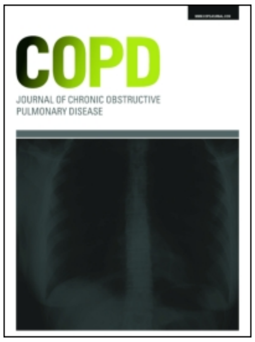

COPD: Journal of Chronic Obstructive Pulmonary Disease

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/icop20

# Comprehensive Cluster Analysis for COPD Including Systemic and Airway Inflammatory Markers

Halehsadat Nekoee Zahraei , Françoise Guissard , Virginie Paulus , Monique Henket , Anne-Françoise Donneau & Renaud Louis

COPD: JOURNAL OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE
https://doi.org/10.1080/15412555.2020.1833853



# Comprehensive Cluster Analysis for COPD Including Systemic and Airway Inflammatory Markers

Halehsadat Nekoee Zahraei[a,b], Françoise Guissard[b], Virginie Paulus[b], Monique Henket[b], Anne-Françoise Donneau[a], and Renaud Louis[b]

[a]Biostatistics Unit, Department of Public Health, University of Liège, Liège, Belgium; [b]Department of Pneumology, GIGA, University of Liège, Liège, Belgium

**ABSTRACT**

Chronic obstructive pulmonary disease (COPD) is a complex, multidimensional and heterogeneous disease. The main purpose of the present study was to identify clinical phenotypes through cluster analysis in adults suffering from COPD. A retrospective study was conducted on 178 COPD patients in stable state recruited from ambulatory care at University hospital of Liege. All patients were above 40 years, had a smoking history of more than 20 pack years, post bronchodilator FEV1/FVC <70% and denied any history of asthma before 40 years. In this study, the patients were described by a total of 84 mixed sets of variables with some missing values. Hierarchical clustering on principal components (HCPC) was applied on multiple imputation. In the final step, patients were classified into homogeneous distinct groups by consensus clustering. Three different clusters, which shared similar smoking history were found. Cluster 1 included men with moderate airway obstruction ($n = 67$) while cluster 2 comprised men who were exacerbation-prone, with severe airflow limitation and intense granulocytic airway and neutrophilic systemic inflammation ($n = 56$). Cluster 3 essentially included women with moderate airway obstruction ($n = 55$). All clusters had a low rate of bacterial colonization (5%), a low median FeNO value (<20 ppb) and a very low sensitization rate toward common aeroallergens (0-5%). CAT score did not differ between clusters. Including markers of systemic airway inflammation and atopy and applying a comprehensive cluster analysis we provide here evidence for 3 clusters markedly shaped by sex, airway obstruction and neutrophilic inflammation but not by symptoms and T2 biomarkers.

**ARTICLE HISTORY**
Received 17 March 2020
Accepted 1 October 2020

**KEYWORDS**
chronic obstructive pulmonary disease (COPD); cluster analysis; phenotypes; factor analysis of mixed data; missing data

## Introduction

Chronic obstructive pulmonary disease (COPD) is a complex and heterogeneous disease with a large number of subtypes and a multifactorial background [1, 2]. It may be caused by different pathophysiologic mechanisms (sometimes referred to as endotypes) but may share similar observed characteristics (phenotypes) [3]. These phenotypes divide all patients into several groups with common features that help patients to receive effective care and achieve better clinical results.

Within the data mining framework one of the most recognized methods for discovering knowledge in multivariate dataset, is clustering. Cluster analysis attempt to find groups of patients such that patients in the same cluster are more similar to each other than to patients in another cluster [4]. In recent years, cluster analysis was applied as a popular method to examine heterogeneity of patients with asthma [5, 6] or COPD [7–12]. Within the clustering framework, missing values, which are present in such large datasets, and how to handle it have not been referred among different studies [13]. As different analysis methods exist with their own criterion for clustering and determination of the number of clusters, hierarchical and nonhierarchical, different results are proposed. Another key point is number of variables in clustering, how select them and this fact that correlation exists between initially selected variables, the strategies of data transformation using principal component analysis solved this point [9, 10, 14–17].

In this paper, we introduce a flexible framework for cluster analysis on multidimensional dataset, which handles missing value by multiple imputation. COPD clustering studies have brought insight to the importance of co-morbidities and systemic inflammation as components accounting for disease variability [18]. There are studies that used demographic variables, symptoms, spirometry, imaging and comorbidities to build the clusters [19]. However, there are not many studies using clustering that investigated the airway inflammatory component and the atopic status in large cohort of COPD. As it is recognized that some COPD may express T2 biomarkers [20], we included serum IgE and FeNO in our standard routine investigation of COPD in order to see whether the T2 trait is frequent and strong enough to shape a cluster in a population of COPD denying any existence of asthma before the age of 40 years.

CONTACT Halehsadat Nekoee Zahraei H.Nekoee@uliege.be Bat.B23 Biostatistics, Quartier Hopital, Avenue de l'Hopital, 3, 4000 Liege, Belgium.
Supplemental data for this article can be accessed at https://doi.org/10.1080/15412555.2020.1833853.

Here we performed clustering analysis on a cohort of 178 COPD patients recruited from ambulatory care in whom detailed lung function, blood and sputum analysis were available.

## Material and methods

The retrospective study was conducted on 178 stable COPD patients recruited from ambulatory care in our COPD clinic in the Pneumology Department of the University hospital of Liege. We have a general agreement from the ethics committee to use clinical data collected from routine practice to make retrospective reports. The protocol was approved by the Hospitalo-Facultaire Universitaire ethics committee, Liege (institutional review board 2005/181). Every patient attending ambulatory clinic care signs an informed consent stating that they accept this principle. Selection criteria to be referred to our COPD clinic were symptomatic patients (including at least one of the three following symptoms: dyspnea, cough and sputum production) with FEV1/FVC ratio post bronchodilation less than 70%, age above 40 years and smoking history of at least 20 pack years. None had an asthma history starting before the age of 40. At the COPD clinic the patients had systematic pre and post bronchodilation spirometry, sputum induction, blood sampling, and completed the self-administered CAT questionnaire. From the clinical data, a comprehensive list of 84 variables was derived and divided into six categories, i.e. demographics, pulmonary function tests, treatment features, blood cell counts and systemic inflammatory markers, atopic status, and sputum cell counts and microbiology. One unavoidable problem in huge dataset is the presence of missing value. The percentage of missing values ranged from 0% to 23% and 75% of patients presented at least one missing value. This matter creates serious problems as most of the classical statistical methods are not designed to handle incomplete data. By default, in statistical methods and software, patients with at least one missing value are discarded from the analysis. In practical research, multiple imputation is a popular and very flexible technique for handling missing value [21]. Multiple imputation replaces each missing value with a set of $m$ ($>1$) plausible values. Therefore, instead of one incomplete dataset, multiple imputation provides $m$ separate complete datasets [22]. In the present study, $m$ was fixed to 100. Then, 100 imputed datasets are analyzed by the method that would have been appropriate if the data had been complete. The derived results from the analyses are then combined to produce the final quantity of interest following Rubin's rules [23]. The number of variables is an important issue in cluster analysis specially in determining the number of clusters. Indeed, large number of variables result in poor discrimination distance and misspecification classification. Therefore, in this study with huge numbers of variables, the percentage of contribution of variables was determined, and then, the strategy of variable reduction was applied. Cluster analysis is a powerful but unsupervised method. The big issue in this unsupervised method is, to evaluate the quality of the clustering framework for classification. Clustering validation was considered to evaluate this issue. Statistical methodology is discussed in the online supplement in more detail.

**Table 1.** Process of factor analysis and cluster analysis to describe phenotypes in COPD.

| |
|---|
| (1) Multiple imputation |
| i) Obtain *100* complete dataset by multiple imputation (MICE) |
| (2) Factor analysis for mixed data (FAMD) |
| i) Determine quantitative and qualitative variable |
| ii) Apply FAMD for each imputed dataset |
| iii) Determine the number of components for each imputed dataset |
| (3) Hierarchical clustering |
| i) Choosing the best number of clusters for each imputed dataset |
| (4) Partitioning Clustering |
| i) Consider the number of clusters in previous step |
| ii) Assign patients to each cluster for each imputed dataset by partitioning clustering |
| (5) Consensus clustering |
| i) Combine all ensemble clustering to get a final best clustering |
| (6) Assign patients to the final result of consensus clustering |
| i) Allocate patients in raw dataset (dataset with missing value) and each imputed dataset to calculate final result of consensus clustering |
| (7) Description of clustering |
| i) Calculate median for raw dataset and overall median for imputed datasets |
| ii) Comparison between cluster (Kruskal-Wallis and Chi-squared tests and Dunn's multiple comparison test) |

## Statistical method

Quantitative variables were summarized using median and interquartile range (P25 - P75); while count and percentage were used for qualitative variables. Outlines of the proposed steps to assign clustering for patients are described in Table 1. Missing values were imputed by draw from the posterior predictive distribution of bayesian model and predictive mean matching (PMM) was used as a robust method to model misspecification in imputing values. Since this study contains quantitative and qualitative variables, factor analysis for mixed data (FAMD) was applied for creating new components. In cluster analysis step, the number of clusters for each imputed dataset was determined by hierarchical clustering and a package of 30 indices for determining the relevant number of clusters, then K-means was applied for assigning clusters to patients. In consensus step, final clustering result was achieved by minimizing the sum of squared distance of existing clustering results. For each clustering output, two indices for internal clustering validation and stability validation were calculated. Output of consensus clustering was considered as the individual final clustering result for raw and all of imputed datasets. Then, median and interquartile range were calculated for all variables in raw dataset and for each imputed dataset. Finally, an overall median with corresponding interquartile range was calculated over all 100 imputed datasets. All variables were compared between the derived clusters using Kruskal-Wallis and Chi-squared tests for quantitative and qualitative variables, respectively. Comparison among clusters was applied according to Dunn's multiple comparison test. Finally, the difference between three groups was displayed by boxplot. All analyses were performed using R statistical software. P values $<0.05$ were considered as statistically significant.

**Table 2.** Characteristics of the COPD cohort.

| | Variable | | Median (IQR)/percentage (frequency) | Percentage (number) of missing value |
|---|---|---|---|---|
| Demographic | Age (year) | | 64.5(57–72) | 0% (0) |
| | Sex (Male) | | 54.49% (97) | 0% (0) |
| | Height (cm) | | 167(160–175) | 0.56% (1) |
| | Weight (kg) | | 67(58–78) | 0.56% (1) |
| | BMI (kg/m$^2$) | | 23.62 (21.22–27.18) | 0.56% (1) |
| | Cigarette Packs (year) | | 37.2 (22.5–50) | 2.25% (4) |
| | Cigarettes (day) | | 20 (10–23.5) | 3.93% (7) |
| | Smoking Duration(year) | | 43 (33–50) | 2.8% (5) |
| | OCS Course | 0 | 60.71% (102) | |
| | | 1 | 23.81% (40) | 5.62% (10) |
| | | ≥2 | 15.48% (26) | |
| | Antibiotic Course | 0 | 36.47% (62) | 4.49% (8) |
| | | 1 | 51.76% (88) | |
| | | ≥2 | 11.76% (20) | |
| | Emergency Room Admission for asthma or COPD | 0 | 84.66% (149) | 1.12% (2) |
| | | 1 | 14.77% (26) | |
| | | ≥2 | 0.57% (1) | |
| | Number of hospitalizations for asthma or COPD | 0 | 85.55% (148) | 2.8% (5) |
| | | 1 | 12.72% (22) | |
| | | ≥2 | 1.73% (3) | |
| Pulmonary | FeNO (ppb) | | 16 (10–25) | 7.30% (13) |
| | FEV1 (mL) | | 1380 (1085–1810) | 0% (0) |
| | FEV1 predicted (%) | | 53 (43–66) | 0% (0) |
| | FEV1 PD (mL) | | 1480 (1180–1930) | 0.56% (1) |
| | FEV1 PD predicted (%) | | 57 (47–71) | 0.56% (1) |
| | Reversibility (%) | | 7 (1–12) | 0.56% (1) |
| | FVC (mL) | | 2505 (2020–3067.5) | 0% (0) |
| | FVC predicted (%) | | 77 (64–89) | 0% (0) |
| | FVC post (mL) | | 2580 (2080–3200) | 0.56% (1) |
| | FVC post (%) | | 80 (68–92) | 0.56% (1) |
| | FEV1/ FVC pre (%) | | 56.4 (48.12–64.27) | 0% (0) |
| | FEV1/ FVC post (%) | | 57.9 (49.4–66.5) | 0.56% (1) |
| | TLC (L) | | 6.45 (5.37–7.17) | 16.29% (29) |
| | TLC predicted (%) | | 109 (96–123) | 16.29% (29) |
| | RV (L) | | 3.58 (3.07–4.55) | 16.29% (29) |
| | RV predicted (%) | | 168 (137–201) | 16.29% (29) |
| | RV/TLC (%) | | 59.58 (52.79–66.21) | 16.29% (29) |
| | DLCO (mmol/kPa.min) | | 3.98 (2.91–5.3) | 17.41% (31) |
| | DLCO predicted (%) | | 49 (37.5–60) | 17.41% (31) |
| | DLCO/VA | | 0.93 (0.71–1.16) | 17.41% (31) |
| | DLCO/VApredicted (%) | | 67 (50–81.5) | 17.41% (31) |
| | sGaw (1/kPa*sec) | | 0.47 (0.34–0.71) | 23.03% (41) |
| | FRC (L) | | 4.91 (4.09–5.89) | 21.34% (38) |
| | FRC predicted (%) | | 158 (138–183) | 21.34% (38) |
| Treatment | Treatment (Yes) | | 61.76% (105) | 4.49% (8) |
| | ICS (Yes) | | 55.11% (97) | 1.12% (2) |
| | OCS (Yes) | | 5.11% (9) | 1.12% (2) |
| | LAMA (Yes) | | 51.13% (90) | 1.12% (2) |
| | LABA (Yes) | | 67.04% (118) | 1.12% (2) |
| | SABA (Yes) | | 40.91% (72) | 1.12% (2) |
| | LTRA (Yes) | | 3.41% (6) | 1.12% (2) |
| | Theophylline (Yes) | | 2.84% (5) | 1.12% (2) |
| CAT score | | | 25 (16–31) | 0% (0) |
| Blood | Leucocytes (µL) | | 7.97 (6.67–9.73) | 7.86% (14) |
| | Neutrophils (%) | | 60.7 (54–67.75) | 8.43% (15) |
| | Lymphocytes (%) | | 27.8 (21.35–34.3) | 8.43% (15) |
| | Monocytes (%) | | 7.9 (6.55–9.45) | 8.43% (15) |
| | Eosinophils (%) | | 2 (1–3.25) | 8.43% (15) |
| | Basophils (%) | | 0.4 (0.3–0.6) | 8.43% (15) |
| | Neutrophils (µL) | | 4840.71 (3782.47–6027.27) | 8.43% (15) |
| | Lymphocytes (µL) | | 2213.4 (1739.08–2710.78) | 8.43% (15) |
| | Monocytes (µL) | | 646.99 (496.08–836.3) | 8.43% (15) |
| | Eosinophils (µL) | | 147.63 (77.46–252.0) | 8.43% (15) |
| | Basophils (µL) | | 31.24 (21.37–48.3) | 8.43% (15) |
| | Fibrinogen (g/l) | | 3.52 (2.96–4.02) | 4.49% (8) |
| | CRP (mg/l) | | 2.45 (1.10–5.75) | 4.49% (8) |
| | Alpha 1 antitrypsin (g/l) | | 1.5 (1.36–1.68) | 12.36% (22) |
| | Calcium (mmol/L) | | 2.42 (2.36–2.47) | 6.74% (12) |
| | 25(OH) Vitamine D (ng/ml) | | 20 (12–30.97) | 13.48% (24) |
| | Phosphate (mmol/L) | | 0.93 (0.79–1.06) | 6.74% (12) |
| Atopy | IgE (KU/L) | | 72.50 (22.75–227.75) | 10.11% (18) |
| | RAST DPT (d1) %>0.35 (KU/L) | | 11.39% (18) | 11.24% (20) |
| | RAST Cat (e1), %>0.35 (KU/L) | | 3.75% (6) | 10.11% (18) |
| | RAST Dog (e5), %>0.35 (KU/L) | | 3.12% (5) | 10.11% (18) |

(*continued*)

**Table 2.** Continued.

| | Variable | Median (IQR)/percentage (frequency) | Percentage (number) of missing value |
|---|---|---|---|
| | RAST Grass (GX3), %>0.35 (KU/L) | 6.96% (11) | 11.24% (20) |
| | RAST microog (MIX1), %>0.35 (KU/L) | 9.37% (15) | 10.11% (18) |
| | RAST Birch (t3), %>0.35 (KU/L) | 1.25% (2) | 10.67% (19) |
| Sputum | Positive Aerobic Sputum Culture | 8.96% (13) | 18.54% (33) |
| | Weight of sputum (g) | 1.72 (1.1–2.98) | 20.78% (37) |
| | Total Cell Counts ($10^6$/g) | 2.31 (0.98–5.6) | 20.78% (37) |
| | Squamous (%) | 10 (3–33) | 20.78% (37) |
| | Viability (%) | 69 (55–84) | 20.78% (37) |
| | Macrophages (%) | 12.3 (5–23.6) | 21.91% (39) |
| | Lymphocytes (%) | 1 (0–2) | 21.91% (39) |
| | Neutrophils (%) | 74.8 (56.87–91.05) | 21.34% (38) |
| | Eosinophils (%) | 1.35 (0.2–4.40) | 21.34% (38) |
| | Epithelial cells (%) | 2.5 (0.6–7.75) | 21.91% (39) |
| | Macrophages ($10^3$/g) | 268.28 (80.04–583.8) | 23.03% (41) |
| | Lymphocytes ($10^3$/g) | 18.48 (0–43.8) | 23.03% (41) |
| | Neutrophils ($10^3$/g) | 1307.5 (503.89–3831.61) | 22.47% (40) |
| | Eosinophils ($10^3$/g) | 28.86 (2.77–232.69) | 22.47% (40) |
| | Epithelial cells ($10^3$/g) | 55.4 (11.25–202.86) | 22.47% (40) |

## Results

Characteristics of the total patients and percentage of missing values before imputing are presented in Table 2. Patients were mostly males (54.49%) with age ranging from 40 to 84 years and displayed a normal weight (median body mass index was 23.62). Patients had a consistent tobacco consumption history with a median pack/year of 37.

Table 3 shows the order and the impact of each variable on clustering. The highest contribution for variables in clustering is for FEV1 (mL), FEV1 PD (mL), FEV1 PD predicted (%), FEV1 predicted (%) and FVC (mL) (Figure 1). Based on all operational processes indicated in Table 1, three distinct clusters with acceptable values for validation were identified weighing similarly regarding the number of patients from this procedure for COPD dataset and output for 100 imputed datasets displayed in Table 4. In this study, two indices for internal clustering validation and two indices for clustering stability validation are reported. Silhouette Width and Dunn Index values for Internal measures are 0.61 and 0.54, respectively. Average Proportion of Non-overlap and Average Distance between Means for Stability measures are 0.02 and 0.01, respectively (For more information, refer to supplementary, clustering validation). There were striking **sex** differences between the clusters with a clear dominance of male in clusters 1 and 2 while cluster 3 was essentially composed of women. Smoking history was similar between clusters and BMI was slightly higher in cluster 1 while still remaining in the normal range. Clusters 2 and 3 received more often courses of OCS the year prior to the visit whereas there was no difference regarding the number of antibiotic course. Clusters 2 and 3 were those in which patients received more maintenance treatment including ICS, LAMA, LABA and also used more often SABA as reliever. Cluster 2 and 3 had also more impaired lung function with more severe airway obstruction, lung hyperdistension and severely reduced diffusing capacity and transfer coefficient (Figure 2). As far as inflammation is concerned, the cluster 2 had more severe systemic and airway neutrophilic inflammation, with slightly raised fibrinogen but not CRP. Circulating lymphocytes were reduced in cluster 2 (Figures 3 and 4). Absolute sputum eosinophil counts were higher in cluster 2 than in cluster 1 while no difference was seen in blood (Figure 5). FeNO levels were similar between clusters and no difference was seen regarding total serum IgE (Figure 5) nor sensitizations to aeroallergens, which were rare in all three clusters. Interestingly, CAT score did not differ between the three clusters despite clear differences in lung function impairment (Figure 6).

## Discussion

In this study, we characterized COPD patients into three distinctly different groups by applying general and flexible statistical computation in dataset with missing values. In the present study, clustering was applied to a large number of variables instead of selecting a limited number of variables. Although missing values are a common and pervasive problem in diverse datasets such as COPD with large number of variables, missing values have not been considered property in the clustering literature. Based on these restrictions, classification on COPD datasets has not been comprehensively investigated. Therefore, in this exhaustive study, phenotypes in COPD dataset were described by imputing missing values, principal components and cluster analysis with many analytical decisions, which overcome limitations, often reported in previous clustering studies.

The concept of treatable trait has become very popular over the last years and it has been suggested to avoid label of asthma or COPD among patients with severe chronic airway diseases [24]. Adopting this taxonomic view our COPD population could be seen as a population featuring the trait of fixed airway obstruction after a significant smoking history and denying any previous diagnosis of asthma before the age of 40. One strength of our study, compared to previous clustering analysis in COPD, is that it included airway inflammatory features, FeNO and atopic status in the parameters subjected to analysis.

We actually found 3 clusters of COPD, strikingly linked to sex with two clusters showing male dominance while the third was essentially a female cluster. There were clear differences between lung function impairment between the clusters whereas quantitative smoking history was quite

**Table 3.** Percentage of the contribution of variables in clustering.

| Order | Variables in order of priority | Percentage of contribution value |
|---|---|---|
| 1 | FEV1 (mL) | 0.795354 |
| 2 | FEV1 PD (mL) | 0.792187 |
| 3 | FEV1 PD predicted (%) | 0.708569 |
| 4 | FEV1 predicted (%) | 0.696766 |
| 5 | FVC (mL) | 0.510238 |
| 6 | FVC post (mL) | 0.49998 |
| 7 | FVC predicted (%) | 0.488792 |
| 8 | FVC post (%) | 0.488091 |
| 9 | RV/TLC (%) | 0.373624 |
| 10 | FEV1/ FVC post (%) | 0.318253 |
| 11 | DLCO (mmol/kPa.min) | 0.290582 |
| 12 | Emergency Room Admission for asthma or COPD | 0.290175 |
| 13 | FEV1/ FVC pre (%) | 0.283042 |
| 14 | Blood_Neutrophils (%) | 0.207177 |
| 15 | Blood_Lymphocytes (%) | 0.185377 |
| 16 | DLCO predicted (%) | 0.167944 |
| 17 | sGaw (1/kPa*sec) | 0.166534 |
| 18 | LAMA (Yes) | 0.162801 |
| 19 | LABA (Yes) | 0.161971 |
| 20 | ICS (Yes) | 0.159877 |
| 21 | Blood_Neutrophils (μL) | 0.14896 |
| 22 | Treatment (Yes) | 0.148225 |
| 23 | RV predicted (%) | 0.143311 |
| 24 | Number of hospitalizations for asthma or COPD | 0.132193 |
| 25 | FRC predicted (%) | 0.12998 |
| 26 | Weight (kg) | 0.115709 |
| 27 | OCS Course | 0.099888 |
| 28 | RV (L) | 0.092256 |
| 29 | Height (cm) | 0.08524 |
| 30 | Blood_Lymphocytes (μL) | 0.076047 |
| 31 | DLCO/VA | 0.073329 |
| 32 | BMI (kg/m$^2$) | 0.067138 |
| 33 | DLCO/VApredicted (%) | 0.066027 |
| 34 | Blood_Leucocyes (μL) | 0.064596 |
| 35 | CAT score | 0.062324 |
| 36 | OCS (Yes) | 0.061563 |
| 37 | Blood_Monocytes (%) | 0.060734 |
| 38 | Total Cell Counts (10$^6$/g) | 0.060281 |
| 39 | Alpha 1 antitrypsin (g/l) | 0.059218 |
| 40 | Fibrinogen (g/l) | 0.057644 |
| 41 | TLC predicted (%) | 0.054684 |
| 42 | Viability (%) | 0.052038 |
| 43 | Sputum_Eosinophils (10$^3$/g) | 0.051767 |
| 44 | SABA (Yes) | 0.046993 |
| 45 | Antibiotic Course | 0.045641 |
| 46 | Age (year) | 0.044547 |
| 47 | Weight of sputum (g) | 0.042916 |
| 48 | Sputum_Neutrophils (10$^3$/g) | 0.04283 |
| 49 | Sex (Male) | 0.040351 |
| 50 | Sputum_Lymphocytes (10$^3$/g) | 0.035228 |
| 51 | Theophylline (Yes) | 0.034857 |
| 52 | Smoking Duration(year) | 0.030539 |
| 53 | CRP (mg/l) | 0.030115 |
| 54 | FRC (L) | 0.028202 |
| 55 | Positive Aerobic Sputum Culture | 0.027121 |
| 56 | Blood_Eosinophils (%) | 0.024198 |
| 57 | RAST Grass (GX3), %>0.35 (KU/L) | 0.020597 |
| 58 | Sputum_Neutrophils (%) | 0.01398 |
| 59 | Positive Aerobic Sputum Culture | 0.013011 |
| 60 | IgE (KU/L) | 0.010926 |
| 61 | 25(OH) Vitamine D (ng/ml) | 0.00967 |
| 62 | Blood_Basophils (%) | 0.008521 |
| 63 | FeNO (ppb) | 0.008171 |
| 64 | RAST Birch (t3), %>0.35 (KU/L) | 0.007232 |
| 65 | Squamous (%) | 0.007092 |
| 66 | Sputum_Epithelial cells (%) | 0.005941 |
| 67 | TLC (L) | 0.005787 |
| 68 | Sputum_Eosinophils (%) | 0.00525 |
| 69 | Cigarettes (day) | 0.005246 |
| 70 | Blood_Monocytes (μL) | 0.004464 |
| 71 | RAST DPT (d1) %>0.35 (KU/L) | 0.002943 |
| 72 | Sputum_Lymphocytes (%) | 0.002614 |
| 73 | RAST microog (MIX1), %>0.35 (KU/L) | 0.002255 |
| 74 | Sputum_Macrophages (10$^3$/g) | 0.001858 |
| 75 | Phosphate (mmol/L) | 0.001491 |
| 76 | RAST Cat (e1), %>0.35 (KU/L) | 0.001486 |
| 77 | LTRA (Yes) | 0.001457 |
| 78 | Sputum_Macrophages (%) | 0.001438 |
| 79 | Blood_Basophils (μL) | 0.001318 |
| 80 | RAST Dog (e5), %>0.35 (KU/L) | 0.001291 |
| 81 | Calcium (mmol/L) | 0.001184 |
| 82 | Blood_Eosinophils (μL) | 0.000887 |
| 83 | Reversibility (%) | 0.00079 |
| 84 | Cigarette Packs (year) | 0.00077 |

similar, pointing to different susceptibility to tobacco among patients. The percentage of contribution of the different variables to the clustering. It appears that functional criteria including airway flow and the degree of airway obstruction and lung hyperdistension and the % of blood lymphocytes are amongst the most important criteria to structure the cluster while variables like smoking history, FeNO and Vitamin D level were rather homogeneous between the subjects.

The cluster 2 is conspicuously the most severe group of patients with marked airway obstruction, lung hyperdistension capacity together with intense neutrophilic inflammation both at the systemic and the airway level, in keeping with the previously reported relationship between the severity of airway obstruction and the neutrophilic inflammation [25]. Cluster 2 and 3 had also impaired diffusing capacity and transfer coefficient pointing to emphysema. Despite severe emphysema the level of α1- antitrypsin was higher in cluster 2, perhaps indicating a response of the body trying to counteract the lung destruction favored by the intense neutrophilic inflammation.

Associated with neutrophilic inflammation, cluster 2, displays a small rise in fibrinogen level even through the median value remained within the normal range and below the threshold of 5.1 g/l, shown to be predictive of an excess of mortality [26]. Of note is the fact that the intensity in neutrophilic airway inflammation is not associated with bacterial colonization identified by classical bacterial culture, which was rather low come close to 10% for the whole cohort. Of course, it does not imply that microbiome may be profoundly disturbed in COPD and more sophisticated microbiological analyses might have revealed differences between the clusters. Altered microbiome may be the consequence of frequent antibiotic courses received by the patients as shown in our cohort since almost two third of the patients had received antibiotics for bronchitis the year prior the visit. It is worth noting that there was no difference between clusters in the number of antibiotic courses. As opposed to exacerbation defined by OCS course, exacerbation defined by antibiotic course was not related the severity of lung function impairment nor to the severity of airway inflammation.

**Table 4.** Characteristics of patients with COPD after imputation, Median (IQR) / Percentage (frequency) in each cluster and comparison between clusters.

| | Variable | | Result for clustering on 100 imputed datasets | | | |
|---|---|---|---|---|---|---|
| | | | Cluster 1 ($n = 67$) | Cluster 2 ($n = 56$) | Cluster 3 ($n = 55$) | P-value |
| Demographic | Age (year) | | 62(55–66)[a] | 67(60.75–74.25)[b] | 67(58.5–73)[b+] | <0.001 |
| | Sex (Male) | | 76.12%(51)[a] | 75%(42)[a] | 7.27%(4)[b] | <0.0001 |
| | Height (cm) | | 173(166.5–178)[a] | 169.5(162–17.25)[a] | 161(156–165)[b] | <0.0001 |
| | Weight (kg) | | 77(65–89)[a] | 67(56.75–76)[b] | 60(53–66)[c] | <0.0001 |
| | BMI (kg/m$^2$) | | 25.59(22.34–30.21)[a] | 23.08(20.55–24.89)[b] | 23.15(19.89–25.55)[b] | <0.001 |
| | Cigarette Packs (year) | | 36.9(21.52–50)[a] | 42.5(24.62–52.5)[a] | 34.5(21.42–43.87)[a] | >0.05 |
| | Cigarettes (day) | | 20(10–25)[a] | 20(10–21.25)[a] | 20(10–20)[a] | >0.05 |
| | Smoking Duration(year) | | 40(31–46)[a] | 46.5(34.25–52.25)[b] | 44(37–50)[b] | <0.05 |
| | OCS Course | 0 | 80.59%(54) [a] | 30.35%(17) [b] | 61.82%(34) [a] | <0.0001 |
| | | 1 | 13.43%(9) | 32.14%(18) | 23.64%(13) | |
| | | ≥2 | 4.48%(3) | 35.71%(20) | 10.91%(6) | |
| | Antibiotic Course | 0 | 43.28%(29) [a] | 32.14%(18) [a] | 32.73%(18) [a] | >0.05 |
| | | 1 | 46.27%(31) | 50%(28) | 58.18%(32) | |
| | | ≥2 | 8.95%(6) | 16.07%(9) | 10.91%(6) | |
| | Emergency Room Admission | 0 | 95.52%(64) [a] | 71.43%(40) [b] | 85.45%(47) [a] | <0.0001 |
| | for asthma or COPD | 1 | 4.48%(3) | 26.78%(15) | 14.54%(8) | |
| | | ≥2 | 0%(0) | 1.78%(1) | 0%(0) | |
| | Number of hospitalizations | 0 | 97.01%(65) [a] | 67.86%(38) [b] | 89.09%(49) [a] | <0.0001 |
| | for asthma or COPD | 1 | 2.98%(2) | 26.78%(15) | 10.91%(6) | |
| | | ≥2 | 0%(0) | 5.36%(3) | 0%(0) | |
| Pulmonary | FeNO (ppb) | | 15(10.5–23)[a] | 18.5(10–30.5)[a] | 14(8–25.5)[a] | >0.05 |
| | FEV1 (mL) | | 1950(1700–2205)[a] | 1085(780–1352.5)[b] | 1240(1075–1390)[b] | <0.0001 |
| | FEV1 predicted (%) | | 66(56–78)[a] | 37(30–45.25)[b] | 55(47.5–66)[c] | <0.0001 |
| | FEV1 PD (mL) | | 2000(1820–2405)[a] | 1190(837.5–1380)[b] | 1310(1135–1470)[b] | <0.0001 |
| | FEV1 PD (%) | | 69(60.5–81.5)[a] | 40(32.75–50)[b] | 59(51.5–69.5)[c] | <0.0001 |
| | Reversibility (%) | | 7(1–10.5)[a] | 6.5(1–11.25)[a] | 8(2–13.5)[a] | >0.05 |
| | FVC (mL) | | 3200(2625–3780)[a] | 2427.5(1705–2767.5)[b] | 2140(1820–2390)[c] | <0.0001 |
| | FVC predicted (%) | | 86(74–99.5)[a] | 65(55.75–74.25)[b] | 77(68.5–85.5)[c] | <0.0001 |
| | FVC post (mL) | | 3320(2805–3870)[a] | 2385(1797.5–2900)[b] | 2130(1870–2445)[c] | <0.0001 |
| | FVC post (%) | | 90(79.5–101)[a] | 68(56.75–76.5)[b] | 80(71.5–91.5)[c] | <0.0001 |
| | FEV1/ FVC pre (%) | | 63.1(56.45–66.65)[a] | 47(40.37–50.42)[b] | 57.1(53.15–64.5)[a] | <0.0001 |
| | FEV1/ FVC post (%) | | 64(58.3–68.55)[a] | 47.5(42.8–53.2)[b] | 59.9(54.5–67.05)[a] | <0.0001 |
| | TLC (mL) | | 6725(5725–7385)[a] | 6830(5847.5–8107.5)[a] | 5440(4975–6005)[b] | <0.0001 |
| | TLC predicted (%) | | 103(93–112)[a] | 113(100.75–125.25)[b] | 115(102–127.5)[b] | <0.001 |
| | RV (mL) | | 3320(2970–3820)[a] | 4515(3690–5277.5)[b] | 3270(2920–3725)[a] | <0.0001 |
| | RV (%) | | 141(128–171.5)[a] | 183.5(163.87–228.5)[b] | 172(148.5–193.25)[c] | <0.0001 |
| | RV/TLC (%) | | 52.99(46.71–57.18)[a] | 66.37(61.28–69.14)[b] | 60.84(57.03–64.06)[c] | <0.0001 |
| | DLCO (mmol/kPa.min) | | 5.2(4.06–5.88)[a] | 3.47(2.67–4.44)[b] | 3.27(2.69–4.03)[b] | <0.0001 |
| | DLCO predicted (%) | | 57(46.5–67)[a] | 41.5(30.75–58)[b] | 45(37.5–53.5)[b] | <0.0001 |
| | DLCO/VA | | 1.03(0.79–1.28)[a] | 0.80(0.63–0.99)[b] | 0.93(0.71–1.15)[b] | <0.0001 |
| | DLCO/VA predicted (%) | | 70(57–88)[a] | 59.5(45.75–79)[b] | 60(50–77.5)[b] | <0.001 |
| | sGaw (1/kPa*sec) | | 0.65(0.44–0.88)[a] | 0.35(0.24–0.45)[b] | 0.47(0.34–0.73)[c] | <0.0001 |
| | FRC (L) | | 4.83(4.08–5.43)[a] | 5.87(4.82–6.49)[b] | 4.24(3.77–4.95)[c] | <0.0001 |
| | FRC predicted (%) | | 143(122–161)[a] | 174(154.75–192)[b] | 160.5(141.5–180.5)[c] | <0.0001 |
| Treatment | Treatment (Yes) | | 44.78%(30)[a] | 78.57%(44)[b] | 67.27%(37)[b] | <0.0001 |
| | ICS (Yes) | | 29.85%(20)[a] | 71.43%(40)[b] | 69.10%(38)[b] | <0.0001 |
| | OCS (Yes) | | 1.49%(1)[a] | 16.07%(9)[b] | 0%(0)[a] | <0.0001 |
| | LAMA (Yes) | | 34.33%(23)[a] | 73.21%(41)[b] | 49.09%(27)[a] | <0.0001 |
| | LABA (Yes) | | 44.78%(30)[a] | 82.14%(46)[b] | 76.36%(42)[a] | <0.0001 |
| | SABA (Yes) | | 25.37%(17)[a] | 50%(28)[b] | 49.09%(27)[b] | <0.001 |
| | LTRA (Yes) | | 2.98%(2)[a] | 0%(0)[a] | 7.27%(4)[a] | >0.05 |
| | Theophylline (Yes) | | 1.49%(1)[a] | 5.36%(3)[a] | 1.82%(1)[a] | >0.05 |
| CAT score | | | 22(15–31)[a] | 26(19–30)[a] | 24(17–30)[a] | >0.05 |
| Blood | Leucocytes (μL) | | 7.71(6.52–9.43)[a] | 8.56(6.87–11.10)[a] | 7.81(6.64–9.2)[a] | >0.05 |
| | Neutrophils (%) | | 58.2(52.75–63.55)[a] | 67.6(60.1–73.22)[b] | 58.3(53.25–64.85)[a] | <0.0001 |
| | Lymphocytes (%) | | 30.3(25.65–35)[a] | 20.95(15.7–27.27)[b] | 29(23.82–35.12)[a] | <0.0001 |
| | Monocytes (%) | | 8.7(7.1–9.8)[a] | 7.8(6.34–9.82)[b] | 7.1(6.4–8.67)[b] | <0.05 |
| | Eosinophils (%) | | 2(1.1–3.25)[a] | 1.9(0.9–3.22)[a] | 2.1(0.95–3.35)[a] | >0.05 |
| | Basophils (%) | | 0.4(0.3–0.6)[a] | 0.3(0.2–0.5)[a] | 0.4(0.3–0.7)[a] | >0.05 |
| | Neutrophils (μL) | | 4529.13(3723.5–5501.7)[a] | 5574.91(4097.7–7842.7)[b] | 4840.71(3393.7–5743.1)[a] | <0.05 |
| | Lymphocytes (μL) | | 2346.8(1844.6–2968.1)[a] | 1847.22(1340.4–2382.2)[b] | 2253.3(1846.6–2710.8)[a] | <0.0001 |
| | Monocytes (μL) | | 668.82(507.24–831.28)[a] | 652.58(511.72–864.82)[a] | 605.82(485.32–748.05)[a] | >0.05 |
| | Eosinophils (μL) | | 138(75.12–240.24)[a] | 168.78(69.23–258.42)[a] | 147.63(88.26–255.97)[a] | >0.05 |
| | Basophils (μL) | | 30.84(21.48–42.58)[a] | 27.63(18.88–46.25)[a] | 32.04(23.95–48.97)[a] | >0.05 |
| | Fibrinogen (g/l) | | 3.43(2.88–3.86)[a] | 3.65(3.08–4.50)[b] | 3.54(2.98–3.80)[b] | <0.05 |
| | CRP (mg/l) | | 2.3(1.25–4.95)[a] | 2.9(0.87–7.7)[a] | 1.8(1–4.75)[a] | >0.05 |
| | Alpha 1 antitrypsin (g/l) | | 1.49(1.34–1.57)[a] | 1.61(1.39–1.76)[b] | 1.49(1.35–1.64)[a] | <0.05 |
| | Calcium (mmol/L) | | 2.42(2.37–2.47)[a] | 2.41(2.35–2.48)[a] | 2.44(2.38–2.47)[a] | >0.05 |
| | 25(OH) Vitamine D (ng/ml) | | 18(12–30.9)[a] | 17.5(10–26.75)[a] | 24(15.5–33)[a] | >0.05 |
| | Phosphate (mmol/L) | | 0.89(0.77–0.95)[a] | 0.88(0.72–1.06)[a] | 1.01(0.93–1.11)[b] | <0.0001 |
| Atopy | IgE (KU/L) | | 52(23–184.5)[a] | 91.5(28.75–304.75)[a] | 73(21–170)[a] | >0.05 |
| | RAST DPT (d1) %>0.35 (KU/L) | | 13.43%(9)[a] | 16.07%(9)[a] | 7.27%(4)[a] | >0.05 |

(*continued*)

**Table 4.** Continued.

| | Variable | Result for clustering on 100 imputed datasets Cluster 1 ($n=67$) | Cluster 2 ($n=56$) | Cluster 3 ($n=55$) | P-value |
|---|---|---|---|---|---|
| | RAST Cat (e1), %>0.35 (KU/L) | 1.49%(1)[a] | 7.14%(4)[a] | 1.82%(1)[a] | >0.05 |
| | RAST Dog (e5), %>0.35 (KU/L) | 2.98%(2)[a] | 5.36%(3)[a] | 0%(0)[a] | >0.05 |
| | RAST Grass (GX3), %>0.35 (KU/L) | 5.97%(4)[a] | 14.28%(8)[a] | 1.82%(1)[a] | >0.05 |
| | RAST microog (MIX1),%>0.35 (KU/L) | 7.46%(5)[a] | 12.5%(7)[a] | 10.91%(6)[a] | >0.05 |
| | RAST Birch (t3), %>0.35 (KU/L) | 0%(0)[a] | 3.57%(2)[a] | 0%(0)[a] | >0.05 |
| Sputum | Positive Aerobic Sputum Culture | 8.95%(6)[a] | 16.07%(9)[a] | 12.73%(7)[a] | >0.05 |
| | Weight of sputum (g) | 1.77(1.28–3.01)[a] | 1.45(0.97–3.05)[a] | 1.65(0.93–3.28)[a] | >0.05 |
| | Total Cell Counts ($10^6$/g) | 1.70(0.90–4.41)[a] | 4.90(1.58–15.96)[b] | 2.65(1.08–5.43)[a] | <0.05 |
| | Squamous (%) | 19(6–42)[a] | 7(2–31.12)[b] | 9.5(2–23.5)[b] | <0.05 |
| | Viability (%) | 68(53.5–85.5)[a] | 69(51–86)[a] | 67(48.5–79.5)[a] | >0.05 |
| | Macrophages (%) | 14(6.7–26.1)[a] | 6.5(2.54–15.15)[b] | 17(7–29.95)[a] | <0.0001 |
| | Lymphocytes (%) | 1.2(0–2.3)[a] | 0.6(0.15–1.5)[a] | 1.2(0–3)[a] | >0.05 |
| | Neutrophils (%) | 71.4(56.35–88.2)[a] | 78.9(58.8–92.19)[a] | 60.4(39.3–81.05)[b] | <0.001 |
| | Eosinophils (%) | 1.4(0.2–4.92)[a] | 2.3(0.2–9.12)[a] | 2(0.7–9.15)[a] | >0.05 |
| | Epithelial cells (%) | 2.4(0.4–8.1)[a] | 3.25(0.57–11.05)[a] | 4.7(1.3–15.2)[a] | >0.05 |
| | Macrophages ($10^3$/g) | 233.55(53.28–691.65)[a] | 301.29(31.37–755.45)[a] | 330.48(58.13–816.67)[a] | >0.05 |
| | Lymphocytes ($10^3$/g) | 14.2(0–43.07)[a] | 23.75(4.39–63.9)[a] | 25.2(3.42–72.5)[a] | >0.05 |
| | Neutrophils ($10^3$/g) | 1090(490.86–2339.64)[a] | 2482.84(883.9–11720.8)[b] | 1110(493.04–3483.82)[a] | <0.05 |
| | Eosinophils ($10^3$/g) | 19.92(0–166.96)[a] | 71.67(0–638.26)[b] | 33.04(5.65–293.7)[a] | <0.001 |
| | Epithelial cells ($10^3$/g) | 38.45(5.66–142.29)[a] | 94(18.07–292.65)[a] | 81.36(16.4–301.24)[a] | >0.05 |

$^{+}$There are no significant differences between two clusters with the same letter.

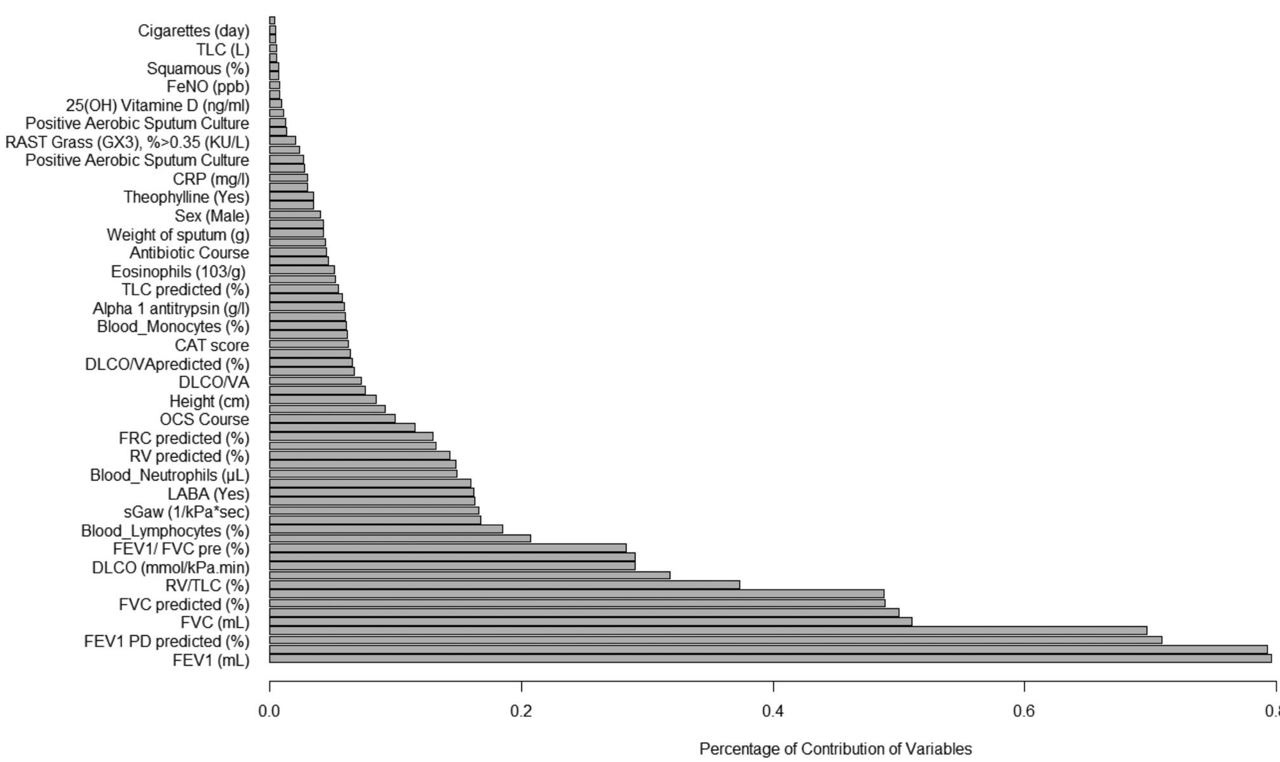


**Figure 1.** Percentage of the contribution of variables in clustering.

We also looked at phosphocalcic metabolism and found, as expected in western Europe, reduced levels of 25 OH Vit D (< 30 ng/ml) in all clusters without any difference between the groups. While Calcium levels, a tightly regulated ion, were normal and similar between the three clusters, there were striking differences in the levels of phosphates, which were clearly lower in cluster 1 and 2. The literature about phosphate level in COPD is virtually absent and the clinical meaning of our finding remains obscure though there might be a sex effect as this demographic trait best differentiates cluster 1 and 2 from cluster 3.

Eosinophilic trait is a marker of response to corticoids in asthma but also in COPD. In our study, blood eosinophilic inflammation does not appear to be discriminant feature between the clusters but cluster 2 shows a greater absolute, but not relative, sputum eosinophil cell count. However, the three clusters had median value of sputum eosinophil counts greater than that we found in a healthy population [27] and,

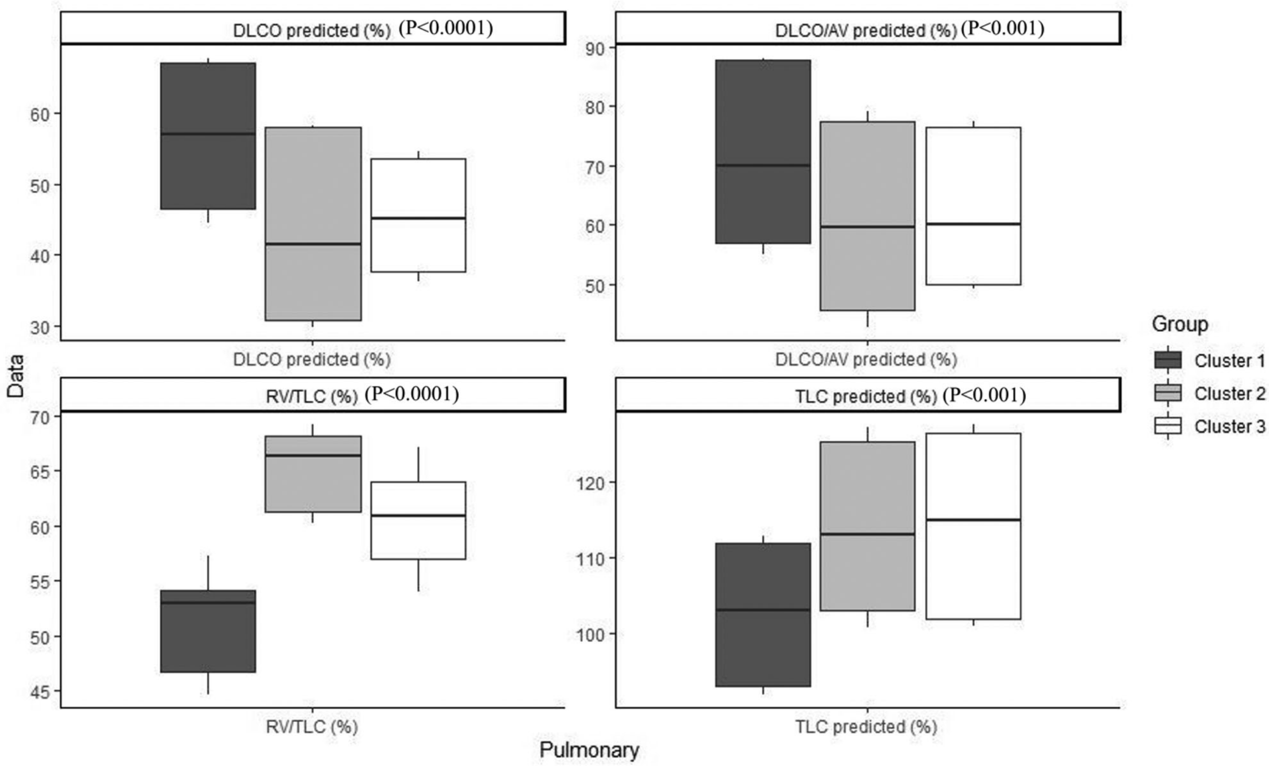


**Figure 2.** Lung function in three clusters.

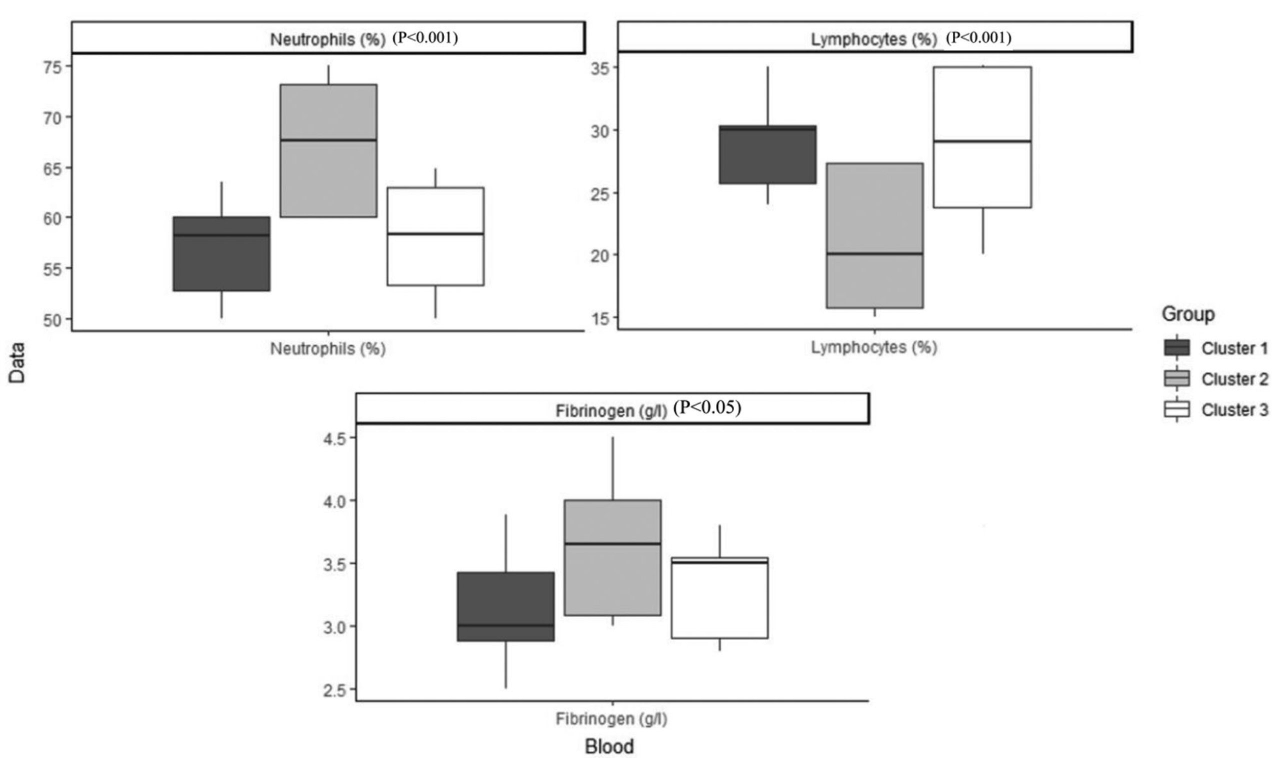


**Figure 3.** Blood in three clusters.

actually, rather similar to what is found in large population of unselected asthmatics [28]. Furthermore, the greater amount of eosinophils present in sputum of COPD was noted despite heavier treatment with ICS in this cluster, which points to some corticoresistance in this cluster.

Atopic status based on positive RAST toward aeroallergens was low in the three clusters but total serum IgE was

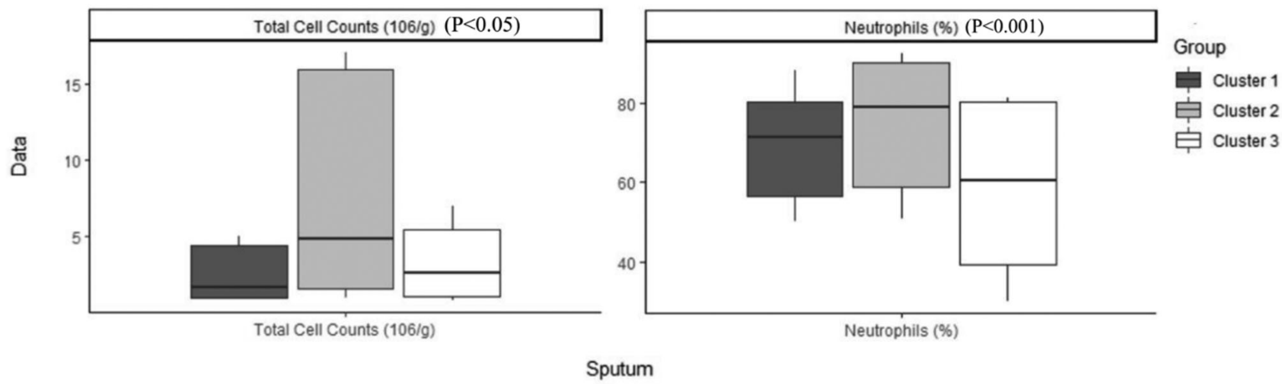


**Figure 4.** Sputum in three clusters.

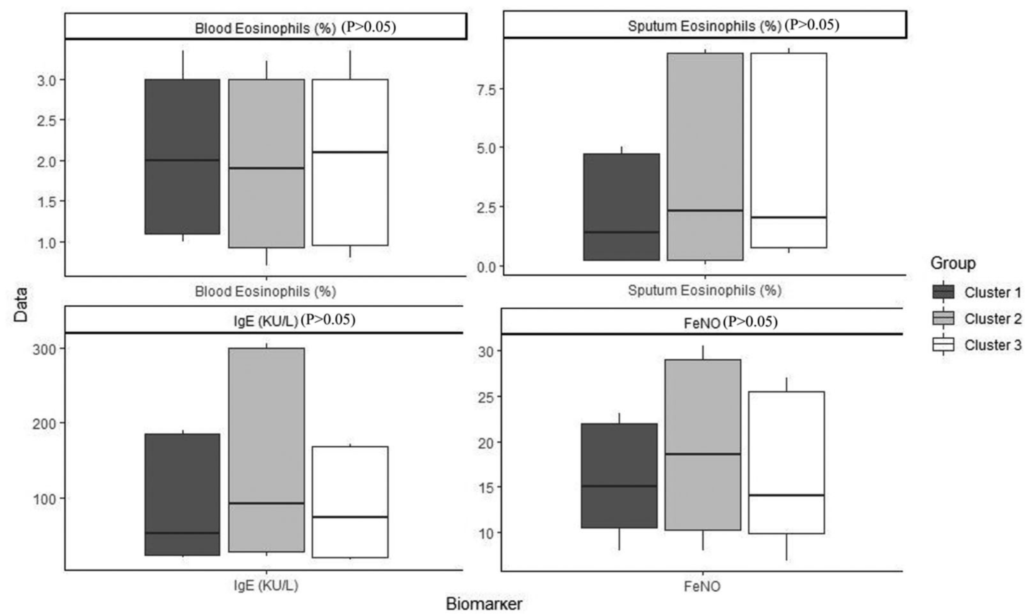


**Figure 5.** T2 biomarkers in three clusters.

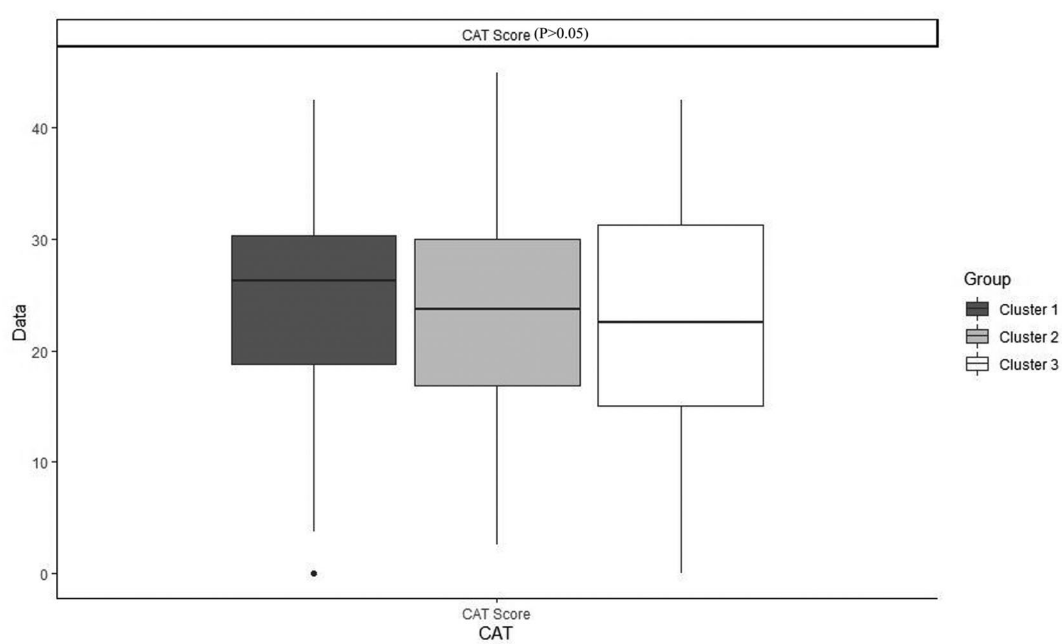


**Figure 6.** Cat Score in three clusters.

**Table 5.** Characteristics of patients with COPD before imputation, Median (IQR) / Percentage (frequency) in each cluster and comparison between clusters.

| | | | Result for clustering on raw dataset | | | |
|---|---|---|---|---|---|---|
| | Variable | | Cluster 1 | Cluster 2 | Cluster 3 | P-value |
| Demographic | Age (year) | | 62(55–66)$^{b}$ | 67(60.75–74.25)$^{a}$ | 67(58.5–73)$^{a}$ | <0.001 |
| | Sex (Male) | | 76.12%(51[67])$^{a}$ | 75%(42[56])$^{a}$ | 7.27%(4[55])$^{b}$ | <0.0001 |
| | Height (cm) | | 174(167–178)$^{a}$ | 169.5(162–175.25)$^{a}$ | 161(156–165)$^{b}$ | <0.0001 |
| | Weight (kg) | | 77(65–89)$^{c}$ | 67(56.75–76)$^{a}$ | 60(53–66)$^{b}$ | <0.0001 |
| | BMI (kg/m$^2$) | | 25.59(22.21–30.21)$^{b}$ | 23.08(20.56–24.89)$^{a}$ | 23.15(19.89–25.55)$^{a}$ | <0.001 |
| | Cigarette Packs (year) | | 36.9(21–50)$^{a}$ | 42.5(23.25–52.5)$^{a}$ | 35(22.05–43.87)$^{a}$ | >0.05 |
| | Cigarettes (day) | | 20(10–25)$^{a}$ | 20(10–21.25)$^{a}$ | 20(10–20)$^{a}$ | >0.05 |
| | Smoking Duration(year) | | 40(32–46.5)$^{c}$ | 47(34.25–52.25)$^{a}$ | 44(37–50)$^{ab}$ | <0.05 |
| | OCS Course | 0 | 81.25%(52[64]) $^{b}$ | 31.37%(16[51]) $^{a}$ | 64.15%(34[53]) $^{b}$ | <0.0001 |
| | | 1 | 14.06%(9[64]) | 35.29%(18[51]) | 24.53%(13[53]) | |
| | | ≥2 | 4.69%(3[64]) | 33.33%(17[51]) | 11.32%(6[53]) | |
| | Antibiotic Course | 0 | 42.86%(27[63]) $^{a}$ | 32.07%(17[53]) $^{a}$ | 33.33%(18[54]) $^{a}$ | >0.05 |
| | | 1 | 47.62%(30[63]) | 50.94%(27[53]) | 57.41%(31[54]) | |
| | | ≥2 | 9.52%(6[63]) | 16.98%(9[53]) | 9.26%(5[54]) | |
| | Emergency Room Admission for asthma or COPD | 0 | 95.45%(63[66]) $^{b}$ | 70.91%(39[55]) $^{a}$ | 85.45%(47[55]) $^{b}$ | <0.0001 |
| | | 1 | 4.54%(3[66]) | 27.27%(15[55]) | 14.54%(8[55]) | |
| | | ≥2 | 0%(0[66]) | 1.82%(1[55]) | 0%(0[55]) | |
| | Number of hospitalizations for asthma or COPD | 0 | 96.97%(64[66]) $^{b}$ | 67.92%(36[53]) $^{a}$ | 88.89%(48[54]) $^{b}$ | <0.0001 |
| | | 1 | 3.03%(2[66]) | 26.41%(14[53]) | 11.11%(6[54]) | |
| | | ≥2 | 0%(0[66]) | 1.89%(1[53]) | 0%(0[54]) | |
| Pulmonary | FeNO (ppb) | | 15(10.5–23)$^{a}$ | 19(10–32.25)$^{a}$ | 15(8–24)$^{a}$ | >0.05 |
| | FEV (mL) | | 1950(1700–2205)$^{b}$ | 1085(780–1352.5)$^{a}$ | 1240(1075–1390)$^{a}$ | <0.0001 |
| | FEV1 predicted (%) | | 66(56–78)$^{c}$ | 37(30–45.25)$^{a}$ | 55(47.5–66)$^{b}$ | <0.0001 |
| | FEV1 PD (mL) | | 2000(1820–2405)$^{b}$ | 1190(837.5–1380)$^{a}$ | 1310(1135–1470)$^{a}$ | <0.0001 |
| | FEV1 PD predicted (%) | | 69(60.5–81.5)$^{c}$ | 40(32.75–50)$^{a}$ | 59(51.5–69.5)$^{b}$ | <0.0001 |
| | Reversibility (%) | | 6(0.5–10.5)$^{a}$ | 6.5(1–11.25)$^{a}$ | 8(2–13.5)$^{a}$ | >0.05 |
| | FVC (mL) | | 3200(2625–3780)$^{c}$ | 2427.5(1705–2767.5)$^{a}$ | 2140(1820–2390)$^{b}$ | <0.0001 |
| | FVC predicted (%) | | 86(74–99.5)$^{c}$ | 65(55.75–74.25)$^{a}$ | 77(68.5–85.5)$^{b}$ | <0.0001 |
| | FVC post (mL) | | 3320(2805–3870)$^{c}$ | 2385(1797.5–2900)$^{a}$ | 2130(1870–2445)$^{b}$ | <0.0001 |
| | FVC post (%) | | 90(79.5–101)$^{c}$ | 68(56.75–76.5)$^{a}$ | 80(71.5–91.5)$^{b}$ | <0.0001 |
| | FEV1/ FVC pre (%) | | 63.1(56.45–66.65)$^{b}$ | 47(40.37–50.42)$^{a}$ | 57.1(53.15–64.5)$^{b}$ | <0.0001 |
| | FEV1/ FVC post (%) | | 63.9(58.3–68.55)$^{b}$ | 47.5(42.8–53.2)$^{a}$ | 59.9(54.5–67.05)$^{b}$ | <0.0001 |
| | TLC (mL) | | 6710(5725–7385)$^{a}$ | 6945(5832.5–8107.5)$^{a}$ | 5430(4890–6005)$^{b}$ | <0.0001 |
| | TLC predicted (%) | | 103(93–112.5)$^{b}$ | 112.5(100–125.25)$^{a}$ | 117(102–127)$^{a}$ | <0.001 |
| | RV (mL) | | 3450(2990–3820)$^{b}$ | 4640(3687.5–5277.5)$^{a}$ | 3250(2845–3690)$^{b}$ | <0.0001 |
| | RV predicted (%) | | 141(128–173)$^{c}$ | 182.5(162.5–228.5)$^{a}$ | 168(145.5–189.5)$^{b}$ | <0.0001 |
| | RV/TLC (%) | | 53.18(47.13–56.82)$^{c}$ | 66.23(61.34–69.03)$^{a}$ | 60.95(56.93–64.04)$^{b}$ | <0.0001 |
| | DLCO (mmol/kPa.min) | | 5.13(3.80–5.87)$^{b}$ | 3.55(2.69–4.45)$^{a}$ | 3.27(2.67–4)$^{a}$ | <0.0001 |
| | DLCO predicted (%) | | 56(43–67)$^{b}$ | 42.5(30.75–57.25)$^{a}$ | 45(38–55)$^{a}$ | <0.0001 |
| | DLCO/AV | | 1(0.79–1.28)$^{b}$ | 0.83(0.63–1.02)$^{a}$ | 0.93(0.69–1.13)$^{a}$ | <0.0001 |
| | DLCO/AV predicted (%) | | 70(56.5–88)$^{b}$ | 66(45.75–79)$^{a}$ | 62(50–75)$^{a}$ | <0.001 |
| | sGaw (1/kPa*sec) | | 0.68(0.47–0.93)$^{c}$ | 0.36(0.24–0.47)$^{a}$ | 0.47(0.35–0.76)$^{b}$ | <0.0001 |
| | FRC (L) | | 4.88(4.09–5.44)$^{c}$ | 5.86(4.88–6.49)$^{a}$ | 4.24(3.77–5.05)$^{b}$ | <0.0001 |
| | FRC predicted (%) | | 143(122–162)$^{c}$ | 171(154–195.75)$^{a}$ | 161(141.5–183.5)$^{b}$ | <0.0001 |
| Treatment | Treatment (Yes) | | 43.75%(28[64])$^{a}$ | 78.84%(41[52])$^{b}$ | 66.67%(36[54])$^{b}$ | <0.0001 |
| | ICS (Yes) | | 30.30%(20[66])$^{a}$ | 70.91%(39[55])$^{b}$ | 69.09%(38[55])$^{b}$ | <0.0001 |
| | OCS (Yes) | | 1.51%(1[66])$^{a}$ | 14.54%(8[55])$^{b}$ | 0%(0[55])$^{a}$ | <0.0001 |
| | LAMA (Yes) | | 34.85%(23[66])$^{a}$ | 72.73%(40[55])$^{b}$ | 49.09%(27[55])$^{a}$ | <0.0001 |
| | LABA (Yes) | | 45.45%(30[66])$^{a}$ | 83.64%(46[55])$^{b}$ | 76.36%(42[55])$^{a}$ | <0.0001 |
| | SABA (Yes) | | 25.76%(17[66])$^{a}$ | 50.91%(28[55])$^{b}$ | 49.09%(27[55])$^{b}$ | <0.001 |
| | LTRA (Yes) | | 3.03%(2[66])$^{a}$ | 0%(0[55])$^{a}$ | 7.27%(4[55])$^{a}$ | >0.05 |
| | Theophylline (Yes) | | 1.51%(1[66])$^{a}$ | 5.45%(3[55])$^{a}$ | 1.82%(1[55])$^{a}$ | >0.05 |
| CAT score | | | 22(15–31)$^{a}$ | 26(19–30)$^{a}$ | 24(17–30)$^{a}$ | >0.05 |
| Blood | Leucocytes (μL) | | 7.44(6.44–9.43)$^{a}$ | 8.58(7.02–10.75)$^{a}$ | 7.77(6.64–9.2)$^{a}$ | >0.05 |
| | Neutrophils (%) | | 58.9(52.75–63.75)$^{b}$ | 68.6(61.27–73.22)$^{a}$ | 57.7(53.25–64.55)$^{b}$ | <0.0001 |
| | Lymphocytes (%) | | 30.1(25.65–35)$^{b}$ | 20.7(15.7–26.85)$^{a}$ | 29(23.85–35.15)$^{b}$ | <0.0001 |
| | Monocytes (%) | | 8.7(7–9.8)$^{b}$ | 7.8(6.37–9.82)$^{a}$ | 7.1(6.4–8.75)$^{a}$ | <0.05 |
| | Eosinophils (%) | | 2.1(1.1–3.25)$^{a}$ | 1.8(0.85–3.12)$^{a}$ | 2.1(0.95–3.35)$^{a}$ | >0.05 |
| | Basophils (%) | | 0.4(0.3–0.6)$^{a}$ | 0.3(0.2–0.52)$^{a}$ | 0.4(0.3–0.7)$^{a}$ | >0.05 |
| | Neutrophils (μL) | | 4503.6(3554.3–5501.7)$^{b}$ | 5574.9(4116.2–7842.7)$^{a}$ | 4840.71(3393.7–5743.1)$^{b}$ | <0.05 |
| | Lymphocytes (μL) | | 2346.79(1853.9–3019.6)$^{b}$ | 1885.8(1294.4–2424.2)$^{a}$ | 2267.98(1846.6–2710.8)$^{b}$ | <0.0001 |
| | Monocytes (μL) | | 668.82(494.15–857.91)$^{a}$ | 707.29(534.38–864.82)$^{a}$ | 605.82(457.16–761.76)$^{a}$ | >0.05 |
| | Eosinophils (μL) | | 150.72(84.57–240.24)$^{a}$ | 169.57(75.89–275.16)$^{a}$ | 147.63(85.47–248.25)$^{a}$ | >0.05 |
| | Basophils (μL) | | 30.84(21.37–42.16)$^{a}$ | 27.63(18.64–47.66)$^{a}$ | 32.04(23.95–49.36)$^{a}$ | >0.05 |
| | Fibrinogen (g/l) | | 3.42(2.87–3.86)$^{b}$ | 3.71(3.08–4.58)$^{a}$ | 3.54(2.91–3.80)$^{a}$ | <0.05 |
| | CRP (mg/l) | | 2.3(1.25–5.6)$^{a}$ | 2.9(0.9–7.7)$^{a}$ | 2.1(1–5.15)$^{a}$ | >0.05 |
| | Alpha 1 antitrypsin (g/l) | | 1.49(1.35–1.57)$^{b}$ | 1.63(1.39–1.76)$^{a}$ | 1.45(1.35–1.65)$^{b}$ | <0.05 |
| | Calcium (mmol/L) | | 2.42(2.37–2.47)$^{a}$ | 2.41(2.35–2.48)$^{a}$ | 2.44(2.38–2.48)$^{a}$ | >0.05 |
| | 25(OH) Vitamine D (ng/ml) | | 19(12–32)$^{a}$ | 16.5(10–24.25)$^{a}$ | 24(13–32.5)$^{a}$ | >0.05 |
| | Phosphate (mmol/L) | | 0.89(0.79–0.95)$^{a}$ | 0.87(0.71–1.03)$^{a}$ | 1.01(0.93–1.12)$^{b}$ | <0.0001 |
| Atopy | IgE (KU/L) | | 52(23–233.5)$^{a}$ | 91.5(25–300.25)$^{a}$ | 73(19–170)$^{a}$ | >0.05 |
| | RAST DPT (d1) %>0.35 (KU/L) | | 13.33%(8)$^{a}$ | 14.28%(7)$^{a}$ | 6.12%(3)$^{a}$ | >0.05 |

(*continued*)

**Table 5.** Continued.

| | Variable | Result for clustering on raw dataset Cluster 1 | Cluster 2 | Cluster 3 | P-value |
|---|---|---|---|---|---|
| | RAST Cat (e1), %>0.35 (KU/L) | 1.64%(1)[a] | 8.16%(4)[a] | 2%(1)[a] | >0.05 |
| | RAST Dog (e5), %>0.35 (KU/L) | 3.28%(2)[a] | 6.12%(3)[a] | 0%(0)[a] | >0.05 |
| | RAST Grass (GX3), %>0.35 (KU/L) | 6.67%(4)[a] | 12.24%(6)[a] | 2.04%(1)[a] | >0.05 |
| | RAST microog (MIX1), %>0.35 (KU/L) | 6.56%(4)[a] | 12.24%(6)[a] | 10%(5)[a] | >0.05 |
| | RAST Birch (t3), %>0.35 (KU/L) | 0%(0)[a] | 4.08%(2)[a] | 0%(0)[a] | >0.05 |
| Sputum | Positive Aerobic Sputum Culture | 5.26%(3)[a] | 13.95%(6)[a] | 8.89%(4)[a] | >0.05 |
| | Weight of sputum (g) | 1.95(1.37–3.88)[a] | 1.46(1.06–3.07)[a] | 1.64(0.93–3.94)[a] | >0.05 |
| | Total Cell Counts ($10^6$/g) | 1.46(0.96–2.81)[b] | 4.15(1.50–15.96)[a] | 1.75(0.98–3.43)[b] | <0.05 |
| | Squamous (%) | 21(6.5–42)[b] | 7.5(2–38.25)[a] | 7(2–14.5)[a] | <0.05 |
| | Viability (%) | 67(50–81)[a] | 69.5(51–86.25)[a] | 61(41–77.5)[a] | >0.05 |
| | Macrophages (%) | 14(5.8–28.15)[b] | 6.25(2.2–15.45)[a] | 18(9.25–35.65)[b] | <0.0001 |
| | Lymphocytes (%) | 1(0–2.1)[a] | 0.6(0.15–1.57)[a] | 1(0–2.6)[a] | >0.05 |
| | Neutrophils (%) | 70.6(53.3–88.2)[a] | 77.5(55.4–92.25)[a] | 52(37.6–80)[b] | <0.001 |
| | Eosinophils (%) | 1.6(0.2–5.45)[a] | 2.3(0.35–10.95)[a] | 2.8(0.9–20.6)[a] | >0.05 |
| | Epithelial cells (%) | 1.8(0.4–7.1)[a] | 2.6(0.57–10.97)[a] | 2.4(1.2–8.35)[a] | >0.05 |
| | Macrophages ($10^3$/g) | 231.57(40.01–564.3)[a] | 317.39(32.25–801.83)[a] | 154(15.65–557.81)[a] | >0.05 |
| | Lymphocytes ($10^3$/g) | 12.84(0–40.65)[a] | 23.75(1.12–66.3)[a] | 14.08(0–46.22)[a] | >0.05 |
| | Neutrophils ($10^3$/g) | 1090(484.54–1788.33)[b] | 2333.6(814.1–12390.5)[a] | 814.05(486.85–2726.72)[b] | <0.05 |
| | Eosinophils ($10^3$/g) | 13.16(0–68.89)[b] | 44.27(0–511.5)[a] | 22(0–209.46)[b] | <0.001 |
| | Epithelial cells ($10^3$/g) | 44.1(5.66–142.29)[a] | 165.44(36.15–302.1)[a] | 67.52(13.7–252.84)[a] | >0.05 |

[+]There are no significant differences between two clusters with the same letter.

measured at higher level than those usually found in a general population of this age, maybe pointing to a role of IgE mediated processes in the pathophysiology of the disease. It has also been suggested that smoking may stimulate IgE production [29]. However, it is a production clearly directed toward something different from the classical aeroallergens encountered in asthma. The reason why atopic status is so low in COPD, clearly lower than in a general population, is likely to be related to the age of the COPD patients. Indeed, it was demonstrated in population studies that specific IgE levels decreased with age [30] and we have shown in a large asthmatic population that the rate of sensitization to aeroallergens was sharply declining with age after 60 years [31].

Of interest, and also perhaps surprising, is the fact the CAT score does not differ between the groups despite clear differences in lung function impairment and extent in airway inflammation. This shows that CAT score cannot capture the airflow limitation nor airway inflammation, indicating that symptoms, lung function and airway inflammation are different domains accounting for the disease variability [32].

Our study has obviously limitations as we have not taken into account comorbidities, exercise capacity and lung imaging in our parameters, which are key variables in phenotyping the COPD patients in clinical practice. Our purpose here was rather to explore variables traditionally linked to asthma such FeNO, IgE and eosinophilic inflammation and to see whether they can play a significant contribution in defining the variability of the disease in patients > 40 years with smoking history and persistent airway obstruction. Our data indicate that FeNO, blood eosinophils and serum IgE, though being significantly different from what is found in a healthy population for total serum IgE, are not able to single out a particular cluster. Therefore, most of the T2 biomarkers had not enough variability among the patients to shape a cluster. However, eosinophilic (together with neutrophilic) airway inflammation is raised in the cluster that shows the most severe lung function impairment. These data may have importance as it has been shown by retrospective post hoc analysis that ICS treatment in eosinophilic COPD might actually slow down the lung function decline [33]. The impact of targeting eosinophilic inflammation in COPD should be given careful consideration in long term clinical trials using not only ICS but also anti-interleukine-5.

Another important limitation of our study is the lack of a validation cohort whereas Castaldi *et al* have shown that reproducibility of COPD clustering across studies was rather modest. The size of our cohort was however too small to split our population and perform meaningful clustering with our extensive set of variables. However, we presented validation with two statistical indices.

In conclusion, in a cohort of COPD we found 3 clusters of patients with similar age and smoking history but very different sex distributions and lung function and inflammatory parameter. In particular, we identified a cluster of male patients with intense granulocytic airway inflammation combined with severe airway flow limitation and lung hyperdistention, who are prone to exacerbate and undergo recurrent hospitalizations. These clusters need to be confirmed in a new cohort of patients, ideally from other centers.

## Disclosure of interest

The author reports no conflicts of interest in this work.

## ORCID

Anne-Françoise Donneau http://orcid.org/0000-0001-8948-7545

## References

1. Rennard SI, Drummond MB. Early chronic obstructive pulmonary disease: definition, assessment, and prevention. Lancet Lond Engl. 2015;385(9979):1778–1788. DOI:10.1016/S0140-6736(15)60647-X

2. Fabbri LM, Luppi F, Beghe B, et al. Complex chronic comorbidities of COPD. Eur Respir J. 2008;31(1):204–212. DOI:10.1183/09031936.00114307
3. Deliu M, Sperrin M, Belgrave D, et al. Identification of asthma subtypes using clustering methodologies. Pulm Ther. 2016;2:19–41. DOI:10.1007/s41030-016-0017-z
4. Johnson RA, Wichern DW. Applied multivariate statistical analysis: international edition. 6th ed. Pearson; ISBN-13 : 978-0135143506 Publisher: Pearson (4 May 2007)
5. Haldar P, Pavord ID, Shaw DE, et al. Cluster analysis and clinical asthma phenotypes. Am J Respir Crit Care Med. 2008;178(3):218–224. DOI:10.1164/rccm.200711-1754OC
6. Moore WC, Meyers DA, Wenzel SE, National Heart, Lung, and Blood Institute's Severe Asthma Research Program, et al. Identification of asthma phenotypes using cluster analysis in the severe asthma research program. Am J Respir Crit Care Med. 2010;181(4):315–323. DOI:10.1164/rccm.200906-0896OC
7. Castaldi PJ, Benet M, Petersen H, et al. Do COPD subtypes really exist? COPD heterogeneity and clustering in 10 independent cohorts. Thorax. 2017;72(11):998–1006. DOI:10.1136/thoraxjnl-2016-209846
8. Vanfleteren LEGW, Spruit MA, Groenen M, et al. Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2013;187(7):728–735. DOI:10.1164/rccm.201209-1665OC
9. Burgel P-R, Roche N, Paillasseur J-L, INITIATIVES BPCO Scientific Committee, et al. Clinical COPD phenotypes identified by cluster analysis: validation with mortality. Eur Respir J. 2012;40(2):495–496. DOI:10.1183/09031936.00228511
10. Burgel P-R, Paillasseur J-L, Caillaud D, Initiatives BPCO Scientific Committee, et al. Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. Eur Respir J. 2010;36(3):531–539. DOI:10.1183/09031936.00175109
11. Garcia-Aymerich J, Gomez FP, Benet M, on behalf of the PAC-COPD Study Group, et al. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. Thorax. 2011;66(5):430–437. DOI:10.1136/thx.2010.154484
12. Fens N, van Rossum AGJ, Zanen P, et al. Subphenotypes of mild-to-moderate COPD by factor and cluster analysis of pulmonary function, CT imaging and breathomics in a population-based survey. COPD J Chronic Obstr Pulm Dis. 2013;10(3):277–285. DOI:10.3109/15412555.2012.744388
13. Basagaña X, Barrera-Gómez J, Benet M, et al. A framework for multiple imputation in cluster analysis. Am J Epidemiol. 2013;177(7):718–725. DOI:10.1093/aje/kws289
14. Pistolesi M, Camiciottoli G, Paoletti M, et al. Identification of a predominant COPD phenotype in clinical practice. Respir Med. 2008;102(3):367–376. DOI:10.1016/j.rmed.2007.10.019
15. Paoletti M, Camiciottoli G, Meoni E, et al. Explorative data analysis techniques and unsupervised clustering methods to support clinical assessment of Chronic Obstructive Pulmonary Disease (COPD) phenotypes. J Biomed Inform. 2009;42(6):1013–1021. DOI:10.1016/j.jbi.2009.05.008
16. Burgel P-R, Nesme-Meyer P, Chanez P, et al. Cough and sputum production are associated with frequent exacerbations and hospitalizations in COPD subjects. Chest. 2009;135(4):975–982. DOI:10.1378/chest.08-2062
17. Burgel P-R, Paillasseur J-L, Peene B, et al. Two distinct chronic obstructive pulmonary disease (COPD) phenotypes are associated with high risk of mortality. Sethi. S, ed. PLoS ONE. 2012;7(12):e51048. DOI:10.1371/journal.pone.0051048
18. Burgel P-R, Paillasseur J-L, Janssens W, et al. A simple algorithm for the identification of clinical COPD phenotypes. Eur Respir J. 2017;50(5):1701034. DOI:10.1183/13993003.01034-2017
19. Burgel P-R, Paillasseur J-L, Roche N. Identification of clinical phenotypes using cluster analyses in COPD patients with multiple comorbidities. BioMed Res Int. 2014;2014:420134–420139. DOI:10.1155/2014/420134
20. Cataldo D, Corhay J-L, Derom E, et al. A Belgian survey on the diagnosis of asthma-COPD overlap syndrome. COPD. 2017; Volume 12:601–613. DOI:10.2147/COPD.S124459
21. Molenberghs G, Kenward M. Missing data in clinical studies. John Wiley & Sons; ISBN:9780470849811 Publisher is: John Wiley & Sons: Chichester, 2007.
22. Buuren S. v. Flexible imputation of missing data. CRC Press; ISBN-13: 978-1439868249 Publisher: Chapman and Hall/CRC; 1st Edition (March 29, 2012).
23. Rubin DB. Multiple imputation for nonresponse in surveys. John Wiley & Sons; ISBN:0470317361, 9780470317365 Publisher: John Wiley & Sons, Inc. 2009.
24. Agusti A, Pavord ID. Do we really need a new classification of airway diseases? Lancet Respir Med. 2018;6(12):891–893. DOI:10.1016/S2213-2600(18)30386-2
25. Moermans C, Heinen V, Nguyen M, et al. Local and systemic cellular inflammation and cytokine release in chronic obstructive pulmonary disease. Cytokine. 2011;56(2):298–304. DOI:10.1016/j.cyto.2011.07.010
26. Celli BR, Locantore N, Yates J, for the ECLIPSE Investigators, et al. Inflammatory biomarkers improve clinical prediction of mortality in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2012;185(10):1065–1072. DOI:10.1164/rccm.201110-1792OC
27. Demarche S, Schleich F, Henket M, et al. Detailed analysis of sputum and systemic inflammation in asthma phenotypes: are paucigranulocytic asthmatics really non-inflammatory? BMC Pulm Med. 2016;16(1):46. DOI:10.1186/s12890-016-0208-2
28. Schleich FN, Chevremont A, Paulus V, et al. Importance of concomitant local and systemic eosinophilia in uncontrolled asthma. Eur Respir J. 2014;44(1):97–108. DOI:10.1183/09031936.00201813
29. O'Connor GT, Sparrow D, Weiss ST. The role of allergy and nonspecific airway hyperresponsiveness in the pathogenesis of chronic obstructive pulmonary disease. Am Rev Respir Dis. 1989;140(1):225–252. DOI:10.1164/ajrccm/140.1.225
30. Amaral AFS, Newson RB, Abramson MJ, et al. Changes in IgE sensitization and total IgE levels over 20 years of follow-up. J Allergy Clin Immunol. 2016;137(6):1788–1795.e9. DOI:10.1016/j.jaci.2015.09.037
31. Manise M, Bakayoko B, Schleich F, et al. IgE mediated sensitisation to aeroallergens in an asthmatic cohort: relationship with inflammatory phenotypes and disease severity. Int J Clin Pract. 2016;70(7):596–605. DOI:10.1111/ijcp.12837
32. Lapperre TS, Snoeck-Stroband JB, Gosman MME, Groningen and Leiden Universities Corticosteroids in Obstructive Lung Disease Study Group, et al. Dissociation of lung function and airway inflammation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2004;170(5):499–504. DOI:10.1164/rccm.200401-112OC
33. Pavord ID, Lettis S, Locantore N, et al. Blood eosinophils and inhaled corticosteroid/long-acting $\beta$-2 agonist efficacy in COPD. Thorax. 2016;71(2):118–125. DOI:10.1136/thoraxjnl-2015-207021