Development of a prediction model to identify undiagnosed chronic obstructive pulmonary disease patients in primary care settings in China

Buyu Zhang[1], Dong Sun[1], Hongtao Niu[2,3,4], Fen Dong[3,4], Jun Lyu[1,5,6], Yu Guo[7], Huaidong Du[8,9], Yalin Chen[10], Junshi Chen[11], Weihua Cao[1], Ting Yang[2,3,4], Canqing Yu[1,5], Zhengming Chen[9], Liming Li[1,5], on behalf of the China Kadoorie Biobank Collaborative Group

[1]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China;
[2]Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China–Japan Friendship Hospital, Beijing 100029, China;
[3]National Center for Respiratory Medicine and National Clinical Research Center for Respiratory Diseases, Beijing 100029, China;
[4]Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Beijing 100007, China;
[5]Peking University Center for Public Health and Epidemic Preparedness and Response, Beijing 100191, China;
[6]Key Laboratory of Molecular Cardiovascular Sciences (Peking University), Ministry of Education, Beijing 100191, China;
[7]National Center for Cardiovascular Diseases, Fuwai Hospital, Chinese Academy of Medical Sciences, Beijing 100037, China;
[8]Medical Research Council Population Health Research Unit at the University of Oxford, Oxford OX3 7LF, UK;
[9]Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK;
[10]Maiji Center for Disease Control and Prevention, Tianshui, Gansu 741020, China;
[11]China National Center for Food Safety Risk Assessment, Beijing 100022, China.

Abstract

Background: At present, a large number of chronic obstructive pulmonary disease (COPD) patients are undiagnosed in China. Thus, this study aimed to develop a simple prediction model as a screening tool to identify patients at risk for COPD.

Methods: The study was based on the data of 22,943 subjects aged 30 to 79 years and enrolled in the second resurvey of China Kadoorie Biobank during 2012 and 2013 in China. We stepwisely selected the predictors using logistic regression model. Then we tested the model validity through P–P graph, area under the receiver operating characteristic curve (AUROC), ten-fold cross validation and an external validation in a sample of 3492 individuals from the Enjoying Breathing Program in China.

Results: The final prediction model involved 14 independent variables, including age, sex, location (urban/rural), region, educational background, smoking status, smoking amount (pack-years), years of exposure to air pollution by cooking fuel, family history of COPD, history of tuberculosis, body mass index, shortness of breath, sputum and wheeze. The model showed an area under curve (AUC) of 0.72 (95% confidence interval [CI]: 0.72–0.73) for detecting undiagnosed COPD patients, with the cutoff of predicted probability of COPD=0.22, presenting a sensitivity of 70.13% and a specificity of 62.25%. The AUROC value for screening undiagnosed patients with clinically significant COPD was 0.68 (95% CI: 0.66–0.69). Moreover, the ten-fold cross validation reported an AUC of 0.72 (95% CI: 0.71–0.73), and the external validation presented an AUC of 0.69 (95% CI: 0.68–0.71).

Conclusion: This prediction model can serve as a first-stage screening tool for undiagnosed COPD patients in primary care settings.

Keywords: Chronic obstructive pulmonary disease; Screening; Prediction model; China Kadoorie Biobank

Introduction

Chronic obstructive pulmonary disease (COPD) is a worldwide public health challenge and induces substantial economic, social and healthcare burdens due to its high prevalence and related disability and mortality.[1] In China, COPD currently ranks the third leading cause of both death and loss of disability-adjusted life years,

Access this article online

Quick Response Code:

Website: www.cmj.org

DOI: 10.1097/CM9.0000000000002448

Correspondence to: Canqing Yu, Department of Epidemiology and Biostatistics, School of Public Health, Peking University, 38 Xueyuan road, Haidian district, Beijing 100191, China
E-Mail: yucanqing@pku.edu.cn

Chinese Medical Journal 2023;136(6)

Received: 01-03-2022; **Online:** 27-03-2023 **Edited by:** Peifang Wei and Xiangxiang Pan

accounting for over one million deaths in 2019.[2] The prevalence of COPD in Chinese people aged 40 years or older reached about 14%.[3,4] At the same time, the disease is commonly underdiagnosed in primary care settings,[5] which usually leaves undiagnosed patients with sustained health damage,[6] and poor outcomes.[7,8] According to a nationwide study in China, only 2.2% of patients with COPD reported that they had been diagnosed with COPD. Only 10.7% of patients had been tested by spirometry before the survey.[4] Although spirometry is considered a "gold standard" for COPD diagnosis, it is often underused in primary care settings due to lack of expertise in performing spirometry, high cost and time consumption of spirometry, and low confidence in spirometry interpretation, etc.[9,10] To solve this problem, the National Heart, Lung, and Blood Institute proposed a three-stage algorithm for COPD case-finding strategy, which recommended using a risk factor/symptom questionnaire as the first-stage screening tool.[11] In view of the large proportion of undiagnosed patients with COPD in need of detection,[3,4] a screening tool with favorable ability to identify previously undiagnosed patients would bring many benefits. Hence, it is of great necessity to develop a prediction model as a simple and economical screening tool to help detect patients with COPD in primary care settings. This tool can serve as a filter to select people at risk to receive further spirometry for accurate diagnosis,[10] and therefore, promote early diagnosis that can improve patient outcomes.[12]

To date, a number of questionnaires have been developed for screening COPD patient, such as the Self-Scored COPD Population Screener Questionnaire, the Chronic Obstructive Pulmonary Disease Assessment Questionnaire (COPD-AQ) and the Lung Function Questionnaire,[13] while most questionnaires were designed in Western populations for individuals with specific risk.[14-17] In China, the validation studies of the screening questionnaires, which were developed in Western populations, were mostly conducted in clinical populations from secondary or tertiary care settings, which could not verify the accuracy of the screening tools in the community populations in primary care settings.[18-22] In recent years, some screening tools were established in the Chinese population from clinical settings,[9,22-24] but they may be inappropriate to be applied in community services. And, few studies have developed the screening tool in the Chinese population based on a relatively large sample from community population.[12]

The present study aimed to develop a simple and question-based prediction model based on the community population from the China Kadoorie Biobank (CKB) to screen people at risk for COPD in primary care settings in China.

Methods

Ethics approval

The CKB study was approved by the Institutional Review Boards at the Chinese Center for Disease Control and Prevention (July 8th, 2004, No. 005/2004, Beijing) and Oxford University (Feb 3rd, 2005, No. 025–04, Oxford, UK). The Enjoying Breathing Program was approved by the China–Japan Friendship Hospital (No. 2019–41-k29). All participants provided written informed consent.

Selection and description of participants

The present study used the data from the second resurvey of the CKB study, which have been described in detail previously.[25-27] Briefly, CKB recruited participants (aged 30–79 years) from ten geographically diverse areas across China, which included five urban areas and five rural areas. The baseline survey was conducted from 2004 to 2008 and involved about 500,000 participants. In the second resurvey, approximately 5% of participants (25,239 people) were randomly selected from the ten areas and resurveyed during 2013 and 2014.

In the present study, we excluded the subjects with any of the following criteria: (1) having self-reported previous diagnosis of COPD, bronchitis, or emphysema ($n = 1191$); (2) having self-reported asthma ($n = 266$); (3) having self-reported tuberculosis at present ($n = 28$); and (4) missing the data of critical variables including the ratio of forced expiratory volume in one second (FEV_1) and the forced vital capacity (FVC), smoking status, or cooking fuel type ($n = 1018$). Since the model was established to identify undiagnosed patients, all the previously diagnosed COPD patients were excluded from the sample. The remaining patients in the sample were previously undiagnosed (defined as the COPD patients who were identified by the spirometry in the CKB study but without any self-reported previous diagnosis of COPD, bronchitis, or emphysema). Finally, a total of 2296 subjects were excluded, and 22,943 subjects were left for analysis.

Predictors and outcome assessment

In the CKB study, all enrolled subjects were asked to finish a questionnaire and to receive physical examination including height, weight and spirometry test. The interviewer-administered questionnaire collected information about demographic characteristics, health-related lifestyle and medical history, etc. Moreover, the FEV_1 and the FVC were measured by spirometry, and body mass index (BMI) was calculated according to height and weight.

In this study, the diagnosis criteria of COPD was defined as the FEV_1/FVC ratio less than 0.70. Clinically significant patients were defined as previously undiagnosed patients with an FEV_1 <60% predicted who are symptomatic or at risk for acute exacerbation of COPD.

The primary aim of the development of this questionnaire-based prediction model is to help find undiagnosed COPD patients in primary care settings, and in order to improve the convenience of using questionnaire, we considered the ease of administration when designing the questions and options. Based on literature reviews and the data of CKB study,[4,5,12,13,24-29] we initially selected 20 items as potential predictors to develop the model, including five categories: socio-demographic characteristics, health-related lifestyle, pollution, medical history, and physical

examination and respiratory symptoms. The socio-demographic items involved age, sex, location (urban/rural), region (defined as the study area where the participants were recorded as permanent residents and which could be identified through official residential records; categorized as northwest, north, northeast, central, southwest, south, and east according to administrative geographical division) and education background (primary school and below, middle school and high school, college and above). The variables of health-related lifestyle included smoking status (current smoker or others), smoking amount (pack-years), and alcohol drinking status (never, former, and current drinker). Then, four items about pollution consisted of duration of passive smoking (years of living with smoker, categorized as 0 years, 1–20 years, 21–40 years and ≥41 years), occupational exposure (to gas/dust/fibers for at least six consecutive months), duration of exposure to air pollution by heating fuel (duration of using wood or coal as heating fuel, and categorized as 0–30 years, 31–60 years and ≥61 years), and duration of exposure to air pollution by cooking fuel (duration of using wood or coal as cooking fuel, and categorized as 0–30 years, 31–60 years and ≥61 years). Two variables about medical history included family history of COPD (family history was defined as at least one immediate family member who has been diagnosed with COPD; missing values were recoded as no family history, and the corresponding sensitivity test presented consistent results) and the history of tuberculosis (with previous diagnosis of tuberculosis and without current tuberculosis). Moreover, the items of physical examination and respiratory symptoms involved BMI (<18.5 kg/m^2, 18.5–23.9 kg/m^2, 24.0–27.9 kg/m^2, and ≥28.0 kg/m^2), cough (coughing frequently for over three months during the past 12 months), sputum (expectoration after getting up in the morning in the past 12 months), shortness of breath (becoming short of breath while walking on level ground), slowing down while walking (slowing down due to chest discomfort while walking on level ground) and wheeze (usually wheezing or having chest whistle sound during the past 12 months).

Statistical analysis

Modeling and evaluation

Mean ± standard deviation (for continuous variables with non-normal distribution), median (Q_1, Q_3) (for continuous variables with non-normal distribution) or *n* (%) (for categorial variables) according to COPD categories (COPD and non-COPD) were used to describe the characteristics of socio-demographic status, health-related lifestyle, pollution, medical history, physical examination, and respiratory symptoms. The univariate logistic regression analysis was used to compare the baseline characteristics of COPD and non-COPD participants. Then, the multiple logistic regression with forward stepwise selection was used to determine the predictors in the model according to their effect on predicting COPD. The variables that contributed significantly to the prediction ($P < 0.05$) were retained in the model.

We evaluated the model in a series of graphic and non-graphic methods. The discrimination of the model was estimated in terms of the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, specificity, positive and negative predictive values, percent correctly classified, positive and negative likelihood ratios. In ROC analysis, we first evaluated the ability of the model to distinguish previously undiagnosed COPD patients (as defined above) from healthy individuals. Then, we additionally tested the ability of the model to identify clinically significant COPD patients from the general population. The cutoff value was determined to obtain a relatively high sensitivity to meet the application requirements of finding more undiagnosed patients, rather than using the point with the max Youden's index. Besides, P–P graph that compared the mean predicted and observed risk of COPD by 10th of predicted risk was used to examine whether the predicted risk fitted well with the observed risk of COPD.

Validation

The validity of the model was further tested by ten-fold cross validation and external validation. The external validation was based on the data collected between 2020 and 2021 in the Enjoying Breathing Program,[30] which involved 29 regions randomly selected nationwide. The participants, aged over 40 years old, were recruited when they visited the primary health care institutions. They received pulmonary function test in the program, with their information about socio-demographic factors, health-related behavior, medical history and symptoms being collected.[30] We used the data of the participants who were not previously diagnosed of COPD, chronic bronchitis, emphysema, asthma, and without active lung tuberculosis ($n = 6011$). The subjects with missing data on key variables or outliers of FEV_1/predicted and FEV_1/FVC were excluded, leaving a total of 3494 individuals for final analysis. This prediction model was used to identify COPD patients and clinically significant patients in the validation sample, and the corresponding AUROC values were reported.

Furthermore, we compared this prediction model with a previous study, which provided a COPD Screening Questionnaire (COPD-SQ) among general population,[13] in external validation group through ROC analysis. In order to compare the accuracy of present model and COPD-SQ, we used both tools to identify COPD patients from the validation population, respectively.

The flow chart of the methodology of the present study was illustrated in Supplementary Figure 1, http://links.lww.com/CM9/B290. All analyses were conducted using STATA 15.1 (StataCorp LP, College Station, TX, USA) and two-tailed tests. Statistical significance was set at $P < 0.05$.

Results

Subject characteristics

Among a total of 22,943 subjects, 5443 (23.7%) were identified as COPD using spirometry. Compared to participants without COPD, the COPD patients were more likely to be older, male, from rural areas, current smoker, and with low BMI [Supplementary Table 1, http://links.lww.com/CM9/B290]. Moreover, of the 5443

COPD patients, a total of 1703 (31.3%) subjects were classified as clinically significant patients.

Model development and performance

In the stepwise multiple logistic regression analysis, 14 variables from the preselected 20 potential predictors were statistically significant ($P < 0.05$), including age, gender, location (urban/rural), region, educational background, smoking status, smoking amount (pack-years), years of exposure to air pollution by cooking fuel, family history of COPD, history of tuberculosis, BMI, shortness of breath, sputum, and wheeze. The multivariable-adjusted odds ratio (95% confidence intervals [CIs]) were presented in Table 1, and the coefficients of predictors were presented in Supplementary Table 2, http://links.lww.com/CM9/B290. No collinearity were observed among the selected variables, with the variance inflation factors of the variables ranging from 1.01 to 4.33.

The AUC value of the prediction model to screen undiagnosed COPD patients was 0.72 (95% CI: 0.72–0.73) [Figure 1]. At the selected cutoff of predicted probability of COPD = 0.22, the corresponding sensitivity, specificity, positive and negative predictive values, correct classification ratio, positive and negative likelihood ratios were 70.13%, 62.25%, 36.62%, 87.01%, 64.12%, 1.86, and 0.48, respectively. Then, to test the ability of the model to identify undiagnosed patients with clinically significant COPD in the whole sample population, the corresponding AUC value was 0.68 (95% CI: 0.66–0.69). Besides, for the calibration of the model, P-P graph reported a regression coefficient of 1.01 (95% CI: 0.92–1.11) and a constant of −0.003 (95% CI: −0.030–0.024) [Figure 2].

Validation

In terms of the internal validation, ten-fold cross validation revealed similar resolving ability (AUROC: 0.72, 95% CI: 0.71–0.73). In the external validation sample of 3494 subjects from the Enjoying Breathing Program, 1637 (46.85%) individuals were classified as COPD patients, and 1535 (43.93%) individuals were classified as clinically significant patients. The present

Table 1: The results of multiple logistic regression of the selected predictors in non-COPD and COPD individuals.

Items	Non-COPD ($n = 17,500$)	COPD ($n = 5443$)	OR (95% CI)	*P* values
Age (years)	57.6 ± 9.6	63.7 ± 10.3	1.06 (1.05–1.06)	<0.001
Gender (female)	11,441 (65.4)	2808 (51.6)	0.70 (0.64–0.76)	<0.001
Location (urban)	7993 (45.7)	1891 (34.7)	0.67 (0.60–0.76)	<0.001
Region*				
Southwest	1673 (9.6)	873 (16.0)	1 (reference)	
South	3223 (18.4)	658 (12.1)	0.59 (0.49–0.71)	<0.001
East	5188 (29.7)	1566 (28.8)	0.68 (0.60–0.77)	<0.001
Central	4223 (24.1)	1323 (24.3)	0.60 (0.54–0.67)	<0.001
Northwest	1683 (9.6)	575 (10.6)	0.66 (0.58–0.76)	<0.001
Northeast	1510 (8.6)	448 (8.2)	1.14 (0.94–1.39)	0.182
Education background				
Primary school and below	8392 (48.0)	3494 (64.2)	1 (reference)	
Middle school and high school	7971 (45.5)	1764 (32.4)	0.79 (0.73–0.86)	<0.001
College graduates and above	1137 (6.5)	185 (3.4)	0.57 (0.47–0.68)	<0.001
Current smoker	3112 (17.8)	1628 (29.9)	1.40 (1.26–1.56)	<0.001
Smoking pack-years	22.0 (6.9, 36.6)	28.5 (12.0, 44.8)	1.00 (1.00–1.01)	<0.001
Years of exposure to air pollution by cooking fuel†				
0–30	13,902 (79.5)	3976 (73.1)	1 (reference)	
31–60	3173 (18.1)	1140 (20.9)	1.12 (1.03–1.22)	0.010
≥61	425 (2.4)	327 (6.0)	1.37 (1.17–1.62)	<0.001
Family history of COPD‡	2276 (13.0)	858 (15.8)	1.22 (1.11–1.33)	<0.001
History of tuberculosis	155 (0.9)	97 (1.8)	1.55 (1.18–2.05)	0.002
BMI (kg/m^2)				
<18.5	460 (2.6)	338 (6.2)	1.35 (1.15–1.59)	<0.001
18.5–23.9	7617 (43.5)	2904 (53.3)	1 (reference)	
24.0–27.9	6930 (39.6)	1686 (31.0)	0.71 (0.66–0.77)	<0.001
≥28.0	2493 (14.3)	515 (9.5)	0.60 (0.54–0.67)	<0.001
Shortness of breath§	1442 (8.2)	656 (12.1)	1.40 (1.26–1.56)	<0.001
Sputum‖	836 (4.8)	418 (7.7)	1.28 (1.12–1.47)	<0.001
Wheeze¶	238 (1.4)	124 (2.3)	1.43 (1.12–1.82)	0.004

Data are presented as *n* (%), median (Q_1, Q_3) or mean ± standard deviation. * Southwest: Sichuan; South: Hainan, Guangxi; East: Shandong, Jiangsu, Zhejiang; Central: Hunan, Henan; Northwest: Gansu; Northeast: Heilongjiang. † Using wood or coal as cooking fuel. ‡ At least one immediate family member had COPD. § Feeling short of breath when walking on level ground. ‖ Expectoration after getting up in the morning in the past 12 months. ¶ Wheezing or having chest whistle sound during the past 12 months. The constant was 0.021. BMI: Body mass index; CI: Confidence interval; COPD: Chronic obstructive pulmonary disease; OR: Odds ratio.

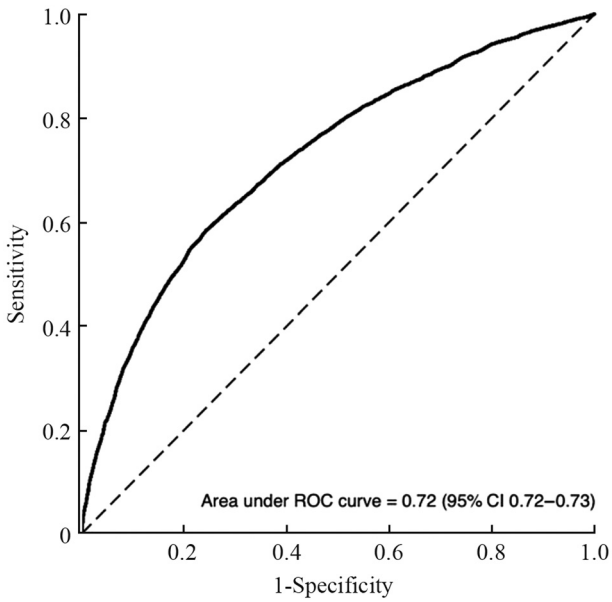

Figure 1: Receiver operating characteristic (ROC) curve for the prediction model to screen undiagnosed COPD patients from community population. CI: Confidence interval; COPD: Chronic obstructive pulmonary disease.

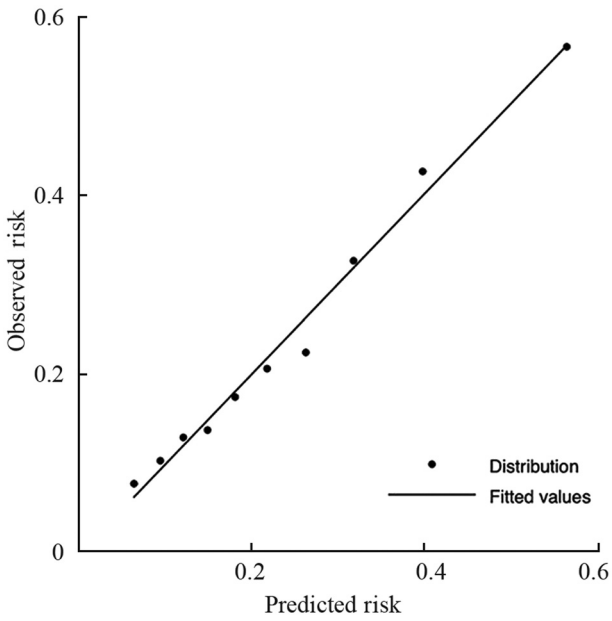

Figure 2: The P–P plot of predicted and observed risk of chronic obstructive pulmonary disease (COPD).

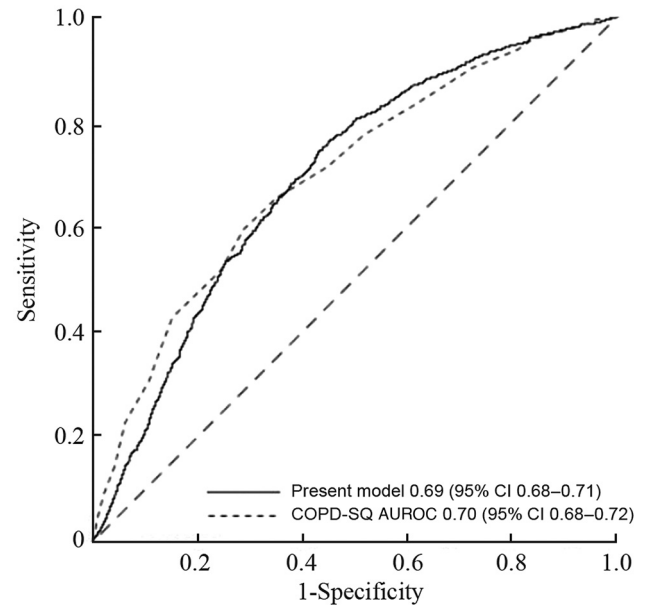

Figure 3: Comparison of receiver operating characteristic (ROC) curve between the present prediction model and the previous COPD-SQ for identifying undiagnosed COPD patients in the validation sample from the Enjoying Breathing Program. AUROC: Area under the receiving operating characteristics; CI: Confidence interval; COPD: Chronic obstructive pulmonary disease; COPD-SQ: COPD screening questionnaire.

prediction model yielded an AUROC value of 0.69 (95% CI: 0.68–0.71) for detecting COPD patients, which was close to the performance in the CKB population, and the sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratio were 55.96%, 71.94%, 63.74%, 64.95%, 1.99, and 0.61, respectively. Besides, the AUROC for identifying clinically significant patients was 0.68 (95% CI: 0.67–0.70). Additionally, the COPD-SQ, for comparison, presented an AUROC of 0.70 (95% CI: 0.68–0.72) for identifying COPD patients [Figure 3], with the sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratio being 71.64%, 55.29%, 61.70%, 65.97%, 1.60, and 0.51, respectively. We tested the difference between the AUCs of the present model and the COPD-SQ, and reported the *P* value of 0.55, which indicated that there was no statistically significant difference between the AUCs of the present model and the COPD-SQ.

Discussion

Key findings

The present study developed a question-based prediction model for screening undiagnosed COPD patients in primary care settings for Chinese population, and it incorporated 14 predictors, including age, sex, location (urban/rural), region, educational background, smoking status, smoking amount (pack-years), years of exposure to air pollution by cooking fuel, family history of COPD, history of tuberculosis, BMI, shortness of breath, sputum, and wheeze. The model presented reasonable validity and reliability in identifying individuals who are likely to have COPD.

Interpretation and implications

According to the course of COPD, delayed diagnosis may result in irreversible impairments of pulmonary function of patients.[31,32] Therefore, the importance of early detection of the disease was emphasized in some consensus and guidelines, such as the Global Initiative for Chronic Obstructive Lung Disease program and a position paper by the American Thoracic Society and the European Respiratory Society.[1,33] Although spirometry test is the gold standard for diagnosing COPD patients, it is usually

insufficiently used in primary care settings. Hence, it seems impractical to promote widespread spirometry testing for early detection without pre-selection of at-risk patients, because of the high cost and the lack of trained personnel.[34-36] Thus, a brief and easy-to-complete screening tool, which is based on patient-reported information, is needed to help screen patients at risk for COPD in primary care settings. This study elaborated the development and initial validation of a simple, reliable, question-based prediction model for COPD that can be easily administrated in primary care settings in community population. Thus, this prediction model can be used for the first-stage screening to detect individuals at risk for COPD, and to prompt their further spirometry assessment for confirmed diagnosis.

Previous studies suggested that a large number of undiagnosed COPD patients needed detection in primary care settings.[11] The most significant therapeutic benefit of treatment for COPD was thought likely to be in symptomatic individuals with an FEV_1 <60% predicted (i.e., clinically significant COPD).[12] Existing evidence has shown that earlier detection of undiagnosed patients with clinically significant COPD in primary care settings could improve short- and long-term outcomes and may acquire great benefit-to-cost ratio.[11] Thus, we tested the ability of the model to identify undiagnosed patients with clinically significant COPD. The acceptable performance of this prediction model in this aspect indicated that it has practical application value for screening the clinically significant patients who are greatly in need of early detection.

A literature, which reviewed 33 COPD screening tools and epidemiologic studies worldwide, reported that the AUCs of the studies conducted in the general population, ranged from 0.65 to 0.79, with sensitivities ranging from 65.8% to 91.7% and specificities from 46.7% to 97.7%.[12] The AUC, sensitivity and specificity of the present model, all within the ranges of previous studies, demonstrated that the validity of the present model is reasonable and acceptable. As reported in previous review, most studies related to COPD case identification adopted a range of factors, including age, gender, smoking status, smoking pack years, BMI, sputum, cough, wheeze, shortness of breath, exposure to pollution, and family history,[12] which were all incorporated in our prediction model. Besides, our model also included other known risk factors, such as education background and history of tuberculosis.[12]

The performance of the present prediction model and the COPD-SQ was not significantly different in the external validation sample. Besides, it should be noted that the grading of respiratory symptoms which was relatively complex and required professional knowledge to understand was included in COPD-SQ as predictors, and might decrease its feasibility for ordinary people in primary care settings. In comparison, the present model incorporated some easily accessible factors as predictors, such as rural/urban areas and the administrative geographical division of China, therefore, it may be not only more convenient for primary care settings, but also more useful for nationwide screening than COPD-SQ.

Strengths and limitations

Compared with previous work in the Chinese population,[9,22-24] some improvements were made in this study, such as the relatively large sample size recruited from community population, the participants randomly selected from ten geographically diverse areas across China, and the external validation in an independent population to further confirm the performance and generalizability of the prediction model.

However, several limitations of this study should be noted. First, the participants of the CKB study did not receive bronchodilators before taking spirometry test, which made it difficult to exclude asthma patients. In the present study, we excluded the participants with self-reported asthma to minimize the misclassification of COPD. Second, the survey of the CKB study did not include some risk factors of COPD, such as childhood lung infection history, which might serve as a predictor in prediction model. In addition, there is one thing needed to be clarify that the prevalence of COPD in our sample (23.7%) was relatively higher than the results of previous nationwide studies (about 14%), which might be the result of the following reasons. (1) The data used in the present study were from ten study areas and could not represent the nationwide prevalence. (2) Our sample included a large proportion of participants from Sichuan province (2546 subjects, 11.1%), where the prevalence of COPD ranks the highest in China.[28] (3) The mean age of our sample (59.0 years) is relatively higher than the previous study that reported a prevalence of 13.6% (54.9 years). (4) Some undiagnosed asthma patients might remain in the sample, which may lead to an overestimated COPD prevalence.

In conclusion, a prediction model was developed as a first-stage screening tool to identify patients with COPD. The model could help detect the large number of undiagnosed patients in primary care settings, and therefore, enhance the early diagnosis of COPD patients. It could also help identify undiagnosed patients with clinically significant COPD who particularly need diagnosis and treatment.

The members of steering committee and collaborative group are listed in the supplementary material.

Funding

This work was supported by the National Key Research & Development Program of China (Nos. 2016YFC1303904 and 2016YFC0900500) and National Natural Science Foundation of China (Nos. 81941018, 91846303, and 91843302). The funders had no role in the study design, data collection, data analysis and interpretation, writing of the report, and the decision to submit the article for publication.

Conflicts of interest

None.

References

1. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease (2022 report), 2021.

Available from: https://goldcopd.org/wp-content/uploads/2021/11/GOLD-REPORT-2022-v1.1-22Nov2021_WMV.pdf. [Accessed on September 6, 2021].
2. Global Health Data Exchange. Global Burden of Disease Study 2019, 2019. Available from: https://vizhub.healthdata.org/gbd-results/. [Accessed on September 6, 2021].
3. Fang LW, Gao P, Bao HL, Tang X, Wang BH, Feng YJ, *et al*. Chronic obstructive pulmonary disease in China: a nationwide prevalence study. Lancet Respir Med 2018;6:421–430. doi: 10.1016/S2213-2600(18)30103-6.
4. Wang C, Xu J, Yang L, Xu Y, Zhang X, Bai C, *et al*. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet 2018;391:1706–1717. doi: 10.1016/S0140-6736(18)30841-9.
5. Soriano JB, Zielinski J, Price D. Screening for and early detection of chronic obstructive pulmonary disease. Lancet 2009;374:721–732. doi: 10.1016/S0140-6736(09)61290-3.
6. 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 Respir Crit Care Med 2017;195:748–756. doi: 10.1164/rccm.201603-0622OC.
7. Martinez CH, Mannino DM, Jaimes FA, Curtis JL, Han MLK, Hansel NN, *et al*. Undiagnosed obstructive lung disease in the United States associated factors and long-term mortality. Ann Am Thorac Soc 2015;12:1788–1795. doi: 10.1513/AnnalsATS.201506-388OC.
8. Labonte LE, Tan WC, Li PZ, Mancino P, Aaron SD, Benedetti A, *et al*. Undiagnosed chronic obstructive pulmonary disease contributes to the burden of health care use data from the CanCOLD study. Am J Respir Crit Care Med 2016;194:285–298. doi: 10.1164/rccm.201509-1795OC.
9. Cui JY, Zhou YM, Tian J, Wang XW, Zheng JP, Zhong NS, *et al*. A discriminant function model as an alternative method to spirometry for COPD screening in primary care settings in China. J Thorac Dis 2012;4:594–600. doi: 10.3978/j.issn.2072-1439.2012.11.06.
10. Stanley AJ, Hasan I, Crockett AJ, van Schayck OCP, Zwar NA. Validation of the COPD diagnostic questionnaire in an Australian general practice cohort: a cross-sectional study. Prim Care Respir J 2014;23:92–97. doi: 10.4104/pcrj.2014.00015.
11. National Heart, Lung, and Blood Institute. A Case-finding Strategy for Moderate-to-Severe COPD in the United States, 2009. Available from: https://www.nhlbi.nih.gov/events/2008/case-finding-strategy-moderate-severe-copd-united-states. [Accessed on September 6, 2021].
12. Han M, Steenrod A, Bacci E, Leidy N, Mannino D, Thomashow B, *et al*. Identifying patients with undiagnosed COPD in primary care settings: insight from screening tools and epidemiologic studies. Chronic Obstr Pulm Dis 2014;2:103–121. doi: 10.15326/jcopdf.2.2.2014.0152.
13. Zhou YM, Chen SY, Tian J, Cui JY, Li XC, Hong W, *et al*. Development and validation of a chronic obstructive pulmonary disease screening questionnaire in China. Int J Tuberc Lung Dis 2013;17:1645–1651. doi: 10.5588/ijtld.12.0995.
14. Yawn BP, Mapel DW, Mannino DM, Martinez FJ, Donohue JF, Hanania NA, *et al*. Development of the Lung Function Questionnaire (LFQ) to identify airflow obstruction. Int J Chronic Obstr 2010;5:1–10. doi: 10.2147/COPD.S7683.
15. Bailey WC, Sciurba FC, Hanania NA, Donohue JF, Ferguson GT, Zibrak JD, *et al*. Development and validation of the Chronic Obstructive Pulmonary Disease Assessment Questionnaire (COPD-AQ). Prim Care Respir J 2009;18:198–207. doi: 10.4104/pcrj.2009.00032.
16. Calverley P, Nordyke R, Halbert RJ, Isonaka S, Nonikov D. Development of a Population-Based Screening Questionnaire for COPD. COPD 2005;2:225–232. doi: 10.1081/COPD-57594.
17. Price DB, Tinkelman DG, Halbert RJ, Nordyke RJ, Isonaka S, Nonikov D, *et al*. Symptom-based questionnaire for identifying COPD in smokers. Respiration 2006;73:285–295. doi: 10.1159/000090142.
18. Pan ZH, Dickens AP, Chi CH, Kong X, Enocson A, G Cooper B, *et al*. Accuracy and cost-effectiveness of different screening strategies for identifying undiagnosed COPD among primary care patients (=40 years) in China: a cross-sectional screening test accuracy study: findings from the Breathe Well group. BMJ Open 2021;11:e051811. doi: 10.1136/bmjopen-2021-051811.
19. Zhou JW, Yu N, Li XM, Wang W. Accuracy of six chronic obstructive pulmonary disease screening questionnaires in the Chinese population. Int J Chronic Obstr 2022;17:317–327. doi: 10.2147/COPD.S341648.
20. Gu YH, Zhang Y, Wen Q, Ouyang Y, Shen YC, Yu H, *et al*. Performance of COPD population screener questionnaire in COPD screening: a validation study and meta-analysis. Ann Med 2021;53:1198–1206. doi: 10.1080/07853890.2021.1949486.
21. Xu XT, Wu FQ, Xu Q. Examination and evaluation of COPD risk seven score scales in early COPD screening. J Mod Med Health 2018;34:3152–3155. doi: 10.3969/j.issn.1009-5519.2018.20.015.
22. Shi Z, Wang SL, Wang HY, Liu GQ. Survey and evaluation of chronic obstructive pulmonary disease screening questionnaire suitable to the Chinese BMI classification (in Chinese). Chin J Gen Pract 2014;13:184–187. doi: 10.3760/cma.j.issn.1671-7368.2014.03.009.
23. Su KC, Ko HK, Chou KT, Hsiao YH, Su VYF, Perng DW, *et al*. An accurate prediction model to identify undiagnosed at-risk patients with COPD: a cross-sectional case-finding study. NPJ Prim Care Respir Med 2019;29:22. doi: 10.1038/s41533-019-0135-9.
24. Ren LP, Zhang Y, Shi P, Wu B, Guo XJ. Preliminary study of symptom-based questionnaire for identifying COPD in China. Int J Respi 2018;38:1766–1770. doi: 10.3760/cma.j.issn.1673-436X.2018.23.002.
25. Li LM, Lv J, Guo Y, Collins R, Chen JS, Peto R, *et al*. The China Kadoorie Biobank: related methodology and baseline characteristics of the participants (in Chinese). Chin J Epidemiol 2012;33:249–255. doi: 10.3760/cma.j.issn.0254-6450.2012.03.001.
26. Chen ZM, Li LM, Chen JS, Collins R, Wu F, Guo Y, *et al*. Cohort profile: the Kadoorie Study of Chronic Disease in China (KSCDC). Int J Epidemiol 2005;34:1243–1249. doi: 10.1093/ije/dyi174.
27. Chen ZM, Chen JS, Collins R, Guo Y, Peto R, Wu F, *et al*. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol 2011;40:1652–1666. doi: 10.1093/ije/dyr120.
28. Wang N, Cong S, Fan J, Bao HL, Wang BH, Yang T, *et al*. Geographical disparity and associated factors of copd prevalence in china: a spatial analysis of national cross-sectional study. Int J Chronic Obstr 2020;15:367–377. doi: 10.2147/COPD.S234042.
29. Yu K, Lv J, Qiu GK, Yu CQ, Guo Y, Bian Z, *et al*. Cooking fuels and risk of all-cause and cardiopulmonary mortality in urban China: a prospective cohort study. Lancet Glob Health 2020;8:430–439. doi: 10.1016/S2214-109X(19)30525-X.
30. Jia CB, Zhang CY, Fang F, Huang K, Dong F, Gu XY, *et al*. Enjoying breathing program: a national prospective study protocol to improve chronic obstructive pulmonary disease management in Chinese primary health care. Int J Chronic Obstr 2020;15:2179–2187. doi: 10.2147/COPD.S258479.
31. Miravitlles M, Soriano JB, García-Río F, Muñoz L, Duran-Tauleria E, Sanchez G, *et al*. Prevalence of COPD in Spain: Impact of undiagnosed COPD on quality of life and daily life activities. Thorax 2009;64:863–868. doi: 10.1136/thx.2009.115725.
32. Rennard SI, Drummond MB. Early chronic obstructive pulmonary disease: definition, assessment, and prevention. Lancet 2015;385:1778–1788. doi: 10.1016/S0140-6736(15)60647-X.
33. Celli BR, MacNee W, Agusti A, Anzueto A, Berg B, Buist AS, *et al*. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J 2004;23:932–946. doi: 10.1183/09031936.04.00014304.
34. Han MLK, Min GK, Mardon R, Renner P, Sullivan S, Diette GB, *et al*. Spirometry utilization for COPD: how do we measure up? Chest 2007;132:403–409. doi: 10.1378/chest.06-2846.
35. Ferguson GT, Enright PL, Buist AS, Higgins MW. Office spirometry for lung health assessment in adults: a consensus statement from the national lung health education program. Chest 2000;117:513–530. doi: 10.1378/chest.117.4.1146.
36. Lee TA, Bartle B, Weiss KB. Spirometry use in clinical practice following diagnosis of COPD. Chest 2006;129:1509–1515. doi: 10.1378/chest.129.6.1509.

How to cite this article: Zhang B, Sun D, Niu H, Dong F, Lv J, Guo Y, Du H, Chen Y, Chen J, Cao W, Yang T, Yu C, Chen Z, Li L. Development of a prediction model to identify undiagnosed chronic obstructive pulmonary disease patients in primary care settings in China. Chin Med J 2023;136:676–682. doi: 10.1097/CM9.0000000000002448