Respiratory Medicine (2013) 107, 1939–1947

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/rmed

Factor analysis in predominantly severe COPD: Identification of disease heterogeneity by easily measurable characteristics

Dirkje S. Postma [a,*], Antonio R. Anzueto [b], Christine Jenkins [c], Barry J. Make [d], Thomas Similowski [e], Ollie Östlund [f], Göran S. Eriksson [g], Peter M. Calverley [h]

[a] University of Groningen, University Medical Center Groningen, Department of Pulmonary Medicine and Tuberculosis, GRIAC Research Institute, AA11, Hanzeplein, PO Box 30001, 9700 RB Groningen, The Netherlands
[b] Department of Pulmonary Medicine and Allergology, University of Texas Health Sciences Center, San Antonio, TX, USA
[c] University of Sydney and Woolcock Institute of Medical Research, Camperdown, Australia
[d] Division of Pulmonary Sciences and Critical Care Medicine, National Jewish Health, University of Colorado, Denver, USA
[e] Service de Pneumologie et Réanimation, Groupe Hospitalier Pitié-Salpêtrière, Paris, France
[f] AstraZeneca R&D, Mölndal, Sweden
[g] Department of Respiratory Medicine and Allergology, University Hospital, Lund, Sweden
[h] Pulmonary and Rehabilitation Research Group, University Hospital Aintree, Liverpool, UK

Received 18 April 2013; accepted 11 July 2013
Available online 15 August 2013

KEYWORDS
Chronic obstructive pulmonary disease;
Exacerbations;
Heterogeneity;
Factor analysis

Summary
Background: The clinical and demographic variables defining the heterogeneity of chronic obstructive pulmonary disease (COPD) are unclear. A *post-hoc* analysis of five randomised studies in patients with a history of previous exacerbations examined the clinical and demographic characteristics describing moderate-to-very-severe COPD.
Methods: Factor analysis was performed on all continuous baseline demographic and clinical data, without variable selection. Analyses were based on the full cohort and on stratifications by pack-years smoked, smoking status, gender, and comorbidities; patient exacerbation history was analysed in two of the five studies.
Findings: 6162 COPD patients were evaluated (70% male; 40% current smokers; mean pre-

* Corresponding author. Tel.: +31 (0)503613532; fax: +31 (0)503619320.
E-mail address: d.s.postma@umcg.nl (D.S. Postma).

0954-6111/$ - see front matter
http://dx.doi.org/10.1016/j.rmed.2013.07.011

bronchodilator forced expiratory volume in 1 s [FEV_1] 35.2% predicted). Baseline clinical and demographic variables loaded differentially on six factors with minimal overlap, explaining 60.4% of the heterogeneity: 1) symptoms (cough, dyspnoea, sleep disturbance), health status, reliever use; 2) pre-bronchodilator FEV_1, FEV_1/forced vital capacity, morning peak expiratory flow (PEF), body mass index (BMI); 3) blood pressure; 4) age, months since first COPD symptoms; 5) PEF variability; 6) pulse, FEV_1 reversibility. Most factors loaded similarly in stratified and exacerbation analyses. BMI loaded with reversibility in females, and with age and months since first COPD symptoms in ex-smokers. Exacerbations loaded to factor 6.
Interpretation: Readily available data can explain ~60% of COPD heterogeneity in a large dataset of predominantly severe COPD patients. Factors were robust over determinants of disease outcome; gender, smoking status, pack-years smoked, and comorbidities. The main factors were largely unchanged by adding exacerbations. Only BMI loaded to other factors.

Background

Chronic obstructive pulmonary disease (COPD) is associated with smoking and biomass smoke exposure, and has a major impact on global healthcare costs [1]. COPD was one of the leading causes of disability in 2010 [2] and is predicted to be the third leading cause of mortality by 2030 [3]. Effective treatment is complicated by the clinical, physiological, and pathological heterogeneity of this condition [4]. Clinical presentation varies, and symptoms (cough, phlegm, and dyspnoea) may occur alone or in combination at every stage. COPD is associated with progressive persistent airflow limitation and accelerated lung function decline [5]. There is considerable variability in lung function decline [6] and heterogeneity in the pathological processes underlying COPD [7].

Accurate assessment of COPD severity is important in guiding management. The Global initiative for chronic Obstructive Lung Disease (GOLD) divides COPD severity into four grades, based on airflow limitation, symptoms, and exacerbation history [5]. COPD management should not be based solely on lung function, since clinical characteristics (e.g. exacerbations and health-related quality of life [HRQL]) can be improved, irrespective of GOLD grade [8,9]. It is unknown how COPD subtypes, defined by clinical and demographic characteristics, are influenced by interventions or how they relate to other clinical characteristics. Identification of COPD phenotypes will allow more focused research on disease biomarkers, and thus more specific treatment strategies [10]. One of the challenges is to develop such phenotypes in an objective way independently of investigator bias.

We examined whether clinical and demographic characteristics of specific COPD patient subsets can be obtained from a dataset of patients in clinical trials. Factor analysis was performed using baseline characteristics from five studies [11—15]. We assessed how patient characteristics potentially influencing disease severity, outcome, and management could be clustered. Analyses were stratified for smoking status, gender, pack-years smoked, and comorbidities. Moreover, since exacerbations predict future outcomes [8], we analysed the two studies with information on past number of exacerbation separately [12,14]. We anticipated that differences would emerge within these groups based on this approach, which is independent of pre-existing biases, providing insight into variables that characterise a large group of COPD patients.

Methods

Baseline data were analysed from five trials involving COPD patients with a history of exacerbation in the past year, pre-bronchodilator FEV_1 $\leq$50% predicted and FEV_1/vital capacity (VC) [11,13] or FEV_1/forced VC (FVC) [12,14,15] $<$70%, and smoking history $\geq$10 pack-years [11—15]. All continuous baseline demographic and clinical data were used, without selection of variables, to ensure an unbiased approach. FEV_1 reversibility was assessed with salbutamol (albuterol; 180—200 μg) [12,14] or terbutaline (two inhalations, 0.5 mg/inhalation) [11,13,15] and expressed as increase in FEV_1 % predicted. Height, weight, systolic/diastolic BP, and pulse were assessed at run-in and St George's Respiratory Questionnaire (SGRQ) [16] scores at run-in completion. Peak expiratory flow (PEF) absolute values, PEF variability ([morning PEF—evening PEF]/morning PEF), symptom scores (cough, breathlessness, nocturnal awakening due to dyspnoea), and reliever use were calculated from patient diaries. Symptoms were scored using a 5-point Likert scale. Comorbidities were derived from patient reports of physician-diagnosed ischaemic cardiovascular disease, arteriosclerosis, hypertension, depression, anxiety, gastroesophageal reflux, and diabetes; these were combined to reduce the parameters. Time (months) since COPD symptom onset was determined from case report forms.

Factor analysis was performed to determine which characteristics were closely related. The covariance structure of baseline variables is described using classical principal component analysis (PCA), which does not require specific distributional properties of data, followed by an orthogonal rotation using the varimax principle [17]. The importance of factors is described by scree plots, and the number of components retained in the rotated structure determined by the eigenvalue one criterion and the proportion of total heterogeneity explained by these factors. For stratified analyses, supplementary rotations provided factor structures with the same number of factors for both sub-groups, allowing direct comparisons. Analyses were stratified by gender, pack-years smoked ($<$40 or $\geq$40 pack-years), smoking status (current or ex-smokers), and

Table 1 Baseline characteristics of patients with COPD participating in the five studies included in the analysis.

Characteristic	Rennard et al. [12] NCT00206167 (n = 1964)	Tashkin et al. [14] NCT00206154 (n = 1704)	Welte et al. [15] NCT00496470 (n = 660)	Calverley et al. [11] [a] (n = 1022)	Szafranski et al. [13] [a] (n = 812)	All (n = 6162)
Age, years	63.1 (9.1)	63.4 (9.1)	62.4 (8.7)	64.0 (9.0)	64.2 (8.9)	63.4 (9.1)
Male, *n* (%)	1255 (63.9)	1161 (68.1)	496 (75.2)	770 (75.3)	641 (78.9)	4323 (70.2)
BMI, kg/m^2	27.0 (5.8)	26.6 (5.6)	26.4 (5.2)	24.3 (4.8)	25.1 (5.1)	26.1 (5.5)
Pulse, bpm	77.2 (10.6)	77.7 (10.6)	78.1 (10.6)	79.5 (9.4)	78.4 (11.9)	78.0 (10.6)
DBP, mmHg	78.5 (9.6)	79.0 (9.7)	79.6 (9.5)	81.0 (10.2)	81.3 (10.2)	79.5 (9.7)
SBP, mmHg	131.3 (15.9)	131.6 (15.7)	132.9 (16.5)	134.8 (16.0)	134.1 (17.6)	132.5 (16.2)
Smoking history:						
Ex-smokers, *n* (%)	1130 (57.5)	980 (57.5)	370 (56.1)	669 (65.5)	532 (65.5)	3681 (59.7)
Current smokers, *n* (%)	834 (42.5)	1049 (61.6)	568 (86.1)	700 (68.5)	393 (38.5)	3922 (63.6)
Smoking history, pack-years	46.9 (27.2)	46.0 (26.3)	40.7 (22.8)	38.7 (22.5)	44.4 (24.8)	44.3 (25.6)
Months since first COPD symptoms	129.7 (86.2)	126.0 (86.3)	88.9 (76.1)	150.0 (111.4)	128.7 (90.9)	127.6 (91.8)
Pre-bronchodilator FEV_1, % predicted	34.4 (9.4)	34.2 (9.5)	37.9 (8.6)	36.1 (9.7)	36.1 (9.7)	35.2 (9.5)
Pre-bronchodilator FEV_1, % VC or FVC	48.3 (10.2)	47.0 (10.3)	47.1 (10.6)	43.5 (11.7)	42.5 (12.0)	46.2 (11.0)
Post-bronchodilator FEV_1, % predicted	39.6 (11.5)	39.8 (11.9)	43.5 (11.3)	42.1 (12.3)	41.8 (11.5)	40.8 (11.8)
GOLD Grade 4, *n* (%)	423 (21.5)	375 (22.0)	75 (11.4)	173 (16.9)	128 (15.8)	1174 (19.1)
GOLD Grade 3, *n* (%)	1187 (60.4)	991 (58.2)	420 (63.6)	569 (55.7)	488 (60.1)	3655 (59.3)
GOLD Grade 2, *n* (%)	349 (17.7)	329 (19.3)	160 (24.2)	275 (26.9)	186 (22.9)	1299 (21.1)
GOLD Grade 1, *n* (%)	2 (0.1)	5 (0.3)	4 (0.6)	2 (0.2)	3 (0.4)	16 (0.3)
FEV_1 reversibility, % predicted	5.3 (5.6)	5.6 (6.0)	5.6 (7.3)	5.8 (6.3)	5.8 (6.4)	5.6 (6.1)
Morning PEF, % predicted	41.7 (13.8)	41.1 (12.9)	35.2 (12.5)	46.1 (14.7)	40.5 (12.4)	41.4 (13.7)
PEF variability	0.1 (0.1)	0.1 (0.1)	0.2 (0.2)	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)
SGRQ scores:						
Total score	55.0 (16.7)	55.5 (16.4)	62.2 (17.6)	48.2 (18.4)	52.4 (16.5)	54.5 (17.4)
Symptom score	66.4 (19.0)	65.3 (19.6)	73.2 (17.5)	57.3 (20.3)	62.0 (18.9)	64.7 (19.7)
Impact score	43.0 (19.5)	44.3 (19.7)	51.8 (21.9)	41.3 (21.9)	43.8 (19.7)	44.1 (20.4)
Activity score	69.9 (18.5)	69.7 (17.4)	73.5 (19.0)	60.0 (21.0)	66.8 (19.3)	68.2 (19.2)
Reliever use:						
During the night	0.7 (1.1)	0.6 (1.1)	1.1 (1.1)	0.4 (0.8)	0.6 (0.9)	0.6 (1.0)
During the day	3.2 (2.8)	2.9 (3.0)	3.1 (2.4)	1.4 (1.8)	3.2 (2.8)	2.8 (2.8)
Comorbidities, *n* (%)	1256 (64.0)	1054 (61.9)	379 (57.4)	365 (35.7)	280 (34.5)	3334 (54.1)

All values expressed as mean (standard deviation), unless otherwise stated. PEF variability is defined as (morning PEF − evening PEF)/morning PEF. GOLD grades include FEV_1%predicted as follows: GOLD Grade 1 $\geq$ 80% predicted, GOLD Grade 2 50−80% predicted, GOLD Grade 3 30−50%predicted, GOLD Grade 4 $\leq$ 30% predicted.

BMI, body mass index; COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; GOLD, Global initiative for chronic Obstructive Lung Disease; PEF, peak expiratory flow; SBP, systolic blood pressure; SGRQ, St George's Respiratory Questionnaire; VC, vital capacity.

[a] These studies have no registration numbers, as they were conducted before clinical trial registration was required.

comorbidities (with or without). We also examined the impact of exacerbations, based on the two of the five studies with data on patient exacerbation history [12,14]. Results tables are presented with rotated factor loadings, with variables sorted by factor with largest absolute loading.

Pre- and post-bronchodilator FEV_1 and FEV_1 reversibility were co-linear, therefore post-bronchodilator FEV_1 was not analysed. Since morning and evening PEF and PEF variability were almost co-linear, evening PEF was omitted. Thus 16 characteristics were examined: age, body mass index (BMI), pre-bronchodilator FEV_1, pre-bronchodilator FEV_1/FVC [11,12,14] or FEV_1/VC [13,15], FEV_1 reversibility, SGRQ total score, months since first COPD symptoms, systolic and diastolic BP, pulse, morning PEF, PEF variability, scores of breathlessness, cough, sleep disturbance, and reliever use.

Patients with missing values were removed from analysis; potential bias was investigated with comparative summaries of included and excluded patients.

Results

Baseline clinical and demographic characteristics of patients were comparable. All patients (n = 6162; 70% male) had moderate-to-very-severe COPD (defined by post-

bronchodilator FEV_1 [GOLD criteria] [5] (Table 1). Included and excluded patients had similar characteristics (Supplementary Table 1).

The 16 demographic and clinical characteristics loaded over six factors with minimal overlap, explaining 60.4% of the total COPD heterogeneity, are described in Table 2.

The importance of factors was determined by scree plots, with the number of components retained in the rotated structure determined by eigenvalue one criterion, variables being sorted according to largest absolute loading. PEF variability is described as factor 5, and FEV_1 reversibility and pulse are described as factor 6 in the sub-analyses.

Stratified analysis

Factor analyses of the total and stratified groups are shown in Fig. 1, Table 4 and Supplementary Tables 2—5, respectively. Variables loading on different factors in the total and stratified analyses are highlighted.

Gender

Females were slightly younger than males (mean 62 vs. 64 years), had smoked less (39 vs. 47 pack-years), were more often current smokers (47% vs. 37%), and had slightly better lung function (post-bronchodilator FEV_1 43% vs. 40% predicted; Supplementary Table 6). SGRQ activity scores were higher (72 vs. 67) and comorbidities were more common (58% vs. 53%) in females than males. BMI and PEF variability were comparable between genders.

In females, BMI did not load with lung function (factor 2, 0.21) as in males (0.43) and the total group (0.40; Supplementary Table 2), instead loading with FEV_1 reversibility. Pulse changed its loading from FEV_1 reversibility (as in the total group) to BP (factor 3, 0.35). Gender differences in SGRQ activity scores necessitated the SGRQ total score being replaced by the sub-domains; this did not change the results (data not shown).

Table 2 Demographic and clinical characteristics loaded over six factors explaining 60.4% of the total COPD heterogeneity.

Factor		
1	16.2%	Breathlessness, cough and sleep disturbance due to respiratory symptoms, SGRQ total score, and reliever use
2	12.0%	Pre-bronchodilator FEV_1, pre-bronchodilator FEV_1/(F)VC, morning PEF, and BMI
3	10.3%	SBP and DBP
4	7.6%	Age and months since first COPD symptoms
5	7.1%	PEF variability
6	7.1%	FEV_1 reversibility and pulse

BMI, body mass index; COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; PEF, peak expiratory flow; SBP, systolic blood pressure; SGRQ, St George's Respiratory Questionnaire.

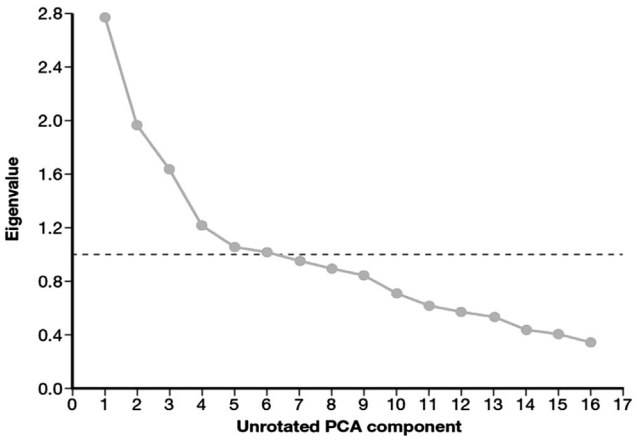

Figure 1 Scree plot of the full data set. PCA, principal component analysis.

Pack-years smoked

There were minimal differences in the demographic and clinical characteristics of patients with a smoking history <40 and ≥ 40 pack-years (Supplementary Table 7). Patient characteristics were distributed across the six factors in a similar manner for the pack-year sub-divisions (Supplementary Table 3) as the total group (Table 3).

In patients with a smoking history <40 pack-years, BMI loaded with lung function (factor 2, 0.48), consistent with the total group (0.40). However, BMI loaded almost equally to lung function and FEV_1 reversibility and pulse (0.36 and 0.38, respectively) in the group with ≥ 40 pack-years. Pulse loaded equally (-0.35) on FEV_1 reversibility and on age and months since first COPD symptoms (factor 4) in the group with <40 pack-years, in contrast to the total group, where pulse loaded with factor 6.

Smoking status

Current and ex-smokers had comparable demographic and clinical characteristics (Supplementary Table 8). Current smokers were slightly younger than ex-smokers (mean 60 vs. 66 years), with worse cough (1.9 vs. 1.4) and SGRQ symptom scores (69.3 vs. 61.6). However, tobacco exposure was similar between the groups (45 vs. 44 pack-years).

In current smokers, BMI loaded with lung function (factor 2, 0.51) but not FEV_1 reversibility and pulse (-0.18), consistent with BMI loadings in the total group (0.40 and 0.24, factor 2 and 6, respectively). In contrast, BMI did not load with lung function in ex-smokers (0.29), instead loading with factor 4 (0.48). Otherwise, loadings were comparable in terms of combinations and values, and when SGRQ sub-domains versus SGRQ total scores were analysed.

Comorbidities

Patients with and without self-reported comorbidities had comparable demographic and clinical characteristics

Table 3 Factor loadings in the total group of patients with COPD in the five studies ($n = 6162$).

Variable	1	2	3	4	5	6
Breath symptom score	***0.79***	−0.10	0.00	0.05	0.00	−0.02
Cough symptom score	***0.73***	0.13	−0.04	−0.12	−0.02	−0.04
SGRQ total score	***0.73***	−0.05	−0.01	−0.02	0.11	−0.11
Sleep symptom score	***0.70***	0.07	−0.00	−0.02	−0.09	0.04
Reliever use	***0.52***	−0.26	0.02	0.05	0.24	0.05
Pre-bronchodilator FEV_1, % predicted	−0.14	***0.81***	−0.02	−0.01	−0.03	0.03
Pre-bronchodilator FEV_1, % VC or FVC	0.06	***0.85***	−0.03	−0.03	0.04	−0.06
Morning PEF, % predicted	−0.23	***0.51***	0.07	−0.03	−0.40	0.37
BMI, kg/m^2	0.15	***0.40***	0.26	−0.01	0.09	0.24
DBP, mmHg	−0.01	0.00	***0.88***	−0.10	−0.03	−0.04
SBP, mmHg	−0.05	0.04	***0.86***	0.18	−0.01	−0.03
Age, years	−0.18	−0.04	0.02	***0.75***	0.01	−0.07
Months since first COPD symptoms	0.11	−0.01	0.04	***0.69***	−0.03	0.02
PEF variability, %	0.02	0.05	−0.01	−0.04	***0.94***	0.05
Pulse, bpm	0.04	−0.08	0.23	−0.27	0.07	***−0.45***
FEV_1 reversibility, % predicted	−0.03	0.01	0.06	−0.17	0.08	***0.84***
Variability explained, %[a]	16.2	12.0	10.3	7.6	7.1	7.1

PEF variability is defined as (morning PEF − evening PEF)/morning PEF.
BMI, body mass index; COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; GOLD, Global initiative for chronic Obstructive Lung Disease; PEF, peak expiratory flow; SBP, systolic blood pressure; SGRQ, St George's Respiratory Questionnaire; VC, vital capacity.
Figures in bold and italic indicate the loading of factors.
[a] Total = 60.4.

Table 4 Principal component analyses for the total group, and by stratifications, among patients with COPD.

Variable	Total	Gender		Smoking history (pack-years)		Smoking status		Absence/ presence of comorbidities	
		Male	Female	<40	≥40	Ex	Current	Absent	Present
Breath symptom	0.79	0.80	0.76	0.79	0.79	0.76	0.82	0.80	0.77
Cough symptom	0.73	0.74	0.73	0.74	0.73	0.74	0.75	0.76	0.72
SGRQ total	0.73	0.73	0.73	0.72	0.74	0.73	0.72	0.74	0.72
Sleep symptom	0.70	0.70	0.69	0.74	0.66	0.68	0.70	0.70	0.69
Reliever use	0.52	0.52	0.51	0.53	0.50	0.51	0.51	0.54	0.50
Pre-bronchodilator FEV_1, % predicted	0.81	0.80	0.84	0.78	0.83	0.84	0.77	0.83	0.83
Pre-bronchodilator FEV_1, % VC or FVC	0.85	0.85	0.82	0.85	0.84	0.83	0.84	0.84	0.85
Morning PEF, % predicted	0.51	0.51	0.49	0.49	0.53	0.55	0.49	0.53	0.49
BMI, kg/m^2	0.40	0.43	0.60[a]	0.48	0.36[d]	0.48[b]	0.51	0.33[f]	−0.30
DBP, mmHg	0.88	0.87	0.87	0.87	0.89	0.87	0.89	0.85	0.89
SBP, mmHg	0.86	0.86	0.85	0.85	0.87	0.86	0.89	0.85	0.89
Age, years	0.75	0.77	0.71	0.72	0.76	0.78	0.74	0.70	0.76
Months since first COPD symptoms	0.69	0.67	0.74	0.72	0.68	0.61	0.60	0.66	0.57
PEF variability	0.94	0.93	0.94	0.93	0.94	0.92	0.92	0.94	0.90
Pulse	−0.45	−0.42	0.35[c]	−0.35[e]	−0.41	−0.47	−0.67	0.79	−0.41
FEV_1 reversibility, % predicted	0.84	0.87	0.77	0.89	0.84	0.75	0.68	−0.50	0.88

BMI, body mass index; COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; GOLD, Global initiative for chronic Obstructive Lung Disease; PEF, peak expiratory flow; SBP, systolic blood pressure; SGRQ, St George's Respiratory Questionnaire; VC, vital capacity.
[a] Loads with FEV_1 reversibility.
[b] Loads with Age and Months since first COPD symptoms.
[c] Loads with blood pressure.
[d] Almost equally loading with reversibility (0.38).
[e] Equally loading on reversibility (−0.35), and on age and months since first COPD symptoms (−0.35).
[f] Loads on lung function (0.33) and on BP (0.40), see Supplementary Tables in the online data repository for further information. PEF variability is defined as (morning PEF − evening PEF)/(morning PEF).

(Supplementary Table 9). Patients without comorbidities versus with comorbidities were more often male (73% vs. 68%) and current smokers (42% vs. 38%). Patients without comorbidities also had a slightly lower BMI (24.8 vs. 27.2 kg/m^2), FEV_1 reversibility (5.4% vs. 5.7%), daily reliever use (3.3 vs. 3.6), breathlessness score (1.9 vs. 2.0), SGRQ total score (52.7 vs. 56.0), and SGRQ activity score (65.6 vs. 70.3), whereas pulse was comparable (78 vs. 78 bpm) between the groups.

BMI loaded comparatively on lung function (factor 2) and BP (factor 3) (0.33 and 0.40, respectively) in the group without comorbidities, and with lung function (factor 2), and age and months since first COPD symptoms (factor 4) in the group with comorbidities (0.30 and −0.36, respectively) (Supplementary Table 5). FEV_1 reversibility loaded similarly with factor 4 (−0.44) and factor 5 (−0.50 in combination with pulse) in the group without comorbidities. Results did not change when analysing SGRQ sub-domains versus the total score.

Exacerbations

Factor analysis on the studies with data on patient exacerbation history [12,14] (Supplementary Tables 10 and 11) showed that factors 1−4 were comparable to the total group; only reversibility shifted from pulse (factor 6) to PEF variability (factor 5). When adding prior exacerbations to the model, exacerbations loaded to factor 6 with pulse.

Discussion

We used a series of clinical and demographic characteristics to investigate the heterogeneity of moderate-to-very-severe COPD in a large group of COPD patients participating in five clinical trials [11−15]. Six clusters of characteristics explained ∼60% of COPD heterogeneity: 1) respiratory symptoms, reliever use, and HRQL, 2) lung function and BMI, 3) BP, 4) age and time since first COPD symptoms, 5) PEF variability, and 6) bronchodilator responsiveness and pulse.

All continuous baseline variables were assessed, without selection of variables, thereby ensuring an unbiased approach. Respiratory symptoms, reliever use, and HRQL loaded together strongly as one factor over all stratified analyses. This was also true for lung function variables, diastolic and systolic BP, age and time since first COPD symptoms, PEF variability, and FEV_1 reversibility. In contrast, BMI loaded with lung function in the total cohort and in males, those with <40 pack-years, current smokers, and those with comorbidities, but with FEV_1 reversibility in females, and with age and months since first COPD symptoms in ex-smokers. This is interesting given the independent relationship of BMI to accelerated lung function decline in similar patients [18]. Smoking history had minimal impact, as similar factor loadings existed in the pack-year stratifications, consistent with previous studies [10]. This does not mean that smoking is unimportant in disease development but that, once established, the prior smoking intensity is not a key variable distinguishing between subgroups of patients. Smoking status affected factor loadings in ex-smokers, with a shift in BMI from lung function to age and months since first COPD symptoms.

Females had different factor loadings to males, as pulse loaded with BP and BMI loaded with FEV_1 reversibility. Patients with and without comorbidities had similar factor loadings to the total group. Comorbidities in COPD patients require complex care, with potential diagnostic difficulties and increased hospitalisation risk. Further analyses are required to determine the risks associated with specific comorbidities. We found that most patient characteristics are robust and independent measures of COPD, while others (e.g. BMI) are influenced by gender and smoking cessation.

Prior exacerbations are an important determinant of future exacerbations, which are also a driver of COPD prognosis [8]. We therefore separately analysed the two cohorts with data on patient exacerbation history. Interestingly, this did not increase the explanation of variance in COPD heterogeneity (60.4% without exacerbations, 57.6% with exacerbations included), and exacerbations loaded on factors with the least explanatory value. Alternative clinically relevant factors may explain the remaining COPD variance, including emphysema type, small airway disease, and exercise capacity. Nevertheless, the current analysis has identified important variables explaining COPD heterogeneity and these were robust over the five cohorts studied.

Consistently with previous studies, we have demonstrated that the clinical manifestations of COPD are variable and that the degree of airflow limitation does not capture disease heterogeneity [4]. Few studies have used factor or cluster analyses in assessing the multidimensional nature of COPD. Celli et al. reported that six factors explained the heterogeneity of severe COPD (lung function, symptoms, health status, FEV_1 reversibility, BMI plus walking distance, and dyspnoea plus reliever use) [19]. In contrast, we have demonstrated that other variables can influence BMI. The differential findings may be related to sample size and gender differences between the studies; the analysed characteristics also differed, as Celli et al. did not examine PEF variability. Another study investigating COPD heterogeneity in patients with all COPD stages identified three factors; 1) SGRQ and modified Medical Research Council dyspnoea score, 2) age and smoking history, and 3) BMI and FEV_1 [20], consistent with our findings that FEV_1 loads with BMI while respiratory symptoms load with HRQL. We have extended previous findings by showing that FEV_1 reversibility and pulse load on one factor, and that PEF variability is independent. Although BMI was a variable factor, most other factors were robust throughout the stratified analyses.

A novel finding was that PEF variability does not cluster with other variables, appearing as a single factor. Circadian PEF variation has a weak relationship with hyperresponsiveness, symptoms, FEV_1, and bronchodilator response in asthma [21−23]. Therefore, these measures may provide different information regarding disease severity. Our data indicate that PEF variability is different from spirometry, FEV_1 reversibility, respiratory symptoms, and reliever use, consistent with reports that COPD symptoms worsen during the morning [24]. Further investigations should assess how increased PEF variability affects the

current and future disease status of patients, and the impact of treatment [25]. However, like bronchodilator reversibility, PEF variability explains only a small proportion of COPD heterogeneity.

This study has important strengths. The patient group was large with variability between patients in all analysed characteristics. Patient-oriented outcomes were included, but also BP and pulse, with potential implications for COPD management and outcomes. All available baseline variables were analysed, including surrogate measures of cardiovascular disease (concomitant medications), therefore the parameters were unbiased. All studies had similar inclusion/exclusion criteria, using the same data acquisition and collection methodology. We could therefore assess PEF variability as a parameter for COPD heterogeneity.

A study limitation is that patients in the current analysis were at the more severe end of the disease spectrum; patients with mild COPD were excluded and those with moderate COPD accounted for 21% of the study population. However, most patients seeking medical advice have more severe airflow limitation and are symptomatic. Our inclusion of patients with moderate-to-very-severe COPD reflects the typical patient population seeking medical advice, those most commonly seen in clinical practice, and those with the largest impact on COPD healthcare costs. COPD subtype classification may be beneficial for such patients, potentially leading to the development of more targeted and effective therapies. We have clustered comorbidities, as these reflect the complexity of disease care and diagnosis, and hospitalisation risk in COPD; this risk increases with more comorbidities [26,27]. Although our estimation of comorbidities was not as rigorous as with the prospective collection of data in a systematic manner, concomitant medications are likely to signal clinically important disease. The lower prevalence of comorbidities in, predominantly male, smokers may suggest underreporting and shows the need for prospective, objective studies. Physical activity data were not available from the five clinical trials we analysed. This parameter may influence the variable clustering variables, but recording such data requires patient attendance at specialist centres. A strength of our study is that simple measures available to many GPs, clinicians, and specialists were used, therefore the study has relevance to healthcare providers across the COPD field.

COPD prevalence in females is increasing, therefore analyses were stratified by gender. Consistent with other reports, females had worse HRQL despite slightly better post-bronchodilator FEV_1 values. Factor analysis showed BMI loading with lung function in males (as in the total group), but loading strongly with reversibility in females. This difference was not related to BMI differences (males 26.1 kg/m^2, females 26.2 kg/m^2), potentially reflecting the higher bronchodilator response in females [28]. Airway hyperresponsiveness has also been associated with female obesity [29]; the reason for this remains unclear.

Recent respiratory symptom measures in conjunction with other objective disease surrogates are needed to assess the complexity of COPD. Composite measures such as the body mass index, airflow obstruction, dyspnoea, and exercise capacity (BODE) index have been developed [30], but this does not fully capture the COPD spectrum. It is important to assess whether interventions aiming to improve these factors will change disease outcomes [25]. Our findings were robust in the five cohorts and the subanalysis in two cohorts with data on patient exacerbation history. Nevertheless the results need replication in large datasets, with patients from GP and specialist care centres, covering the full spectrum of COPD severity. Clinical and biological parameters signifying disease progression should also be examined.

Conclusions

We have identified variables in a large dataset that describes the heterogeneity of COPD. The six factors were robust, explaining ~60% of COPD heterogeneity in a population with predominantly severe but also some moderate disease, and include important determinants of disease outcome, such as gender, smoking, comorbidities, and exacerbations. Only BMI loaded to other factors, particularly in females and ex-smokers. Currently, all COPD patients are treated similarly, despite their clinical heterogeneity. Our study suggests that COPD can be described more comprehensively, using simple measures available to many healthcare professionals. Identification of COPD subtypes may lead to the development of more specific and effective COPD management strategies.

Funding

AstraZeneca.

Author contributions

DSP, ARA, CJ, BJM, TS, OÖ, GSE, and PMC contributed to conceiving, writing, and revising the manuscript. OÖ was responsible for statistical analyses.

Conflicts of interest

The University of Groningen has received honoraria for DSP advising on the conduct and analysis of clinical trial data from AstraZeneca, Nycomed, and Teva, as well as for lectures at meetings supported by AstraZeneca, Chiesi, GlaxoSmithKline, Nycomed, and Teva. The University of Groningen has received money for research by unrestricted educational grants from AstraZeneca and Chiesi; AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, and Nycomed have provided support for travel to meetings.

ARA is a consultant and speaker for AstraZeneca, Bayer Schering Pharma, Boehringer Ingelheim, Dey Pharma, GlaxoSmithKline, and Pfizer, and has received honoraria from these companies. Educational presentations have been developed for AstraZeneca, Bayer Schering Pharma, Boehringer Ingelheim, Dey Pharma, and Pfizer. Support for travel to meetings has also been provided by AstraZeneca.

CJ is a board member for AstraZeneca, GlaxoSmithKline, Merck Limited, and Novartis. Educational presentations have been developed for AstraZeneca and GlaxoSmithKline,

with grants also pending for these companies. Lectures have been presented on behalf of AstraZeneca, GlaxoSmithKline, Hunter Immunology, and Novartis. Support for travel to meetings has been provided by AstraZeneca.

BJM is a board member for AstraZeneca, Forest, Boehringer Ingelheim, Dey Pharma, Embryon, Johnson and Johnson, MedImmune, Novartis, Nycomed, Pfizer, and Respironics, and a consultant for Astellas and Chiesi. Clinical trial data has been reviewed for Spiration, with grants currently pending with AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, MedImmune, Nabi, Pfizer, and Sunovian. Lectures have been presented on behalf of Boehringer Ingelheim, GlaxoSmithKline, and Pfizer, with video presentations developed for Boehringer Ingelheim and questionnaires produced for UBC. Educational presentations and programs have been developed (Circulate WebMD, Creative Educational Concepts, France Foundation, Johns Hopkins University, Medscape, National Jewish Health, and VHA) and BJM has been a speaker for educational programs at Abbott, the American Academy of Family Practice, the American College of Chest Physicians, the American Thoracic Society, Breathe LA, the Cleveland Clinic, and VHA. Support for travel to meetings has also been provided by AstraZeneca.

TS is a board member for AstraZeneca, Boehringer Ingelheim France, Novartis France, Nycomed Corporate, Nycomed France, Pierre Fabre, and a consultant for Pierre Fabre and Rox biomedical. Grants are currently pending with Maquet France, Novartis France, and Pierre Fabre. Lectures have been presented on behalf of AstraZeneca France, Hamilton, Medapharma, Merck Sharp & Dohme France, and Pfizer France, with a manuscript prepared from Boehringer Ingelheim France. Honoraria have also been received from AstraZeneca, Atrostim, and Synapse biomedical. Support for travel to meetings has also been provided by AstraZeneca.

GSE and OÖ were full-time employees of AstraZeneca at the time the analyses were conducted. Both own stocks within the company.

PMC is a board member for Boehringer Ingelheim, the Department of Health Respiratory Programme Board, GlaxoSmithKline, and Nycomed. He has been a consultant for Novartis and provided expert testimony for Forest. PMC has received honoraria for advising on the conduct and analysis of clinical trial data from AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, and Nycomed. He has also spoken at meetings supported by these companies. Support for travel to meetings has been provided by AstraZeneca.

Acknowledgements

AstraZeneca, the study sponsor, was involved in the statistical data analysis, writing and revision of this manuscript, and the final decision to submit for publication. The authors thank Malin Fagerås and Anders Persson of AstraZeneca R&D, Mölndal, Sweden, for their assistance with the development of this manuscript and additional statistical analyses, respectively. Anna Mett and Natasha Hausman of inScience Communications, Springer Healthcare, provided editorial assistance (funded by AstraZeneca).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.rmed.2013.07.011.

References

[1] Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006;3:e442.
[2] 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;380:2163—96.
[3] Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. Lancet 2007;370:765—73.
[4] Agusti A, Calverley PM, Celli B, Coxson HO, Edwards LD, Lomas DA, et al. Characterisation of COPD heterogeneity in the ECLIPSE cohort. Respir Res 2010;11:122.
[5] Global Initiative for Chronic Obstructive Lung Disease. Global strategy for diagnosis, management, and prevention of COPD (2011 update). Available from: http://www.goldcopd.org/uploads/users/files/GOLD_Report_2011_Jan21.pdf; 2011.
[6] Drummond MB, Hansel NN, Connett JE, Scanlon PD, Tashkin DP, Wise RA. Spirometric predictors of lung function decline and mortality in early chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2012;185:1301—6.
[7] Hogg JC, Timens W. The pathology of chronic obstructive pulmonary disease. Ann Rev Pathol 2009;4:435—59.
[8] Hurst JR, Vestbo J, Anzueto A, Locantore N, Mullerova H, Tal-Singer R, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. N Engl J Med 2010;363:1128—38.
[9] Jones PW. Health status measurement in chronic obstructive pulmonary disease. Thorax 2001;56:880—7.
[10] Han MK, Agusti A, Calverley PM, Celli BR, Criner G, Curtis JL, et al. Chronic obstructive pulmonary disease phenotypes: the future of COPD. Am J Respir Crit Care Med 2010;182:598—604.
[11] Calverley PM, Boonsawat W, Cseke Z, Zhong N, Peterson S, Olsson H. Maintenance therapy with budesonide and formoterol in chronic obstructive pulmonary disease. Eur Respir J 2003;22:912—9.
[12] Rennard SI, Tashkin DP, McElhattan J, Goldman M, Ramachandran S, Martin UJ, et al. Efficacy and tolerability of budesonide/formoterol in one hydrofluoroalkane pressurized metered-dose inhaler in patients with chronic obstructive pulmonary disease: results from a 1-year randomized controlled clinical trial. Drugs 2009;69:549—65.
[13] Szafranski W, Cukier A, Ramirez A, Menga G, Sansores R, Nahabedian S, et al. Efficacy and safety of budesonide/formoterol in the management of chronic obstructive pulmonary disease. Eur Respir J 2003;21:74—81.
[14] Tashkin DP, Rennard SI, Martin P, Ramachandran S, Martin UJ, Silkoff PE, et al. Efficacy and safety of budesonide and formoterol in one pressurized metered-dose inhaler in patients with moderate to very severe chronic obstructive pulmonary disease: results of a 6-month randomized clinical trial. Drugs 2008;68:1975—2000.
[15] Welte T, Miravitlles M, Hernandez P, Eriksson G, Peterson S, Polanowski T, et al. Efficacy and tolerability of budesonide/formoterol added to tiotropium in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2009;180:741—50.
[16] Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow

limitation. The St. George's respiratory questionnaire. Am Rev Respir Dis 1992;145:1321—7.
[17] Manly BFJ. Multivariate statistical methods: a primer. London, New York: Chapman & Hall; 1986.
[18] Casanova C, de Torres JP, Aguirre-Jaíme A, Pinto-Plata V, Marin JM, Cordoba E, et al. The progression of chronic obstructive pulmonary disease is heterogeneous. Am J Respir Crit Care Med 2011;184:1015—21.
[19] Celli BR, Calverley PM, Rennard SI, Wouters EF, Agusti A, Anthonisen N, et al. Proposal for a multidimensional staging system for chronic obstructive pulmonary disease. Respir Med 2005;99:1546—54.
[20] Burgel PR, Paillasseur JL, Caillaud D, Tillie-Leblond I, Chanez P, Escamilla R, et al. Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. Eur Respir J 2010;36:531—9.
[21] Kerstjens HA, Brand PL, de Jong PM, Koeter GH, Postma DS. Influence of treatment on peak expiratory flow and its relation to airway hyperresponsiveness and symptoms. The Dutch CNSLD Study Group. Thorax 1994;49:1109—15.
[22] Clark TJ. Diurnal rhythm of asthma. Chest 1987;91:137S—41S.
[23] Oosterhoff Y, Kauffman HF, Rutgers B, Zijlstra FJ, Koeter GH, Postma DS. Inflammatory cell number and mediators in bronchoalveolar lavage fluid and peripheral blood in subjects with asthma with increased nocturnal airways narrowing. J Allergy Clin Immunol 1995;96:219—29.
[24] Kessler R, Partridge MR, Miravitlles M, Cazzola M, Vogelmeier C, Leynaud D, et al. Symptom variability in patients with severe COPD: a pan-European cross-sectional study. Eur Respir J 2011;37:264—72.
[25] Postma D, Anzueto A, Calverley P, Jenkins C, Make BJ, Sciurba FC, et al. A new perspective on optimal care for patients with COPD. Prim Care Respir J 2011;20:205—9.
[26] Mapel DW, Hurley JS, Frost FJ, Petersen HV, Picchi MA, Coultas DB. Health care utilization in chronic obstructive pulmonary disease: a case-control study in a health maintenance organization. Arch Intern Med 2000;160:2653—8.
[27] Soriano JB, Visick GT, Muellerova H, Payvandi N, Hansell AL. Patterns of comorbidities in newly diagnosed COPD and asthma in primary care. Chest 2005;128:2099—107.
[28] Lopez Varela MV, Montes de Oca M, Halbert RJ, Muino A, Perez-Padilla R, Talamo C, et al. Sex-related differences in COPD in five Latin American cities: the PLATINO study. Eur Respir J 2010;36:1034—41.
[29] Sood A, Dawson BK, Eid W, Eagleton LE, Henkle JQ, Hopkins-Price P. Obesity is associated with bronchial hyperresponsiveness in women. J Asthma 2005;42:847—52.
[30] Celli BR, Cote CG, Marin JM, Casanova C, Montes de Oca M, Mendez RA, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. N Engl J Med 2004;350:1005—12.