Journal of Statistics on Health Decision

EXTENDED ABSTRACT

Functional Status in People with COPD: A Cluster Analysis with Bootstrap Resampling

Guilherme Rodrigues[1,2], Joana Antão[1,2,3], Alda Marques[1,2]

[1]Lab3R – Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro (ESSUA), Aveiro, Portugal
[2]iBiMED – Institute of Biomedicine, University of Aveiro, Aveiro, Portugal
[3]Department of Research and Development, Ciro, Horn, the Netherlands

Introduction:

Functional status of people with chronic obstructive pulmonary disease (COPD) is heterogeneous, complex, and highly meaningful to their daily life. It is also a strong predictor of acute exacerbations of COPD (AEOPD), healthcare utilisation and mortality (1, 2). Nevertheless, it has been overlooked by treatment options. Understanding the heterogeneity of this important health domain, might contribute to better personalised care. This study sought to identify clusters based on functional status of people with COPD.

Methods:

A secondary analysis of socio-demographic, lung function, activity-related dyspnoea (modified Medical Research Council (mMRC)), impact of the disease (COPD Assessment Test (CAT)) and functional status data collected between 2017-2021 in GENIAL (PTDC/DTP-PIC/2284/2014), PRIME (PTDC/SAU-SER/28806/2017), 3R (SAICT-POL/23926/2016), and CENTR(AR) (POISE-03-4639-FSE-000597) was conducted.

People (aged ≥ 18 years old) diagnosed with COPD, according to GOLD criteria, and clinically stable in the previous month (i.e., no hospital admissions, AECOPD, or changes in medication) were included. Those with the presence of other respiratory diseases or any clinical condition which could have hindered test performance were excluded.

The six-minute walk test (6MWT), one-minute sit-to-stand test (1-min STS), quadriceps muscle voluntary contraction (QMVC) and handgrip muscle strength have been rescaled using the z-score formula. A principal component analysis was performed on scaled data; components extracted were used to group individuals with the K-means clustering algorithm. Principal components (PCs) were retained until cumulative percentage of total variance reached 70% minimum (3). Multivariate outliers were detected by comparing the Mahalanobis distance to a chi-square distribution (with a critical value set at 0.999) and removed from the analysis. The total within-cluster sum of squares was computed for different values of k; the optimum number of clusters was taken as the inflexion point on the curve.

Functional status differences between clusters were explored using one-way Multivariate Analysis of Variance (MANOVA), followed by one-way Analysis of Variance (ANOVA) with Bonferroni adjusted p-values and post-hoc multiple pairwise comparison tests. Multivariate skew and kurtosis were evaluated with Mardia's test for multivariate normality, and chi-square quantile-quantile plot. Homogeneity of variance-covariance matrices was assessed with Box's M test. Cluster stability was measured by bootstrap resampling methods (999 resampling runs) (4, 5). The Jaccard coefficient was used as a cluster-wise measure of cluster stability, allowing to quantify the quality of the clustering solution (4).

One-way ANOVA was used for continuous variables followed by pairwise comparisons with Bonferroni correction, and the Pearson's Chi-square or Fisher-exact test for categorical variables, to characterise clusters with regard to socio-demographic data, lung function, activity-related dyspnoea, and impact of the disease. Normality and homoscedasticity were assessed with Shapiro-Wilk and Bartlett's test, respectively.

All statistical analysis were performed using R Statistical Software (4.2.0), with a significance level set at 0.05.

Results:

In total, 132 people with COPD (107 (81.06%), 68.15 ± 7.82 yrs, FEV1 56.25 ± 19.27 %predicted) were included for analysis. Two PCs were retained (PC1 56% (2.22); PC2 21% (0.82)), and four clusters were identified (depicted in Fig. 1, 2). Differences between clusters for the combined dependent variables were

Keywords:
Cluster Analysis; COPD; Functional Status

Corresponding author:
Alda Marques
amarques@ua.pt

Conflict of interest:
The authors declare no conflict of interest.

First published: 20JUL2022

Table 1 - Socio-demographic and clinical characteristics of people with chronic obstructive pulmonary disease for the total sample and per cluster.

	Total (n=132)	Cluster 1 (n=45)	Cluster 2 (n=29)	Cluster 3 (n=27)	Cluster 4 (n=31)	p
Age, years	68.15 ± 7.82	71.0 ± 6.60b,d	65.45 ± 6.85a,c	70.89 ± 7.05b,d	64.16 ± 8.58a,c	<0.001
Sex [male]	107 (81.1)	33 (73.3)	27 (93.1)	17 (63.0)	30 (96.8)	0.001
FEV1 %pred	56.25 ± 19.27	54.56 ± 16.36	59.9 ± 18.28	52.0 ± 19.41	59.0 ± 23.44	0.346
mMRC [≥2]	75 (56.8)	26 (57.8)	15 (51.7)	22 (81.5)	12 (38.7)	0.011
CAT [≥10]	87 (65.9)	32 (71.1)	18 (62.1)	23 (85.2)	14 (45.2)	0.011
6MWT, m	408.8 ± 113.22	411.64 ± 62.29c,d	440.38 ± 67.54c,d	247.74 ± 92.41a,b,d	515.39 ± 50.64a,b,c	<0.001
1-min STS, reps	23.30 ± 7.17	23.49 ± 3.85c,d	22.28 ± 3.98c,d	14.11 ± 3.27a,b,d	31.97 ± 4.93a,b,c	<0.001
QMVC, kg/F	31.34 ± 9.94	26.73 ± 5.53b,d	40.64 ± 12.14a,c	23.43 ± 5.83b,d	36.2 ± 4.83a,c	<0.001
Handgrip strength, kg	34.14 ± 9.22	29.53 ± 6.65b,d	42.24 ± 7.27a,c	26.63 ± 7.39b,d	39.77 ± 5.52a,c	<0.001

Continuous variables are presented as mean ±standard deviation and categorical variables as counts and percentages. Legend: 6MWT, 6-minute walk test; 1-min STS, 1-minute sit-to-stand test; QMVC, Quadriceps maximum voluntary contraction. a – p<0.05 vs cluster 1; b – p<0.05 vs cluster 2; c – p<0.05 vs cluster 3; d – p<0.05 vs cluster 4

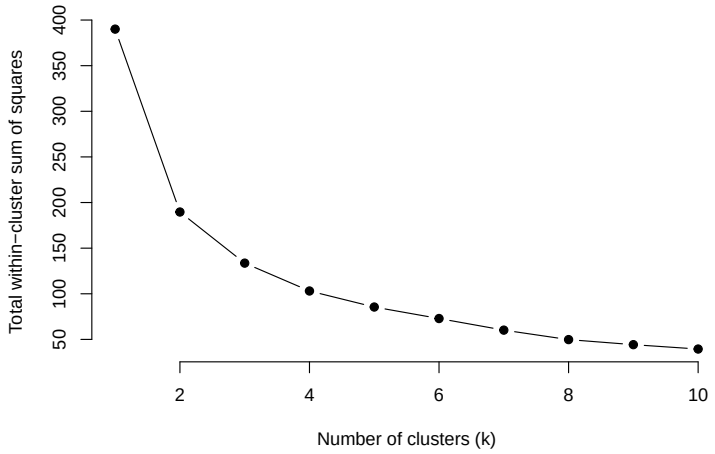

Figure 1 - Scree plot (elbow criterion) for determining the optimum number of clusters (k = 4).

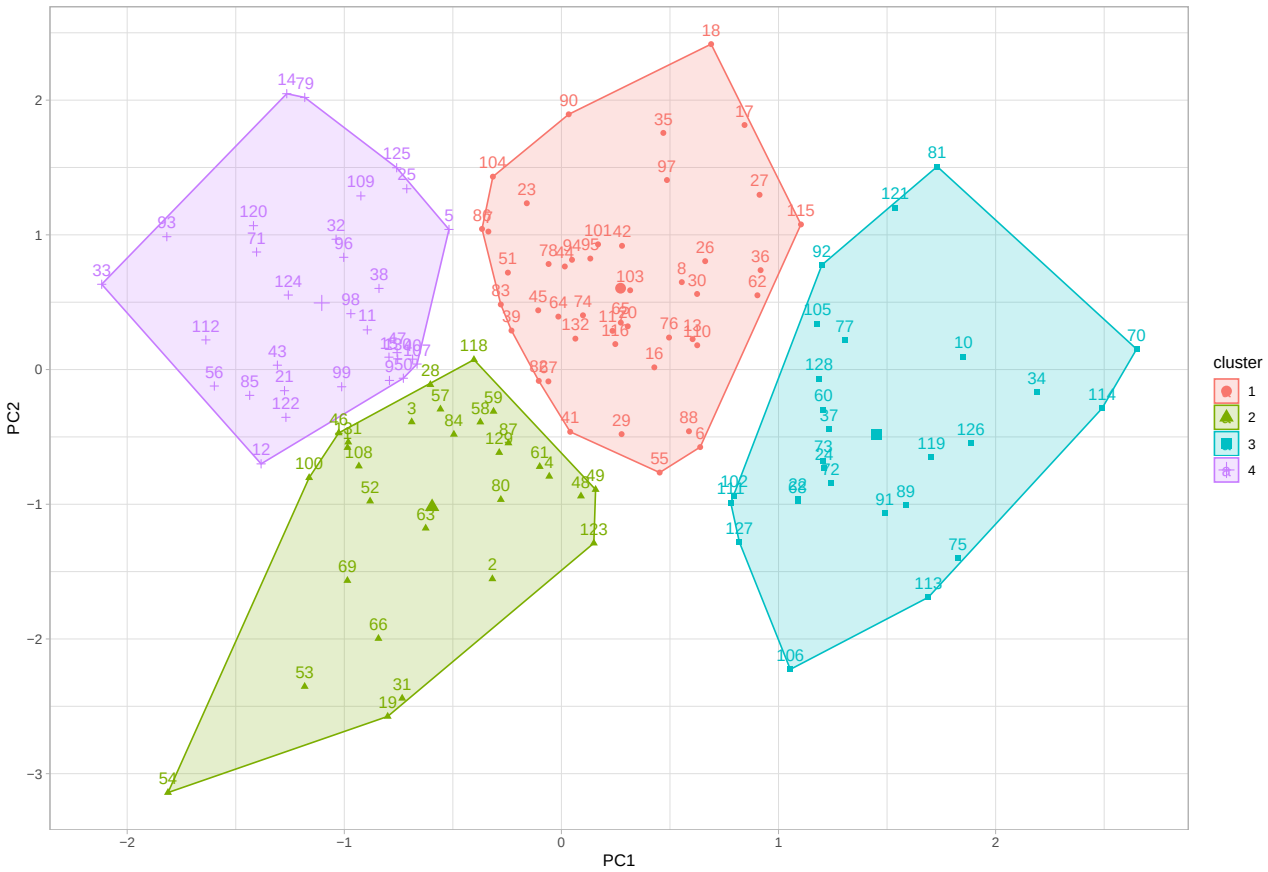

Figure 1 - Two-dimensional, graphical representation of the four clusters based on functional status of people with chronic obstructive pulmonary disease. X-axis or PC1 is the first rank component with the variance percentage of 56%, and Y-axis or PC2 is the second rank component with the variance percentage of 21%.

statistically significant, $F(12, 331.01) = 44.39$, $p < 0.001$; Wilks' $\Lambda = .079$; partial $\eta 2 = .45$. Follow-up univariate ANOVAs showed that all variables were significantly different between clusters, using a Bonferroni adjusted α level of 0.0125: 6MWT ($F(3, 128) = 77.27$, $p < 0.001$; partial $\eta 2 = 0.64$); 1-min STS ($F(3, 128) = 94.17$, $p < 0.001$; partial $\eta 2 = 0.69$); QMVC ($F(3, 128) = 35.62$, $p < 0.001$; partial $\eta 2 = 0.45$); handgrip strength ($F(3, 128) = 39.74$, $p < 0.001$; partial $\eta 2 = 0.48$). Comparisons between clusters are presented in Table 1. The Jaccard coefficients for nonparametric bootstrap were 0.78, 0.59, 0.88, and 0.72, for cluster 1, 2, 3, and 4, respectively.

Conclusion:

Four clusters were identified in people with COPD, which were significantly different in all variables of functional status, activity-related dyspnoea, and impact of the disease. Nevertheless, additional phenotypic characterisation based on treatable clinical traits is needed, which may guide tailored treatment regimens to improve this meaningful outcome. Cluster validity, their behaviour over time and differential response to treatment needs further investigation.

Ethics committee and informed consent:

All studies were approved by several Ethics Committees (CHMA 09/2016-10/2018; ULS Matosinhos 10/CES/JAS 17/02/2017-73/CE/JAS 12/10/2018; CHBV 777638-086892-15/05/2019; HDFF 1807/2017-27/05/2019; ARS Centro 64/2016-73/2016-85/2018-16/2020; UICISA-E P620-10/2019) and subjects gave their informed consent before they were enrolled in the studies.

References:

1. Fan VS, Ramsey SD, Make BJ, Martinez FJ. Physiologic variables and functional status independently predict COPD hospitalizations and emergency department visits in patients with severe COPD. Copd. 2007;4(1):29-39. Epub 2007/03/17. https://doi.org/10.1080/15412550601169430. PubMed PMID: 17364675.
2. Vaes AW, Spruit MA, Koolen EH, Antons JC, de Man M, Djamin RS, et al. "Can Do, Do Do" Quadrants and 6-Year All-Cause Mortality in Patients With COPD. Chest. 2022. Epub 2022/01/14. https://doi.org/10.1016/j.chest.2021.12.657. PubMed PMID: 35026297.
3. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016;374(2065):20150202. https://doi.org/10.1098/rsta.2015.0202.
4. Hennig C. Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis. 2007;52(1):258-71. https://doi.org/10.1016/j.csda.2006.11.025.
5. Hennig C. fpc: Flexible Procedures for Clustering. R package version 2.2-9. 2020.