History- and symptom-based prediction of pulmonary function and the presence of chronic obstructive pulmonary disease: development and validation of a nomogram

Sang Chul Lee  
National Health Insurance Service Ilsan Hospital

Chansik An  (✉ chansikan@gmail.com)  
National Health Insurance Service Ilsan Hospital  https://orcid.org/0000-0002-0484-6658

Jongha Yoo  
National Health Insurance Service Ilsan Hospital

Sungho Park  
National Health Insurance Service Ilsan Hospital

Donggyo Shin  
National Health Insurance Service Ilsan Hospital

Chang Hoon Han  
National Health Insurance Service Ilsan Hospital

Research Article

Keywords: Chronic obstructive pulmonary disease, machine learning, spirometry, primary care

DOI: https://doi.org/10.21203/rs.3.rs-105761/v1

License: ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Early suspicion followed by assessing lung function with spirometry could decrease the underdiagnosis of chronic obstructive pulmonary disease (COPD) in primary care. We aimed to develop a nomogram to predict the FEV$_1$/FVC ratio and the presence of COPD.

**Methods:** We retrospectively reviewed data of 4,241 adult patients who underwent spirometry between 2013 and 2019. By linear regression analysis, variables associated with FEV$_1$/FVC were identified in the training cohort (n=2,969). Using the variables as predictors, a nomogram was created to predict the FEV$_1$/FVC ratio and validated in the test cohort (n=1,272).

**Results:** Older age (Odds ratio [OR], 0.858; 95% confidence interval [CI], 0.833-0.885), male sex (OR, 0.149; 95% CI, 0.064–0.348), current or past smoking history (OR, 0.036; 95% CI, 0.015–0.086), the presence of dyspnea (OR, 0.086; 95% CI, 0.027–0.275), and the absence of overweight (OR, 2.445; 95% CI, 1.210–4.942) were significantly associated with the FEV$_1$/FVC ratio. In the final testing, the developed nomogram showed a mean absolute error of 0.822 (95% CI, 0.789–0.854) between the predicted and actual FEV$_1$/FVC ratios. The overall performance was best when FEV$_1$/FVC < 70% was used as a diagnostic criterion for COPD; the sensitivity, specificity, and balanced accuracy were 82.3%, 68.6%, and 75.5%, respectively.

**Conclusion:** The developed nomogram could be used to identify potential patients at risk of COPD who may need further evaluation, especially in the primary care setting where spirometry is not available.

**Background**

Chronic obstructive pulmonary disease (COPD), the third leading cause of mortality worldwide, is a common and preventable disease characterized by progressive airflow obstruction [1, 2]. One of the main challenges in COPD is its frequent underdiagnosis [3]. People with early or undiagnosed COPD have been shown to have significant morbidity from exacerbations many years before their diagnosis, which can burden healthcare costs [4].

One of the most important factors contributing to the delayed diagnosis of COPD is the low use of spirometry in primary care [3, 5, 6]. People with early or undiagnosed COPD are most likely to encounter the healthcare system in the primary care setting. Therefore, earlier diagnosis of COPD in a primary care setting followed by proper management could significantly improve the prognosis of the disease [7]. However, primary care providers do not always have access, time, or adequate training to use spirometry for patients suspected of having COPD [8].

Alternatively, an easy-to-use tool to predict spirometry results or the presence of COPD would be helpful in enhancing COPD screening in primary care. When a patient is predicted to have COPD, the primary care provider would refer the patient for spirometry. Previous studies have reported prediction models for the diagnosis or prognosis of COPD patients [9–12]. However, to the best of our knowledge, none of these
models was based on information available in the primary care setting alone in predicting spirometry results or the presence of COPD.

Therefore, the purpose of this study was to develop and validate a history- and symptom-based nomogram that can be conveniently used to predict a spirometry result—the forced expiratory volume-one second (FEV\textsubscript{1})/forced vital capacity (FVC) ratio—and the presence or absence of COPD.

**Methods**

**Study population**

We searched our electronic medical record database and found 6,322 adult (≥40 years) patients who underwent pulmonary function tests, including spirometry at a single medical institution in South Korea (hereafter, Korea) between January 2012 and December 2019. Of these, 1,703 patients who were already diagnosed with COPD in 2012 were excluded. Thus, patients included in this study were either first diagnosed with COPD between 2013 and 2019 or were not diagnosed with COPD during the study period. When a patient underwent multiple spirometry measurements, the first test result was used to exclude the possible treatment effect on spirometry results. Of these patients, 378 were excluded because smoking history is missing. Other respiratory ailments, such as asthma, bronchiectasis, interstitial lung disease, were not included in this study. The final study cohort was randomly split into train and test cohorts with a ratio of 7:3 while preserving the same proportion of COPD patients (Fig. 1).

**Measurements of lung function and Definition of COPD**

To measure the lung function, spirometry was performed according to the American Thoracic Society/European Respiratory Society (ATS/ERS) standards by trained research assistants [13]. Dry rolling-seal spirometer (Model 2130; SensoMedics, Yorba Linda, CA, USA) was used for all subjects. All spirometry traces were reviewed by a lung function specialist to determine whether they fulfilled the reproducibility and acceptability criteria of the ATS/ERS Task Force.

The normal predictive values for spirometry data were calculated using a reference equation derived from Korea's general population [14]. A fixed criterion of predicted forced expiratory volume in 1 second per forced vital capacity (i.e., FEV\textsubscript{1}/FVC <0.7) was used to diagnose patients with COPD in accordance with the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines [15].

**Variables**
Outcome variables were the FEV$_1$/FVC ratio and the presence or absence of COPD. Predictors for the outcome were age, sex, overweight (defined as body mass index [BMI] > 25 kg/m$^2$), smoking history, symptoms of dyspnea, cough, or sputum, the presence or absence of underlying hypertension, diabetes, congestive heart failure, coronary vascular disease, stroke, or anemia, and the prior use of salbutamol or antibiotics.

Smoking history could be obtained from both our medical records and the national health screening results, as in Korea, all adults over 40 years old are mandated to undergo the biannual national health screening, which contains a questionnaire about smoking habits. However, there were approximately 5 times more missing values in the health screening records than the medical records. Therefore, we mainly used smoking history from the medical records; only when smoking history was missing in the medical records, we used smoking history, if present, from the health screening database instead.

**Statistical analysis and prediction model**

Continuous or categorical variables were compared between the training and test cohorts using t-test or chi-square tests, respectively. Univariable and multivariable linear regression was performed to determine the association between the risk factors and FEV$_1$/FVC ratio and find independent predictors for our prediction model. In the multivariable regression, only variables with a significant association with the FEV$_1$/FVC ratio in the univariable regression were used.

A linear regression model for predicting the FEV$_1$/FVC ratio was fit in the training cohort and validated in the test cohort using mean absolute error (MAE) as an evaluation metric. In addition, the agreement between the predicted and actual FEV$_1$/FVC values was graphically assessed using the Bland-Altman plot. A nomogram to predict FEV$_1$/FVC was created based on the prediction model fitted in the training cohort.

Using predicted FEV$_1$/FVC values as a diagnostic criterion, the area under receiver operating characteristic curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and balanced accuracy were calculated for discriminating between patients with and without COPD.

Our study cohort was imbalanced, with approximately 9 times more patients in the non-COPD group than in the COPD group. In an imbalanced cohort, it is highly likely that predicted outcome values are biased towards the majority group (i.e., non-COPD group or higher FEV$_1$/FVC ratio in this study). Therefore, when training the model, we used the synthetic minority over-sampling technique (SMOTE) algorithm to create synthetic minority class cases to balance the two classes [16].
All analyses were performed using R 3.6.0. The packages used include 'stats (v3.6.0)' for linear regression, pROC (v1.15.3)' for AUC analysis, ‘epiR (v1.0-15)’ for calculating diagnostic performances, ‘DMwR (v0.4.1)’ for SMOTE, ‘rms (v5.1-3.1)’ for drawing a nomogram, and ‘BlandAltmanLeh (v0.3.1)’ for drawing a Bland-Altman plot. Two-sided probability values of <0.05 were considered statistically significant.

Results

Patient characteristics

The final study cohort comprised 4,241 patients (2,204 men and 2,037 women) with a mean age of 67 (range, 39–98) years. The mean or frequency of all the variables was not significantly different between the training and test cohorts (Table 1). The prevalence of COPD was 10.7% (455/4,241) in the study population. The mean FEV$_1$/FVC ratio was 55.3 (standard deviation [SD], 11.3) in the COPD group, and 79.5 (SD, 5.57) in the non-COPD group.

Factors associated with FEV$_1$/FVC

In the multivariable linear regression analysis, older age (Odds ratio [OR], 0.858; 95% confidence interval [CI], 0.833-0.885), male sex (OR, 0.149; 95% CI, 0.064–0.348), current or past smoking history (OR, 0.036; 95% CI, 0.015–0.086), and the presence of dyspnea (OR, 0.086; 95% CI, 0.027–0.275) were significantly associated with decreased FEV$_1$/FVC, while the presence of overweight (OR, 2.445; 95% CI, 1.210–4.942) was with increased FEV$_1$/FVC (Table 2).

Prediction model

The mean difference between the predicted and actual FEV$_1$/FVC values (i.e., MAE) was 8.858 in the training cohort and 8.721 in the test cohort. For FEV$_1$/FVC in the range between 65 and 75, the MAE was 6.324 in the training cohort and 6.490 in the test cohorts.

In the diagnosis of COPD using the predicted FEV$_1$/FVC ratio, the AUC was 0.832 (95% CI, 0.812–0.845) and 0.822 (95% CI, 0.789–0.854) in the training and test cohorts, respectively. The overall performance was best when the criterion of FEV$_1$/FVC < 70 was used to diagnose COPD; the sensitivity, specificity, PPV, NPV, and balanced accuracy were 82.3%, 68.6%, 25.5%, 96.7%, and 75.5%, respectively (Table 3).
The Bland-Altman plot revealed a trend that our model overestimated FEV$_1$/FVC when an actual FEV$_1$/FVC value was less than 65; in this range, many cases were observed above the upper 95% limit of agreement (Fig. 2). Hence, the effective range of the FEV$_1$/FVC ratio predicted by our nomogram was from 65 to 90; a predicted FEV$_1$/FVC value less than 65 or larger than 90 must be interpreted as 'less than 65' or 'larger than 90', respectively, instead of the value itself (Fig. 3).

**Discussion**

In this study, we developed a multivariable model to identify patients who are expected to have decreased pulmonary function and thus is at risk for COPD. In developing this prediction model, we aimed to create an easy-to-use tool that can help primary care providers decide whether to refer patient suspected of having COPD to a facility where spirometry is available. Thus, we examined variables that are obtainable from simple physical examination and history taking for potential predictors. In our study, the five variables associated with airflow limitation (i.e., decreased FEV$_1$/FVC) were older age, male gender, the absence of overweight, the presence of dyspnea, and ever-smoking history.

Old age, male gender, and smoking are well-known risk factors for COPD. Historically, COPD has been considered a disease of elderly male smokers, although evidence suggests that this historical view is slowly changing [17]. The prevalence and mortality of COPD have increased more rapidly in women than in men during the past two decades, attributed to the changing smoking trends during the past 50 years [18]. Hence, reevaluation of risk stratification by gender is warranted in the future.

Tobacco smoking is the most powerful risk factor for COPD. Although the acquisition of accurate and correct information on the actual smoking habits—duration, amount, and type of cigarette—is of utmost importance, the information in electronic medical records is often quite variable depending on the timing of data entry, visit route (i.e., outpatient, emergency room, or general ward), or medical staff who entered the data. Thus, we processed the smoking data as a binary variable: non-smoker and ever (current or past)-smoker.

In this study, the presence of overweight showed a protective effect, which is in line with previous studies. A study with Asian COPD patients reported that COPD patients with a high BMI have a better pulmonary function [19]. In another study, while underweight was associated with poor survival in COPD, there was a protective effect of overweight and obesity on mortality on COPD patients [20].

GOLD guidelines also support the use of multivariable prediction models to assess the prognostic profile and facilitate follow-up of patients, instead of single predictors such as spirometry or history of exacerbations [15]. Since the occurrence and manifestation of COPD is unique to each race and country [21, 22], we believe that our model could screen more undiagnosed COPD patients in Korea. We wish that we could improve our model as more data are obtained in the future and eventually develop a robust, reliable prediction model that can be used nationwide.
This study has some limitations. First, further external validation is needed, because this model was developed with a retrospective study in a single institution. Second, detailed smoking history (the type of smoking, amount, and duration) was not used in our analysis. In addition to the conventional tobacco smoking, various electronic cigarettes using new nicotine delivery technologies have recently gained popularity in public. Although recent national health screening questionnaires are changing to reflect recent smoking behavior, the data used in this study did not contain it.

**Conclusions**

In conclusion, we developed a nomogram to predict an FEV$_1$/FVC ratio and the presence of COPD based on age, gender, weight, the presence of dyspnea, and smoking history. This nomogram could be used conveniently to screen potential high-risk patients, especially in the primary care setting, where spirometry is not available.

**Abbreviations**

COPD, chronic obstructive pulmonary disease; FEV$_1$, forced expiratory volume-one second; FVC, forced vital capacity; ATS/ERS, American Thoracic Society/European Respiratory Society; GOLD, global initiative for chronic obstructive lung disease; BMI, body mass index; MAE, mean absolute error; AUC, area under curve; PPV, positive predictive value; NPV, negative predictive value; OR, odds ratio; CI, confidence interval

**Declarations**

**Ethics approval and consent to participate**

The Institutional Review Board of National Health Insurance Service Ilsan Hospital (NHIMC 2020-06-005) approved this Health Insurance Portability and Accountability Act-compliant retrospective study and waived the informed consent. All methods were performed in accordance with relevant guidelines and regulations.

**Consent for publication**

Not applicable.

**Availability of data and materials**
Due to the institutional policy, data can only be made available to researchers who subject to a non-disclosure agreement, upon reasonable request.

**Competing interests**

The authors declare that they have no conflict of interest.

**Funding**

This work was supported by the Technology Innovation Program (20011642, CDM based algorithm treatment protocol service system development and spread), funded by the Ministry of Trade, Industry & Energy (Korea).

**Authors' contributions**

SCL, CA, and CHH conceived the study. JY and DS obtained and extracted data. CA and SHP cleaned data. SCL and CA analyzed the data and wrote the paper; CA performed the statistical analysis. All authors have taken due care to ensure the integrity of this work, and all authors read and approved the final manuscript. CA and CHH were in charge of the overall direction.

**References**

1. Mirza S, Clay RD, Koslow MA, Scanlon PD. COPD Guidelines: A Review of the 2018 GOLD Report. Mayo Clin Proc. 2018;93:1488–502.

2. Quaderi SA, Hurst JR. The unmet global burden of COPD. Global Heal Epidemiology Genom. 2018;3:e4.

3. López-Campos JL, Tan W, Soriano JB. Global burden of COPD. Respirology. 2016;21:14–23.

4. Soriano JB, Zielinski J, Price D. Screening for and early detection of chronic obstructive pulmonary disease. Lancet. 2009;374:721–32.

5. Hill K, Goldstein RS, Guyatt GH, Blouin M, Tan WC, Davis LL, et al. Prevalence and underdiagnosis of chronic obstructive pulmonary disease among patients at risk in primary care. Can Med Assoc J. 2010;182:673–8.
6. Sandelowsky H, Ställberg B, Nager A, Hasselström J. The prevalence of undiagnosed chronic obstructive pulmonary disease in a primary care population with respiratory tract infections - a case finding study. Bmc Fam Pract. 2011;12:122.

7. Price D, Freeman D, Cleland J, Kaplan A, Cerasoli F. Earlier diagnosis and earlier treatment of COPD in primary care. Prim Care Resp J. 2011;20:15–22.

8. Walters JA, Hansen EC, Johns DP, Blizzard EL, Walters EH, Wood-Baker R. A mixed methods study to compare models of spirometry delivery in primary care for patients at risk of COPD. Thorax. 2008;63:408.

9. Swaminathan S, Qirko K, Smith T, Corcoran E, Wysham NG, Bazaz G, et al. A machine learning approach to triaging patients with chronic obstructive pulmonary disease. Plos One. 2017;12:e0188532.

10. Martinez FJ, Mannino D, Leidy NK, Malley KG, Bacci ED, Barr RG, et al. A New Approach for Identifying Patients with Undiagnosed Chronic Obstructive Pulmonary Disease. Am J Resp Crit Care. 2016;195:748–56.

11. Ma X, Wu Y, Zhang L, Yuan W, Yan L, Fan S, et al. Comparison and development of machine learning tools for the prediction of chronic obstructive pulmonary disease in the Chinese population. J Transl Med. 2020;18:146.

12. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. Bmj. 2019;367:l5358.

13. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. Eur Respir J. 2005;26:319–38.

14. Choi JK, Paek D, Lee JO. Normal Predictive Values of Spirometry in Korean Population. Tuberc Respir Dis. 2004;58:230–42.

15. Singh D, Agusti A, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: the GOLD science committee report 2019. Eur Respir J. 2019;53:1900164.

16. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res. 2002;16:321–57.

17. Boudewijns EA, Babu GR, Salvi S, Sheikh A, Schayck OC van. Chronic obstructive pulmonary disease: a disease of old age? J Glob Health. 2018;8:020306.

18. Aryal S, Diaz-Guzman E, Mannino DM. COPD and gender differences: an update. Transl Res. 2013;162:208–18.

19. Lim JU, Lee JH, Kim JS, Hwang YI, Kim T-H, Lim SY, et al. Comparison of World Health Organization and Asia-Pacific body mass index classifications in COPD patients. Int J Chronic Obstr. 2017;12:2465–75.

20. Spelta F, Pasini AMF, Cazzoletti L, Ferrari M. Body weight and mortality in COPD: focus on the obesity paradox. Eat Weight Disord - Stud Anorexia Bulimia Obes. 2018;23:15–22.
21. Han MK, Curran-Everett D, Dransfield MT, Criner GJ, Zhang L, Murphy JR, et al. Racial Differences in Quality of Life in Patients With COPD. Chest. 2011;140:1169–76.

22. Eisner MD, Blanc PD, Omachi TA, Yelin EH, Sidney S, Katz PP, et al. Socioeconomic status, race and COPD health outcomes. J Epidemiol Commun H. 2011;65:26.

Tables

Table 1. Patient characteristics
| Variables                                      | Training cohort (n = 2,969) | Test cohort (n = 1,272) | P-value | Total cohort (n = 4,241) |
|------------------------------------------------|-----------------------------|-------------------------|---------|--------------------------|
| **Age (years), mean (SD)**                     | 67.3 (12.0)                 | 67 (12.2)               | 0.493   | 67.2 (12.0)              |
| **Sex, male (vs. female)**                     | 1540 (51.8%)                | 664 (52.1%)             | 0.858   | 2204 (51.9%)             |
| **Overweight**                                 | 1147 (38.6%)                | 475 (37.3%)             | 0.449   | 1622 (38.2%)             |
| **Smoking**                                    | 1126 (37.8%)                | 483 (37.9%)             | 0.996   | 1609 (37.9%)             |
| **Spirometry results, mean (SD)**              |                             |                         |         |                          |
| - **FEV₁ (L)**                                 | 2.7 (0.9)                   | 2.7 (0.9)               | 0.986   | 2.7 (1.0)                |
| - **FEV₁, % pred.**                            | 97.8 (25.0)                 | 96.9 (24.2)             | 0.921   | 97.2 (24.7)              |
| - **FVC (L)**                                  | 2.2 (0.7)                   | 2.2 (0.8)               | 0.901   | 2.2 (0.8)                |
| - **FVC, % pred.**                             | 94.8 (24.1)                 | 94.6 (24.2)             | 0.919   | 94.7 (24.2)              |
| - **FEV₁/FVC (%)**                             | 76.9 (9.8)                  | 76.9 (10.0)             | 0.904   | 76.9 (9.9)               |
| - **FEF₂₅₋₇₅ (L)**                             | 1.6 (1.0)                   | 1.6 (1.0)               | 0.953   | 1.6 (1.0)                |
| - **FEF₂₅₋₇₅, % pred.**                        | 56.0 (29.5)                 | 55.6 (31.0)             | 0.930   | 55.8 (30.2)              |
| **Respiratory symptoms**                       |                             |                         |         |                          |
| - **Dyspnea**                                  | 277 (9.3%)                  | 124 (9.7%)              | 0.708   | 401 (9.4%)               |
| - **Cough**                                    | 300 (10.1%)                 | 123 (9.7%)              | 0.710   | 423 (10.0%)              |
| - **Abnormal sputum**                          | 64 (2.2%)                   | 30 (2.4%)               | 0.765   | 94 (2.2%)                |
| **Comorbidities**                              |                             |                         |         |                          |
| - **COPD**                                     | 328 (11%)                   | 127 (10%)               | 0.334   | 455 (10.7%)              |
| - **CHF**                                      | 66 (2.2%)                   | 30 (2.4%)               | 0.872   | 96 (2.3%)                |
| - **Coronary artery diseases**                 | 181 (6.1%)                  | 75 (5.9%)               | 0.860   | 256 (6.0%)               |
| - **Stroke**                                   | 126 (4.2%)                  | 69 (5.4%)               | 0.108   | 195 (4.6%)               |
| - **Hypertension**                             | 1034 (34.8%)                | 431 (33.8%)             | 0.585   | 1465 (34.5%)             |
| - **Diabetes mellitus**                        | 582 (19.6%)                 | 233 (18.3%)             | 0.355   | 815 (19.2%)              |
| - **Anemia**                                   | 84 (2.8%)                   | 34 (2.7%)               | 0.858   | 118 (2.8%)               |
| - **Acid reflux**                              | 266 (8.9%)                  | 100 (7.8%)              | 0.270   | 366 (8.6%)               |
| **Medication history within 1 year**           |                             |                         |         |                          |
FEV$_1$, forced expiratory volume-one second; FVC, forced vital capacity; FEF, forced expiratory flow at 25/50/75 percent of the FVC curve; COPD, chronic obstructive pulmonary disease; CHF, congestive heart failure.

Table 2. Linear regression results for FEV$_1$/FVC ratio in the training set

|                  | Univariable                      | Multivariable                    |
|------------------|----------------------------------|----------------------------------|
|                  | Odds Ratio (95% CI)              | P-value                          | Odds Ratio (95% CI)              | P-value                          |
| Age              | 0.864 (0.839-0.889)              | <0.001                           | 0.858 (0.833-0.885)              | <0.001                           |
| Sex              | 0.021 (0.011-0.043)              | <0.001                           | 0.149 (0.064-0.348)              | <0.001                           |
| Overweight       | 3.518 (1.705-7.256)              | <0.001                           | 2.445 (1.210-4.942)              | 0.013                            |
| Dyspnea          | 0.076 (0.023-0.254)              | <0.001                           | 0.086 (0.027-0.275)              | <0.001                           |
| Cough            | 4.138 (1.218-14.064)             | 0.023                            | 1.492 (0.460-4.836)              | 0.505                            |
| Sputum           | 0.574 (0.051-6.417)              | 0.652                            |                                  |                                  |
| Current or past smoking | 0.017 (0.008-0.034)              | <0.001                           | 0.036 (0.015-0.086)              | <0.001                           |
| Hypertension     | 0.017 (0.008-0.034)              | 0.021                            | 0.856 (0.400-1.834)              | 0.690                            |
| Coronary vascular disease | 0.177 (0.041-0.768)              | 0.021                            | 1.024 (0.243-4.312)              | 0.974                            |
| Congestive heart failure | 0.030 (0.003-0.335)              | 0.004                            |                                  |                                  |
| Stroke           | 0.791 (0.146-4.285)              | 0.786                            |                                  |                                  |
| Diabetes         | 0.945 (0.383-2.335)              | 0.903                            |                                  |                                  |
| Anemia           | 1.784 (0.209-15.207)             | 0.594                            |                                  |                                  |
| Acid reflux      | 1.688 (0.471-6.051)              | 0.124                            |                                  |                                  |
| One or more salbutamol prescription | 0.136 (0.018-1.023)              | 0.053                            |                                  |                                  |
| One or more antibiotics prescription | 0.218 (0.029-1.656)              | 0.141                            |                                  |                                  |
CI, confidence interval; FEV\textsubscript{1}, forced expiratory volume-one second; FVC, forced vital capacity; CHF, congestive heart failure

**Table 3. Model performance in discrimination according to the cutoff predicted FEV\textsubscript{1}/FVC ratio in the test set**

| Cut-off (%) | Sensitivity | Specificity | Precision (PPV) | NPV | Balanced accuracy |
|-------------|-------------|-------------|-----------------|-----|------------------|
| <65.0       | 60.5%       | 83.1%       | 31.9%           | 94.2% | 71.8%           |
|             | (52.2-68.5) | (80.8-85.3) | (26.5-37.7)     | (92.5-95.5) | (66.5-76.9)     |
| <68.0       | 73.5%       | 75.4%       | 28.1%           | 95.6% | 74.4%           |
|             | (65.6-80.4) | (72.8-77.9) | (23.6-32.8)     | (94.9-6.9) | (69.2-79.1)     |
| <70.0       | 82.3%       | 68.6%       | 25.5%           | 96.7% | 75.5%           |
|             | (75.2-88.1) | (65.8-71.3) | (21.7-29.7)     | (95.3-97.9) | (70.5-79.7)     |
| <72.0       | 87.8%       | 60.6%       | 22.6%           | 97.4% | 74.2%           |
|             | (81.3-92.6) | (57.7-63.5) | (19.2-26.2)     | (96-98.5) | (69.5-78)      |
| <75.0       | 95.9%       | 44.9%       | 18.5%           | 98.8% | 70.4%           |
|             | (91.3-98.5) | (420-47.8)  | (15.8-21.5)     | (97.5-99.6) | (66.6-73.2)     |

FEV\textsubscript{1}, forced expiratory volume-one second; FVC, forced vital capacity; PPV, positive predictive value; NPV, negative predictive value.

**Figures**
Patients ≥ 40 years with spirometry results between 2012 and 2019 (n= 6,322)

2,081 were excluded:
- COPD diagnosis before 2013 (n=1,703)
- Smoking history not available (n=378)

4,241 adult patients:
- First diagnosed with COPD between 2013 and 2019 (n=455)
- No COPD diagnosis (n=3,786)

Random split

Training cohort (n=2,969)
- COPD (n=328)
- No COPD (n=2,641)

Test cohort (n=1,272)
- COPD (n=127)
- No COPD (n=1,145)

Figure 1

Flowchart of the study population. COPD, chronic obstructive pulmonary disease.
Figure 2

Plot of differences between predicted and actual FEV1/FVC ratio vs. the mean of the predicted and actual ratios in the test set. The black line represents the mean of the differences, showing the presence of bias, about -4.7, in the predicted FEV1/FVC ratio. The two dotted lines represent the limits of agreement, ±1.96\(\sigma\).
Figure 3

Nomogram predicting FEV1/FVC ratio. The nomogram is used by first giving each variable a score on the 'Points' scale. The scores for all variables are then added to obtain the total score and a vertical line is drawn from the 'Total Points' row to estimate the FEV1/FVC ratio. A predicted FEV1/FVC value <65 or > 90 must be interpreted as 'less than 65' or 'larger than 90', respectively.