ORIGINAL ARTICLE

Clusters of individuals recovering from an exacerbation of chronic obstructive pulmonary disease and response to in-hospital pulmonary rehabilitation

M. Vitacca[a,*], A. Malovini[b], A. Spanevello[c,d], P. Ceriana[e], M. Paneroni[a], M. Maniscalco[f], B. Balbi[g], L. Rizzello[h], R. Murgia[i], R. Bellazzi[l,b], N. Ambrosino[i]

[a] Respiratory Rehabilitation of the Institute of Lumezzane, Istituti Clinici Scientifici Maugeri IRCCS, Brescia, Italy
[b] Laboratory of Informatics and Systems Engineering for Clinical Research of the Institute of Pavia, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy
[c] Respiratory Rehabilitation of the Institute of Tradate, Istituti Clinici Scientifici Maugeri IRCCS, Varese, Italy
[d] Department of Medicine and Surgery, University of Insubria, Varese, Italy
[e] Respiratory Rehabilitation of the Institute of Pavia, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy
[f] Respiratory Rehabilitation of the Institute of Telese, Istituti Clinici Scientifici Maugeri IRCCS, Benevento, Italy
[g] Respiratory Rehabilitation of the Institute of Veruno, Istituti Clinici Scientifici Maugeri IRCCS, Novara, Italy
[h] Respiratory Rehabilitation of the Institute of Milano Istituti Clinici Scientifici Maugeri IRCCS, Milano, Italy
[i] Respiratory Rehabilitation of the Institute of Montescano, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy
[l] Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

Abstract

Introduction and Objectives: Due to the present low availability of pulmonary rehabilitation (PR) for individuals recovering from a COPD exacerbation (ECOPD), we need admission priority criteria. We tested the hypothesis that these individuals might be clustered according to baseline characteristics to identify subpopulations with different responses to PR.

Methods: Multicentric retrospective analysis of individuals undergone in-hospital PR. Baseline characteristics and outcome measures (six-minute walking test - 6MWT, Medical Research Council scale for dyspnoea -MRC, COPD assessment test —CAT) were used for clustering analysis.

Results: Data analysis of 1159 individuals showed that after program, the proportion of individuals reaching the minimal clinically important difference (MCID) was 85.0%, 86.3%, and 65.6% for CAT, MRC, and 6MWT respectively. Three clusters were found (C1-severe: 10.9%; C2-intermediate: 74.4%; C3-mild: 14.7% of cases respectively). Cluster C1-severe showed the worst conditions with the largest post PR improvements in outcome measures; C3-mild showed the least severe baseline conditions, but the smallest improvements. The proportion of participants reaching the MCID in ALL three outcome measures was significantly different among clusters, with C1-severe

* Corresponding author at: Respiratory Rehabilitation Unit, Istituti Clinici Scientifici Maugeri IRCCS. Institute of Lumezzane, Via G. Mazzini 129, 25065, Lumezzane (Brescia), Italy.
E-mail address: michele.vitacca@icsmaugeri.it (M. Vitacca).

having the highest proportion of full success (69.0%) as compared to C2-intermediate (48.3%) and C3-mild (37.4%). Participants in C2-intermediate and C1-severe had 1.7- and 4.6-fold increases in the probability to reach the MCID in all three outcomes as compared to those in C3-mild (OR = 1.72, 95% confidence interval [95% CI] = 1.2 − 2.49, p = 0.0035 and OR = 4.57, 95% CI = 2.68 − 7.91, p < 0.0001 respectively).

Conclusions: Clustering analysis can identify subpopulations of individuals recovering from ECOPD associated with different responses to PR. Our results may help in defining priority criteria based on the probability of success of PR.

Introduction

Pulmonary rehabilitation (PR) including exercise training is a recognized cornerstone of comprehensive management of chronic obstructive pulmonary disease (COPD). In different settings, this modality improves symptoms such as dyspnoea and fatigue, exercise capacity, and health-related quality of life (HRQL).[1,2] Therefore, guidelines recommend PR for individuals with persistent breathlessness and/or exercise limitation with reduced HRQL.[3] Furthermore early PR after an exacerbation (ECOPD) is associated with reduced prevalence of new ECOPD and longer survival and is cost-effective.[4,5] However, there are barriers to in-hospital programs, such as a high number of candidates, transportation, costs, and geographical obstacles.[6] To increase access to programs, governments should develop policies to increase resources, logistics, availability and expertise of healthcare providers, including tele-rehabilitation.[6,7] An additional approach might be to identify characteristics of individuals most likely to receive benefits from PR (responders) in order not to "waste" resources for candidates with scarce probabilities of success.

Clustering analyses are unsupervised multivariate methods helping to identify subpopulations or "clusters". By labelling observations according to the corresponding cluster, these techniques can help interpret the available data.[8] The aim of this multicentre, retrospective study was to test the hypothesis that individuals recovering from an ECOPD could be clustered according to their baseline characteristics, for early identification of subpopulations with different patterns of response to PR in accepted outcome measures.

Methods

This multicentre, retrospective study evaluated a database of hospital medical records of individuals recovering from ECOPD. The study was approved by the Istituti Clinici Scientifici (ICS) Maugeri Ethics Committee (2555 CE 8 June 2021). As a retrospective study, participants had not provided any specific written informed consent, however, at admission to ICS Maugeri hospitals, they had given − in- advance informed consent for the scientific use of their data. As a retrospective analysis, the study was not registered.

Participants

The study has been conducted on individuals consecutively admitted between July, 1th, 2018 to December, 31th, 2021 to hospitals of the ICS Maugeri network (Lumezzane, Pavia, Tradate, Veruno, Milano, Montescano, Telese, Italy), referral hospitals for PR, diagnosis, and care of chronic diseases. These hospitals share common indications to PR,[1] evaluation, diagnostic and management tools, and protocols. During the pandemic period of study (March 2020-December 2021), only participants with negative swab tests were admitted to the in-patient program.

Only data of individuals with lung function and paired prior and post PR results of outcome measures (exercise tolerance, disease impact, dyspnoea) were analyzed. Part of the data have been published or are in press elsewhere.

The inclusion criteria were:

1. Diagnosis of COPD and post-bronchodilator Forced Expiratory Volume at one second (FEV_1)/Forced Vital Capacity (FVC) ratio < 70%.[3]
2. Persistent breathlessness and/or exercise limitation after (within the previous 30 days) an ECOPD needing acute care hospital admission or after (within the previous 4 weeks) an ECOPD managed at the out-patient clinic.
3. Stable conditions as assessed by the absence of acute worsening in symptoms, i.e. no change in dyspnoea, cough, and/or sputum beyond the day-to-day variability, which would have required a change in management, as compared to the conditions reported at home or at discharge from the referring acute care hospital.
4. Availability of data on lung function and pre and post-assessment of outcome measures.

Exclusion criteria were: severe comorbidities: oncological, neurological disorders, heart failure or recent (less than 4 months) acute ischemic cardiovascular diseases with an instability status; inability or lack of willingness to perform or complete at least 12 sessions of the PR program.

Measurements

The following data had been recorded: demographics, anthropometrics (Body-mass index, BMI), history of ECOPD in the previous 12 months,[9] Comorbidity Index of the Cumulative Illness Rating Scale (CIRS),[10] BMI- airflow obstruction-dyspnoea, and exercise capacity (BODE) index,[11] provenience (hospital or home), length of ICS Maugeri hospitals stay (LoS), occurrence of chronic respiratory failure (CRF), distribution in Global Initiative for Obstructive Lung Disease (GOLD) stages, drug therapy.
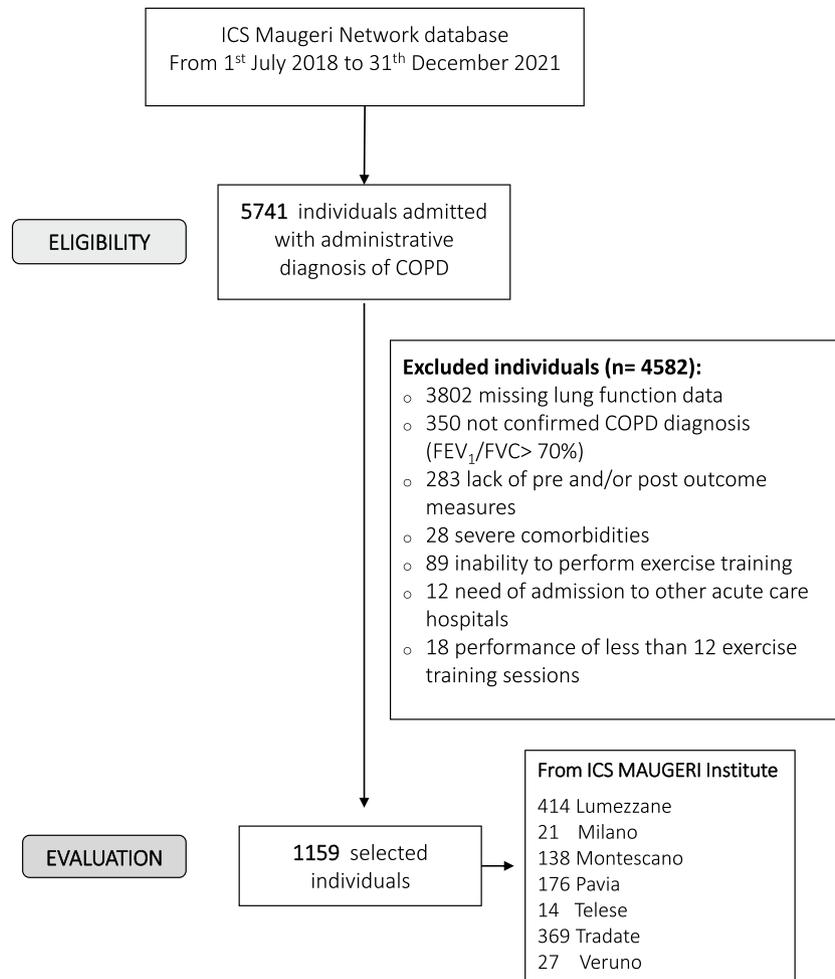
```
┌─────────────────────────────────────┐
│      ICS Maugeri Network database     │
│ From 1st July 2018 to 31th December 2021│
└─────────────────────────────────────┘
```

┌───────────────┐   ┌─────────────────────┐
│ ELIGIBILITY │   │ **5741** individuals admitted │
└───────────────┘   │   with administrative │
                           │   diagnosis of COPD   │
                           └─────────────────────┘

Excluded individuals (n= 4582):
- 3802 missing lung function data
- 350 not confirmed COPD diagnosis (FEV_1/FVC> 70%)
- 283 lack of pre and/or post outcome measures
- 28 severe comorbidities
- 89 inability to perform exercise training
- 12 need of admission to other acute care hospitals
- 18 performance of less than 12 exercise training sessions

From ICS MAUGERI Institute

414 Lumezzane
21   Milano
138 Montescano
176 Pavia
14   Telese
369 Tradate
27   Veruno

┌──────────────┐   ┌─────────────────────┐
│ EVALUATION │   │ **1159** selected   │
└──────────────┘   │   individuals   │
                          └─────────────────────┘

Fig. 1   Flow chart of the study.

Before the PR program the following assessments had been performed:

- Forced expiratory volumes (FEV_1, FVC and FEV_1/FVC, %) according to standards,[12] using the predicted values of Quanjer.[13]
- Dyspnoea by the Italian version of Barthel index Dyspnea (BiD).[14]
- Functional disability by the Barthel index.[15]

Before and after the program, the following outcome measures had been assessed:

- Exercise tolerance by the six-minute walking distance test (6MWT).[16] Data are shown as meters and percent of predicted values.[17] The minimal clinically important difference (MCID) in individuals with COPD has been reported as an improvement in the walked distance by at least 30 meters.[16]
- Dyspnoea by the Medical Research Council (MRC) scale.[18] A one-point reduction in score is considered equivalent to MCID.[19]
- Disease impact by the COPD assessment test (CAT).[20] A two-point reduction in score has been reported as the MCID.[21]

Pulmonary rehabilitation

The ICS Maugeri network hospitals, share a PR program supervised by a multidisciplinary team consisting of chest physicians, nurses, physical therapists, dieticians, and psychologists. The in-hospital multidisciplinary program includes optimization of drug therapy, education, nutritional programs, and psychosocial counseling when appropriate, abdominal, upper, and lower limb muscle activities lifting weights progressively. It includes also supervised cycle exercise training[22] according to Maltais et al[23] until performing 30 min continuous cycling at 50-70% of the maximal load calculated on the basis of the baseline 6MWT according to Luxton et al.[24] Pulse oximetry, arterial blood pressure, and heart rate are monitored during exercise. The total duration of daily activities is 2-3 hours.

During the pandemic period protective measures had been adopted, such as the use of personal protective equipment, increasing distance among individuals (not less than 2 meters during sessions), constant disinfection of tools such as bikes and instruments, frequent air changes, immediate execution of a swab at first harmful signs, and other commonly adopted measures.[25,26]

Statistical analysis

Statistical analyses have been performed by the R statistical software tool version 4.0.5 (www.r-project.org). Post PR changes were dichotomized based on MCID of outcome measures. Numeric variable distribution was described as median (25th, 75th percentiles) since most of them deviated from the normality assumptions based on visual inspection of histograms. Categorical nominal and ordinal variable distributions were described as absolute and relative (%) frequencies. No analysed variable suffered from missing values. The sign test was used to test the null hypothesis of no change (median change = 0) in numeric variable distribution between before and after PR. The Kruskal Wallis test was applied to compare numeric variables distributions among clusters while the Pearson chi-square test with 10,000 simulations was applied to test the null hypothesis of independence between categorical variables and clusters. Multivariate logistic regression was applied to test for association between clusters and the condition of reaching the MCID in all outcome measures including centre as covariate. The significance level was set at α = 0.05, the False Discovery Rate (FDR) correction was applied when appropriate, considering FDR values < 0.10 as statistically significant.

Clustering of patients and machine learning analyses are described in the **Appendix A** section in **Supplementary Methods and Results**.

Results

Fig. 1 shows the flow chart of the study. Data from 1159 individuals were analyzed. The baseline characteristics of participants are shown in Table 1. More than half of the participants were males and included in the most severe GOLD stages.[3]

After the program, all assessed outcome measures improved significantly (Table 2). **Appendix A- Fig. A.1** shows the frequency distribution of outcome measures. The proportion of individuals reaching the MCID was 85.0%, 86.3%, and 65.6% for CAT, MRC, and 6MWT respectively.

The ability of different clustering strategies to identify distinct subpopulations of participants based on baseline characteristics was assessed (**Appendix A- Table A.1** and **Table A.2**). The selected approach allowed to identify three clusters (C) according to the baseline severity: C1-severe (n = 126, 10.9%), C2-intermediate (n = 862, 74.4%), and C3-mild (n = 171, 14.7%).

Table 3 shows the baseline characteristics and outcome measures according to clusters. Individuals in C1-severe were the oldest and the least likely to be treated at home for their ECOPD. These individuals shared also the most severe conditions such as inclusion in GOLD stage D in more than 91% of cases, airway obstruction, high prevalence of CRF, triple inhaler drug use, motor Barthel, BiD, and comorbidities. This subpopulation was characterized also by the worst baseline outcome measures. Individuals in C3-mild showed the least severe conditions, whereas those in C2-intermediate, the most prevalent, may represent an intermediate condition. As also shown in Table 3 among clusters

Table 1 Characteristics of participants.

Variable	Frequency or Median	Min:Max
Age, years	72 (65, 77)	34:93
Gender		
Females	392 (33.8%)	
Males	767 (66.2%)	
BMI, Kg/m^2	26.2 (22.8, 31)	11.7:64.5
BODE index, score	5 (4, 7)	0:10
LoS, days	25 (21, 32)	10:120
Provenience		
Home	851 (73.4%)	
Hospital	308 (26.6%)	
CRF		
No	770 (66.4%)	
Yes	389 (33.6%)	
Inhaler drugs		
Triple	690 (59.5%)	
No triple	469 (40.5%)	
CIRS, score	4 (2, 5)	0:12
Barthel, score	100 (90, 100)	0:100
BiD, score	25 (14, 39)	0:90
FEV$_1$, % prd	44 (34, 56)	12:86
FVC, % prd	70 (58, 81)	27:112
FEV$_1$/FVC, %	46 (42, 56)	21:69
GOLD airflow stages		
1	29 (2.5%)	
2	337 (29.1%)	
3	437 (37.7%)	
4	356 (30.7%)	
GOLD quadrant stages		
A	106 (9.2%)	
B	260 (22.4%)	
C	115 (9.9%)	
D	678 (58.5%)	
CAT, score	18 (12, 24)	0:37
MRC, score	3 (3, 3)	0:4
6MWT, metres	300 (200, 400)	0:635
6MWT % prd	64.1 (43.1, 83)	0:174.3

Legend: Variables distribution is described as absolute and relative frequency (%) or median (25th, 75th percentiles).

Min: max = minimum and maximum values of each numeric variable's distribution. LoS: Length of stay; CRF: Chronic respiratory failure; BMI: Body Mass Index; BiD: Barthel index Dyspnea; CAT: COPD assessment test; MRC: Medical research council; 6MWT: Six-minute walking distance test; FEV$_1$: Forced expiratory volume at one second; FVC: Forced vital capacity; prd: predicted; GOLD: Global Initiative for Obstructive Lung Disease; BODE: body-mass index, airflow obstruction, dyspnoea, and exercise capacity index; CIRS: Comorbidity Index of Cumulative Illness Rating Scale.

there was a statistically significant difference in the proportion of individuals reaching the MCID of 6MWT (p < 0.0001), MRC (p = 0.0002) but not CAT (p = 0.8635).

The proportion of individuals reaching MCID in all three outcome measures was significantly different among clusters (p < 0.0001), C3-mild having a lower probability of full success (37.4% of participants) than C2-intermediate (48.3%) and C1-severe (69.0%). Multivariate logistic regression with

Table 2 Post to Pre-program changes in outcome measures.

Variable	Before	After	Change	p-value	
CAT, score	18 (12, 24)	11 (8, 16)	−5 (−9, −3)	< 0.0001	*
MRC, score	3 (3, 3)	2 (2, 2)	−1 (−2, −1)	< 0.0001	*
6MWT, metres	300 (200, 400)	351 (260, 440)	45 (18, 80)	< 0.0001	*

Legend: Variable = analyzed variable; Data as Median (25th, 75th percentiles).
CAT: COPD assessment test; MRC: Medical research council; 6MWT: Six-minute walking distance test.
 * p-value < 0.05.

adjustment by centre showed that participants in C2-intermediate and C1-severe had 1.7- and 4.6-fold increase in the probability to reach the MCID in all three outcomes as compared to those in C3-mild (OR = 1.72, 95% confidence interval [95% CI] = 1.2 − 2.49, p = 0.0035 and OR = 4.57, 95% CI = 2.68 − 7.91, p < 0.0001 respectively).

Multivariate analyses showed that baseline 6MWT, BiD, inhaler drugs, GOLD stage as well as baseline CAT and MRC represented the subset of variables used for clusters definition with a major influence in discriminating among the three subpopulations (**Appendix A - Fig. A.2**).

The distribution of each outcome measure value was significantly different among clusters before, after PR, and in changes (FDR < 0.10) (Fig. 2).

Table 4 shows the distribution of participants and clusters by centres. C2-intermediate was the most prevalent in all centres, whereas almost 75% of C3-mild was observed in a single centre. **Appendix - Table A.3** reports the outcome variables distribution by cluster according to the participating centres.

Additional and preliminary analyses have been performed to identify decisional rules to classify participants into clusters using baseline variables (**Appendix A− Supplementary Methods and Results**). "Conditional inference trees" was selected as the most informative machine learning algorithm among those tested, reaching a mean classification accuracy (CA) from 10-fold cross validation of 82.49% (majority classifier CA = 74.40%) while mean sensitivity in discriminating between clusters ranging 41.5% - 92.37% (**Appendix A − Table A.4, Table A.5**). When learned on the whole dataset, the conditional inference trees model generated decisional rules to provide a rough distinction of participants into clusters (Fig. 3). According to these rules, an individual could be labelled as belonging to C1-severe if baseline 6MWT ≤ 159 meters AND BiD > 28 points; C2-indermediate if baseline 6MWT ≤ 159 m AND BiD ≤ 28 points or if baseline 6MWT > 159 meters AND GOLD quadrant stages = B, C or D; C3-mild if baseline 6MWT > 159 meters AND GOLD quadrant stage = A (Fig. 3 and **Appendix A − Table A.6**).

Discussion

By clustering analysis, our study distinguished different groups of individuals undergoing in-hospital PR after an ECOPD. These individuals could be characterized by three clusters with different prevalence, baseline characteristics, and responses to PR. Cluster C1-severe had the most severe baseline conditions with the largest improvement, C3-mild showed the best baseline conditions but the smallest change size, whereas C2-intermediate showed conditions and effect size intermediate between C1-severe and C3-mild.

While confirming evidence that individuals with a worse baseline status are good responders to PR,[27] our study also suggests a modality to analyze the characteristics of these individuals in order to define priority criteria for admission to PR of these individuals.

Multivariate analyses identified baseline 6MWT, BiD, inhaler drugs, GOLD stage as well as baseline CAT and MRC as the subset of variables used for clusters definition having a major influence in discriminating among the three subpopulations.

As an additional finding, we were also able to roughly assign participants to the corresponding cluster by conditional inference trees algorithm. Future studies will allow tuning and validating decisional rules on independent data, as well as evaluating the feasibility to implement more accurate models (10-fold cross validation: mean classification accuracy reached by Elastic net logistic regression = 94.41% [data not shown] vs. mean classification accuracy reached by conditional inference trees = 82.49%) into an interactive tool to be used in clinical practice for a more accurate classification.

This is a retrospective study investigating individuals with COPD as shown by lung function,[3] therefore, we excluded all cases without any available lung function data, and those with lung function not confirming COPD (e.g. $FEV_1/FVC > 70\%$). We evaluated individuals admitted to in-hospital PR after an ECOPD. Despite a relevant improvement in COPD treatment, the natural course of ECOPD is unchanged highlighting the importance of prevention. It has been shown that early PR after an ECOPD is cost-effective and results in reduced prevalence of new exacerbations and longer survival.[4,5,28]

The outcome measures assessed in this study (dyspnoea, exercise capacity, disease impact) are widely accepted not only for PR, but are also suggested in an outcome set for clinical trials evaluating the management of ECOPD.[1,29] The present study confirms the benefits of PR including the proportion of responders.[1,2,30] Al Chikhanie et al[31] identified four clusters according to the response of 6MWT to PR. The cluster with the largest proportion of non-responders included older, more severe individuals.[31] However, exercise tolerance is only one of the benefits of PR. Also patient-centered outcomes such as symptoms and disease impact matter. Spruit et al[32] proposed a multidimensional response outcome. In our study, the proportion of individuals reaching the MCID in all three assessed outcome measures was significantly different among clusters.

The high proportion (66.2%) of individuals excluded from this retrospective study due to missing lung function data is not surprising. According to guidelines, lung function might have been not considered as an admission criterion to or an

Table 3 Characteristics and outcome measures according to clusters.

Variable	C1-severe (n = 126, 10.9%)	C2-intermediate (n = 862, 74.4%)	C3-mild (n = 171, 14.7%)	p-value	
Age, years [#]	74 (70, 79)	71 (65, 77)	71 (65, 76)	<0.0001	*
Gender [#]				0.0075	*
Females	57 (45.2%)	286 (33.2%)	49 (28.7%)		
Males	69 (54.8%)	576 (66.8%)	122 (71.3%)		
BMI, Kg/m^2 [#]	24.9 (21.5, 28.4)	26 (22.5, 30.7)	29 (25.3, 33.6)	<0.0001	*
Provenience [#]				<0.0001	*
Home	57 (45.2%)	630 (73.1%)	164 (95.9%)		
Hospital	69 (54.8%)	232 (26.9%)	7 (4.1%)		
CRF [#]				<0.0001	*
No	33 (26.2%)	567 (65.8%)	170 (99.4%)		
Yes	93 (73.8%)	295 (34.2%)	1 (0.6%)		
Inhaler drugs [#]				<0.0001	*
Triple	108 (85.7%)	572 (66.4%)	10 (5.8%)		
No triple	18 (14.3%)	290 (33.6%)	161 (94.2%)		
CIRS, score [#]	5 (5, 6)	4 (2, 5)	3 (2, 3)	<0.0001	*
Barthel, score [#]	80 (55, 94)	100 (90, 100)	100 (90, 100)	<0.0001	*
BiD, score [#]	45 (38, 61)	25 (17, 38)	11 (7, 16)	<0.0001	*
FEV$_1$/FVC, % [#]	45 (40, 48)	46 (40, 56)	55 (48, 60)	<0.0001	*
SpO$_2$, % [#]	95 (93, 96)	94 (92, 96)	95 (94, 96)	<0.0001	*
GOLD stages [#]				<0.0001	*
A	0 (0%)	34 (4%)	72 (42.1%)		
B	6 (4.7%)	189 (21.9%)	65 (38 %)		
C	5 (4%)	89 (10.3%)	21 (12.3%)		
D	115 (91.3%)	550 (63.8%)	13 (7.6%)		
CAT, score					
Before [#]	25 (23, 28)	18 (12, 23)	11 (8, 15)	<0.0001	*
After	16 (12, 23)	11 (8, 16)	6 (3, 9)	<0.0001	*
Change	−8 (−11, −4)	−5 (−10, −3)	−5 (−8, −2)	0.0002	*
Freq. Reach. MCID	109 (86.5%)	730 (84.7%)	146 (85.4%)	0.8635	
MRC, score					
Before [#]	4 (4, 4)	3 (3, 3)	3 (2, 3)	<0.0001	*
After	2 (2, 3)	2 (2, 2)	2 (1, 2)	<0.0001	*
Change	−2 (−2, −1)	−1 (−1, −1)	−1 (−1, −1)	<0.0001	*
Freq. Reach. MCID	118 (93.6%)	750 (87.0%)	132 (77.2%)	0.0002	*
6MWT, metres					
Before [#]	100 (45, 144)	299.5 (215.25, 380)	440 (389, 486.5)	<0.0001	*
After	180 (135, 234)	348 (270, 423)	476 (420, 525)	<0.0001	*
Change	87 (42, 141)	45 (16, 79)	30 (7, 60)	<0.0001	*
Freq. Reach. MCID	106 (84.1%)	563 (65.3%)	91 (53.2%)	<0.0001	*
6MWT% prd					
Before	20.8 (9.8, 30.2)	63.2 (47.0, 78.3)	94.5 (82.3, 102.8)	<0.0001	*
After	42.3 (30.3, 50.8)	74.2 (58.8, 86.9)	100.1 (90.3, 111.5)	<0.0001	*
Change	18.5 (10.2, 30.3)	9.3 (3.4, 16.7)	6.7 (1.3, 13.6)	<0.0001	*

Legend: Variables' distribution by clusters are described as absolute frequency (relative frequency, %) or median (25th, 75th percentiles).
  * p-value < 0.05.
  [#] Variable used for clustering.
C. Cluster; CRF: Chronic respiratory failure; SpO$_2$: pulsed oxygen saturation; BMI: Body Mass Index; BiD: Barthel index Dyspnea; CAT: COPD assessment test; MRC: Medical research council; 6MWT: Six-minute walking distance test; FEV$_1$: Forced expiratory volume at one second; FVC: Forced vital capacity; prd: predicted; GOLD: Global Initiative for Obstructive Lung Disease; Comorbidity Index of CIRS: Cumulative Illness Rating Scale; Freq. Reach. MCID: Frequency of patients reaching the Minimal Clinically Important Difference.

outcome measure for PR.[1] Of course a possible lack of inclusion in the database of results of performed lung function cannot be excluded. However, our study shows that the level of airway obstruction (**Appendix - Fig. A.2**) had a major influence in discriminating the clusters. Therefore, our study

indicates that lung function should be incorporated into the core set of evaluations for admission to PR.

An original and not negligible result of our study is also that BiD[14] can be a reliable outcome measure of PR. Indeed, our study shows that the baseline level of dyspnoea as
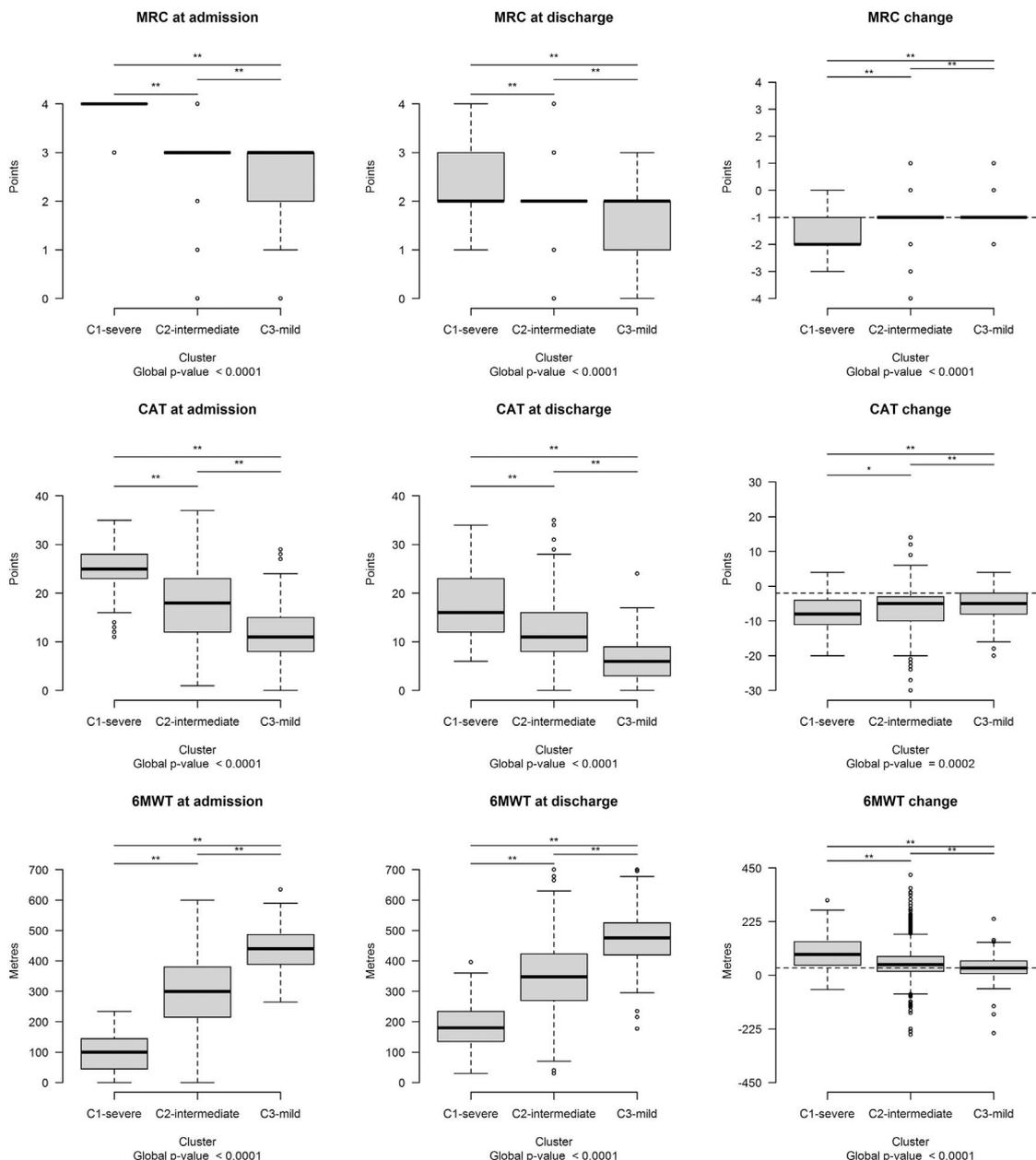
Fig. 2   Outcome measure distribution at admission, discharge, and post program changes. Each boxplot represents (from bottom to top): lowest non-outlier value, 25th, 50th (median value), 75th percentile and highest non-outlier value. Outliers with respect to each variable's distribution are reported as circles. The horizontal bar at the top of each plot indicates that the distribution differs significantly among clusters (** FDR < 0.05; * 0.05 ≤ FDR < 0.10). The horizontal dashed lines in black indicate the Minimal Clinically Important Difference of each outcome change. **Abbreviations** — CAT: COPD Assessment test; MRC: Medical Research Council; 6MWT: Six-minute walking distance test.

assessed by BiD was among the variables used for cluster definition with a major influence in discriminating the three subpopulations (Appendix - Fig. A.2) It has been shown that in-hospital PR results in clinically meaningful improvement in individuals recovering from ECOPD, independent of the severity of dyspnoea as assessed by BID. However, the levels of dyspnoea severity influenced the effect size.[27]

Limitations

This is a retrospective study, with the limitations of such type of studies. However, it represents a real-life condition and its

results are supported by the large sample size in a time when also randomised controlled trials are questioned.[33]

Cluster analysis only applies to the cohort studied and replication is essential in a totally different environment or subsequent cohort.

There were differences in the prevalence of clusters among participating ICS Maugeri centres due to organizational and logistic conditions (**Appendix A - Table A.3**). Participants in C3-mild (about 14% of participants, mainly distributed in a single center- 34.7%) (Table 4) came from their home, were not prescribed inhaler triple therapy, showed better CAT, were included in GOLD A stage, and had a better lung function. The

Table 4 Clusters distribution in the whole sample and by centre.

Cluster	Whole sample (n = 1159, 100%)	Lumezzane (n = 414, 35.7%)	Tradate (n = 369, 31.8%)	Pavia (n = 176, 15.2%)	Montescano (n = 138, 11.9%)	Others (n = 62, 5.4%)
C1-severe	126 (10.9%)	48 (11.6%)	2 (0.5%)	48 (27.3%)	22 (15.9%)	6 (9.7%)
C2-intermediate	862 (74.4%)	325 (78.5%)	239 (64.8%)	128 (72.7%)	116 (84.1%)	54 (87.1%)
C3-mild	171 (14.7%)	41 (9.9%)	128 (34.7%)	0 (0%)	0 (0%)	2 (3.2%)

Legend: variables distribution is described as absolute and relative frequency (%) of patients belonging to the three clusters in the whole sample and by centre.
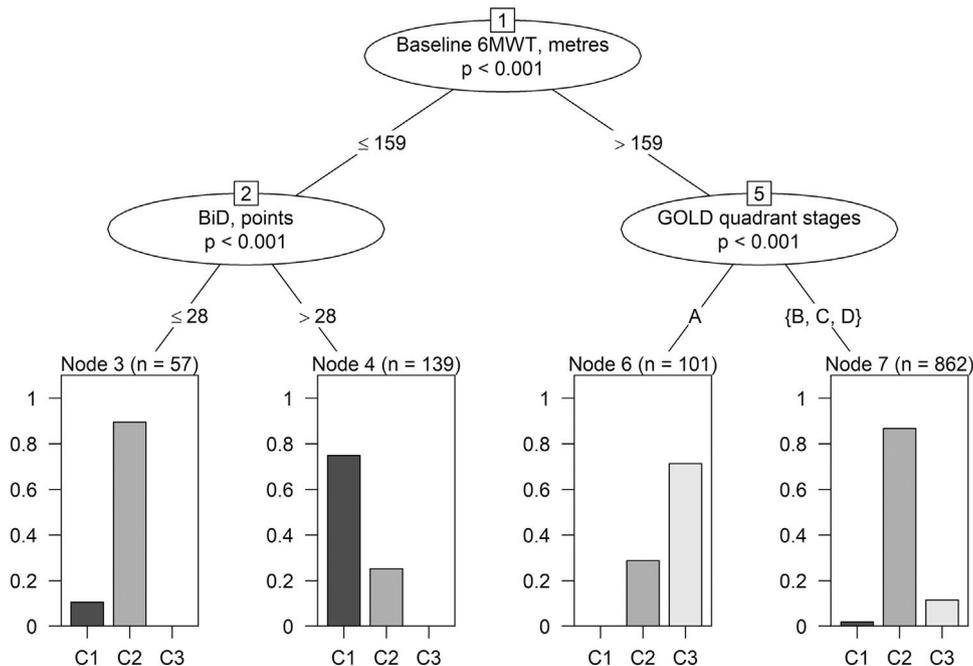
Fig. 3 Conditional inference tree structure. Branches correspond to informative splits in the data leading to the terminal node (leaves). Nodes describe the variable used to split data, branches indicate the splitting values; bar plots represent graphically the relative frequency of patients belonging to the three clusters by terminal node. As an example, approximately 90% of patients with baseline 6MWT values ≤ 159 meters and BiD ≤ 28 points belong to C2-intermediate cluster while about 10% to C1-severe and 0% to C3-mild. Patients within this terminal node are classified as belonging to C2-intermediate (the most frequent cluster within the terminal node). **Abbreviations** −6MWT: Six-minute walking distance test; BiD: Barthel index Dyspnea; GOLD: Global Initiative for Obstructive Lung Disease.

prescription of PR for these mild individuals might be questioned. However, participants were sent by other hospitals or by their GP and, given the retrospective design, we cannot exclude individual decisions by the accepting physicians according to criteria we cannot assess. Actually, all ICS Maugeri hospitals share the same admission criteria, evaluation, and rehabilitation protocols. Therefore, we are confident that these differences have not biased results.

Given the post-acute condition of participants, a control population not performing the program would have clarified whether any improvement in outcome would have been (also) time-dependent. However, given the recognized benefits of PR and the mission of our hospitals, not performing any program would have been unethical.

Conclusions

Our clustering analysis identified subpopulations of individuals recovering from ECOPD characterized by different PR success

rates. These results may help in defining priority criteria based on the probability of success. Our results reflect the specific population of individuals with indications of PR, not comparable and extensible to other populations of individuals with COPD, and should be confirmed by prospective studies.

Authorship contributions

MV and NA conceived and designed the study. MV, AM, and NA contributed to the writing of the manuscript. AM and RB performed formal analysis and visualization. MV, AS, PC, MM, BB, LR, and RM were responsible for investigations. MV, AM, MP, and NA participated in the analysis and discussion of the data, All the authors revised the article critically and approved the final version.

Funding sources

Data availability statement

Data are available from the corresponding author upon reasonable request.

Conflicts of interest

Nicolino Ambrosino is the Chief Editor of Pulmonology. The authors declare they have no conflict of interest.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.pulmoe.2023.01.002.

References

1. Spruit MA, Singh SJ, Garvey C, et al. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. Am J Respir Crit Care Med. 2013;188:e13−64. https://doi.org/10.1164/rccm.201309-1634ST.

2. Paneroni M, Vitacca M, Venturelli M, et al. The impact of exercise training on fatigue in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. Pulmonology. 2020;26:304−13. https://doi.org/10.1016/j.pulmoe.2020.02.004.

3. 2022 Global strategy for prevention, diagnosis and management of COPD. http://goldcopd.org/gold-reports/ Accessed September, 28th.

4. Lindenauer PK, Stefan MS, Pekow PS, et al. Association between initiation of pulmonary rehabilitation after hospitalization for COPD and 1-year survival among Medicare beneficiaries. JAMA. 2020;323:1813−23. https://doi.org/10.1001/jama.2020.4437.

5. Mosher CL, Nanna MG, Jawitz OK, et al. Cost-effectiveness of pulmonary rehabilitation among US adults with chronic obstructive pulmonary disease. JAMA Netw Open. 2022;5:e2218189. https://doi.org/10.1001/jamanetworkopen.2022.18189.

6. Vogiatzis I, Rochester CL, Spruit MA, et al. Increasing implementation and delivery of pulmonary rehabilitation: keymessages from the new ATS/ERS policy statement. Eur Respir J. 2016;47:1336−41. https://doi.org/10.1183/13993003.02151-2015.

7. Paneroni M, Vitacca M, Bernocchi P, Bertacchini L, Scalvini S. Feasibility of tele-rehabilitation in survivors of COVID-19 pneumonia. Pulmonology. 2022;28:152−4. https://doi.org/10.1016/j.pulmoe.2021.03.009.

8. Everitt BS, Landau S, Leese M, et al. Cluster analysis. 5th ed. 352. ISBN: 978-470-3Wiley Series in Probability and Statistics.

9. Kim V, Aaron SD. What is a COPD exacerbation? Current definitions, pitfalls, 8 challenges and opportunities for improvement. Eur Respir J. 2018;52:1801261. https://doi.org/10.1183/13993003.01261-2018.

10. Linn BS, Linn MW, Gurel L. Cumulative illness rating scale. J Am Geriatr Soc. 1968;16:622−6.

11. Celli BR, Cote CG, Marin JM, Casanova C, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. N Engl J Med. 2004;350:1005−12. https://doi.org/10.1056/NEJMoa021322.

12. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. Eur Respir J. 2005;26:319−38. https://doi.org/10.1183/09031936.05.00034805.

13. Quanjer PH, Stanojevic S, Cole TJ, et al. ERS Global Lung Function Initiative. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012;40:1324−43. https://doi.org/10.1183/09031936.00080312.

14. Vitacca M, Paneroni M, Baiardi P, et al. Development of a Barthel Index based on dyspnea for patients with respiratory diseases. Int J Chron Obstruct Pulmon Dis. 2016;11:1199−206. https://doi.org/10.2147/COPD.S104376.

15. Shah S, Vanclay F, Cooper B. Improving the sensitivity of the Barthel Index for stroke rehabilitation. J Clin Epidemiol. 1989;42:703−9. https://doi.org/10.1016/0895-4356(89)90065-6.

16. Holland AE, Spruit MA, Troosters T, et al. An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease. Eur Respir J. 2014;44:1428−46. https://doi.org/10.1183/09031936.00150314.

17. Enright PL, Sherrill DL. Reference equations for the six-minute walk in healthy adults. Am J Respir Crit Care Med. 1998;158:1384−7. https://doi.org/10.1164/ajrccm.158.5.9710086.

18. Fletcher CM. Standardised questionnaire on respiratory symptoms: a statement prepared and approved by the MRC Committee on the aetiology of chronic bronchitis (MRC breathlessness score). Br Med J. 1960;2:1665.

19. de Torres JP, Pinto-Plata V, Ingenito E, et al. Power of outcome measurements to detect clinically significant changes in pulmonary rehabilitation of patients with COPD. Chest. 2002;121:1092−8. https://doi.org/10.1378/chest.121.4.1092.

20. Jones PW, Harding G, Berry P, et al. Development and first validation of the COPD Assessment Test. Eur Respir J. 2009;34:648−54. https://doi.org/10.1183/09031936.00102509.

21. Kon SS, Canavan JL, Jones SE, et al. Minimum clinically important difference for the COPD Assessment Test: a prospective analysis. Lancet Respir Med. 2014;2:195−203. https://doi.org/10.1016/S2213-2600(14)70001-3.

22. Paneroni M, Simonelli C, Vitacca M, Ambrosino N. Aerobic exercise training in very severe chronic obstructive pulmonary disease: a systematic review and meta-analysis. Am J Phys Med Rehabil. 2017;96:541−8. https://doi.org/10.1097/PHM.0000000000000667.

23. Maltais F, LeBlanc P, Jobin J, et al. Intensity of training and physiologic adaptation in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1997;155:555−61. https://doi.org/10.1164/ajrccm.155.2.9032194.

24. Luxton N, Alison JA, Wu J, et al. Relationship between field walking tests and incremental cycle ergometry in COPD. Respirology. 2008;13:856−62. https://doi.org/10.1111/j.1440-1843.2008.01355.x.

25. Ippolito M, Vitale F, Accurso G, Iozzo P, Gregoretti C, Giarratano A, Cortegiani A. Medical masks and Respirators for the Protection of Healthcare Workers from SARS-CoV-2 and other viruses. Pulmonology. 2020;26:204−12. https://doi.org/10.1016/j.pulmoe.2020.04.009.

26. Vitacca M, Carone M, Clini EM, et al. Joint statement on the role of respiratory rehabilitation in the COVID-19 crisis: the Italian

position paper. Respiration. 2020;99:493−9. https://doi.org/10.1159/000508399.

27. Vitacca M, Ambrosino N, Belli S, et al. The severity of acute exacerbations of COPD and the effectiveness of pulmonary rehabilitation. Respir Med. 2021;184:106465. https://doi.org/10.1016/j.rmed.2021.106465.

28. Sánta B, Tomisa G, Horváth A, Balázs T, Németh L, Gálffy G. Severe exacerbations and mortality in COPD patients: a retrospective analysis of the database of the Hungarian National Health Insurance Fund. Pulmonology. 2022. https://doi.org/10.1016/j.pulmoe.2022.11.001. S2531-0437(22)00259-8Epub ahead of print.

29. Mathioudakis AG, Abroug F, Agusti A, et al. ERS statement: a core outcome set for clinical trials evaluating the management of COPD exacerbations. Eur Respir J. 2022;59:2102006. https://doi.org/10.1183/13993003.02006-2021.

30. Garrod R, Malerba M, Crisafulli E. Determinants of success. Eur Respir J. 2011;38:1215−8. https://doi.org/10.1183/09031936.00088611.

31. Al Chikhanie Y, Bailly S, Amroussa I, et al. Clustering of COPD patients and their response to pulmonary rehabilitation. Respir Med. 2022;198:106861. https://doi.org/10.1016/j.rmed.2022.106861.

32. Spruit MA, Augustin IM, Vanfleteren LE, et al. Differential response to pulmonary rehabilitation in COPD: multidimensional profiling. Eur Respir J. 2015;46:1625−35. https://doi.org/10.1183/13993003.00350-2015.

33. Duarte-de-Ara Jo AN, Teixeira P, Hespanhol V, Correia-de-Sousa J. COPD: how can evidence from randomised controlled trials apply to patients treated in everyday clinical practice? Pulmonology. 2022;28:431−9. https://doi.org/10.1016/j.pulmoe.2020.02.003.