ORIGINAL RESEARCH

Identification of subtypes in subjects with mild-to-moderate airflow limitation and its clinical and socioeconomic implications

Jin Hwa Lee,[1] Chin Kook Rhee,[2] Kyungjoo Kim,[2] Jee-Ae Kim,[3] Sang Hyun Kim,[4] Kwang Ha Yoo,[5] Woo Jin Kim,[6] Yong Bum Park,[7] Hye Yun Park,[8] Ki-Suck Jung[9]

[1]Division of Pulmonary and Critical Care Medicine, Department of Medicine, School of Medicine, Ewha Womans University, [2]Department of Internal Medicine, Seoul St Mary's Hospital, The Catholic University of Korea, Seoul, [3]Pharmaceutical Policy Evaluation Research Team, Research Institution, Health Insurance Review and Assessment Service, [4]Big Data Division, Health Insurance Review and Assessment Service, Wonju, [5]Department of Internal Medicine, Konkuk University College of Medicine, Seoul, [6]Department of Internal Medicine and Environmental Health Center, Kangwon National University Hospital, Chuncheon, [7]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Internal Medicine, Hallym University Kangdong Sacred Heart Hospital, [8]Division of Pulmonary and Critical Care Medicine, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, [9]Division of Pulmonary, Allergy and Critical Care Medicine, Department of Internal Medicine, Hallym University Medical Center, Hallym University College of Medicine, Anyang, South Korea

Correspondence: Jin Hwa Lee
Division of Pulmonary and Critical Care Medicine, Department of Medicine, School of Medicine, Ewha Womans University, 1071 Anyangcheon-ro Yangcheon-gu, Seoul 07985, South Korea
Tel +82 2 2650 6007
Fax +82 2 2655 2076
Email jinhwalee@ewha.ac.kr

Purpose: The purpose of this study was to identify subtypes in patients with mild-to-moderate airflow limitation and to appreciate their clinical and socioeconomic implications.

Methods: Subjects who were aged $\geq$20 years and had forced expiratory volume in 1 second (FEV_1) $\geq$60% predicted and FEV_1/forced vital capacity <0.7 were selected from the fourth Korea National Health and Nutrition Examination Survey (KNHANES) in 2007–2012. The data were merged to the National Health Insurance reimbursement database during the same period. k-Means clustering was performed to explore subtypes. For clustering analysis, six key input variables – age, body mass index (BMI), FEV_1% predicted, the presence or absence of self-reported wheezing, smoking status, and pack-years of smoking – were selected.

Results: Among a total of 2,140 subjects, five groups were identified through k-means clustering, namely putative "near-normal (n=232)," "asthmatic (n=392)," "chronic obstructive pulmonary disease (COPD) (n=37)," "asthmatic-overlap (n=893)," and "COPD-overlap (n=586)" subtypes. Near-normal group showed the oldest mean age (72±7 years) and highest FEV_1 (102%±8% predicted), and asthmatic group was the youngest (46±9 years). COPD and COPD-overlap groups were male predominant and all current or ex-smokers. While asthmatic group had the lowest prescription rate despite the highest proportion of self-reported wheezing, COPD, asthmatic-overlap, and COPD-overlap groups showed high prescription rate of respiratory medicine. Although COPD group formed only 1.7% of total subjects, they showed the highest mean medical cost and health care utilization, comprising 5.3% of the total medical cost. When calculating a ratio of total medical expense to household income, the mean ratio was highest in the COPD group.

Conclusion: Clinical and epidemiological heterogeneities of subjects with mild-to-moderate airflow limitation and a different level of health care utilization by each subtype are shown. Identification of a subtype with high health care demand could be a priority for effective utilization of limited resources.

Keywords: phenotype, COPD, asthma, overlap, health care utilization, cluster

Introduction

The prevalence and mortality of chronic obstructive pulmonary disease (COPD) are increasing when compared to those of the other chronic diseases.[1] It is a major burden of current and future society. Meanwhile, it is well known that early detection and management are essential to prevent morbidity and mortality of chronic diseases,[2,3] though adults with early hypertension or early diabetes mellitus do not have any symptoms. Nevertheless, early diagnosis of COPD by spirometry in asymptomatic subjects is not recommended in many COPD guidelines, including the Global initiative

for chronic Obstructive Lung Disease (GOLD),[4] because of negative cost-effectiveness studies.[5–8] On the other hand, forced expiratory volume in 1 second (FEV_1) has been the only biomarker of COPD, though it does not perfectly correlate with several outcomes such as exercise tolerance, health-related quality of life (QOL), and mortality.[9] There have been many attempts to find a novel biomarker of COPD[10,11] and/or to make a combination set, such as the body mass index (BMI) (B), the degree of airflow obstruction (O) and dyspnea (D), and exercise capacity (E) (BODE) index.[12] However, any specific COPD biomarkers except FEV_1 are not widely used. Unfortunately, preserved FEV_1 does not guarantee comfortable respiration because some patients with mild COPD experience quite severe exacerbation.[13] Even smokers without spirometry impairments showed worse QOL and lower exercise capacity compared to never smokers.[14] Several studies describe substantial socioeconomic burden of patients with even mild COPD.[15,16] When limited health care resources impede detection and management of early COPD, identification of subgroups with relatively high socioeconomic burden provides a useful and sensible guide to allocation of health care resources. In order to identify a subtype with high demand of health care utilization, a cluster analysis was performed by using the national database, by linking the fourth Korea National Health and Nutrition Examination Survey (KNHANES) in 2007–2012 with the National Health Insurance (NHI) reimbursement database in the same period. For clustering analysis, six major input variables, which were age, smoking status, smoking pack-years, FEV_1% predicted, BMI, and the presence or absence of reported wheezing, were selected. These last three variables are likely to represent the degree of airflow obstruction, degree of systemic inflammation, and presence or absence of airway hyper-responsiveness, respectively.[12,17–21]

Materials and methods
Data and study population selection
This study utilized two data sources of KNHANES and NHI claims.

KNHANES is a national survey with stratified, multistage, and clustered probability design to ensure a representative sample of non-institutionalized citizens from the Korean population. The data collected from the survey assist the establishment of a national health policy and a knowledge base for the evaluation of health risk factors and health indicators.[22] Details on the research methodology used to gather and analyze the KNHANES data can be found on KNHANES website (https://knhanes.cdc.go.kr/knhanes/eng/index.do [last accessed on December 25, 2016]).

The NHI claims data were provided by the Korean Health Insurance Review and Assessment Service (HIRA), which was established to review claims submitted by health care providers and assess health care in Korea; it is a single agency independent from insurers, providers, and other interested parties. All Koreans are covered under the NHI system, and the HIRA database contains records of health care service utilization such as physicians' office visits, emergency room (ER) visits, hospitalization, and prescriptions and payments for both payers and patients.

A total of 43,864 respondents participated in the KNHANES from 2007 through 2012. A total of 20,970 subjects who were aged >19 years and had performed acceptable spirometry were extracted. From the data, 2,172 subjects who had mild-to-moderate airflow limitation, which was defined as having $FEV_1 \geq 60\%$ predicted[23] and a ratio of FEV_1 to forced vital capacity (FVC) <0.7, were identified. For clustering analysis, six key input variables were selected: age, BMI, FEV_1% predicted, the presence or absence of self-reported wheezing, smoking status, and pack-years of smoking. The presence or absence of self-reported wheezing was determined by the response to the question: "Have you had wheezing or whistling in your chest at any time during the previous year?" Subjects with any missing value of six key input variables (Figure 1) were excluded. K-means clustering was performed to explore subtypes. The data were merged to the NHI reimbursement database for the same period.

Age, sex, BMI, spirometry, smoking status, education, and household income level were included. Lower limit of normal (LLN) for FEV_1/FVC was calculated by using the following prediction equations: $125.77628 - 0.36304 \times$ age (years) $-0.17146 \times$ height (cm) for Korean male and $97.36197 - 0.26015 \times$ age (years) $-0.01861 \times$ height (cm) for Korean female.[24] Subjects' smoking status was divided into current, former, and never smokers. Current smokers are those who have smoked >100 cigarettes in their lifetime and have smoked within the last month. Former smokers are those who have smoked >100 cigarettes in their lifetime but have not smoked in the past month. Never smokers are people who have smoked <100 cigarettes in their lifetime and do not currently smoke.[25] Subjects' education level was categorized into four groups, which were ≥ 6 and <9 years, ≥ 9 and <12 years, ≥ 12 and <16 years, and ≥ 16 years of education. Household income level was divided into four groups based on the average total household income: first quartile, second quartile, third quartile, and the fourth quartile; the
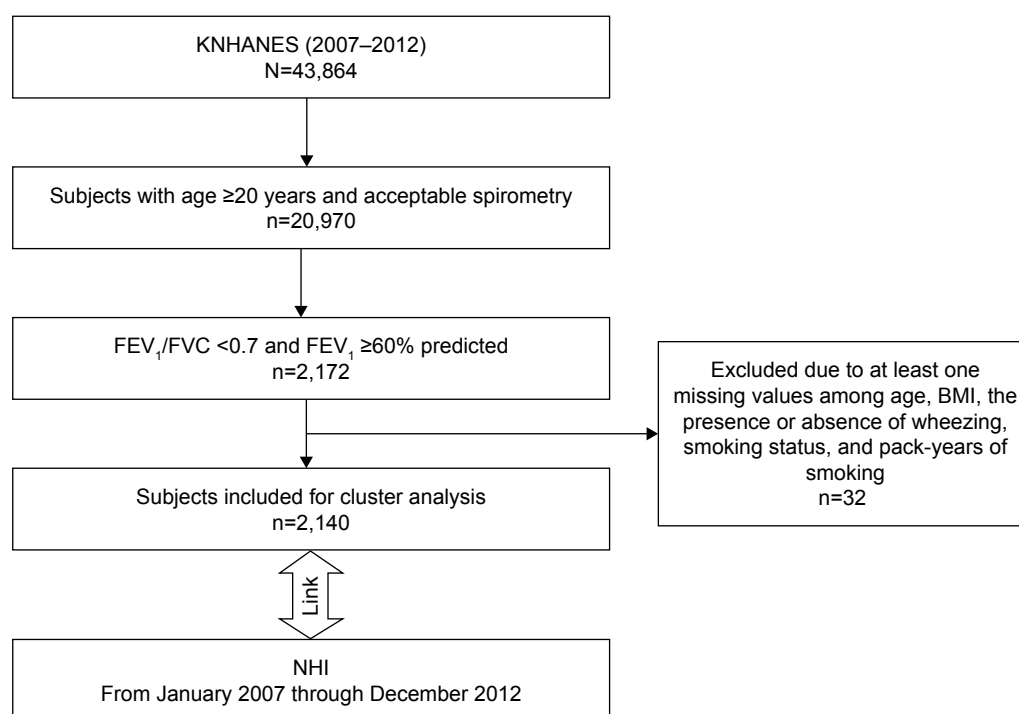
Figure 1 Study flow diagram.
Abbreviations: BMI, body mass index; FEV_1, forced expiratory volume in 1 second; KNHANES, Korea National Health and Nutrition Examination Survey; NHI, National Health Insurance; FVC, forced vital capacity.

first quartile means the lowest income group. The EuroQoL five dimensions questionnaire (EQ-5D) index values, which range between 0 (worst imaginable health state) and 1 (best imaginable health state), were obtained.[26]

Variables on COPD-related health care utilization were extracted from the NHI claims database, which included medical cost, duration of health care use, hospitalization, performed spirometry and radiologic tests, prescribed respiratory medications, and comorbidities. The medical cost was calculated as medical expenses (including insurance benefits and patients' copayments), including outpatient and inpatient fees. The duration of health care use was defined as the length of hospital stay (including admission to an intensive care unit and ER visits) and the number of outpatient visits.

The National Evidence-based Healthcare Collaborating Agency Ethics Committee of Korea approved this study. The requirement for informed consent from the patients was waived due to the retrospective nature of our study by the ethics committee.

Statistical analyses

For analyses, frequencies, percentages, and mean ± standard deviation (SD) were calculated to compare variables among groups. Statistical significance was tested by using the chi-squared test for categorical (discrete) variables and analysis

of variance for continuous variables. The SAS program (ver 9.2 [SAS Institute]) for statistical analyses was used.

Results

Among a total of 2,140 subjects, five groups through k-means clustering – putative "near-normal (n=232)," "asthma (n=392)," "COPD (n=37)," "asthmatic-overlap (n=893)," and "COPD-overlap (n=586)" subtypes – were identified. Asthmatic-overlap means asthma-predominant overlap and COPD-overlap means COPD-predominant overlap. Near-normal group showed the oldest mean age (72±7 years) and highest FEV_1 (102%±8% predicted), and asthmatic group was the youngest (46±9 years). Self-reported wheezing was more prevalent in asthma, asthmatic-overlap, and COPD-overlap groups than in COPD and near-normal groups. COPD and COPD-overlap groups were male predominant (100% and 98%, respectively) and all were current or former smokers. Mean pack-years of smoking was higher in COPD (102.9±14.7 pack-years) and COPD-overlap groups (45.3±12.2 pack-years) than those in the other three groups (<10 pack-years) (Figure 2, Table 1).

When applying the LLN FEV_1/FVC as a criterion for airway obstruction, the asthmatic group showed the highest prevalence of airway obstruction (92%), whereas the lowest prevalence (25%) appeared in the near-normal group. In the
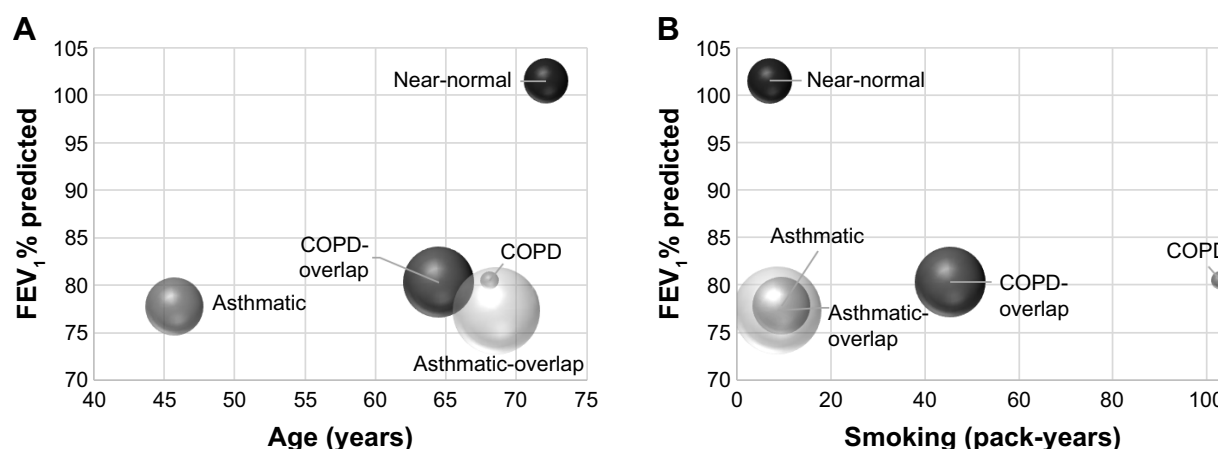
Figure 2 Distribution of five clusters according to mean FEV$_1$% predicted and amount of smoking.
Notes: (**A**) The graph shows the relationship between FEV$_1$% predicted and age of each cluster. (**B**) The graph shows the relationship between FEV$_1$% predicted and amount of smoking of each cluster. Sphere size represents the number of each cluster.
Abbreviations: COPD, chronic obstructive pulmonary disease; FEV$_1$, forced expiratory volume in 1 second.

near-normal group, all subjects had FEV$_1$ ≥80% predicted. History of physician-diagnosed asthma was more common in asthma and asthmatic-overlap groups. Subjects with the longest education duration and those with the highest household income comprise the largest portion of the asthmatic group (Table 2).

While the asthmatic group (7.1%) had the lowest prescription rate despite the highest proportion of self-reported wheezing, the COPD (24.3%), asthmatic-overlap (21.6%),

and COPD-overlap groups (17.1%) showed a high prescription rate of respiratory medicine. The most frequently prescribed drug was theophylline followed by oral beta 2 agonist (Table 3, Figure 3A). Among inhaled medications, short-acting beta 2 agonist (SABA) was the most commonly given, whereas long-acting muscarinic antagonist (LAMA) was the most frequently prescribed in the COPD group. Combinations of inhaled corticosteroids (ICSs) and long-acting beta 2 agonists (LABAs) were the most frequently

Table 1 Epidemiological characteristics according to subtypes of subjects with mild-to-moderate airflow limitation identified by k-means clustering for k=5 clusters

N	All	COPD-overlap	COPD	Asthmatic	Near-normal	Asthmatic-overlap	P-value
	2,140	586	37	392	232	893	
FEV$_1$% predicted	80.9±12.1	80.3±10.9	80.5±13.3	77.8±9.7	101.5±7.9	77.3±8.9	<0.001
BMI, kg/m²	23.6±2.8	23.6±2.7	23.9±2.8	23.4±2.9	23.2±2.5	23.7±2.9	0.047
Self-reported wheezing							0.002
Yes	352 (16)	89 (15)	3 (8)	77 (20)	21 (9)	162 (18)	
No	1,788 (84)	497 (85)	34 (92)	315 (80)	211 (91)	731 (82)	
Age, years	63.7±11.7	64.5±8.7	68.1±6.7	45.7±8.6	72.1±7.2	68.6±6.8	<0.001
≥20 and <30	29 (1)	0 (0)	0 (0)	29 (7)	0 (0)	0 (0)	<0.001
≥30 and <40	49 (2)	1 (0)	0 (0)	48 (12)	0 (0)	0 (0)	
≥40 and <50	179 (8)	22 (4)	0 (0)	157 (40)	0 (0)	0 (0)	
≥50 and <60	410 (19)	148 (25)	4 (11)	158 (40)	15 (7)	85 (10)	
≥60 and <70	710 (33)	233 (40)	17 (46)	0 (0)	59 (25)	401 (45)	
≥70 and <80	667 (31)	167 (29)	15 (41)	0 (0)	125 (54)	360 (40)	
≥80	96 (5)	15 (3)	1 (3)	0 (0)	33 (14)	47 (5)	
Male sex	1,523 (71)	576 (98)	37 (100)	235 (60)	131 (57)	544 (61)	<0.001
Smoking history							<0.001
Current smoker	687 (32)	306 (52)	8 (22)	147 (37)	42 (18)	184 (21)	
Ex-smoker	790 (37)	280 (48)	29 (78)	85 (22)	80 (35)	316 (35)	
Never smoker	663 (31)	0 (0)	0 (0)	160 (41)	110 (47)	393 (44)	
Smoking pack-years	20.3±22.5	45.3±12.2	102.9±14.7	9.4±10.7	6.9±9.6	8.6±10.0	<0.001

Note: Data are presented as mean ± SD or number (%).
Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; SD, standard deviation; FEV$_1$, forced expiratory volume in 1 second.

Table 2 Lung function and socioeconomic characteristics according to subtypes of subjects with mild-to-moderate airflow limitation identified by k-means clustering for k=5 clusters

N	All	COPD-overlap	COPD	Asthmatic	Near-normal	Asthmatic-overlap	P-value
	2,140	586	37	392	232	893	
Spirometry							
FEV$_1$% predicted	80.9±12.1	80.3±10.9	80.5±13.3	77.8±9.7	101.5±7.9	77.3±8.9	<0.001
FVC% predicted	92.5±12.7	92±11.8	89.9±14.5	96.6±11.2	107.1±10.8	87.3±10.5	<0.001
FEV$_1$/FVC	0.64±0.05	0.64±0.06	0.64±0.06	0.66±0.04	0.67±0.03	0.64±0.05	<0.001
FEV$_1$/FVC < LLN	1,226 (57)	309 (53)	16 (43)	362 (92)	57 (25)	482 (54)	<0.001
FEV$_1$ ≥80% predicted	1,060 (50)	276 (47)	17 (46)	157 (40)	232 (100)	378 (42)	<0.001
FEV$_1$ <80% predicted	1,080 (51)	310 (53)	20 (54)	235 (60)	0 (0)	515 (58)	
Past history of asthma	208 (10)	42 (7)	0 (0)	46 (12)	15 (7)	105 (12)	0.002
Past history of tuberculosis	269 (13)	78 (13)	3 (8)	50 (13)	21 (9)	117 (13)	0.43
Education duration, years							<0.001
≥6 and <9	883 (41)	236 (40)	20 (54)	49 (13)	131 (57)	447 (50)	
≥9 and <12	369 (17)	106 (18)	7 (19)	57 (15)	39 (17)	160 (18)	
≥12 and <16	540 (25)	162 (28)	6 (16)	153 (39)	41 (18)	178 (20)	
≥16	344 (16)	82 (14)	4 (11)	133 (34)	21 (9)	104 (12)	
Household income							<0.001
First quartile	536 (25)	127 (22)	12 (32)	24 (6)	93 (40)	280 (31)	
Second quartile	534 (25)	156 (27)	6 (16)	68 (17)	65 (28)	239 (27)	
Third quartile	536 (25)	152 (26)	10 (27)	128 (33)	43 (19)	203 (23)	
Fourth quartile	534 (25)	151 (26)	9 (24)	172 (44)	31 (13)	171 (19)	
EQ-5D index	0.91±0.14	0.92±0.13	0.91±0.11	0.96±0.08	0.87±0.18	0.89±0.16	<0.001

Notes: Data are presented as mean ± SD or number (%). LLN prediction equations: 125.77628–0.36304× age (years) –0.17146× height (cm) for male; 97.36197–0.26015× age (years) –0.01861× height (cm) for female.
Abbreviations: COPD, chronic obstructive pulmonary disease; EQ-5D, The EuroQoL five dimensions questionnaire; SD, standard deviation; LLN, lower limit of normal; FEV$_1$, forced expiratory volume in 1 second; FVC, forced vital capacity.

prescribed inhalers in both the asthma and asthmatic-overlap groups (Table 3, Figure 3B).

Except the asthmatic group, the other four groups showed a high prevalence of comorbidities such as hypertension, diabetes mellitus, hyperlipidemia, and coronary artery disease. Within the asthmatic group, metabolic syndrome and hyperlipidemia were the most common comorbidities (Table 4).

COPD group showed the highest hospitalization rate (16%) and spirometry test (16%) (Table 5). The highest mean in-patient medical cost and the longest total health care used days were observed in the COPD group. Although the COPD group formed only 1.7% of total subjects, their medical cost comprised 5.3% of the total medical cost (Table 6, Figure 4). When calculating a ratio of total medical expense

Table 3 Respiratory medication prescribed according to subtypes of subjects with mild-to-moderate airflow limitation

N	All	COPD-overlap	COPD	Asthmatic	Near-normal	Asthmatic-overlap	P-value
	2,140	586	37	392	232	893	
Any respiratory medicine	365 (17.1)	100 (17.1)	9 (24.3)	28 (7.1)	35 (15.1)	193 (21.6)	<0.001
ICS	35 (1.6)	6 (1.0)	1 (2.7)	3 (0.8)	4 (1.7)	21 (2.4)	0.18
ICS + LABA	70 (3.3)	15 (2.6)	3 (8.1)	5 (1.3)	1 (0.4)	46 (5.2)	<0.001
LAMA	56 (2.6)	18 (3.1)	5 (13.5)	2 (0.5)	3 (1.3)	28 (3.1)	<0.001
SAMA	47 (2.2)	17 (2.9)	2 (5.4)	1 (0.3)	1 (0.4)	26 (2.9)	0.004
SABA	85 (4.0)	33 (5.6)	2 (5.4)	2 (0.5)	6 (2.6)	42 (4.7)	<0.001
Systemic bronchodilators	163 (7.6)	45 (7.7)	4 (10.8)	14 (3.6)	13 (5.6)	87 (9.7)	0.002
Theophylline	197 (9.2)	57 (9.7)	7 (18.9)	13 (3.3)	17 (7.3)	103 (11.5)	<0.001
OCS	96 (4.5)	24 (4.1)	3 (8.1)	9 (2.3)	5 (2.2)	55 (6.2)	0.006
IVCS	99 (4.6)	24 (4.1)	2 (5.4)	4 (1.0)	13 (5.6)	56 (6.3)	0.001
LTRA	52 (2.4)	11 (1.9)	2 (5.4)	3 (0.8)	2 (0.9)	34 (3.8)	0.003

Note: Data are presented as number (%).
Abbreviations: COPD, chronic obstructive pulmonary disease; ICS, inhaled corticosteroids; IVCS, intravenous corticosteroids; LABA, long-acting beta-2 agonists; LAMA, long-acting muscarinic antagonists; LTRA, leukotriene receptor antagonist; OCS, oral corticosteroids; SAMA, short-acting muscarinic antagonists; SABA, short-acting beta-2 agonists.
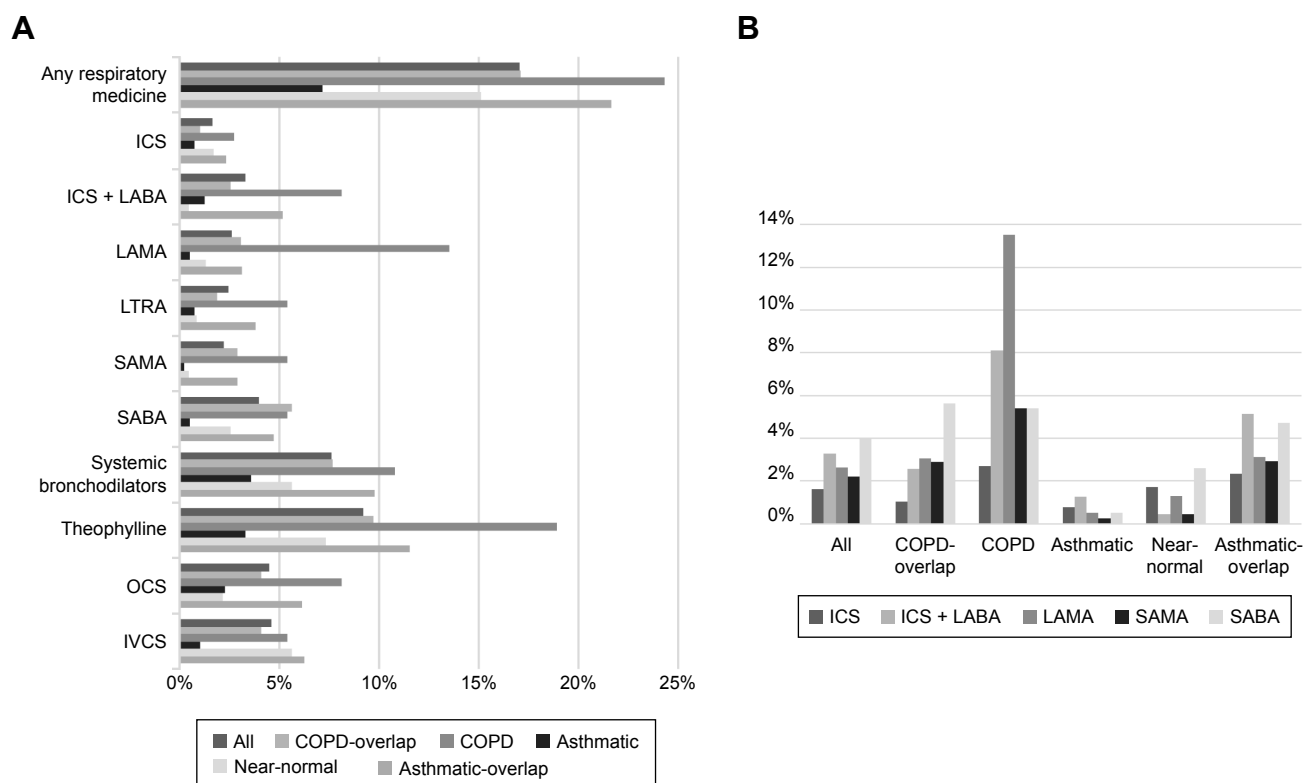
A

B

Figure 3 Prescription patterns of respiratory medication.
Notes: (**A**) Distribution of subtypes of mild-to-moderate airflow limitation according to the respiratory medications. (**B**) Prescription patterns of inhaled medications according to the subtype of mild-to-moderate airflow limitation.
Abbreviations: COPD, chronic obstructive pulmonary disease; ICS, inhaled corticosteroids; IVCS, intravenous corticosteroids; LABA, long-acting beta-2 agonists; LAMA, long-acting muscarinic antagonists; LTRA, leukotriene receptor antagonist; OCS, oral corticosteroids; SAMA, short-acting muscarinic antagonists; SABA, short-acting beta-2 agonists.

Table 4 Comorbidities according to subtypes of subjects with mild-to-moderate airflow limitation

N	All	COPD-overlap	COPD	Asthmatic	Near-normal	Asthmatic-overlap	P-value
	2,140	586	37	392	232	893	
Comorbidities based on KNHANES data							
Hypertension	763 (36)	208 (36)	15 (41)	42 (11)	88 (38)	410 (46)	<0.001
Diabetes mellitus	294 (14)	81 (14)	9 (24)	17 (4)	32 (14)	155 (17)	<0.001
Cerebrovascular disease	59 (3)	12 (2)	2 (5)	4 (1)	8 (3)	33 (4)	0.04
Depression	260 (12)	59 (10)	3 (8)	47 (12)	35 (15)	116 (13)	0.25
Hyperlipidemia	247 (12)	85 (15)	4 (11)	21 (5)	22 (10)	115 (13)	0.002
Lung cancer	7 (0.3)	2 (0.3)	0 (0)	1 (0.3)	0 (0)	4 (0.4)	0.85
Comorbidities based on HIRA data							
Hypertension	1,144 (54)	318 (54)	21 (57)	76 (19)	136 (59)	593 (66)	<0.001
Diabetes mellitus	799 (37)	230 (39)	19 (51)	61 (16)	84 (36)	405 (45)	<0.001
Metabolic syndrome	922 (43)	249 (42)	18 (49)	99 (25)	103 (44)	453 (51)	<0.001
Osteoarthritis	239 (11)	51 (9)	5 (14)	28 (7)	29 (13)	126 (14)	0.001
Osteoporosis	485 (23)	68 (12)	8 (22)	28 (7)	87 (38)	294 (33)	<0.001
Depression	363 (17)	76 (13)	5 (14)	44 (11)	48 (21)	190 (21)	<0.001
Hyperlipidemia	771 (36)	220 (38)	16 (43)	77 (20)	86 (37)	372 (42)	<0.001
Lung cancer	55 (3)	22 (4)	1 (3)	2 (1)	3 (1)	27 (3)	0.02
Coronary artery disease	498 (23)	141 (24)	13 (35)	33 (8)	65 (28)	246 (28)	<0.001

Note: Data are presented as number (%).
Abbreviations: KNHANES, Korea National Health and Nutrition Examination Survey; HIRA, Korea Health Insurance Review and Assessment Service; COPD, chronic obstructive pulmonary disease.

Table 5 Insurance-related data according to subtypes of subjects with mild-to-moderate airflow limitation

N	All	COPD-overlap	COPD	Asthmatic	Near-normal	Asthmatic-overlap	*P*-value
	2,140	586	37	392	232	893	
Insurance type							0.20
Health insurance	2,051 (96)	558 (95)	33 (89)	379 (97)	221 (95)	860 (96)	
Medical aid	89 (4)	28 (5)	4 (11)	13 (3)	11 (5)	33 (4)	
Hospitalization							<0.001
Never	1,903 (89)	534 (91)	31 (84)	377 (96)	201 (87)	760 (85)	
Ever	237 (11)	52 (9)	6 (16)	15 (4)	31 (13)	133 (15)	
ICU admission							0.06
Never	2,110 (99)	573 (98)	36 (97)	392 (100)	228 (98)	881 (99)	
Ever	30 (1)	13 (2)	1 (3)	0 (0)	4 (2)	12 (1)	
Emergency room visit							0.55
Never	2,077 (97)	569 (97)	35 (95)	389 (99)	224 (97)	860 (96)	
Ever	63 (3)	17 (3)	2 (5)	3 (1)	8 (3)	33 (4)	
Chest X-ray							<0.001
Never	2,001 (94)	548 (94)	33 (89)	385 (98)	210 (91)	825 (92)	
Ever	139 (7)	38 (6)	4 (11)	7 (2)	22 (9)	68 (8)	
Spirometry							0.004
Never	1,986 (93)	534 (91)	31 (84)	385 (98)	218 (94)	818 (92)	
Ever	154 (7)	52 (9)	6 (16)	7 (2)	14 (6)	75 (8)	
Computed tomography of the chest							0.37
Never	2,133 (100)	585 (100)	37 (100)	392 (100)	230 (99)	889 (100)	
Ever	7 (0)	1 (0)	0 (0)	0 (0)	2 (1)	4 (0)	

Note: Data are presented as number (%).
Abbreviations: ICU, intensive care unit; COPD, chronic obstructive pulmonary disease.

to household income in order to compare the economic burden of medical cost among the five groups, the mean ratio was the highest in the COPD group (Table 6, Figure 5).

Discussion

This study indicates that clinical and socioeconomic heterogeneity exists among subjects with mild-to-moderate airflow limitation. To identify subtypes with substantial heterogeneity, six key variables that included FEV$_1$% predicted, BMI, the presence or absence of self-reported wheezing, age, smoking status, and pack-years of cigarette smoking, based on previous clinical studies, were chosen.[12,17–21,27,28] FEV$_1$% predicted represents severity of airflow limitation. BMI is known to be a predictor of COPD prognosis and a

Table 6 Health care utilization according to subtypes of subjects with mild-to-moderate airflow limitation

N (%)	All	COPD-overlap	COPD	Asthmatic	Near-normal	Asthmatic-overlap	*P*-value
	2,140 (100)	586 (27.4)	37 (1.7)	392 (18.3)	232 (10.8)	893 (41.7)	
Mean medical cost							
Out-patient clinic, USD	273±685	303±877	762±1,311	117±198	144±245	265±568	0.10
In-patient, USD	3,674±4,380	4,759±5,757	7,024±6,191	1,563±619	3,432±3,663	3,393±3,922	0.03
Total cost, USD	2,045±3,711	2,212±4,621	4,594±5,973	778±921	2,171±3,311	1,991±3,320	0.05
Health care utilization							
Out-patient, days	8.6±15.6	7.5±11.0	9.7±12.1	3.0±2.8	5.3±7.7	10.4±19.4	0.17
In-patient, days	19.3±20.7	19.0±19.6	36.2±33.1	12.0±7.1	17.0±12.8	19.9±22.7	0.18
Total duration, days	15.5±21.5	13.7±18.1	29.5±34.5	7.2±7.2	13.5±14.1	17.4±24.4	0.01
Total medical cost per group							
Out-patient clinic (%)	86,421 (100)	29,042 (33.6)	8,387 (9.7)	2,221 (2.6)	4,324 (5.0)	42,448 (49.1)	
In-patient (%)	870,759 (100)	247,462 (28.4)	42,146 (4.8)	23,448 (2.7)	106,399 (12.2)	451,303 (51.8)	
Total cost (%)	957,180 (100)	276,504 (28.9)	50,533 (5.3)	25,669 (2.7)	110,723 (11.6)	493,751 (51.6)	
Relative medical cost							
Total medical cost/household income	5.08±34.47	9.05±64.75	18.62±39.58	0.35±0.50	4.10±7.86	3.29±6.48	0.33

Note: Data are presented as mean ± SD or number (%).
Abbreviations: USD, US dollars; COPD, chronic obstructive pulmonary disease; SD, standard deviation.
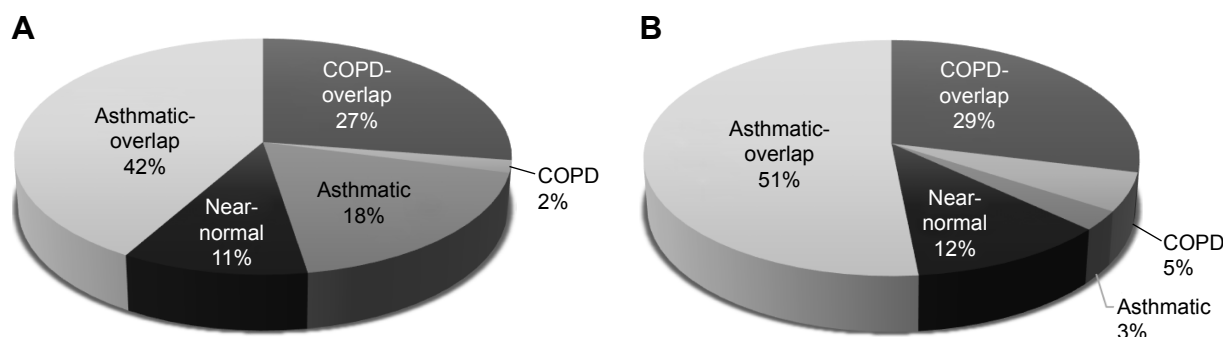
Figure 4 Proportions of subject number and medical cost of five clusters.
Notes: (**A**) Proportions of the number in each cluster relative to the total number of subjects. (**B**) Medical cost of each cluster relative to the total medical cost.
Abbreviation: COPD, chronic obstructive pulmonary disease.

marker of systemic inflammation and may correlate with severity of emphysema.[12,17–19] Wheezing might reflect airway hyper-responsiveness,[20,21] even though it is quite subjective. Aging and smoking are the most important risk factors of airway disease.[27,28]

K-means clustering identifies five distinct groups. A cluster called "COPD" exhibited the highest mean medical cost and health care utilization. Particularly, when considering the household income, the medical cost was much higher than those in the other groups. It was 3.7 times higher than the overall mean of total subjects with mild-to-moderate airflow limitation and 53 times higher than that of the asthmatic group. The COPD cluster comprised all men who were over the age of 50 and all current and ex-smokers, and included a large proportion with lower levels of education and lower economic status. They had the highest rate of hospitalization and prescription implying frequent health care utilization. On the contrary, those in the asthmatic cluster were all aged <60 years. There were 40% of women and 41% of never smokers. Despite the highest proportion of self-reported wheezing and airway obstruction on the basis of the LLN, the highest EQ-5D index and the lowest health resource utilization and prescription rates were found in the asthmatic

group. One third of the asthmatic group reported that they had an education level and 77% belonged to the top half of economic classes, suggesting relatively high socioeconomic status. There were two kinds of overlap clusters, asthma-predominant overlap and COPD-predominant overlap. The main differences between the two overlap clusters were the presence or absence of never-smokers, sex distribution, and the most frequently prescribed inhalers. While ICS/LABA was the most frequently prescribed inhaler in asthmatic-overlap group, LAMA was the most common in COPD-overlap. The mean medical cost was higher in the COPD-overlap group compared to the asthmatic-overlap group, whereas the mean duration of health care utilization was longer in asthmatic-overlap group, which might be related to more women being in the asthmatic-overlap group. Medical economic burden relative to household income was the second highest in the COPD-overlap group. While the COPD-overlap group may be close to COPD with airway hyper-responsiveness, the asthmatic-overlap group might be similar to severe asthma with airway remodeling.

This study has several limitations. The fixed ratio ($FEV_1/FVC < 0.7$) was used as a criteria of airway obstruction, which is likely to lead to an underdiagnosis of airway obstruction for younger people or an overdiagnosis of airway obstruction for older people.[29] It was derived from population-based databases and subjects were not recruited as part of a study of airway disease. Therefore, the key variables for cluster analysis might not be enough to reflect the diverse aspects of airway disease. Nevertheless, our five clusters were very similar to clinical phenotypes in terms of distinct prescription patterns as well as distributions of age, sex, and smoking history. From a different prospective, population-based data were one of the strengths of the study. Since subjects with mild-to-moderate airflow limitation are frequently asymptomatic,[30,31] it is sometimes hard to recruit
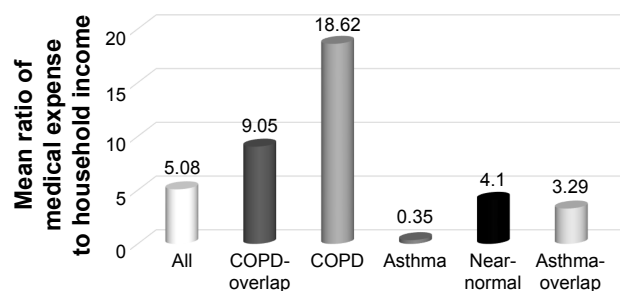

Figure 5 The mean ratio of medical expense to household income according to subtypes of subjects with mild-to-moderate airflow limitation.
Abbreviation: COPD, chronic obstructive pulmonary disease.

them for research purposes. Nationwide randomization sampling and health care utilization data allowed us to conduct research on this population. Moreover, information about education status and household income as well as smoking history, lung function, BMI, wheezing, history of asthma and tuberculosis, and QOL was available in this study, which were not possible to collect when using only claim data. The other limitations were that our FEV_1/FVC was not a post-bronchodilator value and self-reported wheezing was used as an indicator of airway hyper-responsiveness, though these have been largely accepted in previous population-based studies.[21,32]

It was found that clinical heterogeneity exists even in subjects with mild airflow limitation, which may lead to a different level of health care utilization. In particular one group was the main consumer of medical resources. Since health care resources are limited, finding a group that needs lots of medical resources and managing it intensively can be an effective way to utilize all medical resources.

Conclusion

This study suggests that even mild airflow limitation could lead to high socioeconomic burden. Defining a high-risk group such as our COPD group and targeting this group through early intervention will provide a key to ultimately reduce medical cost.

Acknowledgments

Author contributions

Jin Hwa Lee is the guarantor of the accuracy and completeness of the contents of this manuscript, including the data and analysis. Chin Kook Rhee and Kwang Ha Yoo contributed to data acquisition. Jee-Ae Kim and Sang Hyun Kim contributed to data linkage and refinement. Kyungjoo Kim analyzed the data and performed the statistical analysis. All authors critically evaluated the data, revised manuscript, and approved the final version.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Adeloye D, Chua S, Lee C, et al. Global and regional estimates of COPD prevalence: Systematic review and meta-analysis. *J Glob Health*. 2015;5(2):020415.
2. James PA, Oparil S, Carter BL, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311(5):507–520.
3. American Diabetes A. 4. Prevention or delay of type 2 diabetes. *Diabetes Care*. 2016;39(Suppl 1):S36–S38.
4. Global Initiative for Chronic Obstructive Lung Disease. *Global Strategy for the Diagnosis, Management and Prevention of COPD*; 2017. Available from: http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/. Accessed January 20, 2017.
5. Krahn M, Chapman KR. Economic issues in the use of office spirometry for lung health assessment. *Can Respir J*. 2003;10(6):320–326.
6. Qaseem A, Wilt TJ, Weinberger SE, et al. Diagnosis and management of stable chronic obstructive pulmonary disease: a clinical practice guideline update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. *Ann Intern Med*. 2011;155(3):179–191.
7. Force USPST, Siu AL, Bibbins-Domingo K, et al. Screening for chronic obstructive pulmonary disease: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2016;315(13):1372–1377.
8. Guirguis-Blake JM, Senger CA, Webber EM, Mularski RA, Whitlock EP. Screening for chronic obstructive pulmonary disease: evidence report and systematic review for the US Preventive Services Task Force. *JAMA*. 2016;315(13):1378–1393.
9. Vestbo J, Rennard S. Chronic obstructive pulmonary disease biomarker(s) for disease activity needed–urgently. *Am J Respir Crit Care Med*. 2010; 182(7):863–864.
10. Faner R, Tal-Singer R, Riley JH, et al. Lessons from ECLIPSE: a review of COPD biomarkers. *Thorax*. 2014;69(7):666–672.
11. Sin DD, Vestbo J. Biomarkers in chronic obstructive pulmonary disease. *Proc Am Thorac Soc*. 2009;6(6):543–545.
12. Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med*. 2004;350(10):1005–1012.
13. Lee JH, Cho MH, McDonald ML, et al. Phenotypic and genetic heterogeneity among subjects with mild airflow obstruction in COPDGene. *Respir Med*. 2014;108(10):1469–1480.
14. Regan EA, Lynch DA, Curran-Everett D, et al. Clinical and radiologic disease in smokers with normal spirometry. *JAMA Intern Med*. 2015;175(9):1539–1549.
15. Pauwels RA, Rabe KF. Burden and clinical features of chronic obstructive pulmonary disease (COPD). *Lancet*. 2004;364(9434):613–620.
16. Jansson SA, Backman H, Stenling A, Lindberg A, Ronmark E, Lundback B. Health economic costs of COPD in Sweden by disease severity–has it changed during a ten years period? *Respir Med*. 2013; 107(12):1931–1938.
17. Schols AM, Slangen J, Volovics L, Wouters EF. Weight loss is a reversible factor in the prognosis of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 1998;157(6 Pt 1):1791–1797.

18. Vestbo J, Prescott E, Almdal T, et al. Body mass, fat-free body mass, and prognosis in patients with chronic obstructive pulmonary disease from a random population sample: findings from the Copenhagen City Heart Study. *Am J Respir Crit Care Med*. 2006;173(1):79–83.

19. Bhavani S, Tsai CL, Perusich S, et al. Clinical and immunological factors in emphysema progression. Five-year prospective longitudinal exacerbation study of chronic obstructive pulmonary disease (LES-COPD). *Am J Respir Crit Care Med*. 2015;192(10):1171–1178.

20. Burney PG, Chinn S, Britton JR, Tattersfield AE, Papacosta AO. What symptoms predict the bronchial response to histamine? Evaluation in a community survey of the bronchial symptoms questionnaire (1984) of the International Union Against Tuberculosis and Lung Disease. *Int J Epidemiol*. 1989;18(1):165–173.

21. Arif AA, Delclos GL, Lee ES, Tortolero SR, Whitehead LW. Prevalence and risk factors of asthma and wheezing among US adults: an analysis of the NHANES III data. *Eur Respir J*. 2003;21(5):827–833.

22. Kweon S, Kim Y, Jang MJ, et al. Data resource profile: the Korea National Health and Nutrition Examination Survey (KNHANES). *Int J Epidemiol*. 2014;43(1):69–77.

23. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J*. 2005;26(5):948–968.

24. Hwang YI, Kim CH, Kang HR, et al. Comparison of the prevalence of chronic obstructive pulmonary disease diagnosed by lower limit of normal and fixed ratio criteria. *J Korean Med Sci*. 2009;24(4):621–626.

25. Burney PG, Luczynska C, Chinn S, Jarvis D. The European Community Respiratory Health Survey. *Eur Respir J*. 1994;7(5):954–960.

26. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 2001;33(5):337–343.

27. Snider GL. Chronic obstructive pulmonary disease: risk factors, pathophysiology and pathogenesis. *Annu Rev Med*. 1989;40:411–429.

28. Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. *Lancet*. 2015;385(9971):899–909.

29. Affes Z, Rekik S, Ben Saad H. Defining obstructive ventilatory defect in 2015. *Libyan J Med*. 2015;10:28946.

30. Buffels J, Degryse J, Heyrman J, Decramer M, Study D. Office spirometry significantly improves early detection of COPD in general practice: the DIDASCO Study. *Chest*. 2004;125(4):1394–1399.

31. Averame G, Bonavia M, Ferri P, et al. Office spirometry can improve the diagnosis of obstructive airway disease in primary care setting. *Respir Med*. 2009;103(6):866–872.

32. Xuan W, Peat JK, Toelle BG, Marks GB, Berry G, Woolcock AJ. Lung function growth and its relation to airway hyperresponsiveness and recent wheeze. Results from a longitudinal population study. *Am J Respir Crit Care Med*. 2000;161(6):1820–1824.