



Original Article

Validity of an Artificial Neural Network in the Diagnosis of COPD

Alirio Bastidas Goyes¹, *Diana Diaz Quijano¹, Marcela Peralta Forero², Jose Ricardo Cardenas Acosta², Eduardo Tuta Quintero³, Lina Morales Cely³, Lina Martinez³, David Acosta³, Lina Fajardo Latorre³, Christian Labrador López³, Daniel Botero Rosas³

¹Epidemiology Department, Universidad de La Sabana, Chía, Colombia

²Internal Medicine Resident, Universidad de La Sabana, Chía, Colombia

³School of Medicine, Universidad de La Sabana, Chía, Colombia

ABSTRACT

Background/Purpose: Neural networks analyze a large amount of information and are useful in the classification of patients for the diagnosis of chronic obstructive pulmonary disease (COPD). However, its comparative performance with questionnaires for the diagnosis of COPD is unknown. The objective of the study is to evaluate the performance of a neural network against clinical questionnaires in the diagnosis of COPD.

Methods: A cross-sectional study was carried out applying the clinical questionnaires and a perceptron neural network against the spirometric diagnosis of COPD.

Results: A total of 1590 patients were admitted to the study, 13.5% of them were confirmed for COPD diagnosis. In the general population, average age was 67.6 years ($SD = 14.0$), and smoking history was 47.7% (758/1590). The questionnaire with the highest performance was the Could it be COPD with an ACOR of 0.83 (95% CI, 0.81–0.86) ($p < 0.001$), and the lowest performance was the LFQ with an ACOR of 0.66. (95% CI, 0.62–0.70)($p < 0.001$). The ANNs showed an ACOR of 0.89 (95% CI, 0.86–0.91) ($p < 0.001$).

Conclusion: Neural networks show a better diagnostic performance than the usual clinical questionnaires for the diagnosis of COPD.

ISSN 2663-8851/Copyright © 2023, Asian Association for Frailty and Sarcopenia and Taiwan Association for Integrated Care. Published by Full Universe Integrated Marketing Limited.

*Correspondence

Dr. Diana Diaz Quijano
 Epidemiology Department,
 Universidad de La Sabana,
 Chía, Colombia
 E-mail:
 diana.diaz1@unisabana.edu.
 co

Received 24 October 2022

Accepted 10 May 2023

Keywords

COPD, diagnosis,
 questionnaires, neural
 networks.

1. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) represents a worldwide health problem, around 380 million people have the disease, and it is the third cause of death after cardiovascular and neoplastic diseases.¹ The underdiagnosis of COPD ranges from 2.5% to 80% due to various risk factors such as ignorance of the disease on the part of physicians, ignorance of the epidemiology and risk factors involved, misinterpretation of the

clinical characteristics and the results of tests, such as spirometry.^{2,3} Underdiagnosis confers a worse prognosis and leads to a delay in starting adequate treatment.³

The artificial neural networks (ANNs) are machine learning subsets that estimate the probability to obtain certain unknown variables based on previously known values, mimicking the functions of the human brain.⁴⁻⁶ ANNs have been applied for the evaluation and follow-up of different pathologies, including

COPD, improving the precision in the processes of timely diagnosis and prognosis^{4,5}; however, it has not been shown to reduce healthcare costs or modify COPD risk factors.^{6,7}

Despite the use of ANNs in the diagnosis of chronic respiratory diseases, their performance compared to clinical questionnaires for the diagnostic approach to COPD is unknown.⁴⁻⁶ The main objective of this study is to evaluate the performance of validated clinical questionnaires against a neural network for the diagnosis of COPD.

2. METHODS

A cross-sectional study was conducted with diagnostic test analysis in patients attending to the outpatient care of a tertiary clinic in Colombia, between June 1st 2014 to December 30th 2020. The initial evaluation was made by applying the clinical questionnaires "Lung Function Questionnaire (LFQ), COPD Diagnostic Questionnaire (CDQ), COPD Population Screener (COPD PS), the questionnaire PUMA (prevalencia y práctica habitual diagnóstico y tratamiento en población de riesgo de EPOC en Médicos generalistas de 4 países de América Latina), and "Could it be COPD", which are currently validated for COPD screening.

2.1. Study Patients

Eligible patients were ≥ 40 years of age who authorized their participation in the study and underwent pre- and post-bronchodilator spirometry immediately after answering the questionnaires for COPD screening. People who did not complete the questionnaires, who were hemodynamically unstable (acute myocardial infarction, recent pulmonary embolism, unstable angina, thoracic or abdominal aneurysm) and with respiratory tract infection in the last 30 days were excluded from the study.

2.2. Instruments

The LFQ questionnaire is made up of five questions that address the frequency of cough, expectoration, wheezing, dyspnea, and years of exposure to smoking tobacco, with a total score ranging from 25 to 5 points, qualifying a patient with a higher risk of airflow obstruction the lower the score obtained, and with a cutoff at ≤ 18 points. The CDQ questionnaire consists of eight questions related to the patient's age, the total number of years of smoking, body mass index, modification of cough with the weather, presence or absence of cough and expectoration with a cold, wheezing and allergies, with a total score from 0 to 38 points, which according to the score classifies it in 3 categories of probability for COPD, low (< 16.5), medium (16.6–19.5), and high (> 19.5).⁸⁻¹⁰

The COPD-PS questionnaire includes five questions identifying shortness of breath during the last 4 weeks, if they ever cough up something, e.g., mucus or phlegm, reduction in daily activities in the last year due to respiratory problems, if they have smoked at least 100 cigarettes in their life and finally the age in years, with a total score between 0 and 10 points, qualifying a patient with risk of COPD with a score ≥ 4 . The PUMA questionnaire, with seven questions of gender, age, number of pack-year, presence or absence of dyspnea, expectoration, chronic cough and whether spirometry was previously performed, with a total score between 0 to 9 in which more than 4 points qualifies a patient at risk of COPD.⁸⁻¹⁰ The Could it be COPD questionnaire consists of 5 questions that establish the presence of dyspnea, cough with expectoration, limitation in daily life due to dyspnea, history of smoking, and age, with a total score of 0 to 10 points, score > 5 qualifying a patient at risk of COPD.⁸⁻¹⁰

Finally, a perceptron-type neural network with supervised learning was built, whose output variable was a binary response of the presence or absence of COPD. The input information was the clinical characteristics, through 13 entries, with a hidden layer of cinco neurons and a hyperbolic tangent as activation function. The neural network was trained with 70% of the population and validated with the remaining 30%.

2.3. Variables

The outcome of interest was the diagnosis of COPD, defined as airflow obstruction with a post-bronchodilator spirometry ratio of forced expiratory volume in 1 second/forced vital capacity (FEV1/FVC) < 0.7 .⁷ Other variables of interest were the results of spirometry of FEV1, FVC, patient demographics such as age, medical history of respiratory disease, smoking, exposure to wood smoke, history of allergy. In addition, clinical data included in the medical questionnaires.

To minimize biases, the spirometry was performed by trained expert personnel, the equipment was calibrated every day prior to the beginning of each session. Data collection was carried out by people from the research group, previously trained on applied questionnaires, a double check was performed when transcribing the data to the electronic database to reduce the risk of transcription bias.

2.4. Sample Size

The sample size was calculated for a comparison of diagnostic tests, considering a sensitivity for the diagnosis of COPD of 67.3% to 79.8% and a specificity of 51.7% to 66.3%, which were reported in the score validation studies.⁷⁻¹⁰ For a prevalence of the

disease of 14.0%¹⁰, a confidence level of 95%, and a power of 80%, 1408 subjects were required; these subjects were included sequentially throughout the duration of the study.

2.5. Data Analysis

The information was collected in the Excel spreadsheet and analyzed in the SPSS program version 20, licensed by the Universidad de La Sabana. Initially a descriptive analysis was carried out summarizing the quantitative variables in mean and standard deviations if their distribution was normal, or median and interquartile range if it was not normal. The qualitative variables were summarized with frequencies and percentages, a bivariate analysis was performed comparing the qualitative variables with Chi-square and the quantitative variables, depending on their distribution with Student's T-test or Mann-Whitney U test.

According to the cut-off point already established for each questionnaire, the sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio (LR+) and negative likelihood ratio (LR-), number needed to diagnose and number needed to misdiagnose with their respective 95% confidence interval (CI) were calculated. Area under the curve of the receiver operating characteristics (AUROC) was obtained with the scores of each questionnaire and classification values of the neural network for the outcome of presence or absence of COPD; a significant *p*-value less than 0.05 was considered.

2.6. Ethical Considerations

This study was performed within the international ethical standards considered in the Declaration of Helsinki and the national regulations of the Resolution 8430 of 1993, considering it a risk-free investigation, the provisions of Law 1266 of 2008 (Habeas Data) and its Decree Regulation 1377 of 2013, keeping the confidentiality of the data. Patients were not involved in the development of the study, and data were analyzed anonymously.

3. RESULTS

A total of 1590 patients were admitted to the study, only 13.5% of them were confirmed for COPD diagnosis. In the general population, average age was

67.6 years (*SD* = 14.0), and smoking history was 47.7% (758/1590). The pack-year index was 21.8 in subjects with COPD v.s. 12.0 in subjects without the disease (*p* < 0.001). Table 1 describes the demographic characteristics and exposure history.

3.1. Lung Function

FVC, FEV1, and FEV1/FVC post bronchodilator were significantly lower in COPD patients. Table 2 describes the lung function results of the population.

3.2. Clinical Questionnaires Diagnostic and Neural Networks Performance

The questionnaire with the highest performance was

Table 1. Demographic characteristics of the selected patients

	Total population <i>n</i> = 1590	Without COPD <i>n</i> = 1375	COPD <i>n</i> = 215	<i>p</i> -value
Sex <i>n</i> (%)				
Male	728 (45.8)	616 (44.8)	112 (52.1)	0.046
Female	862 (54.2)	759 (55.2)	103 (47.9)	
Age <i>x</i> (<i>SD</i>)	67.6 (14.02)	68.2 (13.57)	63.4 (15.99)	0.008
Weight kg <i>x</i> (<i>SD</i>)	79.9 (9.37)	80.5 (8.88)	76.4 (11.44)	0.007
Height m <i>x</i> (<i>SD</i>)	1.6 (0.09)	1.6 (0.09)	1.6 (0.09)	0.487
BMI <i>x</i> (<i>SD</i>)	30 (4.56)	30.2 (4.43)	29 (5.2)	0.040
Level education <i>n</i> (%)				
Elementary	1726 (32.5)	1428 (32.6)	298 (32.2)	0.040
Middle and high	1373 (25.9)	1128 (25.8)	245 (26.5)	
Professional	1245 (23.5)	1033 (23.6)	212 (22.9)	
Technical	470 (8.9)	397 (9.1)	73 (7.9)	
Specialization	268 (5.1)	219 (5)	49 (5.3)	
Technologist	105 (2.0)	87 (2.0)	18 (1.9)	
Master's degree	79 (1.5)	53 (1.2)	26 (2.8)	
None	23 (0.4)	21 (0.5)	2 (0.2)	
Doctorate	14 (0.3)	11 (0.3)	3 (0.3)	
Has smoked cigarette <i>n</i> (%)	758 (47.7)	543 (39.5)	215 (100)	<0.001
Pack-year <i>x</i> (<i>SD</i>)	13.7 (17.39)	12 (15.87)	21.8 (21.79)	<0.001
Wood smoke exposure <i>n</i> (%)	399 (25.1)	241 (17.5)	158 (73.5)	<0.001
Years wood smoke exposure <i>x</i> (<i>SD</i>)	19.4 (10.27)	14.7 (8.97)	26.6 (7.62)	<0.001

Notes: *x*, mean; *SD*, standard deviation; *n*, number; %, percentage; BMI, body mass index.

Table 2. Lung function

Variables	Total population <i>n</i> = 1590	Without COPD <i>n</i> = 1375	COPD <i>n</i> = 215	<i>p</i> -value
FVC PreB L <i>x</i> (<i>SD</i>)	3.5 (0.9)	3.6 (0.89)	3.3 (0.96)	0.014
FVC PostB L <i>x</i> (<i>SD</i>)	3.6 (0.84)	3.7 (0.84)	3.3 (0.8)	0.005
FEV1 PreB L <i>x</i> (<i>SD</i>)	2.6 (0.79)	2.6 (0.79)	2.4 (0.71)	0.017
FEV1 PostB L <i>x</i> (<i>SD</i>)	2.7 (0.75)	2.8 (0.75)	2.5 (0.66)	0.007
FEV1/FVC PreB L <i>x</i> (<i>SD</i>)	73.6 (13.12)	73.7 (13.35)	72.9 (11.5)	0.641
FEV1/FVC PostB L <i>x</i> (<i>SD</i>)	77.9 (10.86)	81.1 (7)	57.6 (9.07)	<0.001

Notes: *x*, mean; *SD*, standard deviation; *n*, number; %, percentage; FVC, forced vital capacity; L, liters; PreB, pre bronchodilator; PostB, post bronchodilator; FEV1, forced expiratory volume in 1 second.

the Could it be COPD with an ACOR of 0.83 (95% CI, 0.81–0.86) ($p < 0.001$), and the lowest performance was the LFQ with an ACOR of 0.66. (95% CI, 0.62–0.70) ($p < 0.001$). The ANNs showed an ACOR of 0.89 (95% CI, 0.86–0.91) ($p < 0.001$). Full results for performance are presented in Table 3.

4. DISCUSSION

In this study we found that the application of a perceptron-type neural network showed the best diagnostic performance for COPD using information on the clinical characteristics of the subjects who attend to medical care. The questionnaire that showed the highest diagnostic performance was Could it be COPD. Smoking and exposure to wood smoke were found to be related to COPD.

ANNs represent an analogy of the neurons of the human brain with the capacity to process and analyze a large amount of information, being useful in multiple tasks in the medical area.^{11,12} Perceptron type ANNs can recognize and identify patterns of information and generate a classification that can be used to classify healthy or diseased subjects. The neural network used in this study contains an input layer (clinical symptoms and signs of the patient), one or more hidden layers where the different mathematical interrelationships for pattern recognition are generated, and a binary response output layer that gives the final classification.¹³ Our results show that the use of the perceptron-type neural network can contribute to the diagnosis of COPD and that it is better than the different questionnaires (COPD-PS, LFQ, CDQ, PUMA, Could it be COPD) when using clinical information from patients with this disease. Freeman et al., described 5663 data of patients with COPD using of a neural network to recognize their clinical characteristics, this quantitative analytical approach allowed a rapid and precise identification of dyspnea, chronic cough, and expectoration as the main manifestations of the disease, generating new preventive and therapeutic approaches in patients with the COPD.¹⁴

The use of ANNs with supervised learning simulates the clinical learning of doctors in training and in this way the network begins to recognize an information

pattern for a correct diagnosis.^{15,16} Recent studies have allowed the development of probabilistic models for diagnosis through ANNs with supervised learning, favorably influencing the classification of the disease and the recognition of its different radiological phenotypes of COPD, achieving an accuracy of 0.96, and a precision of 1.0.¹⁷ The use of these tools can also improve the underdiagnosis of the disease through the identification of signs and symptoms.¹⁸⁻²⁰ Bodduluri, et al., used an artificial intelligence system to identify and classify structural phenotypes of COPD in 8.980 patients undergoing lung function tests and chest computed tomography.²¹ They found that feeding the ANNs with diagnostic imaging showed better results when compared to traditional pulmonary function tests, and may improve early identification of patients at risk, even before the development of substantial airway obstruction and parenchyma remodeling.

Our results show that the questionnaire with the best performance is the Could it be COPD; data supported by Clavery et al., where they found a good discriminatory capacity for predicting airflow obstruction with a sensitivity of 85% and specificity of 45% in the original validation study.²² However, the performance of this questionnaire is between 60 to 75% in AUROC for the diagnosis of COPD.²² In a systematic review, the validity of the use of the different clinical questionnaires was determined, finding a performance for the sum of all the questionnaires with AUROC: 0.75, for the COPD-PS questionnaires an AUROC: 0.750, LFQ with an AUROC: 0.73 and CDQ with an AUROC: 0.727.²² In our study, the higher performance obtained could be influenced by the addition of the process of neural network classification and the amount of information on the clinical characteristics used in this process, a total of 13, which contrasts with the number of questions in each questionnaire, which ranges between 5 and 8 questions for classification.

Smoking is documented as a risk factor for COPD and other respiratory diseases.^{23,24} Wang et al., conducted a study with 471 patients with COPD and 485 controls, describing the main risk factors for COPD using logistic regression analysis. In patients diagnosed with COPD, smoking index >200 (OR,

Table 3. Clinical questionnaires diagnostic and neural networks performance

	Se (CI 95%)	Sp (CI 95%)	PPV (CI 95%)	NPV (CI 95%)	LR+ (CI 95%)	LR- (CI 95%)	AUCOR (CI 95%)	p-value*
COPD-PS ≥ 4	71.6 (69.4–73.8)	56.4 (53.9–58.8)	20.4 (18.4–22.4)	92.7 (91.4–94)	1.64 (1.48–1.82)	0.50 (0.454–0.558)	0.70 (0.66–0.74)	<0.001
LFQ ≤ 18	50.3 (47.9–52.8)	64.7 (62.3–67)	18.2 (16.3–20.1)	89.3 (87.8–90.8)	1.42 (1.224–1.655)	0.77 (0.661–0.893)	0.66 (0.62–0.70)	<0.001
CDQ ≥ 16	76.3 (74.2–78.4)	44.8 (42.4–47.2)	17.8 (15.9–19.6)	92.4 (91–93.7)	1.38 (1.265–1.51)	0.53 (0.485–0.578)	0.70 (0.66–0.74)	<0.001
PUMA ≥ 5	48.8 (46.4–51.3)	70 (67.8–72.3)	20.3 (18.3–22.3)	89.7 (88.3–91.2)	1.63 (1.39–1.911)	0.73 (0.623–0.856)	0.71 (0.68–0.74)	<0.001
Could it be COPD ≥ 3	86.5 (84.8–88.2)	66 (63.6–68.3)	28.4 (26.2–30.7)	96.9 (96.1–97.8)	2.54 (2.322–2.783)	0.20 (0.187–0.224)	0.84 (0.81–0.87)	<0.001
ANNS	88.1 (83.48–92.71)	90 (86.44–93.56)	86 (81.18–90.91)	91.5 (88.18–94.87)	8.81 (6.25–12.42)	0.13 (0.09–0.19)	0.89 (0.86–0.91)	<0.001

Notes: Se, sensibility; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; LR+, likelihood ratio positive; LR-, likelihood ratio negative; AUROC, area under the curve of receiver operating characteristics.
* Statistical difference values to verify the performance of each questionnaire and neural network.

1.42; 95% CI, 1.28–1.57) and a higher frequency of exposure to wood smoke outdoors (OR, 1.64; 95% CI, 1.47–1.83) and indoors (OR, 2.31; 95% CI, 1.92–2.78) were described.²⁵ In our study, smoking with a high pack-year index and exposure to wood smoke presented a significantly higher frequency in the group of sick patients compared to the healthy ones.

Conducting the study in a tertiary care institution entails a clinical spectrum bias, thus being a limitation of the study. In this type of institution, most patients are in advanced stages of the disease, which influences the results of test sensitivity. Being a single center study, the extrapolation of results to larger populations is limited. However, the number of subjects analyzed is adequate to support the results. Studies using ANNs are needed to corroborate their diagnostic validity.

5. CONCLUSION

Perceptron type neural network show a better diagnostic performance than the usual clinical questionnaires for the diagnosis of COPD. However, more studies are needed to validate the diagnostic performance of ANNs in patients with COPD.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this article.

ACKNOWLEDGEMENTS

Universidad de La Sabana, Clínica Universidad de La Sabana.

FUNDING

Universidad de La Sabana, project code (MED-186-2014).

REFERENCES

1. GBD Chronic Respiratory Disease Collaborators. Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet Respir Med*. 2020;**8**(6):585-596. doi:10.1016/S2213-2600(20)30105-3.
2. Halpin DMG, Criner GJ, Papi A, et al. The 2020 GOLD Science Committee Report on COVID-19 and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2021;**203**(1):24-36. doi:10.1164/rccm.202009-3533SO.
3. Elbeddini A, Tayefehchamani Y. Amid COVID-19 pandemic: Challenges with access to care for COPD patients. *Res Social Adm Pharm*. 2021;**17**(1):1934-1937. doi:10.1016/j.sapharm.2020.06.002.
4. Kriegeskorte N, Golan T. Neural network models and deep learning. *Curr Biol*. 2019;**29**(7):R231-R236. doi:10.1016/j.cub.2019.02.034.
5. Remy-Jardin M, Faivre JB, Kaergel R, et al. Machine learning and deep neural network applications in the thorax: Pulmonary embolism, chronic thromboembolic pulmonary hypertension, aorta, and chronic obstructive pulmonary disease. *J Thorac Imaging*. 2020;**35**(Suppl 1):S40-S48. doi: 10.1097/RTI.0000000000000492.
6. Cao F, Yao K, Liang J. Deconvolutional neural network for image super-resolution. *Neural Netw*. 2020;**132**:394-404. doi:10.1016/j.neunet.2020.09.017.
7. Antuni JD, Barnes PJ. Evaluation of individuals at risk for COPD: Beyond the scope of the global initiative for chronic obstructive lung disease. *Chronic Obstr Pulm Dis*. 2016;**3**(3):653-667. doi:10.15326/jcopdf.3.3.2016.0129.
8. Bastidas-Goyes AR, Cardozo-Niño AO, Quintero-Muñoz E, López-Gómez KA, Suárez-Escobar LP, Hernández-Santos LE. Clinical questionnaires for chronic obstructive pulmonary disease diagnosis: A systematic review and meta-analysis. *Rev Fac Med*. 2021;**69**(1):e88706. doi:10.15446/revfacmed.v69n1.88706.
9. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform*. 2014;**48**:193-204. doi:10.1016/j.jbi.2014.02.013.
10. Montes de Oca M, Zabert G, Moreno D, Laucho-Contreras ME, Lopez Varela MV, Surmont F. Smoke, biomass exposure, and COPD risk in the primary care setting: The PUMA Study. *Respir Care*. 2017;**62**(8):1058-1066. doi:10.4187/respcare.05440.
11. Bugajski A, Lengerich A, Koerner R, Szalacha L. Utilizing an artificial neural network to predict self-management in patients with chronic obstructive pulmonary disease: An exploratory analysis. *J Nurs Scholarsh*. 2021;**53**(1):16-24. doi:10.1111/jnu.12618.
12. Moldogazieva NT, Mokhosoev IM, Zavadskiy SP, Terentiev AA. Proteomic profiling and artificial intelligence for hepatocellular carcinoma translational medicine. *Biomedicines*. 2021;**9**(2):159. doi:10.3390/biomedicines9020159.
13. Pan CH, Hsieh HY, Tang KT. An analog multilayer perceptron neural network for a portable electronic nose. *Sensors (Basel)*. 2012;**13**(1):193-207. doi:10.3390/s130100193.
14. Freeman TCB, Rodriguez-Esteban R, Gottowik J, Yang X, Erpenbeck VJ, Leddin M. A neural network approach for understanding patient experiences of chronic obstructive pulmonary disease (COPD): Retrospective, cross-sectional study of social media content. *JMIR Med Inform*. 2021;**9**(11):e26272. doi:10.2196/26272.
15. Hong YJ, Shim J, Lee SM, Im DJ, Hur J. Dual-energy CT for pulmonary embolism: Current and evolving clinical applications. *Korean J Radiol*. 2021;**22**(9):1555-1568. doi:10.3348/kjr.2020.1512.
16. Ju M, Short AD, Thompson P, et al. Annotating and detecting phenotypic information for chronic obstructive pulmonary disease. *JAMIA Open*. 2019;**2**(2):261-271. doi:10.1093/jamiaopen/ooz009.
17. Thomsen LP, Weinreich UM, Karbing DS, et al. Can computed tomography classifications of chronic obstructive pulmonary disease be identified using Bayesian networks and clinical data? *Comput Methods Programs Biomed*. 2013;**110**(3):361-368. doi:10.1016/j.cmpb.2013.02.001.
18. Lin S, Zhang Q, Chen F, Luo L, Chen L, Zhang W. Smooth Bayesian network model for the prediction of future high-cost patients with COPD. *Int J Med Inform*. 2019;**126**(1):147-155. doi:10.1016/j.ijmedinf.2019.03.017.
19. Schiavi E, Stirbulov R, Hernández Vecino R, Mercurio S, Di Boscio V, Puma Team. COPD screening in primary care in four Latin American countries: Methodology of the PUMA Study. *Arch Bronconeumol*. 2014;**50**(11):469-474. doi:10.1016/j.arbres.2014.03.006.
20. Caballero A, Torres-Duque CA, Jaramillo C, et al. Prevalence of COPD in five Colombian cities situated at low, medium, and

- high altitude (PREPOCOL study). *Chest*. 2008;**133**(2):343-349. doi:10.1378/chest.07-1361.
21. Bodduluri S, Nakhmani A, Reinhardt JM, et al. Deep neural network analyses of spirometry for structural phenotyping of chronic obstructive pulmonary disease. *JCI Insight*. 2020;**5**(13):e132781. doi:10.1172/jci.insight.132781.
 22. Calverley PMA, Nordyke RJ, Halbert RJ, Isonaka S, Nonikov D. Development of a population-based screening questionnaire for COPD. *COPD*. 2005;**2**(2):225-232.
 23. García-Quero C, García-Río F. Smoking-induced small airway dysfunction. An early marker of future COPD? *Arch Bronconeumol (Engl Ed)*. 2021;**57**(1):3-4. doi:10.1016/j.arbr.2020.02.006.
 24. Labaki WW, Rosenberg SR. Chronic obstructive pulmonary disease. *Ann Intern Med*. 2020;**173**(3):ITC17-ITC32. doi:10.7326/AITC202008040.
 25. Wang R, Zhang W, Li Y, et al. Evaluation of risk factors for chronic obstructive pulmonary disease in the middle-aged and elderly rural population of Northeast China using logistic regression and principal component analysis. *Risk Manag Healthc Policy*. 2022;**15**:1717-1726. doi:10.2147/RMHP.S376546.