Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes

Judith Garcia-Aymerich,[1,2,3,4] Federico P Gómez,[5,6] Marta Benet,[1,2,4] Eva Farrero,[7] Xavier Basagaña,[1,2,4] Àngel Gayete,[8] Carles Paré,[9] Xavier Freixa,[9] Jaume Ferrer,[6,10] Antoni Ferrer,[2,6,11,16] Josep Roca,[5,6] Juan B Gáldiz,[12] Jaume Sauleda,[6,13,14] Eduard Monsó,[6,11,15] Joaquim Gea,[2,3,6,16] Joan A Barberà,[5,6] Àlvar Agustí,[5,6,14] Josep M Antó,[1,2,3,4] on behalf of the PAC-COPD Study Group

▶ Additional methods, tables and a figure are published online only. To view these files please visit the journal online (http://thorax.bmj.com).

For numbered affiliations see end of article.

Correspondence to
Judith Garcia-Aymerich, Centre for Research in Environmental Epidemiology (CREAL), Doctor Aiguader 88, 08003 Barcelona, Catalonia, Spain; jgarcia@creal.cat

For author footnote see end of the article.

ABSTRACT

Background Chronic obstructive pulmonary disease (COPD) is increasingly considered a heterogeneous condition. It was hypothesised that COPD, as currently defined, includes different clinically relevant subtypes.
Methods To identify and validate COPD subtypes, 342 subjects hospitalised for the first time because of a COPD exacerbation were recruited. Three months after discharge, when clinically stable, symptoms and quality of life, lung function, exercise capacity, nutritional status, biomarkers of systemic and bronchial inflammation, sputum microbiology, CT of the thorax and echocardiography were assessed. COPD groups were identified by partitioning cluster analysis and validated prospectively against cause-specific hospitalisations and all-cause mortality during a 4 year follow-up.
Results Three COPD groups were identified: group 1 (n=126, 67 years) was characterised by severe airflow limitation (postbronchodilator forced expiratory volume in 1 s (FEV_1) 38% predicted) and worse performance in most of the respiratory domains of the disease; group 2 (n=125, 69 years) showed milder airflow limitation (FEV_1 63% predicted); and group 3 (n=91, 67 years) combined a similarly milder airflow limitation (FEV_1 58% predicted) with a high proportion of obesity, cardiovascular disorders, diabetes and systemic inflammation. During follow-up, group 1 had more frequent hospitalisations due to COPD (HR 3.28, p<0.001) and higher all-cause mortality (HR 2.36, p=0.018) than the other two groups, whereas group 3 had more admissions due to cardiovascular disease (HR 2.87, p=0.014).
Conclusions In patients with COPD recruited at their first hospitalisation, three different COPD subtypes were identified and prospectively validated: 'severe respiratory COPD', 'moderate respiratory COPD', and 'systemic COPD'.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a heterogeneous disease with pulmonary and extrapulmonary manifestations involving a complex array of cellular, organic, functional and clinical events.[1–3] Understanding the phenotypic heterogeneity of COPD and resolving it in well defined subgroups might result in improved prognostic tools and better mechanistic studies,[4–6] and may prevent potentially useful treatments from being discarded on the basis of trials that mixed up various types of COPD.[7]

The interest in understanding COPD heterogeneity is not new. In the past, several COPD subtypes were described based on clinical[8] and epidemiological[9] observations. More recent studies used factor analysis to identify independent factors in COPD, such as exercise capacity and dyspnoea, airflow limitation, hyperinflation, airway inflammation or asthma features,[10–16] or cluster analysis to search for subgroups of COPD.[17 18] These studies, however, have important limitations. First, most of them used a very limited range of variables to describe COPD heterogeneity, and this carries the risk of providing spurious results.[6] Secondly, only one of them[18] included information on the extrapulmonary manifestations of COPD, a domain of the disease with a well established clinical relevance today.[2 3] Finally, and more importantly, none of them assessed the predictive validity of the new subtypes, so their clinical relevance could not be shown.

To address all these limitations, the 'Phenotype and Course of COPD (PAC-COPD)' study sought to identify clinically and epidemiologically meaningful COPD subtypes and to validate them by assessing their relationship with clinically relevant outcomes (hospitalisation and death) during a 4 year follow-up.[19] To this end, the PAC-COPD study adopted a wide multidimensional approach with a comprehensive clinical, functional, biological and imaging characterisation of a well defined cohort of patients with COPD at the time of their first hospital admission because of an exacerbation of the disease.[20]

METHODS

Design

The present study includes a cross-sectional analysis to identify COPD groups and a 4 year prospective assessment of their relationship with cause-specific admissions and all-cause mortality.

Subjects

We recruited all subjects hospitalised for the first time because of a COPD exacerbation in nine teaching hospitals in Spain between January 2004 and March 2006. The diagnosis of COPD was

confirmed by spirometry when the patient had reached clinical stability (see below), according to the criteria of the American Thoracic Society (ATS) and the European Respiratory Society (ERS) (postbronchodilator forced expiratory volume in 1 s (FEV_1) to forced vital capacity (FVC) ratio ≤0.7).[3] The protocol was approved by the Ethics Committees of all the participating hospitals, and written informed consent was obtained from all subjects. Detailed information about recruitment methods has been previously published.[21]

Measurements

Detailed information about the methods, questionnaires, standardisation of the tests and fieldwork supervision is provided in the Supplement material and has been previously reported.[19] Briefly, 3 months after discharge, when clinically stable, patients underwent a comprehensive characterisation that required from two to four visits, on separate days, to the participating hospitals, and included: (1) questionnaires covering sociodemographic and environmental exposure data, smoking habits, dietary habits, self-reported co-morbidities, previous treatments and diagnoses, respiratory symptoms, health-related quality of life, activities of daily living, sleepiness and psychological status; (2) detailed physical examination and Charlson co-morbidity index obtained by a respiratory physician participating in the study; (3) body composition determination by bioelectric impedance; (4) complete lung function tests including forced spirometry, bronchodilator test, body plethysmography, carbon monoxide diffusing capacity, resting arterial blood gases, respiratory and peripheral muscle strength, and night-time pulse oximetry; (5) 6-minute walk test (6MWT); (6) chest radiographs; (7) skin prick tests; (8) induced sputum; (9) blood laboratory analyses, including peripheral blood cell counts, cholesterol, triglycerides and total immunoglobulin E (IgE); (10) measurements of serum inflammatory and oxidative stress markers, centralised in a single laboratory; and (11) Doppler echocardiography evaluation. All these measurements were done in the total sample of patients (n=342). Additionally, due to their technical and logistical requirements, the following measurements could be performed only in some of the participating hospitals: (12) lung density and emphysema quantification from CT by a centralised evaluation using PulmoCT (Siemens, Munich, Germany) (subsample 1: three hospitals, 102 subjects); (13) semiquantitative evaluation of bronchial wall thickness on CT scans (subsample 2: four hospitals, 148 subjects); (14) microbiological culture of sputum (subsample 3: 8 hospitals, 224 subjects); (15) centralised measurement of sputum inflammatory markers and differential cell counts (subsample 4: 7 hospitals, 181 subjects); and (16) a cardiopulmonary incremental exercise test (CPET) with cycloergometer (subsample 5: 6 hospitals, 200 subjects). The overlap between patients in subsamples is shown in Supplementary table 1.

COPD and cardiovascular admissions up to 31 December 2007 were obtained from national administrative databases. Survival status until 31 December 2008 was obtained from direct interview of the patients or their relatives.

Statistical analysis

A detailed version of the statistical analysis and power estimation is available in the Supplement material.

From a total of 536 variables obtained, 224 were considered after excluding those with additive relationships or resulting from categorisations. Since missing values made up a small proportion and were considered to be either completely random or random,[22] multiple imputation through chained equations[23] was used to avoid losing data. Partitioning cluster analysis to group subjects according to distribution of variables (standardised using Z-scores) was done by the k-means method,[24] thus sorting participants into groups in a way that maximises differences between groups and minimises differences within groups. Six cluster analyses were built, one for the total sample and one for each of the five subsamples. Clusters in the subsamples included the 150 variables of the total sample in addition to the specific variables of each subsample. We compared the profile of the distribution of the 150 variables common to cluster analyses in the total sample and in each subsample across the three groups identified separately in each of the six cluster analyses, and tested the individual agreement in the assigned cluster group using Kappa statistics.

To display graphically the COPD groups identified by cluster analysis, we used standardised values of all variables, keeping or changing (multiplying by −1) their sign for all values ranging from +1 (less impairment) to −1 (more impairment), and plotted these values with a colour intensity scale spanning from yellow to red, respectively. To present the relative relevance of each variable to the separation into cluster groups, F values were computed as the ratio of the variance of the group means (between-group variance) over the overall variance of the variable (higher values meaning higher relevance), and was obtained by means of linear regression models using the variable under study as the outcome and the cluster group as the exposure. Sociodemographic characteristics, lifestyle factors and environmental exposures were tested as potential determinants of cluster groups.

To validate the clinical relevance of the groups identified by cluster analysis, admissions and mortality during follow-up were compared across groups using Kaplan—Meier curves and log-rank tests. The influence of COPD severity in the association between cluster groups and longitudinal outcomes was assessed including either ATS/ERS severity stages[3] or the U-BODE (updated body mass index, airflow obstruction, dyspnoea and exercise capacity) index[20] in multivariate models. As sensitivity analyses, we repeated cluster analyses excluding subjects with pneumonia as the concomitant cause for their first COPD hospitalisation, as well as restricting it to males. We also implemented a previously published variable selection procedure that allows dropping noisy, non-informative or redundant variables, thus keeping a smaller number of variables that were used to reconduct the cluster analysis.[25] Analyses were performed with Stata release 10.0 (2008, StataCorp LP), and R 2.6.2 (2008, R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Among 604 elegible patients with COPD, the PAC-COPD cohort included a total of 342 (57%) participants. There were no relevant differences between participants and non-participants. Supplementary figure 1 shows details about exclusions and non-participation. PAC-COPD subjects were mainly men (93%), with a mean age of 68 years and a mean postbronchodilator FEV_1 during clinical stability of 52% predicted (table 1).

The number of cluster groups when using k-means needs to be prespecified. Since we had no a priori information for a rational choice, we looked at the results of two and three cluster groups. Figure 1 presents the distribution of variables according to both options. Classifying participants in three groups maximised the differences for systemic inflammation, nutritional status, cardiovascular disorders and lung density, while keeping a similar contrast for the remaining dimensions. Classification in

Table 1 Characteristics of the study participants (n=342) at their first chronic obstructive pulmonary disease (COPD) admission and during clinical stability*

	At first admission	During clinical stability at least 3 months after the admission
Gender: female, n (%)	24 (7.0)	—
Age (years), mean (SD)	67.5 (8.5)	67.9 (8.6)
Low (IV, V) socioeconomic status, n (%)	259 (81.7)	—
Active workers, n (%)	61 (17.8)	54 (15.8)
Smoking status: current smoker, n (%)	145 (42.4)	109 (32.9)
Charlson index of co-morbidity (score 0—30), median (P25—P75)	2 (1—2)	2 (1—3)
Modified Medical Research Council dyspnoea scale (score 0—5), median (P25—P75)	3 (2—4)	2 (2—3)
Postbronchodilator forced expiratory volume in 1 s (% predicted), mean (SD)	49.2 (15.7)†	52.4 (16.2)
Arterial oxygen partial pressure (mm Hg), mean (SD)	64.8 (12.0)†	74.3 (10.6)
Any respiratory drug treatment, n (%)	193 (56.4)	288 (85.2)

*Subjects were recruited during their first hospital admission because of a COPD exacerbation. The assessment of phenotype heterogeneity (through the several tests and questionnaires) was done when the patient had reached clinical stability and at least 3 months after discharge (see text for more details).
†Measured at hospital discharge.

four or more groups did not result in meaningful patterns or statistical advantage. Thus, separation of the study subjects into three different COPD groups, including 126 (37%), 125 (36%) and 91 (27%) patients, was considered the most clinically meaningful option.

Group 1 was characterised by a higher prevalence and severity of respiratory symptoms, poorer quality of life, worse lung function, lower exercise capacity, lower lung density and thickened bronchial walls (table 2). Group 2 exhibited less overall respiratory impairment than group 1 except for lung density, which was similar in both groups. Group 3 showed a respiratory profile similar to group 2 but with a greater prevalence of overweight, systemic inflammation, cardiovascular diseases and diabetes. The complete distribution of values for the 224 variables is provided in Supplementary table 3. Sensitivity analyses showed very similar results after excluding patients with pneumonia as the concomitant cause of their first COPD hospitalisation, or after restricting the analysis to males. The variable selection procedure identified that 10 variables (dyspnoea (modified Medical Research Council scale), St George's Respiratory Questionnaire-Activity, prebronchodilator and postbronchodilator FEV_1 (% predicted), thoracic gas volume (TGV; % predicted), inspiratory to total lung capacity (IC/TLC) ratio, PaO_2, peripheral blood neutrophil count, body weight and body mass index (BMI)) could be used to obtain the same cluster groups with a mean concordance of 81% (table 2).

No relevant differences in potential risk factors were observed between the three COPD groups (table 3). Remarkably, all groups were of similar age. Smoking status was very similar among them, but smoking history showed that group 2 quit smoking at a younger age (a mean of 3 years earlier), and that group 3 smoked more pack years.

During follow-up, group 1 had more frequent hospitalisations due to COPD, and the highest all-cause mortality, whereas group 3 had more admissions due to cardiovascular disease (figure 2). After adjusting for disease severity, the observed differences in admissions were unchanged (risk of COPD admission in group 1 compared with group 2: unadjusted HR=3.28, ATS/ERS stages adjusted HR=2.89 and U-BODE adjusted HR=2.45; risk of cardiovascular admission in group 3 compared with group 2: unadjusted HR=2.87, ATS/ERS stages adjusted HR=2.88 and U-BODE adjusted HR=2.27). The differences in mortality remained similar but lost statistical

significance (risk of mortality in group 1 compared with group 2: unadjusted HR=2.36, ATS/ERS stages adjusted HR=2.01 and U-BODE adjusted HR=1.41) (Supplementary table 3).

DISCUSSION

In this cohort of 342 patients recruited at the time of their first COPD admission three different groups of COPD patients were identified. Because each group included a large proportion of subjects, and the groups exhibited marked clinical differences, and were associated with different longitudinal outcomes, we propose that they are clinically relevant COPD subtypes. Subtype 1 showed the worst status in most of the respiratory domains of the disease and exercise capacity. Subtype 2 was characterised by a milder respiratory status than subtype 1, and subtype 3 also by a milder respiratory status but a higher prevalence of obesity, cardiovascular disease and diabetes, and higher levels of systemic inflammatory markers.

Previous literature has suggested that part of the phenotypic heterogeneity in COPD is due to a divergent distribution of bronchial airway (chronic bronchitis) and parenchymal disease (emphysema).[1] In our study this was observed only in subtype 2, with a substantial degree of emphysema in the absence of increased wall thickness, both according to CT measures. We also found that asthma-like variables did not contribute to phenotypic heterogeneity, contrasting with other publications.[9] Our results cannot be directly compared with the studies applying factor analysis in COPD,[10–17] since these studies identified a relatively large number of independent disease components from a reduced number of variables. Cluster analysis, instead, allows segregating subjects in different groups.[27] Its recent use[18] in patients with COPD identified four groups, one of them including obesity and chronic heart failure, in keeping with our subtype 3.

Most of the features of the so-called metabolic syndrome, including obesity, higher levels of triglycerides, diabetes, ischaemic heart disease, arterial hypertension and elevated serum levels of C-reactive protein and fibrinogen, were clustered in subtype 3. Fabbri and Rabe[28] proposed that a 'chronic systemic inflammatory syndrome' may contribute to the co-morbidities that frequently co-exist in patients with COPD. Our results give preliminary support to this concept because, first, we found that subtype 3 was characterised by both elevated
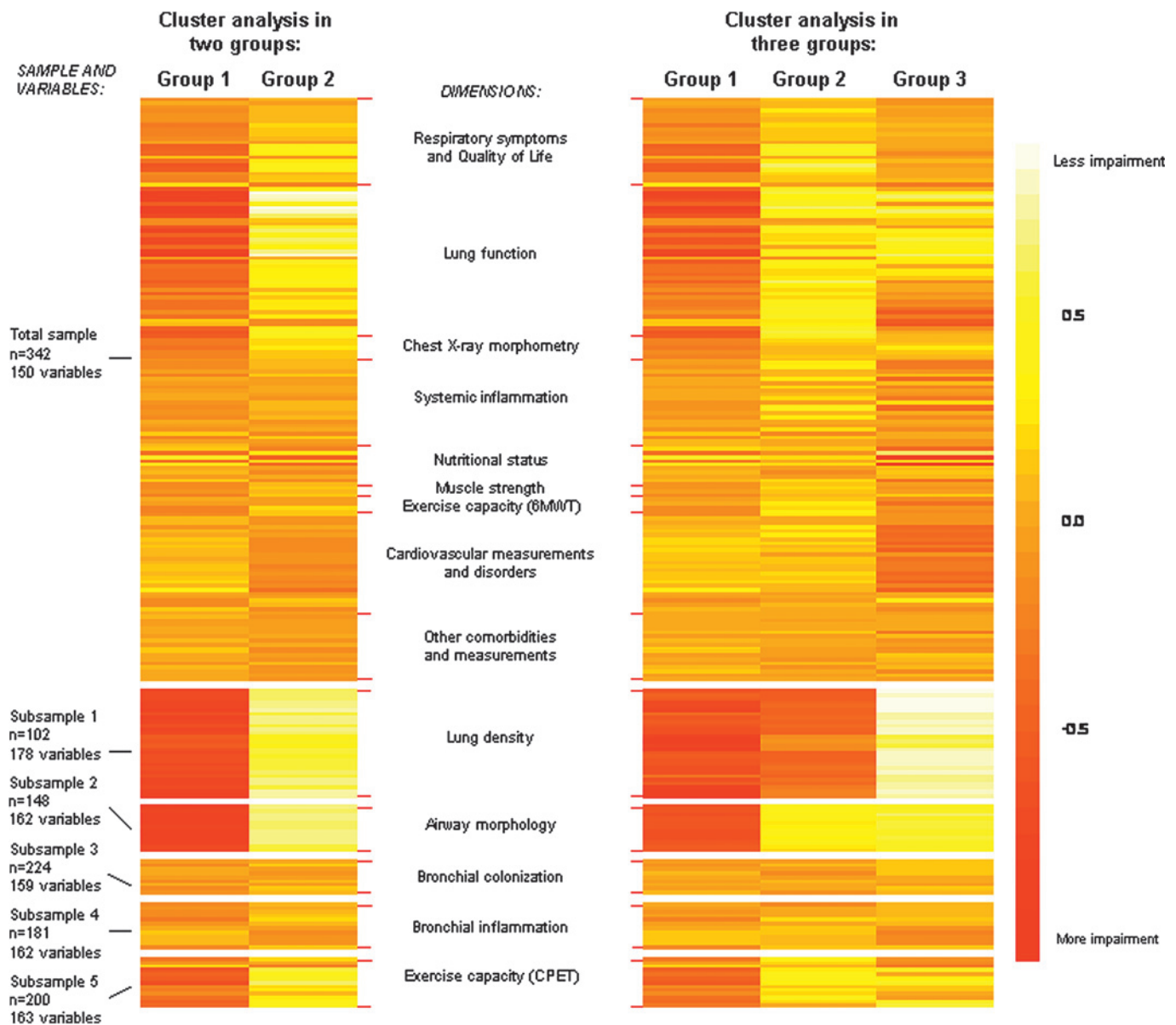
Figure 1 Graphical description of chronic obstructive pulmonary disease (COPD) groups identified by the cluster analysis. Each coloured line represents a variable, whose level of impairment ranges from −1 (more impairment, in red) to +1 (less impairment, in yellow). Values from −1 to +1 correspond to standardised Z-scores, where 0 corresponds to the mean of the total population included in each sample or subsample. Six cluster analyses were built, and, given their high concordance, their results are displayed using the same group categories (see Supplementary table 3 for complete data of all cluster results and their concordance). To make comparisons easier to interpret, we grouped variables under dimensions, as previously published.[26]

markers of systemic inflammation and a high prevalence of co-morbidities; of note, 17% of patients in subtype 3 had congestive heart failure, a figure that contrasts markedly with that seen in subtypes 1 (5%) and 2 (1%). Secondly, obesity might have contributed to systemic inflammation in subtype 3, as previously suggested.[29] Finally, consistently with the hypothesis of the 'chronic systemic inflammatory syndrome',[28] subtype 3 had a higher risk of COPD and cardiovascular admissions than subtype 2. In contrast, the fact that we did not identify significant differences among subtypes in bronchial inflammatory markers and, nonetheless, subtype 3 showed more systemic inflammation than the other two, does not support previous claims that pulmonary inflammation could 'spill-over' from the lungs.[30]

Despite the fact that airflow limitation in subtype 3 was moderate and similar to that seen in subtype 2, patients in the former subtype showed more dyspnoea and poorer health-related quality of life and exercise capacity (actually quite similar to those seen in subtype 1 with much more severe airflow limitation). This is in agreement with the currently accepted apparent paradox that lung function and other clinical parameters are weakly correlated in COPD,[31][32] and might be explained by the combined cardiorespiratory morbidity in subtype 3. As a consequence, it is possible that in clinical trials, this subtype may respond poorly to interventions targeting airflow limitation.[7] Equally important is that, in clinical practice, these patients may benefit from an integrated cardiorespiratory assessment and therapeutic plan.

One salient feature of our study is the large differences in airflow limitation seen across the three identified subtypes, since all patients were hospitalised for the first time due to a COPD exacerbation and had similar age and smoking trajectories. The

Table 2 Description* of the three chronic obstructive pulmonary disease (COPD) groups identified by cluster analysis

	All subjects n=342	Group 1 n=126	Group 2 n=125	Group 3 n=91	F†
Cough day or night, %	49	62	38	44	8.19
Sputum day or night, %	53	63	46	51	4.05
Asthma (self-reported), %	9	11	8	9	0.22
Dyspnoea (modified Medical Research Council scale, 0–5), mean	2.6	3.2	1.9	2.8	33.53
St George Respiratory Questionnaire-Symptoms (0–100), mean	49	54	44	47	11.35
St George Respiratory Questionnaire-Activity (0–100), mean	47	61	32	49	60.01
St George Respiratory Questionnaire-Impacts (0–100), mean	27	36	17	26	42.62
Prebronchodilator forced expiratory volume in 1 s (FEV_1) (% predicted), mean	49	35	58	55	139.74
Postbronchodilator FEV_1 (% predicted), mean	52	38	63	58	148.99
Postbronchodilator FEV_1 to forced vital capacity ratio, mean	0.53	0.44	0.57	0.61	101.49
FEV_1 bronchodilation change (% of prebronchodilator FEV_1), mean	9.0	9.9	9.1	7.6	1.04
Thoracic gas volume (% predicted), mean	137	161	129	114	52.98
Inspiratory capacity/total lung capacity ratio, mean	0.31	0.23	0.34	0.37	95.26
Carbon monoxide diffusing capacity (% predicted), mean	65	54	73	70	13.67
Arterial oxygen partial pressure (mm Hg), mean	74	70	81	72	47.39
Time with arterial oxygen saturation <90% when sleeping (%), mean	19	26	6	27	23.88
Minimum arterial oxygen saturation during 6-minute walk test (%), mean	89	86	92	90	33.13
C-reactive protein in serum (mg/dl), mean	0.86	0.99	0.36	1.37	8.46
Fibrinogen in serum (g/l), mean	4.13	4.23	3.77	4.50	11.40
Tumour necrosis factor in serum (pg/ml), mean	0.87	0.88	0.72	1.05	1.37
Interleukin 6 in serum (pg/ml), mean	1.45	1.52	1.12	1.81	5.88
Peripheral blood neutrophil count ($n\times10^9$/l), mean	4.62	4.76	3.94	5.35	23.55
Skin prick test (number of positive allergens), mean	0.18	0.13	0.18	0.25	1.53
Body weight (kg), mean	76	69	74	90	87.04
Height (cm), mean	1.64	1.62	1.64	1.66	11.24
Body mass index (kg/m^2), mean	28.2	26.0	27.2	32.4	78.30
Fat free mass index (kg/m^2), mean	19.6	18.6	19.3	21.6	29.53
6-Minute walk total distance (m), mean	434	414	466	417	8.52
Myocardial infarction (doctor diagnosed), %	11	6	7	23	9.40
Congestive heart failure (doctor diagnosed), %	7	5	1	17	11.14
Left ventricle telediastolic diameter (mm), mean	50.5	49.6	50.1	52.3	5.34
Ejection fraction (%), mean	59.3	59.7	58.8	59.5	0.25
Tricuspid insufficiency, %	58	66	61	43	4.63
Diabetes (doctor diagnosed), %	19	16	14	32	6.11
Density less than −950 Hounsfield units in both lung infracarinal thin sections (%)‡, mean	14	24	16	6	34.03
Bronchial wall thickness extension in right inferior lobe (score 0–3)‡, mean	1.9	2.7	1.5	1.1	47.26
Potentially pathogenic microorganism colonisation‡, %	34	36	35	26	0.79
Interleukin 8 ($\times10^3$) in sputum supernatant (pg/ml)‡, mean	10.11	11.90	8.54	8.98	2.72
Neutrophil count in sputum ($n\times10^6$/ml)‡, mean	370	250	290	570	1.59
Maximal oxygen uptake at cardiopulmonary incremental exercise test peak (% predicted)‡, mean	62	53	69	68	19.93

*Non-standardised values of a selection of 40 out of 224 variables in figure 1, including those obtained with the variable selection procedure (**bold** text) and variables selected according to clinical experts' knowledge. Complete data are provided in Supplementary table 3.

†F values correspond to the ratio of the variance of the group means (between-group variance) over the overall variance of the variable (higher values meaning higher relevance of the variable for separating cluster groups), and were obtained by means of linear regression models using each variable as the outcome, and the cluster group as the exposure (see Supplement for details).

‡Variables obtained in a subsample of patients (see text for details).

possibility that patients in subtype 1 might be poor perceivers of dyspnoea, a hypothesis that has not been formally explored in COPD but is well established in asthma,[33] should not be excluded. The fact that patients in subtype 2 require hospitalisation despite much milder airflow limitation may be explained by a remarkable degree of emphysema—according to CT scan—which could result in a different clinical presentation at the emergency room, leading to an admission. This subtype reported the lowest proportion of respiratory diagnosis and treatment before recruitment (data not shown) and could represent the largest window of opportunity for improving early diagnosis and management of COPD.

Although smoking is the main risk factor for COPD,[2] we found no evidence that it may be an important source of COPD heterogeneity, though the larger number of pack-years reported by subtype 3 may have contributed to their higher prevalence of cardiovascular and metabolic disease. Likewise, other exposures to occupational or environmental factors were not significantly different among COPD phenotypes although, admittedly, limited information was available. Finally, it is worth noting that patients in subtype 1 were, on average, 4 cm shorter than those in subtype 3. Since height is considered a marker of in utero and childhood lung growth, we hypothesise that subjects in subtype 1 might have had impaired lungs since early life and, for the same amount of smoking, developed a more severe respiratory status at the time of their first hospital admission. The latter observation is consistent with recent reports on the early origin of COPD.[34] As height was included in the cluster analysis, its interpretation as a potential risk factor can be argued.

Some potential limitations of our study also need discussion. The proportion of women included in the study was low, so its

Table 3 Differences in sociodemographic characteristics, lifestyle factors and environmental exposures across the three chronic obstructive pulmonary disease (COPD) groups identified by cluster analysis

	Group 1 n=126	Group 2 n=125	Group 3 n=91	p Value
Sociodemographic characteristics				
Gender: female, n (%)	8 (6.4)	11 (8.8)	5 (5.5)	0.601
Age (years), mean (SD)	67.4 (8.8)	68.8 (8.1)	67.4 (9.0)	0.346
Low (IV, V) socioeconomic status, n (%)	94 (79.0)	97 (82.2)	68 (85.0)	0.552
Lifestyle factors				
Smoking status:				
Never, n (%)	0 (0)	2 (1.6)	0 (0)	0.418
Former, n (%)	77 (63.1)	84 (68.9)	59 (67.8)	
Current, n (%)	45 (36.9)	36 (29.5)	28 (32.2)	
Age at onset of smoking (years), mean(SD)	15.3 (4.4)	16.2 (4.8)	16.9 (6.9)	0.104
Age at quitting smoking (years), mean (SD)	60.7 (9.1)	57.0 (11.6)	60.3 (10.1)	0.068
Time since quitting smoking (years), median (P25—P75)	1 (0—11)	1 (0—15)	1 (0—10)	0.291
Pack-years smoking, mean (SD)	68.6 (39.4)	59.4 (36.8)	79.4 (39.8)	0.001
Smoking duration (years), mean (SD)	45.8 (10.0)	42.8 (13.5)	44.5 (11.4)	0.121
Smoking intensity (pack/day), mean (SD)	1.2 (0.9)	1.1 (0.9)	1.4 (1.0)	0.098
Usual physical activity (kcals/week), median (P25—P75)	4508 (2412—8640)	6270 (4060—8757)	6366 (2655—9900)	0.020
Glucid intake (%), median (P25—P75)	39.9 (34.9—44.4)	38.7 (34.6—43.1)	37.9 (31.9—42.6)	0.276
Protein intake (%), median (P25—P75)	19.7 (17.5—22.2)	19.1 (17.5—22.3)	19.8 (17.6—21.5)	0.878
Lipid intake (%), median (P25—P75)	36.0 (32.0—39.4)	34.1 (31.1—38.3)	34.1 (30.1—38.3)	0.141
Fruit intake (g/d), median (P25—P75)	223 (147—327)	269 (176—373)	238 (160—350)	0.093
Vegetable intake (g/d), median (P25—P75)	291 (218—398)	295 (219—369)	243 (196—345)	0.204
Environmental exposures				
Childhood exposure to wood or charcoal smoke, n (%)	98 (77.8)	106 (84.8)	68 (74.7)	0.160
Occupational exposure to vapours, gases, dust and fums				
No/low exposure, n (%)	51 (46.4)	34 (30.9)	25 (22.7)	0.060
Moderate exposure, n (%)	28 (28.0)	42 (42.0)	30 (30.0)	
High exposure, n (%)	30 (30.9)	38 (39.2)	29 (29.9)	
Distance to the nearest high traffic road:				
0—50 m, n (%)	99 (38.1)	93 (35.8)	68 (26.2)	0.881
51—200 m, n (%)	19 (35.2)	21 (38.9)	14 (25.9)	
>201 m, n (%)	6 (27.3)	9 (40.9)	7 (31.8)	

results should be extrapolated to females with caution. The intensive characterisation of this study was not feasible outside a hospital setting, so identification and validation of COPD subtypes in different COPD populations is needed. Because of such intensive characterisation, we did not include patients with severe co-morbidities, although we consider it unlikely that this would have distorted the observed subtypes given that they were of similar ages. Our study was restricted to patients who

Figure 2 Patterns of hospital admissions and mortality during follow-up, according to the three chronic obstructive pulmonary disease (COPD) groups identified by cluster analysis.
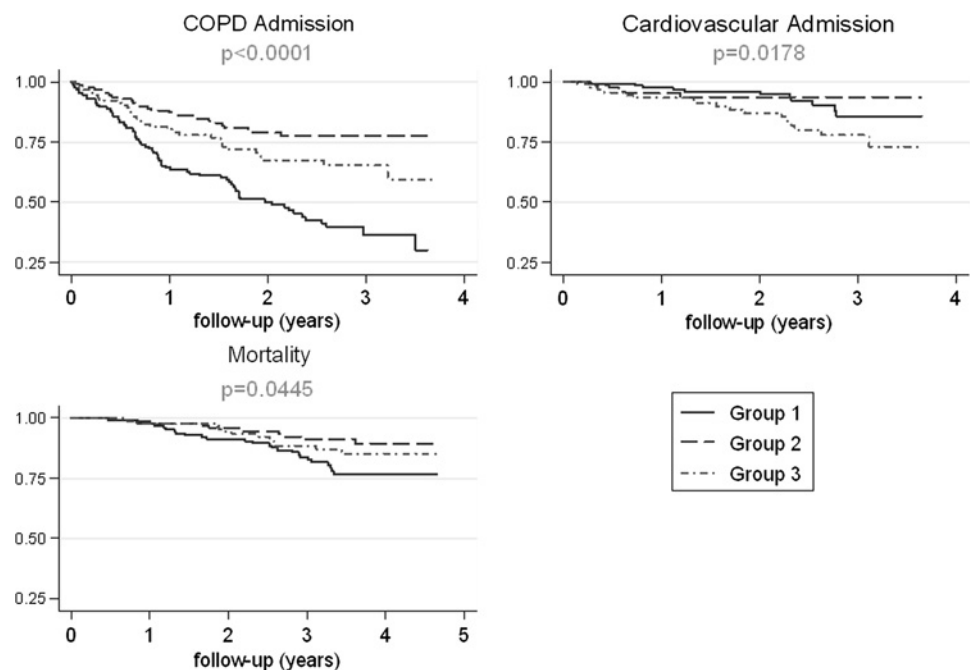
were hospitalised for COPD and thus our findings may be considered to apply only to relatively severe COPD. On the other hand, the selection of patients at their first hospital admission due to COPD should be seen as a strength that allows for a more valid study of the phenotypic heterogeneity of COPD.[35] Finally, our cluster analysis was cross-sectional, and assessing the temporal stability of the identified COPD subtypes remains an important issue.

The most obvious clinical implication of our study is that the first COPD hospitalisation offers an opportunity for a wider than usual characterisation of patients with COPD. Actually, that the longitudinal association between subtype 1 and COPD admissions remained after adjusting for the two currently used criteria to assess COPD severity[3 32] suggests that our subtype classification builds relevant prognostic information that is not accounted for in these criteria. Importantly, only 10 variables that are usually collected in clinical practice would be enough to separate patients into the described cluster groups. Further research aiming to develop an allocating algorithm is needed. Of course, experimental studies will be needed to assess whether therapeutic strategies tailored to these subtypes can modify the natural history of the disease.

In conclusion, in patients with COPD recruited at their first hospitalisation, three different COPD subtypes have been identified and prospectively validated, which we propose to label as 'severe respiratory COPD', 'moderate respiratory COPD', and 'systemic COPD'.

Author affiliations
[1]Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
[2]Municipal Institute of Medical Research (IMIM-Hospital del Mar), Barcelona, Spain
[3]Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain
[4]CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain
[5]Department of Pneumology, Institut del Tòrax, Hospital Clínic-Institut D'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, Barcelona, Spain
[6]CIBER de Enfermedades Respiratorias (CIBERES), Bunyola, Spain
[7]Department of Pneumology, Hospital Universitari de Bellvitge, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), L'Hospitalet de Llobregat, Spain
[8]Department of Radiology, Hospital del Mar-IMIM, Barcelona, Spain
[9]Department of Cardiology, Hospital Clínic-Institut D'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, Barcelona, Spain
[10]Department of Pneumology, Hospital General Universitari Vall D'Hebron, Barcelona, Spain
[11]Department of Pneumology, Hospital de Sabadell, Corporació Parc Taulí, Sabadell, Spain
[12]Department of Pneumology, Hospital Universitario de Cruces, UPV, Barakaldo, Spain
[13]Department of Pneumology, Hospital Universitari Son Dureta, Palma de Mallorca, Spain
[14]Fundació Caubet-Cimera, Bunyola, Spain
[15]Department of Pneumology, Hospital Universitari Germans Trias i Pujol, Badalona, Spain
[16]Department of Pneumology, Hospital del Mar-IMIM, Barcelona, Spain

Author footnote
FPG and MB contributed equally to this work.

The 'Phenotype and Course of COPD (PAC-COPD)' Study Group: Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Josep M Antó (Principal Investigator), Judith Garcia-Aymerich (project coordinator), Marta Benet, Jordi de Batlle, Ignasi Serra, David Donaire-Gonzalez, Stefano Guerra; Hospital del Mar-IMIM, Barcelona, Joaquim Gea (centre coordinator), Eva Balcells, Àngel Gayete, Mauricio Orozco-Levi, Ivan Vollmer; Hospital Clínic-Institut D'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Joan Albert Barberà (centre coordinator), Federico P Gómez, Carles Paré, Josep Roca, Robert Rodriguez-Roisin, Àlvar Agustí, Xavier Freixa, Diego A Rodriguez, Elena Gimeno, Karina Portillo; Hospital General Universitari Vall D'Hebron, Barcelona, Jaume Ferrer (centre coordinator), Jordi Andreu, Esther Pallissa, Esther Rodríguez; Hospital de la Santa Creu i Sant Pau, Barcelona, Pere Casan (centre coordinator), Rosa Güell, Ana Giménez; Hospital Universitari Germans Trias i Pujol, Badalona, Eduard Monsó (centre coordinator), Alicia Marín, Josep Morera; Hospital Universitari de Bellvitge, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), L'Hospitalet de Llobregat, Eva Farrero (centre coordinator), Joan Escarrabill; Hospital

de Sabadell, Corporació Parc Taulí, Institut Universitari Parc Taulí (Universitat Autònoma de Barcelona), Sabadell, Antoni Ferrer (centre coordinator); Hospital Universitari Son Dureta, Palma de Mallorca, Jaume Sauleda (centre coordinator), Bernat Togores; Hospital Universitario de Cruces, UPV, Barakaldo, Juan Bautista Gáldiz (centre coordinator), Lorena López; Instituto Nacional de Silicosis, Oviedo, José Belda.

REFERENCES
1. **Calverley PM,** Walker P. Chronic obstructive pulmonary disease. *Lancet* 2003;**362**:1053—61.
2. **Global Initiative for Chronic Obstructive Lung Disease.** Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (updated 2009). http://www.goldcopd.org (accessed 30 Mar 2010).
3. **Celli BR,** MacNee W; ATS/ERS Task Force. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004;**23**:932—46.
4. **Wenzel SE.** Asthma: defining of the persistent adult phenotypes. *Lancet* 2006;**368**:804—13.
5. **Anon.** A plea to abandon asthma as a disease concept. *Lancet* 2006;**368**:705.
6. **Thornton-Wells TA,** Moore JH, Haines JL. Genetics, statistics and human disease: analytical retooling for complexity. *Trends Genet* 2004;**20**:640—7.
7. **Reilly JJ.** COPD and declining FEV1—time to divide and conquer? *N Engl J Med* 2008;**359**:1616—18.
8. **Dornhorst AC.** Respiratory insufficiency. *Lancet* 1955;**268**:1185—7.
9. **Burrows B,** Bloom JW, Traver GA, et al. The course and prognosis of different forms of chronic airways obstruction in a sample from the general population. *N Engl J Med* 1987;**317**:1309—14.
10. **Ries AL,** Kaplan RM, Blumberg E. Use of factor analysis to consolidate multiple outcome measures in chronic obstructive pulmonary disease. *J Clin Epidemiol* 1991;**44**:497—503.
11. **Mahler DA,** Harver A. A factor analysis of dyspnea ratings, respiratory muscle strength, and lung function in patients with chronic obstructive pulmonary disease. *Am Rev Respir Dis* 1992;**145**:467—70.
12. **Wegner RE,** Jorres RA, Kirsten DK, et al. Factor analysis of exercise capacity, dyspnoea ratings and lung function in patients with severe COPD. *Eur Respir J* 1994;**7**:725—9.
13. **Kostianev SS,** Hodzhev VA, Todorov IT, et al. A factor analysis of dyspnea indexes and lung function parameters in patients with chronic obstructive pulmonary disease. *Folia Med (Plovdiv)* 2001;**43**:27—31.
14. **Fuchs-Climent D,** Le Gallais D, Varray A, et al. Factor analysis of quality of life, dyspnea, and physiologic variables in patients with chronic obstructive pulmonary disease before and after rehabilitation. *Am J Phys Med Rehabil* 2001;**80**:113—20.

15. **Lapperre TS,** Snoeck-Stroband JB, Gosman MM, et al. Dissociation of lung function and airway inflammation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2004;**170**:499—504.
16. **Roy K,** Smith J, Kolsum U, et al. COPD phenotype description using principal components analysis. *Respir Res* 2009;**10**:41.
17. **Weatherall M,** Travers J, Shirtcliffe PM, et al. Distinct clinical phenotypes of airways disease defined by cluster analysis. *Eur Respir J* 2009;**34**:812—18.
18. **Burgel PR,** Paillasseur JL, Caillaud D, et al. Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. *Eur Respir J* 2010;**36**:531—9.
19. **Garcia-Aymerich J,** Gómez FP, Antó JM. [Phenotype characterisation and course of chronic obstructive pulmonary disease (COPD) in the PAC-COPD Study: design and methods] (in Spanish). *Arch Bronconeumol* 2009;**45**:4—11.
20. **Puhan MA,** Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet* 2009;**374**:704—11.
21. **Balcells E,** Antó JM, Gea J, et al. Characteristics of patients admitted for the first time for COPD exacerbation. *Respir Med* 2009;**103**:1293—302.
22. **Rubin DB.** Inference and missing data. *Biometrika* 1976;**63**;581—92.
23. **van Buuren S,** Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;**18**:681—94.
24. **Steinley D.** K-means clustering: a half-century synthesis. *Br J Math Stat Psychol* 2006;**59**:1—34.
25. **Fraiman R,** Justel A, Svarc M. Selection of variables for cluster analysis and classification rules. *J Am Stat Assoc* 2008;**103**:1294—303.
26. **Garcia-Aymerich J,** Agustí A, Barberà JA, et al. [Phenotypic heterogeneity of chronic obstructive pulmonary disease] (in Spanish). *Arch Bronconeumol* 2009;**45**:129—38.
27. **Wardlaw AJ,** Silverman M, Siva R, et al. Multi-dimensional phenotyping: towards a new taxonomy for airway disease. *Clin Exp Allergy* 2005;**35**:1254—62.
28. **Fabbri LM,** Rabe KF. From COPD to chronic systemic inflammatory syndrome? *Lancet* 2007;**370**:797—9.
29. **Sood A.** Obesity, adipokines and lung disease. *J Appl Physiol* 2010;**108**:744—53.
30. **Calverley PM,** Scott S. Is airway inflammation in chronic obstructive pulmonary disease (COPD) a risk factor for cardiovascular events? *COPD* 2006;**3**:233—42.
31. **Ferrer M,** Alonso J, Morera J, et al. Chronic obstructive pulmonary disease stage and health-related quality of life. The Quality of Life of Chronic Obstructive Pulmonary Disease Study Group. *Ann Intern Med* 1997;**127**:1072—9.
32. **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**:1005—12.
33. **Miravitlles M,** Anzueto A, Legnani D, et al. Patient's perception of exacerbations of COPD—the PERCEIVE study. *Respir Med* 2007;**101**:453—60.
34. **Svanes C,** Sunyer J, Plana E, et al. Early life origins of chronic obstructive pulmonary disease. *Thorax* 2010;**65**:14—20.
35. **Laupacis A,** Wells G, Richardson WS, et al. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. *JAMA* 1994;**272**:234—7.