peak expiratory flow.

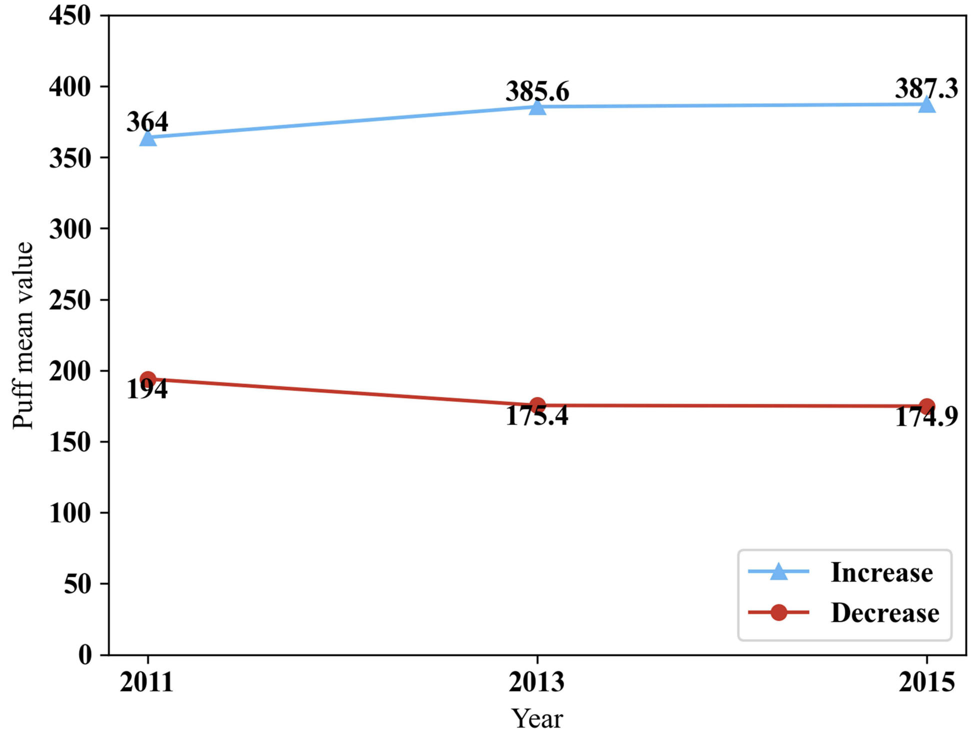

Figure 3 Trajectory graph of PEF in patients with CLDs. Different potential trajectories for PEF among patients with CLDs, which were monitored over the course of the study.
Abbreviation: CLDs, Chronic lung diseases.

samples per leaf node at 3, resulting in an area under the ROC curve (AUC) value of 0.77. Based on the SHAP method, the most significant contributors to PEF prediction were grip strength, age, educational attainment, gender, and asthma (Figure 4).

Table 2 Baseline Characteristics of CLD Patients Grouped by Changes in PEF

Variables	Total (N = 553)	PEF Increase Group (N = 158)	PEF Decrease Group (N = 395)	P-value
Age (years)	61.2 ± 8.9	57.4 ± 7.8	62.8 ± 8.9	<0.001[a]
Male	330 (60%)	123 (78%)	207 (52%)	< 0.001[b]
BMI, kg/m²	23.1 ± 4.3	23.7 ± 3.6	22.8±4.5	0.032[a]
Marital status	479 (87%)	143 (91%)	336 (85%)	0.089[b]
Residence area (Urban /Rural)	150/403	48/110	102/293	0.276[b]
Educational attainment				
No formal education	256 (46.3%)	45 (28.5%)	211 (53.4%)	
Primary school	160 (28.9%)	52 (32.9%)	108 (27.3%)	
Middle school	90 (16.3%)	37 (23.4%)	53 (13.4%)	
High school or above	47 (8.5%)	24 (15.2%)	23 (5.8%)	<0.001[b]
Household income	8380 (21250)	12825 (27574)	6460 (20880)	0.009[d]
Current smoker	208 (37.6%)	68 (43.0%)	140 (35.4%)	0.096[b]
Current drinker	193 (34.9%)	78 (49.4%)	115 (29.1%)	<0.00[b]
Pains	267 (48.3%)	72 (45.6%)	195 (49.4%)	0.419[b]
Sleep duration, h	6 (3)	6 (3)	6 (3)	0.333[d]
Social activities	278 (50.3%)	79 (50%)	199 (50.4%)	0.936[b]
Disabilities	125 (22.6%)	29 (18.4%)	96 (24.3%)	0.131[b]

(*Continued*)

Table 2 (Continued).

Variables	Total (N = 553)	PEF Increase Group (N = 158)	PEF Decrease Group (N = 395)	P-value
Physical activity	508 (91.9%)	150 (94.9%)	358 (90.6%)	0.094[b]
Health Insurance	520 (94%)	145 (91.8%)	375 (94.9%)	0.156[b]
Chronic diseases				
Hypertension	142 (25.7%)	27 (17.1%)	115 (29.1%)	0.003[b]
Diabetes	36 (6.5%)	11 (7.0%)	25 (6.3%)	0.785[b]
Cancer	3 (0.5%)	1 (0.6%)	2 (0.5%)	>0.999[c]
Heart disease	110 (19.9%)	29 (18.4%)	81 (20.5%)	0.567[b]
Dyslipidemia	50 (9.0%)	17 (10.8%)	33 (8.4%)	0.373[b]
Asthma	154 (27.8%)	20 (12.7%)	134 (33.9%)	<0.001[b]
Hospitalization frequency	0 (0)	0 (0)	0 (0)	0.035[d]
Dominant handgrip, kg	32.6 ± 9.8	37.5 ± 9.4	30.6 ± 9.2	<0.001[a]
CESD-10 rating	9 (9)	7 (9)	10 (10)	0.007[d]
ADL scores	0 (0)	0 (0)	0 (0)	0.010[d]
IADL Scores	0 (0)	0 (0)	0 (1)	0.005[d]
Overall cognitive scores	11.5 ± 3.3	12.3 ± 3.0	11.2 ± 3.3	<0.001[a]
Self-health assessment	3 (1)	3 (1)	3 (1)	0.003[d]

Notes: Data are presented as mean ± standard deviation, M (interquartile range), or n (%). [a]Statistical tests include independent-sample *t*-test, [b]Chi-squared test, [c]Fisher's exact test, and [d]mann–Whitney *U*-test.
Abbreviations: BMI, body mass index; CESD-10, 10-item Center for Epidemiologic Studies Depression Scale; ADL, Activities of daily living; IADL, Instrumental activities of daily living; CLD, chronic lung disease; LCGM, latent class growth model; PEF, peak expiratory flow.

Evaluation of the Model Demonstrates High Predictive Accuracy and Clinical Relevance

The performance of the model was evaluated by inputting data from the test set, resulting in the generation of ROC and calibration curves. The ROC curve analysis yielded an AUC of 0.77, indicating good discriminatory ability (Figure 5A). The calibration curve showed that the predictive capacities of the model closely aligned with the ideal prediction line

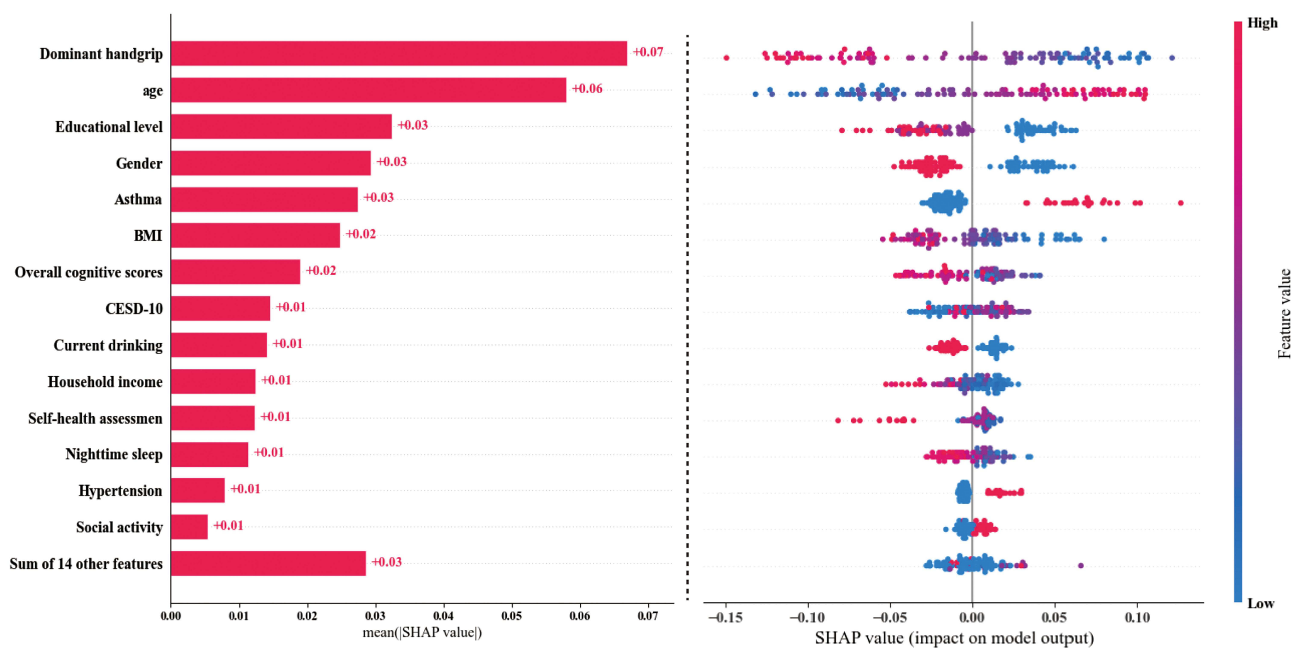

Figure 4 SHAP-based importance of features in the RF model. Ranked visualization of feature importance according to SHAP values, assessing each feature's influence on the predictive performance of the model concerning CLDs.
Abbreviations: CLDs, Chronic lung diseases; RF, Random forest.

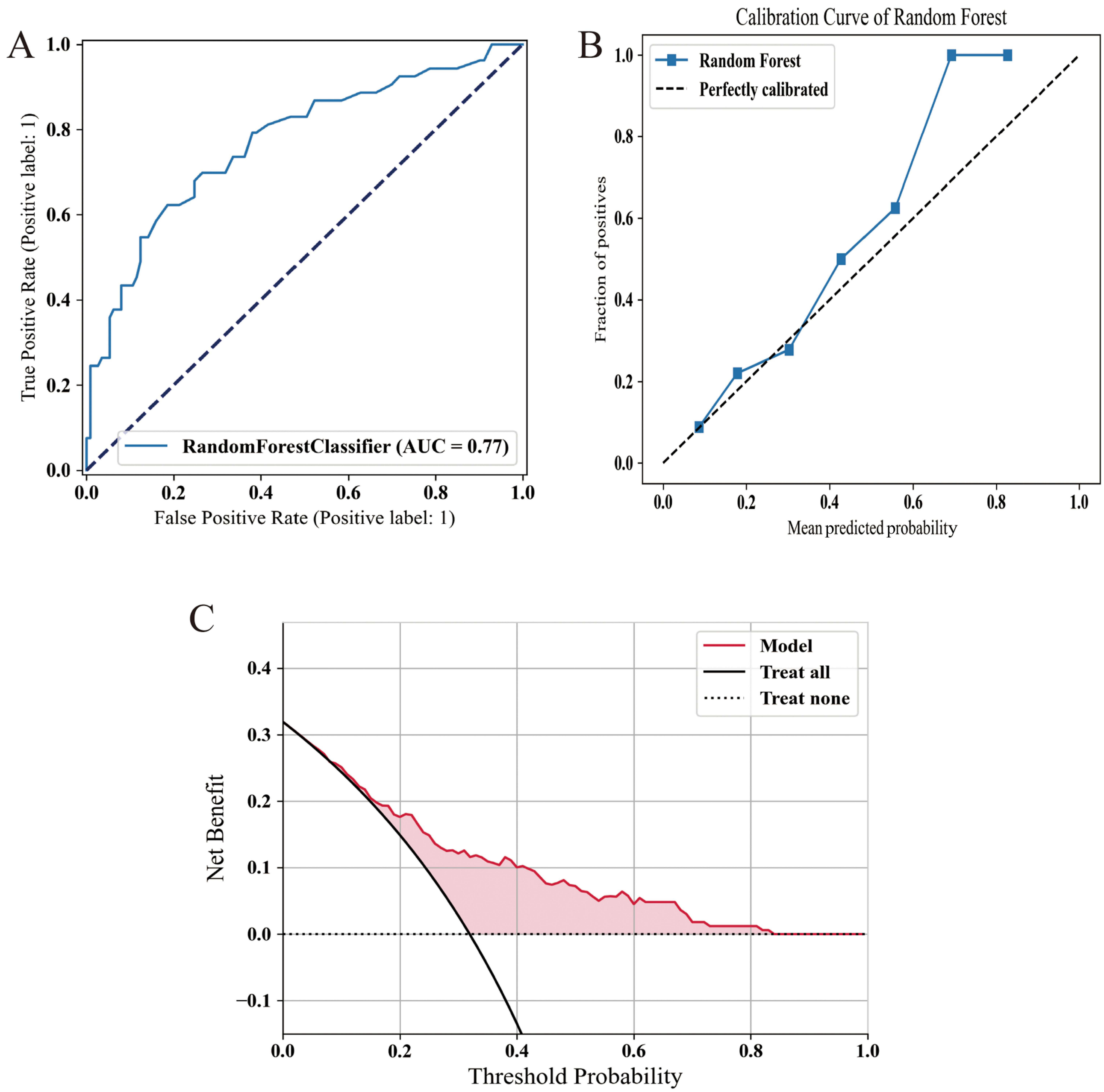

Figure 5 Evaluation of the predictive model. (**A**) ROC curve outlining the model's diagnostic accuracy with an AUC of 0.77. (**B**) Calibration curve showing the model's calibration in terms of predicted versus actual outcomes. (**C**) Decision curve indicating the clinical utility of the model at various decision thresholds.
Abbreviations: ROC, receiver operating characteristic; AUC, Area under the ROC curve.

(slope = 1), suggesting a high degree of fit (Figure 5B). Decision curve analysis was also performed, revealing that the model provided a net benefit for patients within the probability threshold range of 10% to 83%, indicating clinical effectiveness within this range (Figure 5C).

Discussion

This study conducted an in-depth analysis of data from the CHARLS using LCGM and RF models to explore the key risk factors influencing the deterioration of pulmonary function in patients with CLDs over a five-year period. The findings highlight that grip strength, age, education level, gender, and the presence of asthma are significant predictors of lung function deterioration.

Grip strength, recognized as a key measure of overall muscle strength,[17,18] was identified in this study as the foremost risk factor for lung function decline in patients with CLDs. A reduction in grip strength not only signals a general decline in overall health but also highlights the critical connection between muscle function and the prognosis of CLD patients. This conclusion aligns with previous research findings.[19] Hence, management strategies for patients with CLDs should incorporate specific muscle strength training and regular monitoring of BMI. Age emerged as the second most significant predictor, a finding that is consistent with the progressive nature of CLDs and corroborates the results of multiple epidemiological studies.[20–22] As individuals age, a natural decline in lung function occurs, often accompanied by an increase in comorbid conditions. This condition makes patients with CLDs more susceptible to accelerated deterioration in lung function. Educational attainment, identified as the third most significant factor influencing lung function decline in patients with CLDs, plays a pivotal role in shaping patients' health literacy and self-management capabilities. This finding highlights the significant impact of socioeconomic status on chronic disease management and prognosis. Patients with lower education levels pose barriers to health knowledge acquisition, disease management, and accessing high-quality healthcare. A cohort study corroborated this finding, revealing a significant association between the education level of adult offspring and increased risks of readmission and mortality in older patients with COPD.[23] Therefore, when devising comprehensive management plans for patients with CLDs, healthcare providers should not only focus on clinical treatment but also strive to enhance their educational levels through educational interventions, particularly in areas of disease knowledge and self-care skills. This study revealed significant gender differences in the risk of lung function decline among patients with CLDs, with a higher proportion of females in the PEF decrease group compared to the PEF increase group (48% vs 22%, $p < 0.001$; Table 2), which ranked fourth in the risk model. This difference might be attributed to the higher prevalence of asthma in females, as well as other factors such as hormonal levels and lifestyle.[24,25] Asthma emerged as the fifth most influential factor in the progression of CLDs according to the predictive model, thus emphasizing the heterogeneous nature of COPD.[5] Patients with asthma-COPD overlap syndrome (ACOS) may be likely to experience a more rapid decline in lung function.[26,27] There are several potential reasons. ACOS patients have more frequent and severe exacerbations than those with COPD alone, leading to faster lung function deterioration. The chronic inflammation of both conditions may have a synergistic effect, causing more severe airway remodeling and faster progression of airway obstruction.[26,27] ACOS patients have different inflammatory patterns and responses to treatment, making disease management more challenging, underscoring the importance of managing comorbidities.

The analysis of baseline characteristics uncovered that patients in the PEF decrease group had a heavier burden of comorbidities, such as hypertension. Moreover, this group showed significantly higher CESD-10, ADL, and IADL scores compared to the PEF increase group ($p < 0.05$), while their cognitive ability scores were notably lower (11.2 ± 3.3 vs 12.3 ± 3.0, $p < 0.001$). These findings suggest that the PEF decrease group encounters more severe issues in cardiovascular health, respiratory conditions, and mental health, thereby requiring more intensive attention and management. The reduced cognitive ability scores might be linked to a decline in brain function, which potentially impacts their comprehension and implementation of disease management strategies.[14,15] Notably, despite the lack of significant differences between the two groups in smoking status, pain levels, sleep duration, social activities, disability status, and exercise habits, these factors are still essential in the management of CLDs.[28,29]

This study uses extensive longitudinal data and advanced statistical techniques to offer a dynamic view of CLDs. Nonetheless, it has several limitations. The cross-sectional design means that some potentially significant factors, such as environmental exposures and genetic influences, were not accounted for. The study sample was confined to middle-aged and older populations in mainland China, which may limit the generalizability of the findings. In conclusion, this study has elucidated the key risk factors affecting the decline in lung function among patients with CLDs and highlighted the necessity of a multifactorial assessment approach. These insights can aid clinicians in more accurately identifying high-risk patients and developing tailored prevention and intervention strategies, thereby enhancing the long-term outcomes for patients with CLDs.

Conclusion

This research utilized longitudinal data analysis and multivariate modeling to pinpoint the critical risk factors influencing lung function decline in patients with CLDs over a five-year period. It also developed a corresponding risk prediction model. These insights are crucial for implementing early interventions and tailoring personalized treatments in clinical settings to enhance the management of CLDs. Future research should focus on validating these findings in diverse populations and exploring the integration of additional variables, such as genetic and environmental factors, to enhance the predictive accuracy of the model.

Institutional Review Board Statement

This study used publicly available summary data, Ethical approval for this research was granted by the Institutional Review Board (IRB) of Peking University, with the approval number IRB00001052-11015. The study was approved by the Ethics Committee of the Third Affiliated Hospital of Naval Medical University (EHBHKY2024-K004-P001).

Abbreviations

CLD, Chronic lung disease; COPD, Chronic obstructive pulmonary disease; CHARLS, China Health and Retirement Longitudinal Study; LCGM, Latent class growth model; PEF, Peak expiratory flow; ROC, Receiver operating characteristic; BMI, Body mass index; ADL, Activities of daily living; IADL, Instrumental activities of daily living; CESD-10, 10-item Center for Epidemiologic Studies Depression Scale; AIC, Akaike information criterion; BIC, Bayesian information criterion; saBIC, Sample-size adjusted BIC; LMR, Lo-Mendell-Rubin; BLRT, Bootstrapped likelihood ratio test; RF, Random forest; SHAP, SHapley Additive exPlanation; AUC, Area under the receiver operating characteristic curve.

Data Sharing Statement

The data sets used and/or analyzed during this study are available from the CHARLS repository (http://charls.pku.edu.cn), accessible at the following URL: https://charls.pku.edu.cn/info/1131/1812.htm.

Informed Consent Statement

All participants in the survey signed or marked (if illiterate) the informed consent forms. All methods were carried out in accordance with relevant guidelines and regulations.

Acknowledgments

We extend our gratitude to the Institute of Social Science Survey at Peking University for organizing the China Health and Retirement Longitudinal Study (CHARLS). Our thanks also go to all participants, investigators, and research assistants involved in CHARLS for their valuable contributions.

Author Contributions

All authors made a significant contribution to the work report, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

There is no funding to report.

Disclosure

The authors declare no conflict of interest.

References

1. GBD 2019 Chronic Respiratory Diseases Collaborators. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019. *EClinicalMedicine*. 2023;59:101936. doi:10.1016/j.eclinm.2023.101936
2. Adeloye D, Song P, Zhu Y, et al. NIHR RESPIRE Global Respiratory Health Unit. 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*. 2022;10(5):447–458. doi:10.1016/S2213-2600(21)00511-7

3. Zhao Y, Hu Y, Smith JP, et al. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). *Int J Epidemiol*. 2014;43(1):61–68. doi:10.1093/ije/dys203
4. Soriano JB, Polverino F, Cosio BG. What is early COPD and why is it important? *Eur Respir J*. 2018;52(6):1801448. doi:10.1183/13993003.01448-2018
5. Christenson SA, Smith BM, Bafadhel M, Putcha N. Chronic obstructive pulmonary disease. *Lancet*. 2022;399(10342):2227–2242. doi:10.1016/S0140-6736(22)00470-6
6. Yang IA, Jenkins CR, Salvi SS. Chronic obstructive pulmonary disease in never-smokers: risk factors, pathogenesis, and implications for prevention and treatment. *Lancet Respir Med*. 2022;10(5):497–511. doi:10.1016/S2213-2600(21)00506-3
7. Cavaillès A, Brinchault-Rabin G, Dixmier A, et al. Comorbidities of COPD. *Eur Respir Rev*. 2013;22(130):454–475. doi:10.1183/09059180.00008612
8. Hansen EF, Vestbo J, Phanareth K, et al. Peak flow as predictor of overall mortality in asthma and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2001;163(3 Pt 1):690–693. doi:10.1164/ajrccm.163.3.2006120
9. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56–67. doi:10.1038/s42256-019-0138-9
10. Nguena Nguefack HL, Pagé MG, Katz J, et al. Trajectory modelling techniques useful to epidemiological research: a comparative narrative review of approaches. *Clin Epidemiol*. 2020;12:1205–1222. doi:10.2147/CLEP.S265287
11. Zhu L, Wang Y, Li J, et al. Depressive symptoms and all-cause mortality among middle-aged and older people in China and associations with chronic diseases. *Front Public Health*. 2024;12:1381273. doi:10.3389/fpubh.2024.1381273
12. Katz S. The index of ADL: a standardized measure of biological and psychosocial function. *JAMA*. 1963;185:914. doi:10.1001/jama.1963.03060120024016
13. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9:179–186. doi:10.1093/geront/9.3_Part_1.179
14. Li H, Li C, Wang A, et al. Associations between social and intellectual activities with cognitive trajectories inChinese middle-aged and older adults: anationa lly representative cohort study. *Alzheimers Res Ther*. 2020;12(1):115. doi:10.1186/s13195-020-00691-6
15. Chen S, Ling Y, Zhou F, et al. Trajectories of cognitive function among people aged 45 years and older living with diabetes in China: results from a nationally representative longitudinal study (2011–2018). *PLoS One*. 2024;19(5):e0299316. doi:10.1371/journal.pone.0299316
16. Herle M, Micali N, Abdulkadir M, et al. Identifying typical trajectories in longitudinal data: modelling strategies and interpretations. *Eur J Epidemiol*. 2020;35(3):205–222. doi:10.1007/s10654-020-00615-6
17. Norman K, Stobäus N, Gonzalez MC, et al. Hand grip strength: outcome predictor and marker of nutritional status. *Clin Nutr*. 2011;30(2):135–142. doi:10.1016/j.clnu.2010.09.010
18. Leong DP, Teo KK, Rangarajan S, et al. Prospective Urban Rural Epidemiology (PURE) study investigators. Prognostic value of grip strength: findings from the Prospective Urban Rural Epidemiology (PURE) study. *Lancet*. 2015;386(9990):266–273. doi:10.1016/S0140-6736(14)62000-6
19. Celis-Morales CA, Welsh P, Lyall DM, et al. Associations of grip strength with cardiovascular, respiratory, and cancer outcomes and all cause mortality: prospective cohort study of half a million UK Biobank participants. *BMJ*. 2018;361:k1651. doi:10.1136/bmj.k1651
20. Raherison C, Girodet PO. Epidemiology of COPD. *Eur Respir Rev*. 2009;18(114):213–221. doi:10.1183/09059180.00003609
21. López-Campos JL, Tan W, Soriano JB. Global burden of COPD. *Respirology*. 2016;21(1):14–23. doi:10.1111/resp.12660
22. Fang L, Gao P, Bao H, et al. Chronic obstructive pulmonary disease in China: a nationwide prevalence study. *Lancet Respir Med*. 2018;6(6):421–430. doi:10.1016/S2213-2600(18)30103-6
23. Sloth MMB, Neble Larsen E, Godtfredsen NS, et al. Impact of offspring and their educational level on readmission and death among older adults with chronic obstructive pulmonary disease: a nationwide cohort study using multistate survival models. *J Epidemiol Community Health*. 2023;77(9):558–564. doi:10.1136/jech-2022-220243
24. Chowdhury NU, Guntur VP, Newcomb DC, Wechsler ME. Sex and gender in asthma. *Eur Respir Rev*. 2021;30(162):210067. doi:10.1183/16000617.0067-2021
25. Scicluna V, Han M. COPD in women: future challenges. *Arch Bronconeumol*. 2023;59(1):3–4. doi:10.1016/j.arbres.2022.06.008
26. Tho NV, Park HY, Nakano Y. Asthma-COPD overlap syndrome (ACOS): a diagnostic challenge. *Respirology*. 2016;21(3):410–418. doi:10.1111/resp.12653
27. Alshabanat A, Zafari Z, Albanyan O, et al. Asthma and COPD Overlap Syndrome (ACOS): a systematic review and meta analysis. *PLoS One*. 2015;10(9):e0136065. doi:10.1371/journal.pone.0136065
28. Spruit MA, Pitta F, McAuley E, et al. Pulmonary rehabilitation and physical activity in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2015;192(8):924–933. doi:10.1164/rccm.201505-0929CI
29. Neunhäuserer D, Reich B, Mayr B, et al. Impact of exercise training and supplemental oxygen on submaximal exercise performance in patients with COPD. *Scand J Med Sci Sports*. 2021;31(3):710–719. doi:10.1111/sms.13870