ORIGINAL ARTICLE
COPD

# Phenotypic characterisation of early COPD: a prospective case–control study

Borja G. Cosío[1,2], Sergi Pascual-Guardia[2,3,4], Alicia Borras-Santos[2,5], Germán Peces-Barba[2,6], Salud Santos[2,7,8], Laura Vigil[2,9], Juan José Soler-Cataluña[2,10], Cristina Martínez-González[11], Ciro Casanova[12], Pedro J. Marcos[13], Carlos J. Alvarez[2,14,15], José Luis López-Campos[2,16], Joaquim Gea[2,3,4], Judith Garcia-Aymerich[4,5,17], Jesús Molina[18], Miguel Román[19], Jorge Moises[2,20,21,22], Viktoria Szabo[22], Elizabeth A. Reagan[23], Raúl San José Estépar[24], George Washko[25], Alvar Agustí[2,20,21,22] and Rosa Faner[2,22] on behalf of the EARLY COPD study group

ABSTRACT The phenotypic characteristics of chronic obstructive pulmonary disease (COPD) in individuals younger than 50 years of age (early COPD) are not well defined. This prospective, multicentre, case–control study sought to describe these characteristics and compare them with those of smokers (⩾10 pack-years) of similar age with normal spirometry (controls).

We studied 92 cases (post-bronchodilator forced expiratory volume in 1 s ($FEV_1$)/forced vital capacity (FVC) <0.7) and 197 controls. Results were contrasted with participants with similar inclusion criteria recruited into the ECLIPSE and COPDGene cohorts.

Cases had moderate airflow limitation ($FEV_1$ 71.3±20.8%) but were often symptomatic, used healthcare resources frequently, had air trapping (residual volume 150.6±55.5% ref.), had reduced diffusing capacity (84.2±20.7% ref.) and had frequent evidence of computed tomography (CT) emphysema (61%). Of note, less than half of cases (46%) had been previously diagnosed with COPD. Interestingly, they also often reported a family history of respiratory diseases and had been hospitalised because of respiratory problems before the age of 5 years more frequently than controls (12% *versus* 3%, p=0.009). By and large, these observations were reproduced when available in the ECLIPSE and COPDGene cohorts.

These results show that early COPD is associated with substantial health impact and significant structural and functional abnormalities, albeit it is often not diagnosed (hence, treated). The fact that a sizeable proportion of patients with early COPD report a family history of respiratory diseases and/or early-life events (including hospitalisations before the age of 5 years) renders further support to the possibility of early-life origin of COPD.

@ERSpublications
**Early COPD is associated with substantial health impact, and structural and functional abnormalities, albeit it is often not diagnosed and hence, not treated. It is frequently associated with family history of respiratory diseases and early-life events.** https://bit.ly/2ZtoRkp

**Cite this article as:** Cosío BG, Pascual-Guardia S, Borras-Santos A, *et al.* Phenotypic characterisation of early COPD: a prospective case–control study. *ERJ Open Res* 2020; 6: 00047-2020 [https://doi.org/10.1183/23120541.00047-2020].

This article has supplementary material available from openres.ersjournals.com

Received: 29 April 2020 | Accepted after revision: 26 June 2020

Copyright ©ERS 2020. This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.

## Introduction

Chronic obstructive pulmonary disease (COPD) has been traditionally understood as a self-inflicted disease caused by tobacco smoking and characterised by an accelerated decline of lung function with age (as determined by changes in the forced expiratory volume in 1 s ($FEV_1$)) [1]. Yet, recent research has shown that only about half of COPD patients follow this trajectory, whereas the rest develop COPD with a normal rate of lung function decline starting from a low peak lung function in early adulthood [2]. Importantly, these latter patients show a higher prevalence and a decade earlier incidence of comorbid diseases and die prematurely [3]. Because, collectively, these observations open new opportunities for prevention and earlier intervention [4, 5], interest in better understanding the characteristics and risks factors of "early" COPD is growing [6, 7]. The clinical concept of "early COPD", meaning a period of time at the beginning of its natural history, should be differentiated from that of "mild COPD", which refers to the degree of airflow limitation – defined by $FEV_1$ >80% – that can appear at any time during the natural history of the disease. Here we present the cross-sectional results of the first prospective case–control study specifically aimed at describing the clinical, functional, biological and imaging phenotypic characteristics of patients with early COPD (defined operationally by a post-bronchodilator $FEV_1$/forced vital capacity (FVC) <0.7 in smokers (⩾10 pack-years) younger than 50 years of age [6]) as compared with smokers of similar age and normal spirometry [8].

## Methods

The methodology of this study has been detailed elsewhere [8] and is only summarised below.

### *Study design and ethics*

This multicentre, prospective, case–control (1:2, respectively) study (NCT02352220) was approved by the ethics committee of all recruiting centres [8], and all participants signed their informed consent.

### *Participants*

Participants were recruited from 12 tertiary hospitals (by local advertisement) and associated primary care centres in Spain (from an automatically generated list of smokers), which are listed in the Appendix. All participants were 35–50 years of age, of Caucasian origin and were current or former smokers (>10 pack-years) [8]. Cases were defined by the presence of airflow limitation after bronchodilation ($FEV_1$/FVC<0.7) [9], and controls had normal spirometry. Exclusion criteria were: $\alpha_1$-antitrypsin deficiency, conditions that could potentially limit future follow-up (*e.g.* foreseen changes of residence, psychiatric diseases), chronic inflammatory or autoimmune diseases, severe bronchiectasis, active tuberculosis or cancer [8]. Exclusion criteria for controls also included a previously confirmed diagnosis of asthma. All participants had been free of any acute respiratory condition for 8 weeks prior to the baseline study visit.

### *Measurements*

Standardised questionnaires were used to record demographics, family history, early-life events and clinical data [8]. Physiological measures included forced spirometry (before and after bronchodilation), body

**Affiliations**: [1]Servicio de Neumología, Hospital Universitario Son Espases-IdISBa, Palma de Mallorca, Spain. [2]CIBER de Enfermedades Respiratorias, Madrid, Spain. [3]Servei de Pneumologia, Hospital del Mar – IMIM, Barcelona, Spain. [4]Universitat Pompeu Fabra, Barcelona, Spain. [5]ISGlobal, Barcelona, Spain. [6]Fundación Jiménez Díaz, Hospital Universitario, Grupo Quirón, Madrid, Spain. [7]Servicio de Neumología, Hospital Universitario de Bellvitge, Barcelona, Spain. [8]IDIBELL, Universidad de Barcelona, Hospitalet de Llobregat, Barcelona, Spain. [9]Servei de Pneumologia, Hospital Universitari Parc Taulí, Sabadell, Spain. [10]Servicio de Neumología, Hospital Arnau de Vilanova-LLiria, Valencia, Spain. [11]Servicio de Neumología, Hospital Universitario Central de Asturias, Oviedo, Spain. [12]Servicio de Neumología-Unidad de Investigación Hospital Universitario La Candelaria, Universidad de La Laguna, Tenerife, Spain. [13]Dirección de Procesos Asistenciales, Servicio de Neumología, Instituto de Investigación Biomédica de A Coruña (INIBIC), Complejo Hospitalario Universitario de A Coruña, Area Sanitaria de A Coruña, Sergas, Universidade da Coruña, A Coruña, Spain. [14]Servicio Neumología, HU 12 de Octubre, Madrid, Spain. [15]Departamento de Medicina, Facultad Medicina UCM, Madrid, Spain. [16]Unidad Médico-Quirúrgica de Enfermedades Respiratorias, Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío/Universidad de Sevilla, Seville, Spain. [17]CIBER de Epidemiología y Salud Pública, Madrid, Spain. [18]Centro de Salud Francia, DAO, Madrid, Grupo de Respiratorio semFYC, Madrid, Spain. [19]Centro de Salud de Son Pisa, Unidad de investigación en enfermedades respiratorias crónicas en atención primaria, IdISBa, Palma de Mallorca, Spain. [20]Institut Respiratori, Hospital Clinic, Barcelona, Spain. [21]Universitat de Barcelona, Barcelona, Spain. [22]Institut d'Investigacio August Pi i Sunyer, Barcelona, Spain. [23]National Jewish Health, Denver, CO, USA. [24]Applied Chest Imaging Laboratory, Dept of Radiology, Brigham and Women's Hospital, Boston, MA, USA. [25]Division of Pulmonary and Critical Care, Brigham and Women's Hospital, Boston, MA, USA.

**Correspondence**: Borja G. Cosío, Servicio de Neumología, Hospital Universitario Son Espases, Ctra. de Valldemossa 79, 07120 Palma de Mallorca, Spain. E-mail: borja.cosio@ssib.es

plethysmography, single-breath diffusing capacity of the lung for carbon monoxide ($D_{LCO}$) and transfer coefficient of the lung for carbon monoxide ($K_{CO}$), exhaled nitric oxide fraction ($F_{eNO}$) and 6-min walk test (6MWT), all measured according to international recommendations [8]. Reference values were those of the Global Lung Function Initiative (GLI) [10]. A low-dose computed tomography (CT) of the chest was obtained in all participants, and the presence/absence of emphysema was determined qualitatively in the coordinating centre (Hospital Universitario Son Espases-IdISBa (HUSE), Palma de Mallorca) by an experienced radiologist who was blinded to patient/control status. Full blood count, C-reactive protein and fibrinogen levels were determined in each centre [8].

### *Data analysis*

Results are presented as mean±standard deviation, proportion or odds ratio and 95% confidence intervals as appropriate. Cases and controls were compared using unpaired t-test or Chi-squared tests. The Spearman test was used to explore pair-wise correlations between the presence of early COPD and variables determined in the study, whereas a stepwise multiple logistic regression analysis was used to identify characteristics independently associated with the presence of early COPD. Finally, we used Cytoscape (www.cytoscape.org) [11] to present graphically the relationships between variables determined in the study using a network layout.

## Results

### *Phenotypic characterisation of participants*

We studied 289 individuals (92 cases and 197 controls). Table 1 compares their main phenotypic characteristics, and figure 1 presents the frequency distribution of some relevant variables.

#### *Demographics*

The COPD group included a lower proportion of females compared with controls (35% *versus* 50%, p=0.016), and subjects were slightly older (45.8±3.5 *versus* 43.3±4.4 years, p<0.001) and had a higher body mass index (28.1±6.4 *versus* 26.6±4.8 kg·m$^{-2}$, p=0.047); their educational level was lower than the controls, and fewer were active workers (table 1).

#### *Exposure history*

Cases reported a slightly higher cumulative smoking exposure (31.6±16.3 *versus* 24.8±13.7 pack-years, p=0.001) than controls (figure 1), albeit the proportion of current smokers and the age at smoking initiation were not significantly different (table 1). There was a higher proportion of males among the COPD cases (65% *versus* 35%). All women included had a similar smoking history irrespective of whether they had COPD or not, but the proportion of active smokers *versus* former smokers was higher in the control group (80.8% *versus* 19.2%) than in the COPD group (59.4% *versus* 40.6% respectively, p=0.026) Likewise, there were no significant differences between cases and controls with respect to the number of individuals exposed to some of the high-risk jobs for COPD described recently [12].

#### *Family history and early-life events*

Cases reported a higher proportion of several respiratory conditions in their parents, which reached statistical significance for maternal and parental bronchitis as well as for maternal cardiac disease (table 1).

There were no significant differences between patients and controls in the proportion of individuals born prematurely (<37 weeks), with low birth weight (<2.5 kg), mother age at delivery, smoking exposure during pregnancy or living in an urban environment during the first 5 years of life (table 1). Yet, it was interesting to note that 12% of cases reported having been hospitalised for respiratory causes before the age of 5 years *versus* 3% of controls (p=0.009).

#### *Symptoms, previous diagnosis and use of healthcare resources*

Cases were more symptomatic than controls, as shown by a higher proportion of cases having a modified Medical Research Council (mMRC) dyspnoea scale value ⩾2 (20% *versus* 6%, p=0.001; COPD Assessment Test (CAT) 12.5±6.8 *versus* 8.6±6.7, p<0.001), and a higher proportion of cases having chronic bronchitis (25.84% *versus* 10.15%, p=0.00112). Likewise, cases were significantly more likely to have been diagnosed with a respiratory disease (particularly asthma) and use respiratory medications (including oral corticosteroids and antibiotics) than controls. By contrast, the proportion of individuals diagnosed with a cardiovascular disease or that had ever used cardiovascular medications was similar in both groups (table 1). Cases were also more likely to have used healthcare resources for respiratory (but not for other) reasons more often too (table 1).

Of note, only 36 of the 78 cases (46%) in whom this information was available had been previously diagnosed with COPD, albeit they often used respiratory medications over the counter. Participants with

TABLE 1 Main characteristics of cases and controls

| | Cases | Controls | p-value |
|---|---|---|---|
| **Subjects n** | 91 | 107 | |
| **Demographics** | | | |
| Females | 35% | 50% | **0.016** |
| Age years | 45.8±3.5 | 43.3±4.4 | **<0.001** |
| Height cm | 168.7±8.6 | 168.1±9.0 | 0.579 |
| Weight kg | 80.3±20.5 | 75.4±15.9 | **0.043** |
| Body mass index kg·m$^{-2}$ | 28.1±6.4 | 26.6±4.8 | **0.047** |
| High educational level[#] | 52% | 77% | **<0.001** |
| Active worker (last year) | 70% | 86% | **0.002** |
| **Exposure history** | | | |
| Cumulative smoking exposure pack-years | 31.6±116.3 | 24.8±13.7 | **0.001** |
| Current smokers | 66% | 76% | 0.089 |
| Age at smoking initiation years | 16.0±2.8 | 16.6±3.7 | 0.121 |
| High-risk job[¶] | 29% | 20% | 0.129 |
| **Family history** | | | |
| Maternal asthma | 17% | 11% | 0.221 |
| Maternal bronchitis | 16% | 8% | **0.051** |
| Maternal cardiac disease | 31% | 17% | **0.014** |
| Maternal arterial hypertension | 52% | 44% | 0.282 |
| Maternal cerebrovascular disease | 11% | 11% | 1 |
| Maternal diabetes | 18% | 17% | 0.860 |
| Paternal asthma | 13% | 6% | 0.068 |
| Paternal bronchitis | 46% | 27% | **0.005** |
| Paternal cardiac disease | 31% | 25% | 0.430 |
| Paternal arterial hypertension | 43% | 37% | 0.375 |
| Paternal cerebrovascular disease | 18% | 9% | 0.083 |
| **Early-life events** | | | |
| Birth <37 weeks | 5% | 4% | 0.754 |
| Birth weight <2.5 kg | 4% | 6% | 0.760 |
| Mother age at delivery years | 29.6±7.0 | 28.3±5.7 | 0.128 |
| Mother smoked during pregnancy | 7% | 13% | 0.210 |
| Urban residence before the age of 5 | 65% | 74% | 0.120 |
| Hospital admission for respiratory disease before 5 years of age | 12% | 3% | **0.009** |
| **Symptoms, previous diagnosis and use of healthcare resources** | | | |
| mMRC ⩾2 | 29% | 6% | **0.001** |
| CAT | 12.5±6.8 | 8.6±6.7 | **<0.001** |
| Ever diagnosed of a respiratory problem | 63% | 39% | **0.001** |
| Ever diagnosed of asthma | 30% | 9% | **<0.001** |
| Ever use of respiratory medications | 79% | 28% | **<0.001** |
| Oral corticosteroids use in the last 12 months | 20% | 5% | **<0.001** |
| Antibiotic use in the last 12 months | 31% | 21% | 0.073 |
| Ever diagnosed of a cardiovascular disease | 26% | 28% | 0.751 |
| Ever use of cardiovascular medications | 24% | 34% | 0.229 |
| Ambulatory (GP) visits | | | |
| due to respiratory reasons | 60% | 29% | **<0.001** |
| due to non-respiratory reasons | 46% | 51% | 0.438 |
| Emergency room visits | | | |
| due to respiratory reasons | 14% | 1% | <0.001 |
| due to non-respiratory reasons | 27% | 37% | 0.141 |
| Hospital admissions | | | |
| due to respiratory reasons | 11% | 1% | **<0.001** |
| due to non-respiratory reasons | 12% | 10% | 0.686 |
| **Physiology** | | | |
| $FEV_1$ (pre-bd) % ref. | 64.7±20.1 | 95.3±13.9 | **<0.001** |
| FVC (pre-bd) % ref. | 90.4±18.8 | 100.3±14.6 | **<0.001** |
| $FEV_1$/FVC (pre-bd) % | 57.2±10.9 | 77.5±5.9 | **<0.001** |
| $FEF_{25-75\%}$ (pre-bd) % | 39.5±48.4 | 81.8±33.9 | **<0.001** |
| $FEV_1$ (post-bd) % ref. | 71.3±20.8 | 97.7±14.9 | **<0.001** |
| FVC (post-bd) % ref. | 96.5±18.6 | 98.9±15.0 | 0.280 |
| $FEV_1$/FVC (post-bd) % | 58.5±10.4 | 79.9±5.3 | **<0.001** |

Continued

TABLE 1 Continued

| | Cases | Controls | p-value |
|---|---|---|---|
| $FEF_{25-75\%}$ (post-bd) % | 39.1±30.9 | 92.5±35.5 | **<0.001** |
| Total lung capacity, % ref. | 112.7±17.1 | 103.1±15.7 | **<0.001** |
| Residual volume, % ref. | 150.6±55.5 | 113.6±40.9 | **<0.001** |
| $D_{LCO}$ % ref. | 84.16±20.73 | 91.6±13.51 | 0.003 |
| $K_{CO}$ % ref. | 80.53±19.73 | 93.25±12.57 | **<0.001** |
| 6MWD test | | | |
| Distance walked m | 555±83 | 570±82 | 0.152 |
| Dyspnoea at end of test | 2.3±2.0 | 1.5±1.8 | **0.001** |
| $S_{aO_2}$ at end of test | 94.8±4.4 | 96.3±2.5 | **0.003** |
| **Imaging** | | | |
| Presence of CT emphysema (%) | 61% | 32% | **<0.001** |
| **Biomarkers** | | | |
| $F_{eNO}$ ppb | 16.9±13.3 | 17.6±17.0 | 0.741 |
| Haematocrit % | 44.6±3.3 | 43.5±3.6 | **0.015** |
| Leukocytes $\times 10^9 \cdot L^{-1}$ | 8.7 ±2.2 | 7.8±2.3 | **0.003** |
| Neutrophils % | 60.7±8.1 | 57.6±9.3 | **0.005** |
| Lymphocytes % | 28.9±6.8 | 32.4±8.0 | **<0.001** |
| Eosinophils % | 2.5±1.5 | 2.8±1.9 | 0.156 |
| Eosinophils >300 cells$\cdot\mu L^{-1}$ | 23% | 20% | 0.635 |
| Fibrinogen $mg \cdot dL^{-1}$ | 390.8±106.3 | 368.4±81.0 | 0.084 |
| C-reactive protein >3 $mg \cdot /dL^{-1}$ (%) | 0% | 3% | 0.327 |

Results are presented as mean±SD or proportion. p-values <0.05 are highlighted in bold text. mMRC: modified Medical Research Council scale; CAT: COPD Assessment Test; $FEV_1$: forced expiratory volume in 1 s; pre-/post-bd: pre-/post-bronchodilator; % ref.: % of reference value; FVC: forced vital capacity; $FEF_{25-75\%}$: forced expiratory flow at 25–75% of FVC; $D_{LCO}$: diffusing capacity of the lung for carbon monoxide; $K_{CO}$: transfer coefficient of the lung for carbon monoxide; 6MWD: 6-min walk test; $S_{aO_2}$: arterial oxygen saturation; CT: computed tomography; $F_{eNO}$: exhaled nitric oxide fraction. #: High education level includes a Baccalaureate or University degree; ¶: as identified in [12].

undiagnosed COPD were referred to their primary care physician for appropriate treatment and follow-up. Supplementary table S1 shows that most characteristics were similar in cases with and without a previous diagnosis of COPD except for a lower $FEV_1$/FVC ratio in the former (55.2±11.0% *versus* 61.7±9.5%, p=0.007). Of note too, four controls had been diagnosed (erroneously) with COPD before entering our study.

### *Physiology*

By design, cases had airflow limitation, which ranged from mild to severe (figure 1), both before and after bronchodilation (Global Initiative for Chronic Obstructive Lung Disease (GOLD) grades 1–4: 42%, 39%, 15% and 3%, respectively), whereas spirometry was normal in controls (table 1). Mid-expiratory airflow (forced expiratory flow at 25–75% of FVC ($FEF_{25-75\%}$)) was reduced in cases (39.2±30.9% ref.) and also in some controls (92.6±35.5% ref., p<0.0001). GOLD group distribution in cases was 64% A, 15% B, 14% C and 7% D.

Total lung capacity and residual volume was higher in cases than in controls (table 1). $D_{LCO}$ was often abnormal in cases and, interestingly, also in some controls (figure 1). Exercise capacity (6MWD) was normal in cases and controls but, at the end of the test, dyspnoea (Borg score) was higher and arterial oxygen saturation lower in cases (table 1).

### *Imaging*

CT of the thorax could be obtained and interpreted in 231 participants (79.9%): 80 cases (87%) and 151 (77%) controls. In these participants, emphysema was present in 61% of cases and, of note, in 32% of controls too (p<0.001).

### *Biomarkers*

The $F_{eNO}$ was similar in cases and controls. The proportion of circulating leukocytes and neutrophils was higher, and that of lymphocytes lower in cases, whereas circulating eosinophil levels were similar in both groups (table 1). Fibrinogen and C-reactive protein levels were also similar in both groups (table 1).

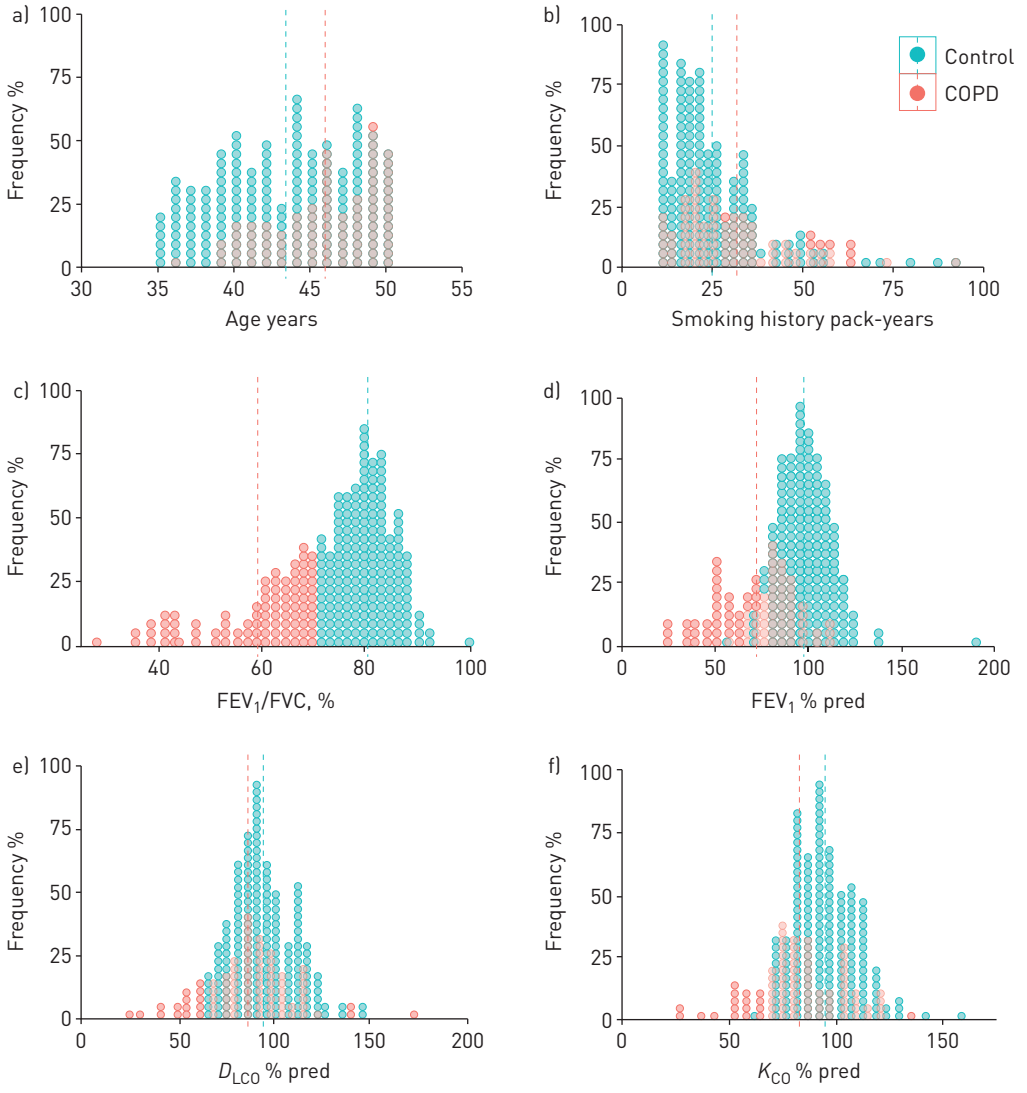



FIGURE 1 Frequency distribution of age (years), cumulative smoking exposure (pack-years), $FEV_1$/FVC (%), $FEV_1$ (% ref.), diffusing capacity of the lung for carbon monoxide ($D_{LCO}$) (% ref.) and KLCO (% ref.) in patients (red; n=92) and controls (green; n=197). Dotted vertical lines represent mean values. For further explanations, see text.

### *Reproducibility of observations*

#### *Age- and sex-matched population analysis*

Because there were small but significant differences between age and sex distribution between cases and controls, we used the R library "*MatchIt*" to individually match 92 controls for sex and age with the 92 cases included in the study. Supplementary table S2 compares the main characteristics of these two groups. By and large, differences observed in the original population (table 1) were maintained in these sex- and age-matched populations (supplementary table S2).

#### *Lower limit of normal* versus *fixed $FEV_1$/FVC ratio*

Using its lower limit of normal (LLN) instead of a fixed $FEV_1$/FVC <0.7 ratio, three participants (1%) were reclassified: one control was now classified as a patient, and two cases were now classified as controls. This is not surprising because the mean LLN of $FEV_1$/FVC for a population of this age is quite close to 0.7 (0.699±0.001). Importantly, the main results did not change when the LLN was used instead of the fixed ratio (supplementary table S3).

#### *External reproducibility in selected ECLIPSE and COPDGene participants*

To explore the external reproducibility of our observations, we accessed the ECLIPSE [13] and COPDGene [14] databases (dbGAP phs001252.v1.p1). Although these two studies were not designed to investigate early

TABLE 2 Main characteristics of patients (forced expiratory volume in 1 s ($FEV_1$)/forced vital capacity (FVC) <0.7) and controls ($FEV_1$/FVC⩾0.7) included in the ECLIPSE study

| | Patients | Controls | p-value |
|---|---|---|---|
| **Subjects n** | 119 | 124 | |
| **Demographics and exposure history** | | | |
| Age years | 47.3±2.7 | 45.7±3.0 | **<0.001** |
| Males | 61% | 52% | 0.162 |
| Body mass index kg·m$^{-2}$ | 26.5±6.9 | 26.1±4.6 | 0.650 |
| Cumulative smoking exposure pack-years | 34.5±17.0 | 24.7±11.7 | **<0.001** |
| Current smokers | 38% | 23% | **0.014** |
| **Family history and early-life events** | | | |
| Familiar asthma | 24% | 19% | 0.342 |
| Familiar bronchitis | 34% | 15% | **0.001** |
| Familiar emphysema | 29% | 8% | **<0.001** |
| Familiar hypertension | 49% | 48% | 1 |
| **Symptoms** | | | |
| mMRC score ⩾2 | 13% | 1% | **0.001** |
| **Previous treatments and use of healthcare resources** | | | |
| Oral Corticosteroid use | 2% | 0% | 0.460 |
| Number of exacerbations | 0.7± 0.9 | NA | |
| **Physiology** | | | |
| $FEV_1$/FVC (post-bd) % | 49.0± 11.8 | 80.1±4.9 | **<0.001** |
| $FEV_1$ (post-bd) % ref. | 49.0± 18.2 | 103.9±10.8 | **<0.001** |
| FVC (post-bd) % ref. | 82.5± 22.1 | 112.8±12.6 | **<0.001** |
| Total lung capacity % ref. | 114.0± 17.5 | 102.79±11.4 | **<0.001** |
| 6MWD test | | | |
| Distance walked m | 425.6±137.7 | NA | |
| **Imaging** | | | |
| Presence of CT emphysema n (%) | 60% | 8% | **<0.001** |
| **Biomarkers** | | | |
| Basophils (%) | 0.33± 0.23 | 0.32± 0.24 | 0.527 |
| Eosinophils (%) | 2.6± 2.1 | 2.55± 1.58 | 0.464 |
| Haematocrit | 0.44±0.04 | 0.43± 0.04 | **0.009** |
| Lymphocytes (%) | 28.04± 8.61 | 30.22± 7.11 | **0.024** |
| Monocytes (%) | 5.96± 2.49 | 5.76± 2.23 | 0.682 |
| Segmented neutrophils (%) | 63.06± 9.52 | 61.15± 8.29 | 0.102 |
| Total neutrophils (total ANC) | 5.23± 2.2 | 4.71± 1.95 | **0.050** |
| Total neutrophils (%) | 63.06± 9.52 | 61.15± 8.29 | 0.102 |
| White blood cells ×10$^9$/L | 8.11± 2.49 | 7.56± 2.37 | 0.057 |

Results are presented as mean±SD or proportion. p-values <0.05 are highlighted in bold text. mMRC: modified Medical Research Council scale; post-bd: post-bronchodilator; % ref.: % of reference value; 6MWD: 6-min walk test; CT: computed tomography; ANC: total absolute neutrophil count.

COPD, we identified 119 cases and 124 controls in the ECLIPSE study (table 2), and 205 cases and 560 controls in the COPDGene study (table 3) that fulfilled the same inclusion criteria used in our study. As summarised in figure 2, our observations were, by and large, reproduced in these two subpopulations of the ECLIPSE and COPDGene cohorts.

### *Bivariate correlations*

Figure 3 presents all the significant (p<0.05) bivariate associations between the presence of early COPD and a number of demographic, clinical, functional, imaging and biomarker variables determined in the study.

### *Multiple regression analysis*

Unadjusted multiple regression logistic analysis identified six independent characteristics associated with the presence of early COPD in the population studied here (table 4). Interestingly, the two with the highest odds ratios referred to past history of respiratory events, such as hospital admission for respiratory disease before the age of 5 years (OR 9.292) and emergency room visits due to respiratory reasons (OR 9.160). Other clinical characteristics associated with the presence of early COPD, albeit with a much lower odds ratio, included symptoms (mMRC ⩾2), cumulative smoking exposure (packs-years), low diffusing capacity ($K_{CO}$, % ref.) and not being an active worker.

TABLE 3 Main characteristics of patients (forced expiratory volume in 1 s ($FEV_1$)/forced vital capacity (FVC) <0.7) and controls ($FEV_1$/FVC⩾0.7) included in the COPDGene study

| | Patients | Controls | p-value |
|---|---|---|---|
| **Subjects n** | 205 | 560 | |
| **Demographics and exposure history** | | | |
| Age years | 47.8±1.5 | 47.5±1.5 | **0.007** |
| Females | 50% | 51% | 0.178 |
| Body mass index kg·m$^{-2}$ | 28.4±7.0 | 29.4±6.4 | 0.051 |
| Cumulative smoking exposure pack-years | 38.2±15.0 | 32.6±16.4 | **<0.001** |
| Current smokers | 76% | 71% | 0.154 |
| High educational level[#] | 47% | 58% | **0.011** |
| Active worker (last year) | 33% | 50% | **<0.001** |
| **Family history and early-life events** | | | |
| Maternal asthma | 14% | 11% | 0.438 |
| Maternal bronchitis | 16% | 16% | 0.997 |
| Maternal emphysema | 23% | 16% | **0.045** |
| Paternal asthma | 7% | 4% | 0.285 |
| Paternal bronchitis | 8% | 10% | 0.765 |
| Paternal emphysema | 26% | 17% | **0.024** |
| Mother smoked during pregnancy | 52% | 47% | 0.261 |
| **Symptoms** | | | |
| mMRC score | 1.9±1.5 | 0.8±1.2 | **<0.001** |
| **Previous treatments and use of healthcare resources** | | | |
| Oral corticosteroid use | 4% | 1% | **0.015** |
| Number of ambulatory (GP) visits due to respiratory reasons (exacerbations) | 0.44±1.02 | 0.11±0.44 | **<0.001** |
| Number of hospital admissions due to respiratory reasons (exacerbations) | 0.20±0.71 | 0.03±0.21 | **<0.001** |
| **Physiology** | | | |
| $FEV_1$/FVC (post-bd) % | 59±12 | 79±1 | **<0.001** |
| $FEV_1$ (post-bd) % ref. | 66±22 | 92±14 | **<0.001** |
| FVC (post-bd) % ref. | 87±19 | 92±14 | **<0.001** |
| 6MWD test | | | |
| Distance walked m | 435.4±114.3 | 476.2±103.2 | **<0.001** |
| Shortness of breath at end of test | 76% | 51% | **<0.001** |
| **Imaging** | | | |
| Presence of CT emphysema n (%) | 29% | 9% | **<0.001** |

Results are presented as mean±SD or proportion. P-values <0.05 are highlighted in bold text. mMRC: modified Medical Research Council scale; post-bd: post-bronchodilator; % ref.: % of reference value; 6MWD: 6-min walk test; CT: computed tomography. [#]: High education level includes a Baccalaureate or University degree.

## Discussion

The results of this study show that patients with early COPD have the following characteristics. 1) They are often symptomatic and use healthcare resources frequently and have remarkable structural (emphysema) and physiological impairment, yet they are frequently undiagnosed and untreated. 2) Many different demographic, clinical, imaging and functional variables are associated with the presence of early COPD (figure 3). Of particular interest is the observation that these patients frequently report a family history of respiratory diseases and hospitalisations because of respiratory problems before 5 years of age, further supporting the potential relevance of early-life events in the pathogenesis of the disease [3, 5, 15]. 3) Despite the fact that the controls had normal spirometry, they showed remarkable functional and structural abnormalities that often overlapped with cases (figure 1). Collectively, these observations contribute to a better understanding of early COPD and may open novel opportunities for prevention and treatment of the disease [4].

### *Previous studies*

Interest on early COPD has gained momentum over the past few years since the publication of a joint analysis of three independent cohorts (Framingham Offspring Cohort (FOC), Copenhagen City Heart Study cohort and the Lovelace cohort), which showed that only about half of adult patients with COPD followed the traditional trajectory described by Fletcher and Peto >40 years ago, characterised by an

accelerated decline of lung function with age [1]; the other half never reach normal peak lung function in early adulthood and develop COPD later in life with a normal rate of lung function decline [2]. Since then, an operational definition of early COPD has been proposed [6] (which we used here), and several studies have investigated risk factors and clinical characteristics of early COPD in existing databases. For instance, Allinson *et al.* [16] used the Medical Research Council National Survey of Health and Development, a nationally representative British cohort followed since birth in 1946, to show that chronic mucus hypersecretion represents an early developmental phase of COPD, and that smoking impairs pulmonary development during adolescence or early adulthood, thus likely facilitating the development of COPD [17]. Likewise, Kalhan *et al.* [18] used the CARDIA Lung Study database to show that persistent respiratory symptoms in young adults are associated with accelerated lung function decline and greater odds of future radiographic emphysema. Finally, very recently, Çolak *et al.* [19] used a large Danish contemporary population-based cohort followed-up for 14.4 years to investigate the prevalence, characteristics and prognosis of individuals with early COPD in the general population. Using the same definition of early COPD used here [6], these authors estimated a prevalence of early COPD in the general

| **Demographics and exposure history** | Spain | Eclipse | COPDgene |
|---|---|---|---|
| Age years | ↑ | ↑ | ↑ |
| Females | ↓ | ~ | ~ |
| Body mass index kg·m$^{-2}$ | ↑ | ~ | ~ |
| Cumulative smoking exposure pack-years | ↑ | ↑ | ↑ |
| Current smokers | ~ | ↑ | ~ |
| High educational level | ↓ | ~ | ↓ |
| Active worker (last year) | ↓ | ~ | ↓ |
| **Family history and early life events** | | | |
| Familiar asthma | ~ | ~ | ~ |
| Familiar bronchitis | ↑ | ↑ | ~ |
| Familiar emphysema | NA | ↑ | ↑ |
| Hospital admission for respiratory disease before the age of 5 years | ↑ | NA | NA |
| **Symptoms** | | | |
| mMRC≥2 | ↑ | ↑ | ↑ |
| **Treatment and use of healthcare resources in the previous 12 months** | | | |
| Oral corticosteroids use in the last 12 months | ↑ | ~ | ↑ |
| Moderate exacerbations | ↑ | ? | ↑ |
| Hospitalised exacerbations | ↑ | ? | ↑ |
| **Physiology** | | | |
| $FEV_1$/FVC post-BD % | ↓ | ↓ | ↓ |
| $FEV_1$ post-BD % pred | ↓ | ↓ | ↓ |
| FVC post-BD % pred | ~ | ↓ | ↓ |
| Total lung capacity % pred | ↑ | ↑ | NA |
| $D_{LCO}$ % pred | ↓ | NA | NA |
| 6MWD test | | | |
| Distance walked | ~ | NA | ↓ |
| Dyspnoea at end of test | ↑ | NA | ↑ |
| $S_{aO_2}$ at end of test | ↓ | NA | NA |
| **Imaging** | | | |
| Presence of CT emphysema, n (%) | ↑ | ↑ | ↑ |
| **Biomakers** | | | |
| Leukocytes x10$^9$/L | ↑ | ↑ | NA |
| Neutrophils % | ↑ | ↑ | NA |
| Lymphocytes % | ↓ | ↓ | NA |
| Eosinophils % | ~ | ~ | NA |
| Haematocrit % | ↑ | ↑ | NA |

FIGURE 2 Comparison of the main characteristics determined in patients with early COPD (*versus* controls) in the three cohorts studied here (our own (SPAIN), ECLIPSE and COPDGene). ↓ Indicates reduced in cases, ↑ indicates increased in cases, ~ indicates similar in patients and controls, and NA means not available in that cohort. Orange, pink and yellow rows indicate, respectively, reproducibility of findings in three out of the three cohorts, two out of two cohorts or two out of three cohorts where this information was available. For further explanations, see text.

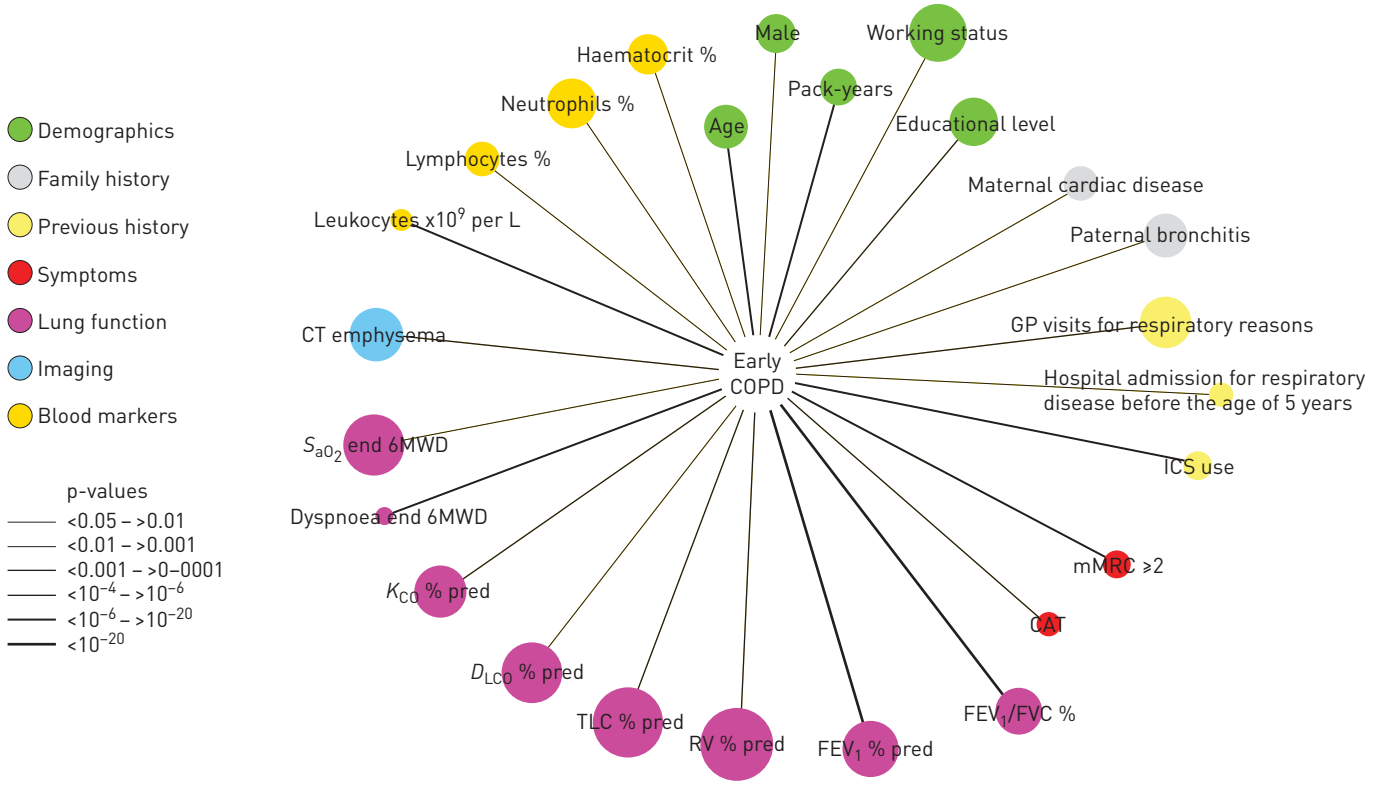



FIGURE 3 Variables associated (p<0.05) with the presence of early COPD in the population studied. Node colour indicates variable category (see legends), node size is proportional to the prevalence of abnormal values of that particular variable and edge (link) width is proportional to p-values (see legends). For further explanations, see text.

population of 15% and showed that, like in our study reported here, patients with early COPD often report chronic respiratory symptoms and have significant lung function impairment [19]. Importantly, ÇOLAK *et al.* [19] also observed that, during follow-up, patients with early COPD had increased risk of respiratory-related hospital admissions and early death, as we had also reported before in the FOC [3]. Collectively, these epidemiological studies provide important information stemming from existing data bases that were collected for different purposes. By contrast, to our knowledge, ours is the first case–control study to investigate prospectively the clinical, physiological, imaging and biological phenotypic characteristics of patients with early COPD at a much more detailed granularity level.

### *Interpretation of findings*

Our study provides several observations of interest. First, it shows that early COPD is associated with a significant, but mostly occult (54% undiagnosed and untreated), disease burden, as illustrated by the level

TABLE 4 Independent clinical characteristics associated with the presence of early chronic obstructive pulmonary disease (COPD) identified by multiple regression logistic analysis (by odds ratio value)

| | OR (95% CI) | p-value |
|---|---|---|
| **Hospital admission for respiratory disease before the age of 5 years** | 9.292 (2.57–36.10) | <0.001 |
| **Emergency room visits due to respiratory reasons** | 9.160 (1.43–180.74) | <0.05 |
| **mMRC ⩾2** | 2.336 (1.36–4.14) | <0.01 |
| **Packs-years** | 1.030 (1.00–1.06) | <0.05 |
| **$K_{CO}$ % ref.** | 0.950 (0.92–0.98) | <0.001 |
| **Active worker** | 0.391 (0.16–0.99) | <0.05 |

mMRC: modified Medical Research Council scale; $K_{CO}$: transfer coefficient of the lung for carbon monoxide; % ref.: % of reference value.

of symptoms and use of healthcare resources, as well as by the presence of significant structural damage (emphysema) and functional impairment. These findings are in line with recent findings by Woodruff *et al.* [20] who showed that (older) smokers with normal spirometry often present significant respiratory symptoms, activity limitation and even episodes of exacerbations. Likewise, they support the importance of diagnosing COPD as early as possible, when currently available therapeutic interventions may be more effective [15]. In this setting, it is of note that Martinez *et al.* [21] have recently reported that the combined use of CAPTURE (COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk), a simple, five-item, patient-completed questionnaire, and peak expiratory flow (with an inexpensive, easy-to-use mechanical device and interpretive thresholds) can discriminate cases and controls precisely in primary care, suggesting that it may be potentially useful for screening of early COPD.

Second, traditionally, COPD has been understood as a self-inflicted disease by tobacco smoking, which, in susceptible individuals, induces an abnormal inflammatory reaction that damages the lung parenchyma and causes airflow limitation [5]. Our results here provide some further support to this paradigm since cumulative smoking exposure was higher in cases (31.6±16.3 pack-years) than in controls (24.8±13.7 pack-years, p=0.001), with some evidence of a more intense and different inflammatory response (more circulating leukocytes and neutrophils and less lymphocytes) in cases than in smoker controls (table 1 and figure 3). On the other hand, however, our results also render support to recent findings indicating that abnormal lung development and early-life events can play a significant role in the pathogenesis of COPD [2, 3, 5, 15, 22, 23], since patients with early COPD in our study often reported a family history of respiratory diseases and were more frequently hospitalised because of respiratory problems before the age of 5 years. This is a notable observation that could not be reproduced in the ECLIPSE or COPDGene studies, because this variable was not recorded in these two studies. Also, the fact that individuals with COPD were more frequently unemployed and had lower educational level could support the role of poverty in contributing to the development of COPD early in life, as has previously been demonstrated [24].

Third, figure 1 clearly shows that there is significant overlap between patients and controls for most of the variables investigated except (by definition) the $FEV_1$/FVC ratio. In fact, a sizeable proportion of controls (*i.e.* smokers with normal spirometry) in our study had reduced $D_{LCO}$ (figure 1) and evidence of CT emphysema (32%), thus showing clear evidence of lung damage. Although these participants were classified as "controls" because their $FEV_1$/FVC was >0.7, differences between groups would have been likely enlarged if they had been classified as "patients". In any case, this observation confirms that smokers with normal spirometry may have emphysema [25] and strengthens the need for a careful reconsideration of the diagnosis and taxonomy of COPD [5, 26]. Of note, the prevalence of emphysema determined in our patients with early COPD was very high (61%), albeit this figure is similar to that observed in ECLIPSE (60%) but larger than that determined in COPDGene (29%). Differences may be driven by technical factors since we (and in ECLIPSE [27], but not in COPDGene [28]) used low-dose radiation (which can overestimate the amount of quantitative emphysema). Alternatively, it is also possible that early COPD patients could be particularly prone to parenchymal damage [29] and/or that abnormal alveoli development can also contribute to early COPD significantly [5, 15, 22].

Finally, it is important to note that some of the factors identified by multiple regression analysis in relation to the presence of early COPD (table 4) may truly be "risk" factors for the early occurrence in COPD (*e.g.* early-life events and smoking exposure) whereas others (*e.g.* symptoms, reduced lung diffusion or not working actively) probably represent "consequences" of the disease. In any case, they contribute to better define the phenotypic characteristics of early COPD and to identify in which young smokers the disease is more likely to be present.

### *Strengths and limitations*

Our study has two major strengths. First, it is the first study to prospectively contrast carefully phenotyped young smokers (<50 years of age) with (early COPD) and without airflow limitation. Second, results were largely reproduced in two independent, international cohorts [13, 14]. Albeit these two cohorts were not originally designed to investigate early COPD, by employing the inclusion criteria used here we identified a significant number of participants with similar characteristics, so available variables in ECLIPSE and COPDGene could be contrasted with our results. Our study has some potential limitations that deserve discussion. First, its sample size is much smaller than that of previous epidemiological studies [2, 3, 16–19]. Yet, the granularity of phenotypic characterisation in our study is much deeper. Second, we cannot exclude a recruitment bias towards more symptomatic cases of early COPD since we recruited patients from primary care centres and tertiary hospital out-patient clinics, and not from the general population. Third, cases were slightly but significantly older than the control group and included fewer

females. However, an individually age- and sex-matched subpopulation confirmed results observed in the total population. Finally, a diagnosis of asthma was considered an exclusion criterion in controls but not in cases. The reason for this was that a "diagnosis of asthma" (not necessarily the disease) is frequently associated with abnormal lung development [15, 22], and the latter is now a well-recognised cause of COPD [2]; by contrast, we were more strict with controls to avoid the inclusion of patients with true asthma.

### *Conclusions*

Early COPD is associated with significant disease burden and use of healthcare resources due to substantial structural and functional abnormalities, which are frequently undiagnosed (hence untreated). Further, a significant proportion of these patients report a family history of respiratory diseases and/or previous respiratory early-life events, supporting the emerging concept that COPD can start very early in life [15], thus opening new opportunities for prevention, diagnosis and treatment of this disease [4].

Acknowledgements: The authors thank all participants in the study for their willingness to contribute to medical research, and all field workers for their dedicated, high quality, daily work.

Full list of field participating investigators in the study: Hospital Universitario Son Espases (Mallorca): Borja G. Cosio; Rocío Cordova Diaz; María Magdalena Pan Naranjo; Joan Palmer Sancho; Miguel Román Rodríguez; Hospital Clínic (Barcelona): Alvar Agustí; Rosa Faner Canet; Joan Albert Barberà; Josep Roca Torrent; Yolanda Torralba Garcia; Jorge Moises Lafuente; Anna Maria Pedro Pijoan; Amparo Hervas Docón; Carmen Herranz; Núria Sanchez Ruano; Hospital del Mar (Barcelona): Joaquim Gea; Diego A; Rodríguez Chiaradía; Anna Rodó-Pin; Clara Martín-Ontiyuelo; Mireia Admetlló; Concepción Ballano Castro; Laura Gutiérrez Martín; José Ignacio Aoiz Linares; Sergi Pascual-Guardia; Marta Mourelo Cereijo; Fundación Jiménez Díaz (Madrid): Germán Peces-Barba Romero; José Fernández Arias; Carolina Gotera Rivera; Manuel Martin Bernal; Guillermo Gallardo Madueño; Andrés Alcázar Peral; Carmelo Palacios Miras; Maria Teresa Pinedo Moraleda; Maria Belén Torres Labandeira; Mercedes Colomo Rodríguez; María Concepción Rodríguez Gallego; Carmen Lobon Agundez; Mónica Nácher Conches; María José Mansilla; Rosario Serrano Martín; Hospital 12 Octubre (Madrid): Carlos J. Álvarez Martínez; Marta Padilla Bernáldez; Jesús Molina París; Hospital Parc Taulí (Sabadell): Laura Vigil Giménez; Eduard Monsó Molas; Laia Seto Gort; Montserrat Baré Mañas; Anna Maria Fabra Noguera; Hospital Virgen del Rocío (Sevilla): José Luís López Campos; Carmen Calero Acuña; Laura Carrasco Hernández; Hospital Universitario de Bellvitge (Hospitalet de Llobregat; Barcelona): Salud Santos Perez; Montserrat Navarro; Elisabeth Serra; Ferran Ferrer Keysers; Damaris Batallé; M; Dolores Peleato Catalan; Albert Dorca; Javier Burgos; Hospital Arnau de Vilanova (Valencia): Juan José Soler-Cataluña; Noelia González García; Lourdes Sánchez Sánchez; Hospital Universitario Central de Asturias (Oviedo): Cristina Martínez González; Amador Prieto Fernández; Susana Martínez González; Hospital Candelaria (Canarias): Ciro Casanova Macario; Delia Mayato; Hospital Universitario de A Coruña: Pedro J Marcos Rodriguez; Luis Domínguez Juncal; Rosario Timiraos Carrasco; Rosa Garcia Palenzuela.

Support statement: In part, by CIBERES, SEPAR, FIS (CP16/00039, PI17/00369, PI18/01008) and an unrestricted grant from Boehringer-Ingelheim. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: B.G. Cosío reports speaker fees from AstraZeneca, grants from Boehringer, grants and speaker fees from Novartis, grants and speaker fees from Chiesi, speaker fees from Rovi, grants from Menarini, and advisory board fees from Esteve, outside the submitted work. S. Pascual-Guardia has nothing to disclose. A. Borras-Santos has nothing to disclose. G. Peces-Barba has nothing to disclose. S. Santos reports personal fees from Almirall, Boehringer Ingelheim, AstraZeneca, Gebro Pharma, GlaxoSmithKline, Novartis, Faes, Grifols and Menarini outside the submitted work. L. Vigil reports to have received lectures from Boehringer-Ingelheim, GSK, Esteve, Menarini, Novartis, Chiesi and Pfizer. J.J. Soler-Cataluña reports personal fees from AstraZeneca, personal fees and non-financial support from Boehringer-Ingelheim, personal fees from Bial, grants and personal fees from GSK, grants, personal fees and non-financial support from Novartis, personal fees from Ferrer, personal fees and non-financial support from Menarini, personal fees from Teva, grants and personal fees from Esteve, outside the submitted work. C. Martínez-González reports in the last three years, a grant from Roche Pharma. C. Casanova declares in the last three years to have received lectures and/or scientific advice from Laboratorios Bial, Boehringer-Ingelheim, Gebropharma, GSK, Esteve, Menarini, Novartis and Rovi. P.J. Marcos reports personal fees from GSK, personal fees from Boehringer Ingelheim, personal fees from Ferrer, personal fees from Menarini and personal fees from Esteve, outside the submitted work. C.J. Alvarez has nothing to disclose. J.L. López-Campos reports personal fees and non-financial support from AstraZeneca, grants, personal fees and non-financial support from Boehringer Ingelheim, grants, personal fees and non-financial support from Chiesi, personal fees and non-financial support from CSL Behring, grants, personal fees and non-financial support from Esteve, personal fees from Ferrer, grants, personal fees and non-financial support from Gebro Pharma, grants, personal fees and non-financial support from GlaxoSmithKline, grants, personal fees and non-financial support from Grifols, grants, personal fees and non-financial support from Menarini, grants, personal fees and non-financial support from Novartis, grants, personal fees and non-financial support from Rovi, personal fees from Teva, outside the submitted work. J. Gea reports grants from Menarini, and advisory board fees from AstraZeneca, GSK and Boehringer Ingelheim, outside the submitted work. J. Garcia-Aymerich has nothing to disclose. J. Molina reports personal fees from Astra-Zeneca, personal fees from Boehringer-Ingelheim, personal fees from Chiesi, grants and personal fees from GSK, personal fees from Menarini, personal fees from Mundipharma, personal fees from Novartis, personal fees from Pfizer, outside the submitted work. M. Román reports personal fees from AstraZeneca, personal fees from Boehringer-Ingelheim, personal fees from Chiesi, grants and personal fees from GSK, personal fees from Menarini, personal fees from Mundipharma, personal fees from Novartis, personal fees from Pfizer, personal fees from Teva, personal fees from Bial, outside the submitted work. J. Moises has nothing to disclose. V. Szabo has nothing to disclose. E.A. Regan has no conflicts. R. San José Estépar reports grants from NHLBI, personal fees from Toshiba, personal fees

from Boehringer Ingelheim, personal fees from Eolo Medical, personal fees from Leuko Labs, outside the submitted work; and he is also a founder and co-owner of Quantitative Imaging Solutions which is a company that provides image based consulting and develops software to enable data sharing. G. Washko reports grants from the NIH; a grant from, and consultancy and advisory board membership for Boehringer Ingelheim; that he is a founder and co-owner of Quantitative Imaging Solutions, which provides image-based consulting and develops software to enable data sharing; consultancy and chairing a DSMB for PulmonX; a grant from BTG Interventional Medicine; a grant from and consultancy for Janssen Pharmaceuticals; and consultancy for GlaxoSmithKline, all outside the submitted work. Dr Washko's spouse works for Biogen, which is focused on developing therapies for fibrotic lung disease. A. Agustí reports personal fees from AstraZeneca, grants and personal fees from Menarini, personal fees from Chiesi, grants and personal fees from GSK, personal fees from Nuvaira, outside the submitted work. R. Faner reports grants and personal fees from GSK, a speaker fee from Chiesi, and grants from Menarini, outside the submitted work.

## References


1 Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J* 1977; 1: 1645–1648.
2 Lange P, Celli B, Agusti A, *et al.* Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med* 2015; 373: 111–122.
3 Agustí A, Noell G, Brugada J, *et al.* Lung function in early adulthood and health in later life: a transgenerational cohort analysis. *Lancet Respir Med* 2017; 5: 935–945.
4 Agusti A, Faner R. COPD beyond smoking: new paradigm, novel opportunities. *Lancet Respir Med* 2018; 6: 324–326.
5 Agusti A, Hogg JC. Update on the pathogenesis of chronic obstructive pulmonary disease. *N Engl J Med* 2019; 381: 1248–1256.
6 Martinez FJ, Han MK, Allinson JP, *et al.* At the root: defining and halting progression of early chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2018; 197: 1540–1551.
7 Soriano JB, Polverino F, Cosio BG. What is early COPD and why is it important? *Eur Respir J* 2018; 52: 1801448.
8 Borràs-Santos A, Garcia-Aymerich J, Soler-Cataluña JJ, *et al.* Determinants of the appearance and progression of early-onset chronic obstructive pulmonary disease in young adults. A case-control study with follow-up. *Archivos de Bronconeumología* 2019; 55: 312–318.
9 Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease, 2020. www.goldcopd.org Date last accessed: June, 10th 2020. Date last updated: June, 10th 2020.
10 Quanjer PH, Stanojevic S, Cole TJ, *et al.* Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012; 40: 1324–1343.
11 Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13: 2498–2504.
12 De Matteis S, Jarvis D, Darnton A, *et al.* The occupations at increased risk of COPD: analysis of lifetime job-histories in the population-based UK Biobank Cohort. *Eur Respir J* 2019; 54: 1900186.
13 Vestbo J, Anderson W, Coxson HO, *et al.* Evaluation of COPD longitudinally to identify predictive surrogate end-points (ECLIPSE). *Eur Respir J* 2008; 31: 869–873.
14 Regan EA, Hokanson JE, Murphy JR, *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010; 7: 32–43.
15 Agusti A, Faner R. Lung function trajectories in health and disease. *Lancet Respir Med* 2019; 4: 358–364.
16 Allinson JP, Hardy R, Donaldson GC, *et al.* The presence of chronic mucus hypersecretion across adult life in relation to chronic obstructive pulmonary disease development. *Am J Respir Crit Care Med* 2016; 193: 662–672.
17 Allinson JP, Hardy R, Donaldson GC, *et al.* Combined impact of smoking and early life exposures on adult lung function trajectories. *Am J Respir Crit Care Med* 2017; 196: 1021–1030.
18 Kalhan R, Dransfield MT, Colangelo LA, *et al.* Respiratory symptoms in young adults and future lung disease. The CARDIA Lung Study. *Am J Respir Crit Care Med* 2018; 197: 1616–1624.
19 Colak Y, Afzal S, Nordestgaard BG, *et al.* Prevalence, characteristics, and prognosis of early COPD: the Copenhagen General Population Study. *Am J Respir Crit Care Med* 2020; 201: 671–680.
20 Woodruff PG, Barr RG, Bleecker E, *et al.* Clinical significance of symptoms in smokers with preserved pulmonary function. *N Engl J Med* 2016; 374: 1811–1821.
21 Martinez FJ, Mannino D, Leidy NK, *et al.* A new approach for identifying patients with undiagnosed chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2017; 195: 748–756.
22 Martinez FD. Early-life origins of chronic obstructive pulmonary disease. *N Engl J Med* 2016; 375: 871–878.
23 Simpson SJ, Turkovic L, Wilson AC, *et al.* Lung function trajectories throughout childhood in survivors of very preterm birth: a longitudinal cohort study. *Lancet Child Adolesc Health* 2018; 2: 350–359.
24 Raju S, Keet CA, Paulin LM, *et al.* Rural residence and poverty are independent risk factors for chronic obstructive pulmonary disease in the United States. *Am J Respir Crit Care Med* 2019; 199: 961–969.
25 Alcaide AB, Sanchez-Salcedo P, Bastarrika G, *et al.* Clinical features of smokers with radiological emphysema but without airway limitation. *Chest* 2017; 151: 358–365.
26 Celli BR, Agustí A. COPD: time to improve its taxonomy? *ERJ Open Res* 2018; 4: 00132–02017.
27 Coxson HO, Dirksen A, Edwards L, *et al.* The presence and progression of emphysema in COPD as determined by CT scanning and biomarker expression: a prospective analysis from the ECLIPSE study. *Lancet Respir* 2013; 1: 129–136.
28 Regan EA, Lynch DA, Curran-Everett D, *et al.* Clinical and radiologic disease in smokers with normal spirometry. *JAMA Intern Med* 2015; 175: 1539–1549.
29 McDonough JE, Yuan R, Suzuki M, *et al.* Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. *N Engl J Med* 2011; 365: 1567–1575.