Identification of clinical phenotypes of patients with Chronic Obstructive Pulmonary Disease (COPD)

Ana Helena Tavares[1,2], PhD
Vera Enes[1], MSc
Vera Afreixo[2], PhD
Ana Machado[3], MSc
Alda Marques[3], PhD

[1] Institute of Biomedicine (iBiMED), University of Aveiro.
[2] Center for Research & Development in Mathematics and Applications (CIDMA), University of Aveiro.
[3] Respiratory Research and Rehabilitation Laboratory (Lab3R), School of Health Sciences (ESSUA), University of Aveiro.

Corresponding author: Ana Helena Tavares, PhD, ahtavares@ua.pt

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a condition characterized by progressive and persistent airflow limitation resulting from a chronic inflammatory response of the airways and lungs in response to inhaled harmful gases and particles. Clinical diagnosis is based on airflow obstruction (assessed with lung function test-spirometry) and symptoms. Its prognosis depends on several factors including acute exacerbations (define as worsening of symptoms that result in additional therapy), environmental exposures, comorbidities and genetic predisposition [1]. COPD is burdensome not only for health, economic and social systems but most importantly to patients since it significantly affect their quality of life. It is known that the disease does not progress in the same way in all patients and that lung function, symptoms and reduction of quality of life may not be correlated. In fact, the interplay between patients' clinical characteristics and different disease phenotypes is not fully understood.

This study aimed to enhance our knowledge on the clinical characteristics of patients with COPD. We retrospectively reviewed 394 patients with COPD referred from hospitals and primary care centers in the north and center regions of Portugal. The clinical data were acquired in the hospitals, primary care centers or in the Respiratory Research and Rehabilitation Laboratory (Lab3R), of the School of Health Sciences, University of Aveiro, under the project GENIAL.

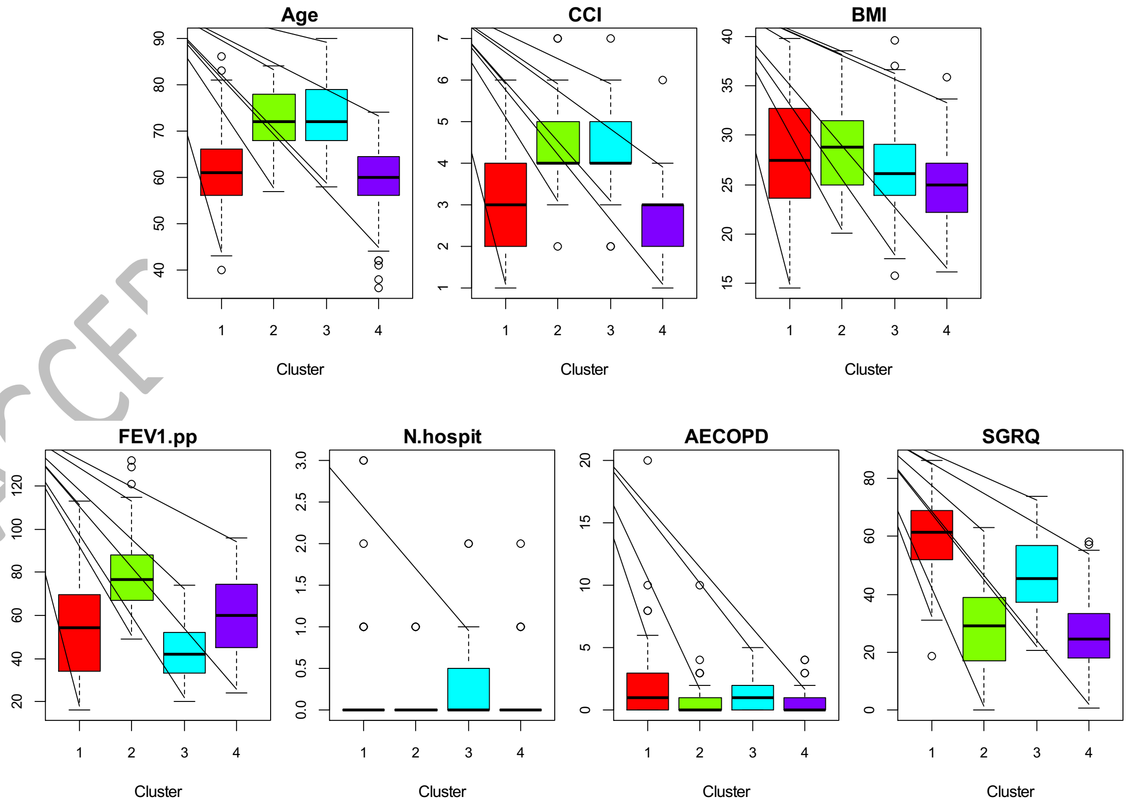
Methods

A clustering procedure is designed to stratify patients with COPD. From the 70 registered variables, we focus on the most commonly assessed clinical variables: age, the Charlson comorbidity index (CCI), body mass index (BMI), forced expiratory volume in one second (FEV1.pp), hospitalization by respiratory cause (N.hospit), number of acute exacerbations (AECOPD), St. George's Respiratory Questionnaire (SGRQ), the modified British Medical Research Council questionnaire (mMRC.score), Hospital Anxiety and Depression Scale (HADS.anx and HADS.depr), COPD assessment test total score (CAT.total), quadriceps muscle strength (QMS.pp), maximal inspiratory pressure (MIP.pp), maximal expiratory pressure

(MEP.pp), and 1-minute sit-to-stand test (1STS.pp). The suffix "pp" denotes percentage predicted, according to gender, age and weight. Clustering aims to find groups in a dataset. Since k-medoids looks for spherical clusters, it works best when the input variables are uncorrelated and have similar scales. We apply Principal Component (PC) Analysis on these vectors, obtaining a set of linearly uncorrelated variables. The number of components to retain is selected such that at least a given percentage of the variance is explained. Scores associated to those first PCs yield a data matrix, on which a partition clustering algorithm was applied. The result of the partition algorithm depends on the number of clusters *n*, which is often hard to choose a priori. Therefore it is common practice to run the method for several values of *n*, and then select the "best" value of *n* as the one which optimizes a certain criterion called a validity index. Many such indices have been proposed in the literature. Here we consider the GAP statistic [2].

Results

In our dataset several variables were strongly correlated, e.g., the Pearson correlation between CAT and SGRQ was 0.79. The PC allows obtaining a set of linearly uncorrelated variables. Our procedure retained 5 principal components that explained around 70% of the total variance of the dataset. Carrying out k-medoids clustering for different numbers of clusters and evaluating the obtained validation indices it appears that four clusters are appropriate (GAP = 0.697± 0.015, for 500 bootstrap samples). By looking at the composition of each cluster we define a patient profile of each cluster. It was observed that older patients present more comorbidities than younger patients and that marked differences exist in patients' symptoms, functionality and quality of life independently of patient's age and comorbidities (Figure 1).



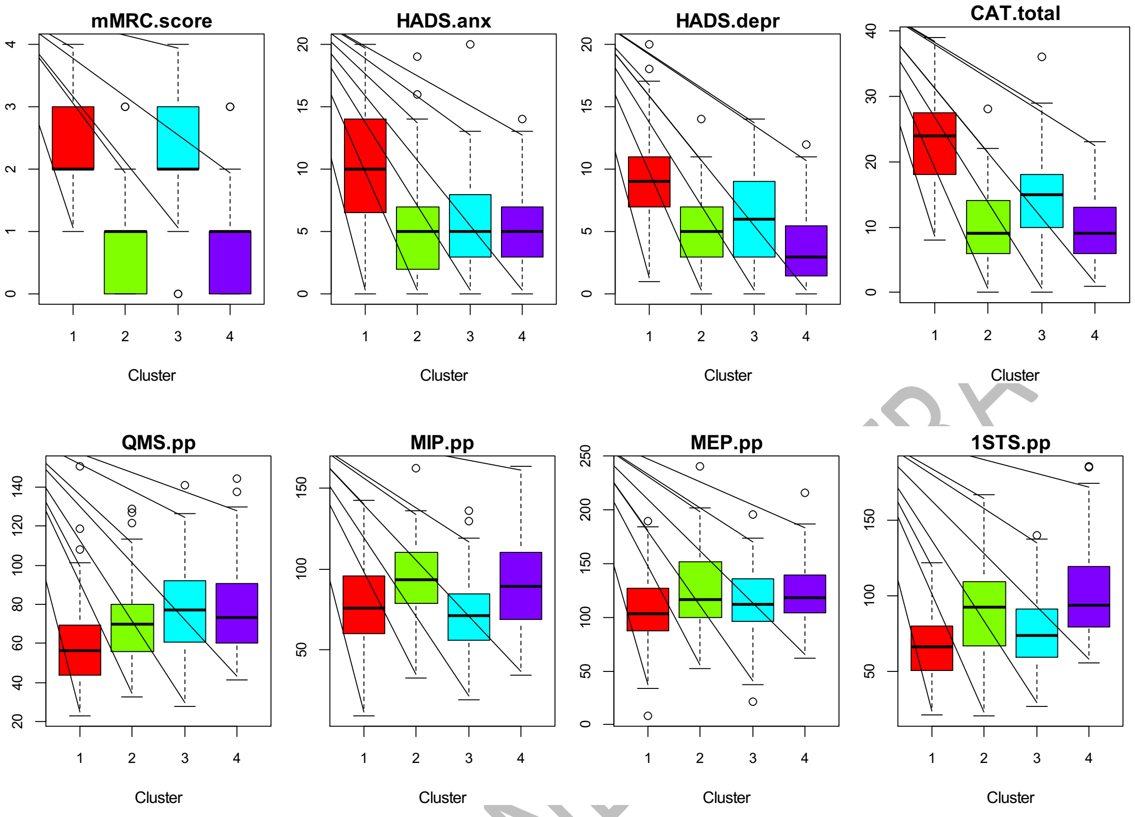


Figure 1. Clinical characteristics of patients allocated to each of the four clusters. The thick line within each box points the median of each cluster of patients in the following variables: age, the Charlson comorbidity index (CCI), body mass index (BMI), forced expiratory volume in one second (FEV1.pp), hospitalization by respiratory cause (N.hospit), number of acute exacerbations (AECOPD), St. George's Respiratory Questionnaire (SGRQ), the modified British Medical Research Council questionnaire (mMRC.score), Hospital Anxiety and Depression Scale (HADS.anx and HADS.depr), COPD assessment test total score (CAT.total), quadriceps muscle strength (QMS.pp), maximal inspiratory pressure (MIP.pp), maximal expiratory pressure (MEP.pp), and 1-minute sit-to-stand test (1STS.pp). The suffix “pp” denotes percentage predicted.

Discussion and conclusions

We found four clusters of COPD patients. There are two clusters of older patients and two of younger patients and within these clusters two opposite profiles of very symptomatic and not very symptomatic. It is now intended to validate the clinical profiles with an external sample of patients that has not been used to create the clusters.

Acknowledgements

This work was funded by Programa Operacional de Competitividade e Interna-cionalização – POCI, through Fundo Europeu de Desenvolvimento Regional - FEDER (POCI-01-0145-FEDER-007628 and POCI-01-0145-FEDER-028806), Fundação para a Ciência e Tecnologia (PTDC/DTP-PIC/2284/2014 and PTDC/SAU-SER/28806/2017). Moreover, the costs resulting from the FCT hirings is funded by national funds (OE), through FCT, I.P., in the scope of the framework contract foreseen in the numbers 4, 5 and 6 of the article 23, of the Decree-Law

57/2016, of August 29, changed by Law 57/2017, of July 19. The work of VA and AT is partially funded by FCT under project UID/MAT/04106/2019.

References

1. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (2019 report), 2019. Available from: https://goldcopd.org/

2. Robert T, Guenther W, Trevor H. Estimating the number of clusters in a data set via the gap statistic*. Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2001; 63(2): 411-423.