

Original Article
Respiratory Diseases



Word Embedding Reveals Cyfra 21-1 as a Biomarker for Chronic Obstructive Pulmonary Disease

Jeongwon Heo ,^{1,2} Da Hye Moon ,^{1,2} Yoonki Hong ,^{1,2} So Hyeon Bak ,³ Jeeyoung Kim ,^{2,4} Joo Hyun Park ,^{2,5} Byoung-Doo Oh ,⁶ Yu-Seop Kim ,⁶ and Woo Jin Kim ^{1,2,4}

¹Department of Internal Medicine, Kangwon National University Hospital, Chuncheon, Korea

²Department of Internal Medicine, School of Medicine, Kangwon National University, Chuncheon, Korea

³Department of Radiology, School of Medicine, Kangwon National University, Chuncheon, Korea

⁴Environmental Health Center, Kangwon National University Hospital, Chuncheon, Korea

⁵Department of Internal Medicine, Soonchunyang University Bucheon Hospital, Bucheon, Korea

⁶Department of Convergence Software, Hallym University, Chuncheon, Korea



Received: Feb 18, 2021

Accepted: Jul 25, 2021

Address for Correspondence:

Woo Jin Kim, MD, PhD

Department of Internal Medicine,
Kangwon National University Hospital, 156
Baengnyeong-ro, Chuncheon 24341, Republic
of Korea.

E-mail: pulmo2@kangwon.ac.kr

© 2021 The Korean Academy of Medical
Sciences.

This is an Open Access article distributed
under the terms of the Creative Commons
Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>)
which permits unrestricted non-commercial
use, distribution, and reproduction in any
medium, provided the original work is properly
cited.

ORCID iDs

Jeongwon Heo

<https://orcid.org/0000-0001-7694-6730>

Da Hye Moon

<https://orcid.org/0000-0001-5602-9572>

Yoonki Hong

<https://orcid.org/0000-0002-1607-6777>

So Hyeon Bak

<https://orcid.org/0000-0003-1039-7016>

Jeeyoung Kim

<https://orcid.org/0000-0002-0132-6630>

Joo Hyun Park

<https://orcid.org/0000-0002-6382-0642>

Byoung-Doo Oh

<https://orcid.org/0000-0002-9426-6541>

Yu-Seop Kim

<https://orcid.org/0000-0003-1645-4173>

ABSTRACT

Background: Although patients with chronic obstructive pulmonary disease (COPD) experience high morbidity and mortality worldwide, few biomarkers are available for COPD. Here, we analyzed potential biomarkers for the diagnosis of COPD by using word embedding.

Methods: To determine which biomarkers are likely to be associated with COPD, we selected respiratory disease-related biomarkers. Degrees of similarity between the 26 selected biomarkers and COPD were measured by word embedding. And we infer the similarity with COPD through the word embedding model trained in the large-capacity medical corpus, and search for biomarkers with high similarity among them. We used Word2Vec, Canonical Correlation Analysis, and Global Vector for word embedding. We evaluated the associations of selected biomarkers with COPD parameters in a cohort of patients with COPD.

Results: Cytokeratin 19 fragment (Cyfra 21-1) was selected because of its high similarity and its significant correlation with the COPD phenotype. Serum Cyfra 21-1 levels were determined in patients with COPD and controls (4.3 ± 5.9 vs. 3.9 ± 3.6 ng/mL, $P = 0.611$). The emphysema index was significantly correlated with the serum Cyfra 21-1 level (correlation coefficient = 0.219, $P = 0.015$).

Conclusion: Word embedding may be used for the discovery of biomarkers for COPD and Cyfra 21-1 may be used as a biomarker for emphysema. Additional studies are needed to validate Cyfra 21-1 as a biomarker for COPD.

Keywords: Chronic Obstructive Pulmonary Disease; Biomarker; Word Embedding; Cyfra 21-1

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a pulmonary disorder characterized by irreversible airflow limitation, with a partially reversible component.¹ COPD is caused by airway and pulmonary parenchymal damage, which can be prevented and treated. Notably, COPD has a high prevalence and mortality rate worldwide, such that it represents a very large socioeconomic burden.²⁻⁴ The Global Burden of Disease, revised in 2010, reported that 300 million people worldwide have COPD; moreover, 3 million deaths worldwide occur each

Woo Jin Kim <https://orcid.org/0000-0003-2927-370X>**Funding**

This work was supported by an Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No.2019-O-01343, Regional strategic industry convergence security core talent training business)

Disclosure

The authors have no potential conflicts of interest to disclose.

Footnote

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see: <http://www.textcheck.com/certificate/U4pQlx>

Author Contributions

Conceptualization: Heo J, Moon DH, Hong Y, Bak SH, Park JH, Kim YS, Kim WJ. Data curation: Heo J, Kim J, Oh BD, Kim YS. Formal analysis: Heo J, Kim J, Oh BD, Kim YS. Funding acquisition: Kim WJ. Investigation: Heo J, Bak SH, Kim J, Park JH, Oh BD, Kim YS. Methodology: Heo J, Moon DH, Hong Y, Bak SH, Kim J, Park JH, Oh BD, Kim YS, Kim WJ. Software: Oh BD, Kim YS. Supervision: Heo J, Hong Y, Bak SH, Park JH, Kim YS. Validation: Heo J, Kim YS. Visualization: Heo J, Kim YS. Writing - original draft: Heo J, Park JH, Oh BD, Kim YS, Kim WJ. Writing - review & editing: Heo J, Moon DH, Hong Y, Bak SH, Park JH, Kim YS, Kim WJ.

year, due to COPD.^{5,6} The reported prevalence of COPD significantly differs among studies; however, according to research using the Global Initiative for Obstructive Lung Disease guidelines and forced expiratory volume in 1 sec/forced vital capacity (FEV₁/FVC) < 0.7, the overall prevalence is 10%.

A biomarker is a measurable substance or molecule that reflects the state and course of a disease.⁷ Suitable biomarkers can be used to identify patients early and improve the diagnosis of acute or chronic diseases.^{8,9} FEV₁ is relatively easy to determine in patients with COPD; it is also highly repeatable and is used to track health outcomes. However, FEV₁ is difficult to change during treatment, does not reflect disease activity, and is not adequately related to mortality and hospitalization; thus, it is not an ideal biomarker.¹⁰ Accordingly, there is an ongoing effort to identify meaningful biomarkers in patients with COPD. However, no ideal biomarkers have been identified.

Word embedding is a method for learning a vector representation of all words in a given corpus. Inferences are added by measuring the similarity among multiple words and performing vector calculations by vectorizing the semantics. In the present study, we investigated the relationships of COPD with its potential biomarkers by using word embedding models, such as Canonical Correlation Analysis, Word2vec, and Global Vector (GloVe).^{10,11}

A few studies have attempted to identify new biomarkers for COPD from the literature.¹² Word embedding offers the possibility of identifying new biomarkers that may exhibit robust associations with COPD, in addition to known biomarkers. However, word embedding cannot be used to verify the usefulness of a biomarker, because of the lack of clinical trials. In the present study, potential biomarkers were selected using word embedding, then measured in patients with COPD and assessed to determine their correlations with various COPD parameters.

METHODS

Cohort

Patients and controls were selected from the COPD in the Dusty Areas study cohort, which was established to observe the clinical outcomes of patients living near cement plants. The design of this cohort has been detailed in other previous studies.^{13,14} Registered patients were assessed by medical interview, physical examination, spirometry, laboratory testing, and computed tomography. COPD was defined as FEV₁/FVC < 0.7 after bronchodilator aspiration.¹⁵ COPD grade was defined as grade 2 (50% ≤ FEV₁ < 80%), grade 3 (30% ≤ FEV₁ < 50%), or grade 4 (FEV₁ < 30%), according to the 2017 Global Initiative for Obstructive Lung Disease rating system.

Patients and study design

The COPD and control groups were matched for age, sex, and smoking history. In total, 130 patients were enrolled by one-to-one matching in the cohort; they were classified as 65 patients with COPD and 65 controls (**Table 1**). The blood concentrations of biomarkers identified by word embedding were examined. Eight patients in the COPD and control groups were excluded because of insufficient blood samples. Thus, 63 patients with COPD and 59 controls were included in the analysis.

Table 1. Basic characteristics of the participants in this study

Characteristics	Values (n = 130)
Sex	
Male	88 (67.7)
Female	42 (32.3)
Age, yr	71.80 ± 7.02
BMI, kg/m ²	23.81 ± 3.09
Pack-years	26.80 ± 22.12
Smoking	
Current smoker	20 (15.4)
Ex-smoker	52 (40.0)
Never smoker	58 (44.6)
mMRC score	
0	33 (25.4)
1	48 (36.9)
2	18 (13.8)
3	22 (16.9)
4	9 (6.9)
CAT score	17.71 ± 10.07
Post-BD FEV ₁ , L	2.00 ± 0.69
Post-BD FVC, L	3.03 ± 0.83
Post-BD FEV ₁ /FVC, % predicted	65.93 ± 12.25
Emphysema index, %	5.56 ± 6.88
MWA%, %	68.49 ± 5.61
IL-6, ng/mL	2.53 ± 3.12
IL-8, ng/mL	18.14 ± 20.59

Data are shown as number (%) or mean ± standard deviation.

BMI = body mass index, mMRC = modified Medical Research Council, CAT = COPD assessment test, Post-BD = post-bronchodilator, FEV₁ = forced expiratory volume in 1 second, FVC = forced vital capacity, MWA = mean wall area, IL = interleukin.

The following variables were recorded for the included patients: age, sex, height, weight, smoking history, modified Medical Research Council Dyspnea Scale (mMRC) score, COPD Assessment Test (CAT) score, emphysema index (EI), and mean wall area% (MWA%).

Word embedding and Cyfra 21-1 measurements

To determine the degrees of similarity between COPD and its biomarkers, 26 biomarkers known to be associated with respiratory diseases were selected (**Supplementary Table 1**).^{16,17} It was difficult to set the criteria for selecting these biomarkers to find new biomarkers of COPD, so we searched the papers for biomarkers related to other lung diseases and chronic disease. And we randomly selected 26 biomarkers for the word embedding. Information related to the biomarkers was extracted from Google searches. Similarities between several words were measured and vector computations with vectorized semantics were performed to enable additional inferences. Relationships between COPD and respective biomarkers were investigated using the Canonical Correlation Analysis, Word2vec, and GloVe word-embedding models. To replace the clinical evaluation, the titles and abstracts of papers retrieved from Google Scholar were analyzed and quantified to estimate the performance of the word embedding models.¹²

Serum Cyfra 21-1 concentrations were determined using the Cyfra 21-1 enzyme-linked immunosorbent assay kit (Demeditec Diagnostics GmbH, Kiel, Germany), in accordance with the manufacturer's protocol. Enzyme-linked immunosorbent assay plates were analyzed using a microplate reader (Molecular Devices, Sunnyvale, CA, USA).

Quantitative analysis of CT was performed using an Aview® system (Coreline Soft Inc., Seoul, Korea). Emphysema index was calculated as the volume fraction (%) of the lung below -950 HU at full inspiration. MWA% [defined as $WA/(WA + \text{lumen area}) \times 100$], were measured near the origin of the right apical and left apicoposterior segmental bronchi.

Statistical analysis

Statistical analysis was performed using Pearson's χ^2 test and the *t*-test to compare characteristics between the COPD and control groups. Correlation analysis was used to determine the relationships of respective variables with the serum Cyfra 21-1 concentration. Multivariate logistic regression analysis was used to calibrate each variable, including age, sex, smoking, and body mass index (BMI); the correlation between serum Cyfra 21-1 concentration and EI was independently assessed. A *P* value < 0.05 was considered statistically significant. Data analysis was performed using SAS software, version 9.4 (SAS Institute, Cary, NC, USA).

Ethical statement

Written informed consent was given by each participant. The study was reviewed by the Institutional Review Board (IRB) of Kangwon National University Hospital (IRB No. KNUH-2012-06-007).

RESULTS

Table 1 summarizes the characteristics of the 130 participants in this study. The proportion of men in the cohort was twofold greater than the proportion of women (67.7% vs. 32.3%). The mean age was 71.8 years, mean BMI was 23.81 kg/m².

Table 2 shows the similarities between markers with the highest correlation values to COPD in the Word2vec 100 and GloVe 100 dimensions. 'Title_both' is the frequency at which the biomarker (eg, 'CC-16') appears simultaneously with COPD in the title of a medical paper. And 'Abs_both' is the frequency at which the biomarker (e.g., 'CC-16') appears simultaneously with COPD in the abstract of a medical paper. The Clara cell secretory protein (recommended by Word2vec) and adiponectin and leptin (recommended by GloVe) were

Table 2. Biomarkers exhibiting highest similarities with chronic obstructive pulmonary disease

Biomarkers	Similarity	Title_both	Abs_both	Words
Word2vec				
CC-16	0.57279	11	15	1
Eotaxin-1	0.48002	1	3	4
SP-D	0.45937	7	8	2
Cyfra 21-1	0.40179	0	4	8
IP-10	0.35346	4	10	5
Average		4.6	8	4
GloVe				
Adiponectin	0.15436	11	17	1
CEA	0.14091	1	8	1
PAI-1	0.11407	1	5	1
SAA	0.10816	4	7	1
Leptin	0.10062	13	17	1
Average		6	10.8	1

CC-16 = Clara cell protein-16, SP-D = surfactant protein-D, Cyfra 21-1 = cytokeratin 19 fragment, IP-10 = interferon-inducible protein 10, GloVe = Global Vector, CEA = carcinoembryonic antigen, PAI-1 = plasminogen activator inhibitor-1, SAA = serum amyloid A.

involved in large amounts of research related to COPD. In contrast, eotaxin-1 and Cyfra 21-1 (recommended by Word2vec) and carcinoembryonic antigen and serum amyloid A (recommended by GloVe) were not involved in large amounts of research related to COPD. Notably, Cyfra 21-1 had a high correlation with COPD, but has had minimal research interest.

Table 3 shows the demographic and clinical characteristics of the matched COPD and control groups. Significant differences between the two groups were observed in the mMRC score, CAT score, post-bronchodilator (post-BD) FEV₁, EI, and MWA. The mMRC and CAT scores were higher in the COPD group than in the control group; the EI and MWA were also significantly higher in the COPD group. No differences were found in the serum levels of interleukin-6 or interleukin-8. No significant differences were detected in serum levels of Cyfra 21-1 (4.3 ± 5.9 vs. 3.9 ± 3.6 , $P = 0.611$) between the COPD and control groups.

Correlations of demographic and clinical characteristics with serum Cyfra 21-1 concentration were analyzed (**Table 4**). BMI ($P = 0.044$) and EI ($P = 0.015$) were significantly correlated with serum Cyfra 21-1 concentration. Multivariate regression analysis confirmed these correlations: a negative correlation ($B = -10.282$, $P = 0.030$) between Cyfra 21-1 and BMI and a positive correlation ($B = 8.44$, $P = 0.046$) between Cyfra 21-1 and EI (**Supplementary Table 2**).

Table 3. Basic characteristics of participants in COPD and control groups

Characteristics	COPD group (n = 63)	Controls (n = 59)	P value
Sex			0.583
Male	42 (66.7)	43 (72.9)	
Female	21 (33.3)	16 (27.1)	
Age, yr	71.5 ± 7.0	71.6 ± 7.2	0.925
BMI, kg/m ²	24.1 ± 3.2	23.8 ± 3.1	0.591
Pack-years	28.1 ± 27.0	23.7 ± 15.3	0.431
Smoking			0.821
Current smoker	9 (14.3)	10 (16.9)	
Ex-smoker	25 (39.7)	25 (42.4)	
Never smoker	29 (46.0)	24 (40.7)	
mMRC score			0.003
0	8 (12.7)	23 (39.0)	
1	26 (41.3)	19 (32.2)	
2	11 (17.5)	5 (8.5)	
3	10 (15.9)	11 (18.6)	
4	8 (12.7)	1 (1.7)	
Post-BD FEV ₁ , L	1.7 ± 0.6	2.4 ± 0.6	< 0.001
Post-BD FVC, L	3.0 ± 0.8	3.2 ± 0.8	0.173
Post-BD FEV ₁ /FVC, % predicted	57.1 ± 8.9	75.3 ± 4.7	< 0.001
CAT score	19.3 ± 9.8	15.7 ± 10.1	0.049
Emphysema index, %	7.3 ± 8.1	3.6 ± 3.7	0.002
MWA%, %	70.4 ± 5.8	66.8 ± 4.9	< 0.001
IL-6, ng/mL	2.3 ± 2.9	2.9 ± 3.5	0.323
IL-8, ng/mL	17.3 ± 16.0	19.1 ± 25.6	0.647
CRP, mg/dL	0.25 ± 0.51	0.30 ± 0.69	0.636
Neutrophil-platelet ratio	27.45 ± 8.70	24.82 ± 7.76	0.081
Lymphocyte-platelet ratio	13.43 ± 4.87	14.58 ± 6.33	0.265
Eosinophil, ×10 ⁹ /L	0.24 ± 0.24	0.23 ± 0.15	0.729
Cyfra 21-1, ng/mL	4.3 ± 5.9	3.9 ± 3.6	0.611

Data are shown as number (%) or mean ± standard deviation.

P value from Pearson's χ^2 test.

The neutrophil-platelet ratio was calculated by neutrophil count ($\times 10^9/L$) $\times 1,000$ /platelet count ($\times 10^9/L$).

The lymphocyte-platelet ratio was calculated by lymphocyte count ($\times 10^9/L$) $\times 1,000$ /platelet count ($\times 10^9/L$).

COPD = chronic obstructive pulmonary disease, BMI = body mass index, mMRC = modified Medical Research Council, Post-BD = post-bronchodilator, FEV₁ = forced expiratory volume in 1 second, FVC = forced vital capacity, CAT = COPD assessment test, MWA = mean wall area, IL = interleukin, CRP = C-reactive protein, Cyfra 21-1 = cytokeratin 19 fragment.

Table 4. Correlations of cytokeratin 19 fragment with clinical and demographic variables

Variables	Correlation coefficient (P value)
Age, yr	-0.100 (0.271)
BMI, kg/m ²	-0.182 (0.044)
Smoking, pack-years	-0.074 (0.555)
mMRC score	-0.050 (0.581)
CAT score	0.018 (0.840)
Post-BD FVC, liters	-0.016 (0.860)
Post-BD FVC, % predicted	0.013 (0.891)
Post-BD FEV ₁ , liters	-0.044 (0.627)
Post-BD FEV ₁ , % predicted	-0.046 (0.611)
Post-BD FEV ₁ /FVC	-0.055 (0.543)
Emphysema index, %	0.219 (0.015)
Wall area percentage, %	-0.109 (0.230)
IL-6, ng/mL	0.072 (0.431)
IL-8, ng/mL	0.064 (0.485)
CRP, mg/dL	-0.018 (0.847)
Neutrophil-platelet ratio	0.098 (0.284)
Lymphocyte-platelet ratio	-0.019 (0.834)
Eosinophil, ×10 ⁹ /L	0.048 (0.596)

Bold font indicates statistically significant results.

BMI = body mass index, mMRC = modified Medical Research Council, CAT = COPD assessment test, Post-BD = post-bronchodilator, FVC = forced vital capacity, FEV₁ = forced expiratory volume in 1 second, IL = interleukin, CRP = C-reactive protein.

DISCUSSION

In this study, we selected potential biomarkers using word-embedding techniques. We compared the serum concentrations of Cyfra 21-1 between a COPD group and a matched control group in the COPD in the Dusty Areas cohort, with the aim of characterizing Cyfra 21-1 as a potential biomarker for COPD. We found a higher EI and lower BMI in patients with a high serum concentration of Cyfra 21-1. Notably, we found that the serum Cyfra 21-1 concentration was not associated with other COPD clinical indicators, such as dyspnea or pulmonary function.

Word embedding is a method for digitization of a word constituting text. Word2vec was developed by Mikolov, who implemented continuous bag-of-words and Skip-Gram word embedding.¹⁰ Words can be embedded based on their meaning in the context of surrounding words. A sufficient number of documents is collected in the target field; embedding is then performed in a low-dimensional space using a representative technique (i.e., Word2vec). The embedded space is generated by reproduction of the contextual meaning of the corresponding vocabulary; if a particular word is very positive, it can be inferred that words near the corresponding word are also likely to have a positive meaning.¹⁸ Various studies have used word embedding; recent research on cancer and biomarkers has provided a new method to study potential biomarkers.¹⁹

COPD is a disease with high prevalence and mortality rates worldwide.²⁰ Pulmonary function tests is relatively easy to determine from patients with COPD, with repeatability and ease of tracing; however, it is not an ideal biomarker because it does not reflect disease activity and is not associated with acute hospitalization, mortality, or quality of life.²¹ Ideal biomarkers improve patient care by allowing early detection of clinical conditions, thereby improving the diagnosis of acute or chronic diseases; they also enable prediction of patient risk, determination of optimal treatment for a given patient, observance of disease progression, and observance of the response to treatment.²² Many studies have been conducted to identify ideal biomarkers of COPD.²³⁻²⁵

Cyfra 21-1 is a cytokeratin 19 segment that was identified as a marker of lung cancer in 1993. It has been repeatedly validated as a marker for non-small cell lung cancer.^{26,27} The Cyfra 21-1 serum tumor marker was related to the proximity of COPD using the Word2vec technique.¹² In our study, serum Cyfra 21-1 levels were significantly correlated with BMI and EI values. In a study of the clinical utility of the tumor marker Cyfra 21-1 in health assessments, factors affecting Cyfra 21-1 were age, sex, smoking status, blood pressure, diabetes, and hepatitis B infection; no analysis of height or weight was performed in that study.²⁷

To the best of our knowledge, the present study is the first to reveal associations of Cyfra 21-1 with BMI and the EI. Thus, in addition to its role as a marker for cancer, Cyfra 21-1 constitutes a marker for benign pulmonary disease. Several studies have shown increased Cyfra 21-1 levels in patients with interstitial pulmonary fibrosis, collagen disease-associated pulmonary fibrosis, and autoimmune pulmonary alveolar proteinosis.^{22,26,27}

There were some limitations in our study. First, we did not consider additional variables associated with Cyfra 21-1 (e.g., hepatitis B infection, diabetes, and blood pressure), which were identified in previous studies. In particular, serum Cyfra 21-1 levels are known as adjuvants for the diagnosis and prognosis of lung squamous cell carcinoma.²⁸ Additionally, Cyfra 21-1 has recently been reported to be elevated in idiopathic pulmonary fibrosis or head and neck squamous cell carcinoma.²⁹ In our cohort, COPD patients did not have enough patients with underlying diseases to analyze these underlying diseases (**Supplementary Table 3**).³⁰ Second, word embedding study was designed to review other potential biomarkers, in addition to Cyfra21-1. However, only one marker was identified due to research cost limitations. Third, COPD is a heterogeneous disease caused by a combination of small airway disease (chronic bronchitis) and parenchymal destruction (emphysema).³¹ However, there is a limitation that cannot be analyzed because there is no information according to the type of COPD in our cohort. Finally, even in the analysis based on the GOLD group, there were only 2 patients with stage 3 or higher, which could not be sufficient for analysis (**Supplementary Table 4**).

Our study is the first to identify a biomarker for COPD using word embedding. We measured serum Cyfra 21-1 concentrations in the COPD and control groups. Correlations of BMI and EI with Cyfra21-1, which were not identified in previous studies, were confirmed. Active research regarding the use of word embedding in various medical fields is expected to be performed in the future. Cyfra 21-1, BMI, and EI should be assessed in patients at risk of COPD, in addition to the conventional parameters of age, diabetes, cancer, and infection.

In conclusion, Cyfra 21-1 concentration was correlated with EI and BMI. Word embedding is expected to be involved in the discovery of new biomarkers for COPD and other medical conditions.

SUPPLEMENTARY MATERIALS

Supplementary Table 1

Twenty-six selected biomarkers

[Click here to view](#)

Supplementary Table 2

Multivariate stepwise regression analysis of the BMI and emphysema index

[Click here to view](#)**Supplementary Table 3**

Analysis according to comorbidity in COPD and control groups

[Click here to view](#)**Supplementary Table 4**

Characteristics of chronic obstructive pulmonary disease patients by Global Initiative for Chronic Obstructive Lung Disease stage

[Click here to view](#)**REFERENCES**

1. Kim HK, Lee SD. Pathophysiology of chronic obstructive pulmonary disease. *Tuberc Respir Dis (Seoul)* 2005;59(1):5-13.
[CROSSREF](#)
2. Lopez AD, Shibuya K, Rao C, Mathers CD, Hansell AL, Held LS, et al. Chronic obstructive pulmonary disease: current burden and future projections. *Eur Respir J* 2006;27(2):397-412.
[PUBMED](#) | [CROSSREF](#)
3. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3(11):e442.
[PUBMED](#) | [CROSSREF](#)
4. Kim C, Kim Y, Yang DW, Rhee CK, Kim SK, Hwang YI, et al. Direct and indirect costs of chronic obstructive pulmonary disease in Korea. *Tuberc Respir Dis (Seoul)* 2019;82(1):27-34.
[PUBMED](#) | [CROSSREF](#)
5. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2163-96.
[PUBMED](#) | [CROSSREF](#)
6. Abubakar I, Tillmann T, Banerjee A; GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;385(9963):117-71.
[PUBMED](#) | [CROSSREF](#)
7. Cazzola M, MacNee W, Martinez FJ, Rabe KF, Franciosi LG, Barnes PJ, et al. Outcomes for COPD pharmacological trials: from lung function to biomarkers. *Eur Respir J* 2008;31(2):416-69.
[PUBMED](#) | [CROSSREF](#)
8. Morrow DA, de Lemos JA. Benchmarks for the assessment of novel cardiovascular biomarkers. *Circulation* 2007;115(8):949-52.
[PUBMED](#) | [CROSSREF](#)
9. Singh D. Blood eosinophil counts in chronic obstructive pulmonary disease: a biomarker of inhaled corticosteroid effects. *Tuberc Respir Dis (Seoul)* 2020;83(3):185-94.
[PUBMED](#) | [CROSSREF](#)
10. Sin DD. Chronic obstructive pulmonary disease: reactive past, preventive future. *Proc Am Thorac Soc* 2009;6(6):523-523.
[PUBMED](#) | [CROSSREF](#)
11. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv*. Forthcoming 2013.
12. Yoon BH, Kim YS. Correlation analysis of chronic obstructive pulmonary disease (COPD) and its biomarkers using the word embeddings. *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Taipei, Taiwan: IJCNLP; 2017, 337-42.

13. Hahm CR, Lim MN, Kim HY, Hong SH, Han SS, Lee SJ, et al. Implications of the pulmonary artery to ascending aortic ratio in patients with relatively mild chronic obstructive pulmonary disease. *J Thorac Dis* 2016;8(7):1524-31.
[PUBMED](#) | [CROSSREF](#)
14. Koo HK, Hong Y, Lim MN, Yim JJ, Kim WJ. Relationship between plasma matrix metalloproteinase levels, pulmonary function, bronchodilator response, and emphysema severity. *Int J Chron Obstruct Pulmon Dis* 2016;11:1129-37.
[PUBMED](#) | [CROSSREF](#)
15. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. *Am J Respir Crit Care Med* 2017;195(5):557-82.
[PUBMED](#) | [CROSSREF](#)
16. Ahn J, Cho J. Current serum lung cancer biomarkers. *J Mol Biomark Diagn* 2013;4:2.
17. Srikanthan K, Feyh A, Visweshwar H, Shapiro JI, Sodhi K. Systematic review of metabolic syndrome biomarkers: a panel for early detection, management, and risk stratification in the West Virginian population. *Int J Med Sci* 2016;13(1):25-38.
[PUBMED](#) | [CROSSREF](#)
18. Fabbri LM, Hurd S. Global strategy for the diagnosis, management and prevention of COPD: 2003 update. *Eur Respir J* 2003;22(1):1-2.
[PUBMED](#) | [CROSSREF](#)
19. Nakayama M, Satoh H, Ishikawa H, Fujiwara M, Kamma H, Ohtsuka M, et al. Cytokeratin 19 fragment in patients with nonmalignant respiratory diseases. *Chest* 2003;123(6):2001-6.
[PUBMED](#) | [CROSSREF](#)
20. Arai T, Inoue Y, Sugimoto C, Inoue Y, Nakao K, Takeuchi N, et al. CYFRA 21-1 as a disease severity marker for autoimmune pulmonary alveolar proteinosis. *Respirology* 2014;19(2):246-52.
[PUBMED](#) | [CROSSREF](#)
21. Han Y, Heo Y, Hong Y, Kwon SO, Kim WJ. Correlation between physical activity and lung function in dusty areas: results from the chronic obstructive pulmonary disease in dusty areas (CODA) cohort. *Tuberc Respir Dis (Seoul)* 2019;82(4):311-8.
[PUBMED](#) | [CROSSREF](#)
22. Stieber P, Bodenmüller H, Banauch D, Hasholzner U, Dessauer A, Ofenloch-Hähnlé B, et al. Cytokeratin 19 fragments: a new marker for non-small-cell lung cancer. *Clin Biochem* 1993;26(4):301-4.
[PUBMED](#) | [CROSSREF](#)
23. Zemans RL, Jacobson S, Keene J, Kechris K, Miller BE, Tal-Singer R, et al. Multiple biomarkers predict disease severity, progression and mortality in COPD. *Respir Res* 2017;18(1):117.
[PUBMED](#) | [CROSSREF](#)
24. Kim DK, Cho MH, Hersh CP, Lomas DA, Miller BE, Kong X, et al. Genome-wide association analysis of blood biomarkers in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;186(12):1238-47.
[PUBMED](#) | [CROSSREF](#)
25. Brusselle G, Pavord ID, Landis S, Pascoe S, Lettis S, Morjaria N, et al. Blood eosinophil levels as a biomarker in COPD. *Respir Med* 2018;138:21-31.
[PUBMED](#) | [CROSSREF](#)
26. Park SY, Lee JG, Kim J, Park Y, Lee SK, Bae MK, et al. Preoperative serum CYFRA 21-1 level as a prognostic factor in surgically treated adenocarcinoma of lung. *Lung Cancer* 2013;79(2):156-60.
[PUBMED](#) | [CROSSREF](#)
27. Kim J, Jung H, Kim D, Lee S, Kim M, Park K. Lack of clinical utility for CYFRA 21-1 in medical screening. *Korean J Fam Pract* 2018;8(1):73-9.
[CROSSREF](#)
28. Wieskopf B, Demangeat C, Purohit A, Stenger R, Gries P, Kreisman H, et al. Cyfra 21-1 as a biologic marker of non-small cell lung cancer. Evaluation of sensitivity, specificity, and prognostic role. *Chest* 1995;108(1):163-9.
[PUBMED](#) | [CROSSREF](#)
29. Simpson JK, Maher TM, Bentley J, Braybrooke R, Carter P, Costa MJ, et al. CYFRA-21-1 as a biomarker with prognostic potential in idiopathic pulmonary fibrosis: an analysis of the PROFILE cohort. *Am J Respir Crit Care Med* 2017;195:A6791.
30. Joo H, Park J, Lee SD, Oh YM. Comorbidities of chronic obstructive pulmonary disease in Koreans: a population-based study. *J Korean Med Sci* 2012;27(8):901-6.
[PUBMED](#) | [CROSSREF](#)
31. Park TS, Lee JS, Seo JB, Hong Y, Yoo JW, Kang BJ, et al. KOLD Study Group. Study design and outcomes of Korean obstructive lung disease (KOLD) cohort study. *Tuberc Respir Dis (Seoul)* 2014;76(4):169-74.
[PUBMED](#) | [CROSSREF](#)