STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. 11, March 2023, pp 299–331.
Published online in International Academic Press (www.IAPress.org)

A Multiobjective Optimization Approach to Pulmonary Rehabilitation Effectiveness in COPD

Jorge Cabral [1,*], Vera Afreixo [1], Cristiana J. Silva [1], Ana Helena Tavares [1,2], Alda Marques [3]

[1]Center for Research and Development in Mathematics and Applications (CIDMA), University of Aveiro, Aveiro, Portugal

[2]Águeda School of Technology and Management, University of Aveiro, Aveiro, Portugal
[3] Respiratory Research and Rehabilitation Laboratory (Lab3R), School of Health Sciences and Institute of Biomedicine (iBiMED), University of Aveiro, Aveiro, Portugal

Abstract Chronic obstructive pulmonary disease (COPD) is a common disease that accounts for a significant individual and societal burden. Pulmonary rehabilitation (PR) is a key management strategy but it is highly inaccessible, making prioritisation highly needed. This study aimed to determine and optimize predictive models of PR outcomes and build a tool to help healthcare professionals in their clinical decision-making about PR prioritisation. Data from patients who performed a 12-week community-based PR programme were analysed. Exercise capacity with the six-minutes walk test distance (6MWD), isometric quadriceps muscle strength with the handheld dynamometry (QMS) and dyspnoea with the modified Medical Research Council dyspnoea scale (mMRC) were assessed before and after PR. Multiple linear regression models were determined based on the Akaike information criteria and a cross-validation method. The resultant multiobjective problem was solved using the Nondominated Sorting Genetic Algorithm-II. R Shiny package was used to create a web-based user interface. Data from 95 patients with COPD (median age of 69 years, 19 female and generally overweight), resulted in linear predictive models for the post-pre difference of the 6MWD, QMS and mMRC with cross-validation R^2 of 0.49, 0.53 and 0.51, respectively. 6MWD and mMRC were common statistically significant predictors. Pareto front patients were obese ex-smoker women that do not do long-term oxygen therapy and that performed PR. The distance to the Pareto front along with the estimates given by our models are easily obtained using the designed R Shiny interface and may help healthcare professionals decide on the prioritisation to PR programmes.

Keywords COPD, mMRC, Multiobjective Optimization, NSGA-II, Pulmonary Rehabilitation, R Shiny

AMS 2010 subject classifications 58E, 62P

DOI: 10.19139/soic-2310-5070-1505

1. Introduction

Chronic obstructive pulmonary disease (COPD) is a common, progressive, preventable and treatable disease that is characterized by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases [1].
COPD kills more than 3 million people worldwide every year [2] and is highly prevalent, especially among older people. It was projected to be the seventh leading cause of disability-adjusted life-years (DALYs) by 2030 [3] but in 2019 was already the sixth leading cause of DALYs of all ages, the fourth leading cause in the age group 50–74 years and the third in the age group 75 years and older [4]. It accounts

*Correspondence to: Jorge Cabral (Email: jorgecabral@ua.pt). Department of Mathematics, University of Aveiro. Campus Universitário de Santiago, 3810-193 Aveiro, Portugal.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

therefore for a great individual, economic and societal burden [5]. Pulmonary rehabilitation (PR) is a non-pharmacological, comprehensive and interdisciplinary intervention, which includes, but is not limited to, exercise training, education and behaviour change designed to improve the physical and psychological condition of people with chronic respiratory disease and to promote the long-term adherence to health-enhancing behaviors [6]. PR has been compared to a Swiss army knife, a multitargeted approach, in which patients' physical, emotional and social treatable traits (i.e., characteristics that are clinically relevant, identifiable and modifiable/treatable) are identified and addressed [7]. It is considered a key management strategy for people with chronic respiratory disease namely to those with COPD since it improves symptoms, physical, emotional, psychosocial functioning quality of life and well-being by a clinically important amount [8-10]. It thus strengthens and maximises the individual maximal level of autonomy and functioning in the community [11].

Nevertheless, insufficient funding, limited resources and suboptimal use of PR, makes PR largely inaccessible when compared with pharmacological treatments [12] and insufficient to fill the needs of communities, contributing for the gap between the established benefits of PR and its actual delivery [10].

Predictive models are extremely useful for forecasting outcomes. One of the most used statistical technique for relating a set of variables that can be used for predictive modelling is the multiple linear regression model [13]. Given the limited resources and infrastructure for PR, predicting which patients will most benefit from PR is important, so those can be prioritised [14]. Currently, no clear baseline demographic, clinical or physiological features can be used as formalized or standardized selection criteria for PR, as individual factors do not reliably predict successful multidimensional PR outcomes [15].

In the real world, there are many problems having two or more multidimensional outcomes or objectives, often conflicting, that we aim to optimize at the same time. They are called multiobjective optimization problems and their solution has been a challenge to researchers for a long time [16-17]. Think, for instance, of the optimization of a vehicle design, where the weight, acceleration characteristics and toe-board intrusion are considered as design objectives [18], the energy production where three conflicting objectives are considered (power production, production costs and carbon dioxide emissions) [19] and the problem of portfolio selection where typically return and risk are the two decision criteria [20].

Health-related problems can also be approached from a multiobjective optimization perspective. Consider the problem of where healthcare facilities should be located to improve the equity of accessibility, raise the accessibility to the entire population, reduce the population that falls outside the coverage range, and decrease the cost of building new facilities [21], the problem of multiple resource allocation of operating rooms which resulted in the proposal of a surgery scheduling scheme [22] or the biobjective optimization problem of the amount of drug usage and the quality of treatment of Human Immunodeficiency Virus infection [23].

Because of the conflict among the objectives, solving a multiobjective optimization problem produces a set of solutions representing the best possible trade-offs among the objectives (i.e., solutions in which one objective cannot be improved without worsening another one). Such solutions constitute the Pareto optimal set and the image of this set (i.e., the corresponding objective function values) form the so-called Pareto front [16].

To solve multiobjective optimization problems, multiobjective evolutionary algorithms (MOEAs) have shown to be important tools due to their effectiveness, flexibility and applicability [17,24]. They use paradigms from natural evolution, such as selection, recombination and mutation to steer a population towards optimal or near-optimal solutions, and typically they are designed to gradually approach sets of Pareto optimal solutions that are well-distributed across the Pareto front [25].

One of the most popular Pareto-based MOEAs is the elitist Nondominated Sorting Genetic Algorithm-II (NSGA-II) [47], which uses a ranking scheme called non-dominated sorting, and adopts a mechanism called crowded comparison operator, which does not require any parameters, as its density estimator [16].

The main objective of this study was to determine and optimize predictive models of PR outcomes and use that information to sort patients who were proposed for PR programmes. A second objective was to

build a R Shiny [53] tool that aims to help healthcare professionals in their clinical decision-making on the inclusion of patients with COPD in PR programmes.

2. Material and methods

2.1. Study design and participants

We have retrospectively analysed data from a prospectively maintained database of patients with COPD who participated in three different studies of PR (3R - SAICT-POL/23926/2016, PRISMA - OHM-E/2018/ 1912 and PRIME - PTDC/SAU-SER/28806/2017) at the Respiratory Research and Rehabilitation Laboratory, at School of Health Sciences of the University of Aveiro, between October 2017 and April 2019. Only people diagnosed with COPD according to the GOLD criteria [1] were included in the analyses. Participants with other respiratory diseases, signs of cognitive impairment (e.g., dementia) or presence of a significant or unstable cardiovascular (e.g., symptomatic ischaemic cardiac disease), neurological (e.g., neuromuscular dystrophy disease) or musculoskeletal disease (e.g., important kyphoscoliosis) or any other clinical condition that may affect data interpretation were excluded.
Patients were divided into two groups. One group served as control and received standard treatment, i.e., optimisation of pharmacological treatment. The other group consisted of those who underwent a 12-weeks community-based PR programme composed of exercise training twice a week and education and psychosocial support once every other week. Nutritional and psychological counselling was also provided, if needed [26].

2.2. Data collection

Sociodemographic (age; gender; educational level – EdLevel), anthropometric (height and weight to calculate the body mass index - BMI) and general clinical (smoking habits - SmokSt; pack-years – PY; long-term oxygen therapy - LTOT; number of acute exacerbations - AECOPD; comorbidities to calculate the Charlson Comorbidity Index – CCI [27]) data were first collected with a structured questionnaire to characterise the sample. Lung function (Forced expiratory volume in one second – FEV_1 and Forced Vital Capacity – FVC) was then assessed with spirometry (MicroLab 3535, CareFusion, Kent, UK) as recommended by the American Thoracic Society and the European Respiratory Society [28].
The following outcomes/outcome measures were collected as described below, at baseline and at three months in both groups.
Dyspnoea. The modified medical research council questionnaire (mMRC) assesses the activities limitation due to dyspnoea in 5 statements, rated from 0 (no troubles with breathlessness) to 4 (too breathless to leave the house) [29-30]. A change of 1 unit is commonly used as minimal clinical important difference [30].
Impact of the disease. The COPD Assessment Test (CAT) is an 8-item questionnaire each assessed with a 6-point Likert scale [31-32]. Scores range from 0-40, and are interpreted as 10 low impact, 11-20 medium, 21-30 high and 31-40 very high impact, with 5 representing the upper limit of normal in healthy non-smokers [33]. A change of 2.9 units is commonly used as minimal clinical important difference [34]. Muscle strength. Quadriceps isometric muscle strength (QMS) were measured with a handheld dynamometer in kilogram-force (microFET2, Hoggan Health, The best Salt Lake City, Utah). Measurements were taken at the dominant side, as previously described [35-38] and the best of three acceptable measurements (less than 10% of variation) was used for analysis.
Changes of 5.2 kgf were used as minimal clinically important differences for the isometric quadriceps muscle strength, measured with the handheld dynamometer [39]. Functional status. The one-minute sit to stand test (1minSTS) replicates a common daily living activity, sitting and standing from a chair. Patients were asked to sit and stand from a 46-48 cm height chair as many times as possible during one minute [40-41]. A change of 3 repetitions is commonly used as minimum clinically important difference [42].

Exercise capacity. The six-minute walk test (6MWT) is a self-paced test of walking capacity. Patients were asked to walk as far as possible during 6 minutes along a flat 30 meters corridor [43]. A change of 25 m is commonly used as minimum clinically important difference [44].
Minimal clinically important difference has been defined as the smallest difference in a measurable clinical parameter that indicates a meaningful change in the condition for better or for worse, as perceived by the patient, clinician, or investigator [45].

2.3. Data optimization and analysis

Multiobjective optimization is an area of multiple-criteria decision-making, concerning mathematical optimization problems involving more than one objective function to be optimized simultaneously [46].
In this work, we consider the multiobjective optimization problem of the type:

$$minimize\ f(x) := [f_1(x), f_2(x), \ldots, f_m(x)] \tag{1}$$

subject to

$$g_i(x) \leq 0,\ i = 1, 2, \ldots, p \tag{2}$$

$$h_j(x) = 0,\ j = 1, 2, \ldots, q \tag{3}$$

where $x = [x_1, x_2, \ldots, x_n]^T$ is the vector of decision variables, $f_k : \mathbb{R}^n \to \mathbb{R},\ k = 1, \ldots, m$ are the objective functions and $g_i, h_j : \mathbb{R}^n \to \mathbb{R},\ i = 1, \ldots, p,\ j = 1, \ldots, q$ are the constraint functions of the problem.
Consider the following definitions:
Definition 1 Given two vectors $x, y \in \mathbb{R}^k$ we say that $x \leq y$ if $x_i \leq y_i$ for $i = 1, \ldots, k$, and that x dominates y (denoted by $x \prec y$) if $x \leq y$ and $x \neq y$.
Definition 2 A vector of decision variables $x \in X \subset \mathbb{R}^n$ is nondominated with respect to X, if there does not exist another $x' \in X$ such that $f(x') < f(x)$.
Definition 3 A vector of decision variables $x^* \in \mathcal{F} \subset \mathbb{R}^n$ ($\mathcal{F}$ is the feasible reagion) is Pareto-optimal if it is nondominated with respect to $\mathcal{F}$.
Definition 4 The Pareto-optimal set $\mathcal{P}^*$ is defined by $\mathcal{P}^* = \{x \in \mathcal{F} \mid x\ is\ Pareto - optimal\}$.
Definition 5 The Pareto front $\mathcal{PF}^*$ is defined by $\mathcal{PF}^* = \left\{f(x) \in \mathbb{R}^k \mid x \in \mathcal{P}^*\right\}$.
Therefore, our aim was to obtain the Pareto optimal set from the set $\mathcal{F}$ of all the decision variable vectors that satisfy (2) and (3) [16].
The NSGA-II is a heuristic algorithm based on an elite principle that adopts an explicit diversity retention mechanism and emphasizes non-dominated solutions in which good solutions can be obtained even in mixed discrete-continuous variables problems [47-49]. Initial populations of chosen size are generated randomly, and then a first-generation offspring population is obtained by the selection, crossover, and mutation of the genetic algorithm after non-dominant sequencing. It mainly combines the populations of the parent and the offspring, and after they are quickly non-dominated, it calculates the crowding degree of the individual in the non-dominant layer. The appropriate individuals are selected by non-dominant relationships and the results of individual crowding degrees to form a new population. At last, a new subpopulation is created by the basic operation of the genetic algorithm. Then these operations are repeated until the conditions for program termination are satisfied [22].
In this work, a 3-objective mixed-variable multiobjective optimization with constraint is performed using nsga2 function from the mco R package [50]. The number of generations used was 30, with an initial population of 10000, a crossover probability of 0.7 and a mutation probability of 0.2. The Pareto front is determined.
Data analysis was performed using RStudio Version 1.3.1056 [51] running R version 4.0.2 [52]. Figure 1 illustrates the procedures followed in our study.
The statistical quantitative variables were summarized using median values, interquartile ranges and amplitude ranges and the categorical statistical variables through count values and percentages. Shapiro-Wilk test was used to assess the assumption of normality. Bartlett test was used to assess homoscedasticity. Wilcoxon rank-sum test was used to evaluate the difference in the median of quantitative variables

with Benjamini & Hockberg correction. Statistically linear dependence was tested using Pearson's and Spearman's correlation tests PR outcomes (i.e., lung function - FEV_1 and FVC; isometric quadriceps muscle strength - QMS; functional capacity - 1minSTS; dyspnoea - mMRC; impact of the disease - CAT; exercise capacity - six-minute walking distance (6MWD)) were defined as the difference between the post and the pre-PR values. Multiple linear regression models for PR outcomes were determined from the set of predictor variables, by entering and removing predictors based on Akaike information criteria (AIC). A further fine tuning of the predictors was conducted based on the adjusted R square (R^2_{adj}) and the R square obtained by a 10-fold, repeated 3 times, cross-validation method (R^2_{CV}). Multicollinearity was measured through the Variance inflation factor (VIF) and a value greater than 5 represented the existence of multicollinearity. Patients with absolute values of standardized residuals greater than 1.96 were considered outliers. The assumptions of the multilinear regression were assessed with the Shapiro-Wilk test, the studentized Breusch-Pagan test and the Durbin-Watson test. Patients that achieved or overcame the MCID were classified as "responders" and the others "non-responders".
A min-max normalization of the outcomes was conducted to determine the mean Euclidean distance of predicted and observed outcomes to all the simulated patients in the Pareto front.
A Bland-Altman assessment for agreement was used to compare the two distances. A range of agreement was defined as mean bias ± 1.96 standard deviation (SD).
Kruskal-Wallis test was used to evaluate differences of distances between groups defined by categorical variables. Statistically linear dependence between the distance to the Pareto front and quantitative variables was assessed using Spearman's correlation tests.
A significance level of 0.05 was considered.

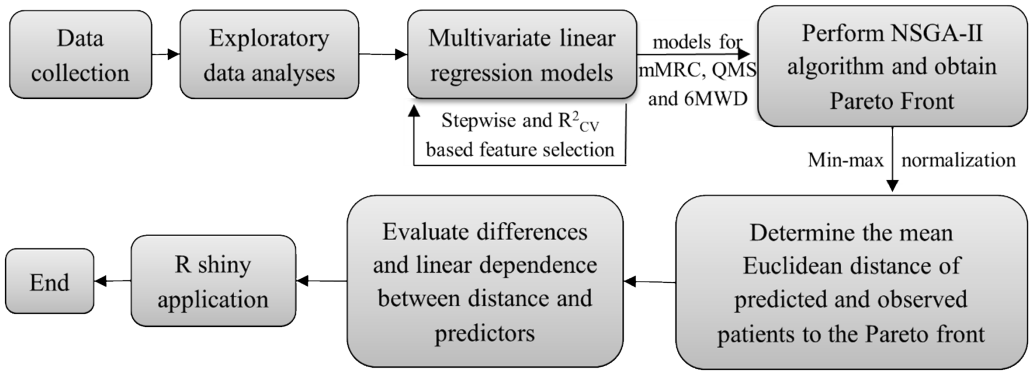

Figure 1. Flowchart of the study procedure. mMRC - Modified Medical Research Council Questionnaire; QMS – Quadriceps Muscle Strength; 6MWD – Six-Minutes Walking Distance.

3. Results

The database consisted of 95 patients with a median age of 69 years, 19 (20.0%) of whom were female. From those, 64 (67.4%) participated in the community-based PR intervention. Statistically significant differences were found in all variables measured Pre-PR and Post-PR except for FEV_1. Table 1 presents the detailed characteristics of all variables considered.

Table 1. Descriptive statistics of collected variables pre- and post-pulmonary rehabilitation of patients with chronic obstructive pulmonary disease (n=95).

Variable	Pre-PR			Post-PR			
	All (n=95)	No-PR (n=31)	PR (n=64)	All (n=95)	No-PR (n=31)	PR (n=64)	No-
Age, years	69 (63.5,72.0)	68 (61,69)	69 (65.5,75.0)	-	-	-	
BMI (x_7), kg/m^2	25.97 (23.44,29.23)	26.93 (24.39,29.58)	25.47 (22.79,28.79)	-	-	-	
PY (x_3)	35.2 (12.1,62.0)	40 (19.38,73.00)	31.2 (10.0,57.0)	-	-	-	
CCI	4 (3,5)	3 (2.5,4)	4 (3,5)	-	-	-	
AECOPD	0 (0,1)	0 (0,1)	1 (0,1)	-	-	-	
FEV_1, litres	1.30 (0.97,1.57)	1.45 (1.29,1.82)	1.19 (0.91,1.43)	1.23 (0.98,1.54)	1.34 (1.08,1.77)	1.20 (0.93,1.41)	0.0
FVC, litres	2.58 (2.11,3.15)	2.84 (2.53,3.71)	2.44 (2.06,2.9)	2.54 (2.03,2.98)	2.61 (2.05,3.27)	2.47 (2.04,2.85)	0.0
CAT (x_{10})	14 (10,21)	11 (8,14)	17 (12,22)	10 (6,16)	8 (5,11)	13 (7,17)	0.0
QMS (x_{11}), kgf	31.20 (26.40,36.75)	31.40 (27.00,37.00)	31.00 (26.38,36.30)	35.10 (28.95,38.75)	34.20 (28.00,37.05)	35.80 (28.98,39.98)	0.3
mMRC (x_8)	2 (1,3)	1 (1,2)	2 (1,3)	1 (1,2)	2 (1,2)	1 (1,2)	0.2
1minSTS	25.0 (20.0,31.0)	27.0 (23.0,32.5)	23.0 (18.8,29.5)	27.0 (21.5,34.0)	28.0 (22.0,36.5)	27.0 (21.8,34.0)	0.3
6MWD (x_9), metres	419.6 (336.4,494.9)	421.0 (392.6,497.6)	413.6 (332.4,493.4)	453.0 (388.2,524.9)	446.4 (401.7,505.0)	460.6 (374.3,531.0)	0.4
Group (PR) (x_1)	64 (67.4%)	0 (0%)	64 (100%)	-	-	-	
Gender (x_2) (female)	19 (20.0%)	7 (22.6%)	12 (18.8%)	-	-	-	
EdLevel (>=5th)	44 (46.3%)	12 (38.7%)	32 (50.0%)	-	-	-	
SmokSt (Ex-smoker) (x_4)	64 (67.4%)	21 (67.7%)	43 (67.2%)	-	-	-	
SmokSt (Smoker) (x_5)	16 (16.8%)	6 (19.4%)	10 (15.6%)	-	-	-	
LTOT (yes) (x_6)	11 (11.6%)	3 (9.7%)	8 (12.5%)	-	-	-	
Dif FEV_1, litres	-	-	-	0.0 (-0.1,0.1)	-0.1 (-0.2,0.0)	0.0 (-0.1,0.1)	
Dif FVC, litres	-	-	-	-0.1 (-0.4,0.1)	0.2 (-0.4,0.0)	0.0 (-0.3,0.2)	
Dif mMRC	-	-	-	0 (-1,0)	0 (0,1)	-1 (-1,0)	
Dif 1minSTS	-	-	-	2 (0,7)	0 (-1,2.5)	4 (0,7)	
Dif 6MWD, metres	-	-	-	31.5 (-0.7,65.6)	5.0 (-13.5,25.6)	49.9 (17.7,88.3)	
Dif CAT				-4 (-7,-1)	-3 (-5,-1)	-5 (-7,-1)	
Dif QMS, kgf	-	-	-	2.5 (-1.0,7.8)	1.0 (-3.5,7.5)	2.8 (0.5,8.2)	

Data are presented as n (%) or median (Interquartile Range); $p - value_{adj}$ bold represents significant statistical difference ($\alpha = 0.05$) between pre- and post-pulm rehabilitation (PR) medians with a p-value adjusted for False Discovery Rate. COPD - Chronic Obstructive Cardiopulmonary Disease; PR – Pulmonary Rehabili BMI – Body Mass Index; PY – Pack Years; CCI - Charlson Comorbidity Index; AECOPD - Acute Exacerbations of COPD; FEV_1 – Forced Expiratory Volume in FVC - Forced Vital Capacity; CAT – COPD Assessment Test; QMS – Quadriceps Muscle Strength; mMRC - Modified Medical Research Council Questionnaire; 1minSTS - One-minute Sit To Stand test; 6MWD – Six-Minutes Walking Distance; EdLevel – Educational Level; SmokSt – Smoking Status; LTOT – Long-Term Dif – Difference between post- and pre-pulmonary rehabilitation measures.

Sample was generally overweight (≈26 kg/m^2) with a minimum BMI of 17.6 kg/m^2 and a maximum of 38.2 kg/m^2. Maximum walking distance was 652 meters and minimum 75 meters in the 6MWT. Maximum QMS value was 55.6 kgf whereas the minimum value was 13.3 kgf. Ex-smokers had a minimum PY of 0.2 whereas actual smokers had a minimum PY of 2.7 with an equal maximum of 180. mMRC assumed all possible values while CAT ranged from 2 to 35. Individual analysis of the outcomes considering the MCID found 37 (38.9%) responders in the mMRC from which 89.2% attended PR programmes. QMS minimal clinical important difference was achieved by 35 (36.8%) patients, 71.4% of them enrolled in PR. The improvement in the distance walked by 52 (54.7%) patients was at least equal to the MCID with 84.6% attending PR programmes. Only 8 (8.4%) patients were considered responders if all outcomes' measures were considered simultaneously. The multilinear regression model obtained for the difference between the post-PR mMRC and the Pre-PR mMRC included 6 predictors, 5 of them statistically significant and no multicollinearity was observed. The adjusted R^2_{adj} had a value of 0.49 and the 10-fold cross-validation R^2_{CV} was slightly higher (Table 2).

From the 7 predictors included in the multilinear regression model for the difference between the post-PR QMS and the pre-PR QMS, 3 were common to the previous model. All predictors were statistically significant and multicollinearity was not observed. Seven patients were removed. The adjusted R^2_{adj} had a value of 0.51 and the 10-fold cross-validation R^2_{CV} was 0.02 higher (Table 3).

The R^2_{adj} and R^2_{CV} for the multilinear regression model obtained for the 6MWD difference between the post and the pre-PR were the lowest of them all and the number of predictors included was the highest (Table 4).

In Sections A, B and C of the appendix we present the univariate linear regression results for the difference between Post- and Pre-outcomes, the observed values versus the predicted values and the

Table 2. Summary of multiple linear regression for the difference of modified medical research council questionnaire in patients with chronic obstructive pulmonary disease (n=95).

	Coeff	95% CI	p-value	VIF
Intercept	2.2477	[1.414;3.081]	<0.001	-
Group (PR) (x_1)	-0.6646	[-0.961;-0.368]	<0.001	1.13
PY (x_3)	-0.0035	[-0.007;-0.0004]	0.029	1.09
LTOT (yes) (x_6)	0.4992	[0.047;0.951]	0.031	1.22
mMRC (x_8)	-0.6450	[-0.820;-0.470]	<0.001	1.99
6MWD (x_9)	-0.0028	[-0.004;-0.001]	<0.001	1.59
CAT (x_{10})	0.0184	[-0.003;0.039]	0.085	1.61

R^2=0.53; R^2_{CV}=0.51; R^2_{adj}=0.49.

Coeff - Regression coefficient estimate; 95% CI - 95% confidence interval; VIF – Variance inflation factor;

PY – Pack years; LTOT – Long-term oxygen therapy; mMRC - Modified Medical Research Council Questionnaire;

PR – Pulmonary rehabilitation; 6MWD – Six-minutes walking distance; CAT – COPD assessment test.

Table 3. Summary of multiple linear regression for the difference of quadriceps muscle strength in patients with chronic obstructive pulmonary disease (n=88).

	Coeff	95% CI	p-value	VIF
Intercept	0.5405	[-8.882;9.963]	0.909	-
Group (PR) (x_1)	4.9981	[3.009;6.987]	<0.001	1.10
SmokSt (Ex-smoker) (x_4)	6.1571	[3.440;8.874]	<0.001	1.98
SmokSt (Smoker) (x_5)	4.9780	[1.558;8.398]	0.005	2.01
BMI (x_7)	0.3736	[0.136;0.612]	0.003	1.27
mMRC (x_8)	-1.5448	[-2.607;-0.482]	0.005	1.51
6MWD (x_9)	0.0169	[0.006;0.028]	0.003	1.85
QMS (x_{11})	-0.6241	[-0.771;-0.477]	<0.001	1.49

R^2=0.55; R^2_{CV}=0.53; R^2_{adj}=0.41.

Coeff - Regression coefficient estimate; 95% CI - 95% confidence interval; VIF – Variance inflation factor;

SmokSt – Smoking status; BMI – Body mass index; mMRC - Modified Medical Research Council Questionnaire;

PR – Pulmonary rehabilitation; 6MWD – Six-minutes walking distance; QMS – Quadriceps muscle strength.

Table 4. Summary of multiple linear regression for the difference of six minutes walking distance in patients with chronic obstructive pulmonary disease (n=89).

	Coeff	95% CI	p-value	VIF
Intercept	81.2005	[28.973;133.428]	0.003	
Group (PR) (x_1)	46.3837	[28.852;63.915]	<0.001	1.12
Gender (female) (x_2)	23.4841	[1.976;44.993]	0.033	1.27
PY (x_3)	-0.2637	[-0.515;-0.012]	0.040	1.85
SmokSt (Ex-smoker) (x_4)	59.0927	[31.607;86.578]	<0.001	2.80
SmokSt (Smoker) (x_5)	24.2149	[-3.350;51.779]	0.084	1.83
LTOT (Yes) (x_6)	-33.5306	[-61.156;-5.905]	0.018	1.24
mMRC (x_8)	-11.5711	[-21.078;-2.064]	0.018	1.59
6MWD (x_9)	-0.2047	[-0.291;-0.119]	<0.001	1.55

R^2=0.50; R^2_{CV}=0.49; R^2_{adj}=0.45.

Coeff - Regression coefficient estimate; 95% CI - 95% confidence interval; VIF – Variance inflation factor;

SmokSt – Smoking status; LTOT – Long-term oxygen therapy; mMRC - Modified Medical Research Council Questionnaire;

PY – Pack years; PR – Pulmonary rehabilitation; 6MWD – Six-minutes walking distance.

analysis of the residuals.
Multilinear regression models obtained for the differences between the post and pre 1minSTS, CAT, FEV_1 and FVC were not considered since the R^2_{CV} and the R^2_{adj} were classified as low or the PR coefficient was not statistically significant.
To formulate our minimization multiobjective optimization we considered the following objectives.
Objective 1: minimize f_1 defined as the difference between the post-mMRC and the pre-mMRC values (dif_mMRC).
Objective 2: minimize f_2 defined as the difference between the pre-QMS and the post-QMS values (dif_QMS).
Objective 3: minimize f_3 defined as the difference between the pre-6MWD and the post-6MWD values (dif_6MWD).
Our multiobjective optimization problem is given by

$$min\ f(x) = \begin{cases} f_1(x) = +2.2477 - 0.6646x_1 - 0.0035x_3 + 0.4992x_6 - 0.6450x_8 - 0.0028x_9 + 0.0184x_10 \\ f_2(x) = -0.5405 - 4.9981x_1 - 6.1571x_4 - 4.9780x_5 - 0.3736x_7 + 1.5448x_8 - 0.0169x_9 + 0.6241x_11 \\ f_3(x) = -81.2005 - 46.3837x_1 - 23.4841x_2 + 0.2637x_3 - 59.0927x_4 - 24.2149x_5 + 33.5306x_6 + 11.5711x_8 + 0.2047x_9 \end{cases}$$

where $x = [x_1, x_2, \ldots, x_{11}]^T$ are subject to

$$\begin{gathered} x_3 - 180 \leq 0 \\ x_4 = 1 \Rightarrow 0.2 - x_3 \leq 0 \\ x_5 = 1 \Rightarrow 2.7 - x_3 \leq 0 \\ 17.6 \leq x_7 \leq 38.2 \\ 75 \leq x_9 \leq 652 \\ 13.3 \leq x_{11} \leq 55.6 \\ x_4 + x_5 - 1 \leq 0 \\ x_1, x_2, x_4, x_5, x_6 \in \{0, 1\} \\ x_8 \in \{0, 1, 2, 3, 4\} \\ x_{10} \in \{2, \ldots, 35\} \\ f_1 + 1 \leq 0;\ f_2 + 5.2 \leq 0;\ f_3 + 25 \leq 0 \end{gathered}$$

The Pareto front obtained by the multiobjective optimization method NSGA-II is represented in Figure 2.

We only considered patients in the Pareto front who achieved the MCID for all outcomes resulting in 2073 simulated patients with dif_mMRC ranging from -2.856 to -1, dif_QMS from -27.134 to -5.200 and dif_6MWT from -148.41 and -25.05. These simulated patients are Ex-smoker women that do not do LTOT and 99.90% performed pulmonary rehabilitation. BMI ranged from 17.65 to 38.24, with a median of 36.63 (interquartile range [32.45;37.81]). PY ranged from 0.20 to 157.53 with a median of 1.82 (interquartile range [0.74;5.05]). mMRC assumed values from 1 to 4 with median, 1st and 3rd quartiles equal to 4. The minimum distance walked in the 6MWT was 75 and the maximum 652, with a median equal to 440 (interquartile range [261;590]). Values for QMS ranged from 13.30 to 39.95 with a median of 15.16 (interquartile range [13.72;20.58]) (Figure 3).

In Section D of the appendix the characteristics of the patients in the Pareto front are coloured according to the different outcomes.
The predicted and the observed outcomes are represented in Figure 4 along with the Pareto front. In Sections E and F of the appendix the 2-dimension projections of the Pareto front, predicted and observed outcomes are plotted.

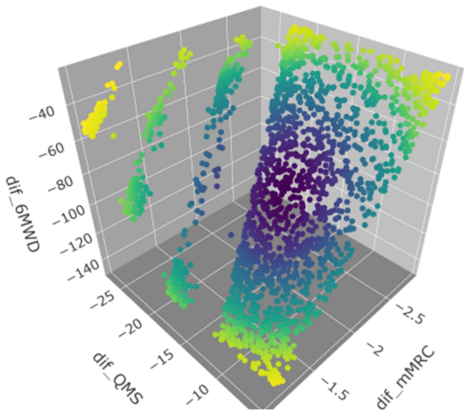

Figure 2. Pareto front. Points are coloured according to the mean Euclidean distance to the remaining points. Colour scale ranges from Yellow, that represents a greater distance, to Purple, that represents a smaller distance. dif_mMRC - difference between the post- and the pre- Modified Medical Research Council Questionnaire; dif_QMS - difference between the pre- and the post-Quadriceps Muscle Strength; dif_6MWD - difference between the pre- and the post-Six-Minutes Walking Distance.

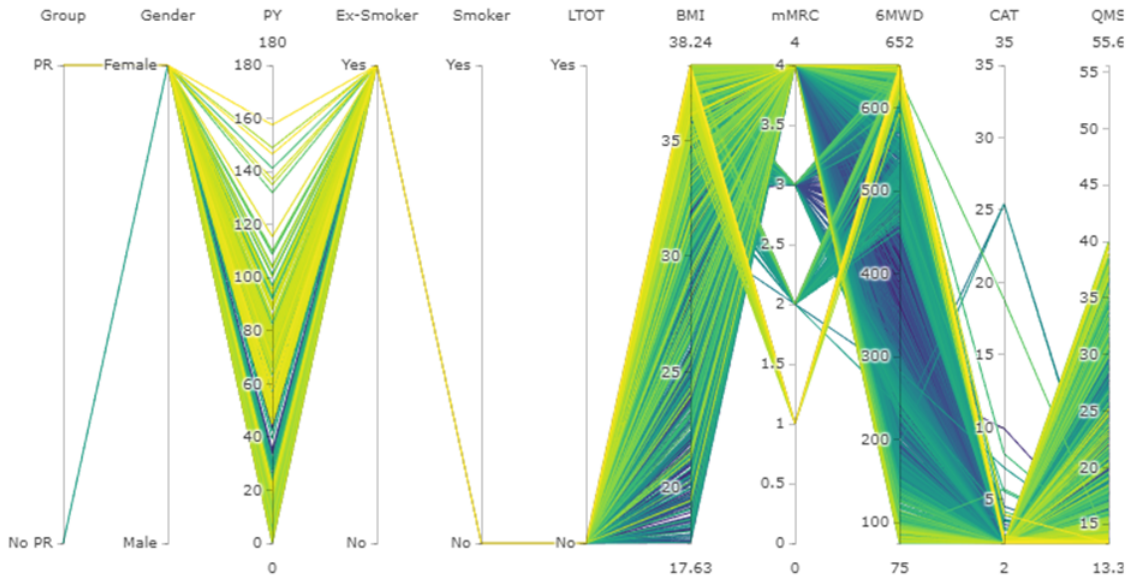

Figure 3. Baseline characteristics of patients with chronic obstructive pulmonary disease in the Pareto front. Lines are coloured according to the mean Euclidean distance of the patients to the remaining patients of the Pareto front. Colour scale ranges from Yellow, that represents a greater distance, to Purple, that represents a smaller distance. PR – Pulmonary Rehabilitation; BMI – Body Mass Index; PY – Pack Years; CAT – Chronic Obstructive Pulmonary Disease Assessment Test; QMS – Quadriceps Muscle Strength; mMRC - Modified Medical Research Council Questionnaire; 6MWD – Six-Minutes Walking Distance; LTOT – Long-Term Oxygen Therapy.

A min-max normalization of each of the three outcomes using the range of predicted values and the Pareto front values was conducted. The Euclidean distance between each three-dimensional point (that corresponds to a patient) and all the simulated points of the Pareto front in both predicted and observed situation was determined and the mean value, by patient, was defined as the predicted/observed distance to the Pareto front.

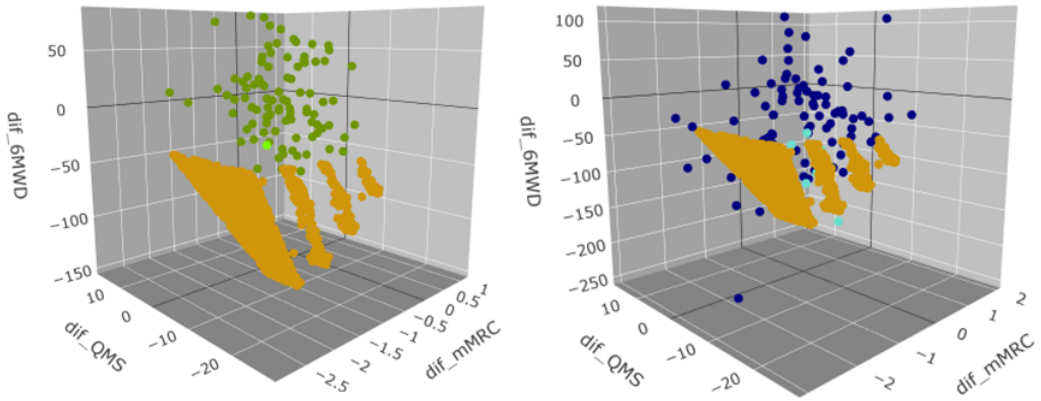

Figure 4. Pareto front and predicted (left) and observed (right) outcomes of patients with chronic obstructive pulmonary disease (light colours represent responder patients). dif_mMRC - difference between the post- and the pre- Modified Medical Research Council Questionnaire; dif_QMS - difference between the pre- and the post-Quadriceps Muscle Strength; dif_6MWD - difference between the pre- and the post-Six-Minutes Walking Distance.

The median of the predicted distance to the Pareto front was 0.70 (interquartile range [0.62;0.85]) and the observed distance had a median of 0.63 (interquartile range [0.54;0.78]) (Figure 5).

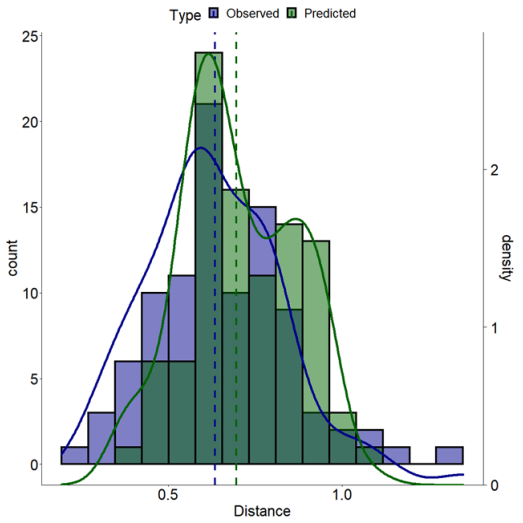

Figure 5. Histogram and density curve of observed (blue) and predicted (green) distance to Pareto front (dashed lines represent the median values) of patients with chronic obstructive pulmonary disease.

The Bland-Altman analysis showed that the distance predicted and observed to the Pareto front were statistically different (p-value<0.001), with a bias of 0.063 (95% confidence interval [0.027;0.099]). The upper limit of agreement (LOA) was 0.411 (95% confidence interval [0.349;0.472]) and the lower LOA was -0.284 (95% confidence interval [-0.346;-0.222]) (Figure 6).

The predicted responder patients had a distance to Pareto front equal to 0.43. The median distance of the non-responder patients was equal to 0.69 (interquartile range [0.63;0.85]). A statistically significant negative correlation was found between the distance and both mMRC (r=-0.39, p-value<0.001) and CAT

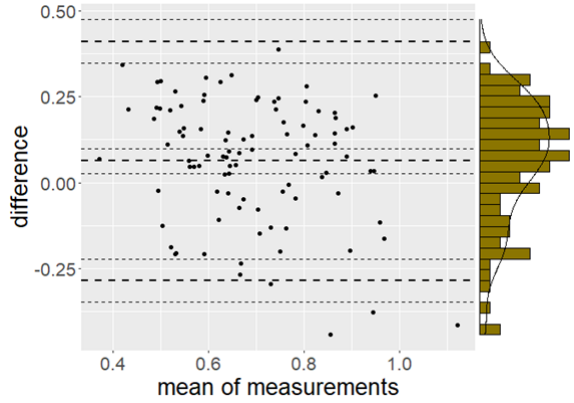

Figure 6. Bland-Altman plot for the measurements of the distance predicted and observed to the Pareto front of patients with chronic obstructive pulmonary disease. Bias and limits of agreement are represented with thick dashed lines, and the 95% confidence interval, with thin dashed lines. Histogram and density plot of the difference of the measurements are plotted on the right margin of the plot.

(r=-0.30, p-value=0.003). A statistically significant positive correlation was found between the distance and both PY (r=0.23, p-value=0.025) and QMS (r=0.36, p-value<0.001). The correlation between the predicted distance and 6MWD was not statistically significant (r=0.19, p-value=0.059). Further analysis of distances is presented in Table 5. In Section G of the appendix a correlogram including all the predictors, outcomes and distances is presented.

Table 5. Predicted distance to the Pareto front by variable.

	Med (IQR)	p-value*
Responder	0.43 -	- **
Non- responder	0.70 [0.63,0.85]	
PR	0.64 [0.59,0.70]	<0.001
No PR	0.90 [0.86,0.94]	
Female	0.61 [0.46,0.80]	0.012
Male	0.71 [0.63,0.87]	
No LTOT	0.67 [0.61,0.85]	0.018
LTOT	0.77 [0.74,0.90]	
EdLevel (>=5th)	0.67 [0.63,0.83]	0.401
EdLevel (<5th)	0.74 [0.60,0.89]	
Non-smoker	0.65 [0.46,0.76]	
Smoker	0.69 [0.63,0.86]	0.111
Ex-smoker	0.70 [0.62,0.86]	

Data are presented as median - Med (Interquartile Range - IQR).

PR – Pulmonary rehabilitation; LTOT – Long-term Oxygen Therapy; EdLevel – Educational Level.

* p-value obtained using Mann-Whitney-Wilcoxon and Kruskal-Wallis tests;

** not enough individuals in the Responder group to perform test.

The R Shiny web-based application created is divided in Seven tabs. Home, Help and About tabs welcome the user, provide general information and explain the different tabs, inputs, charts and results (Figure 7).

mMRC difference predicted, QMS difference predicted and 6MWT difference predicted tabs allows the

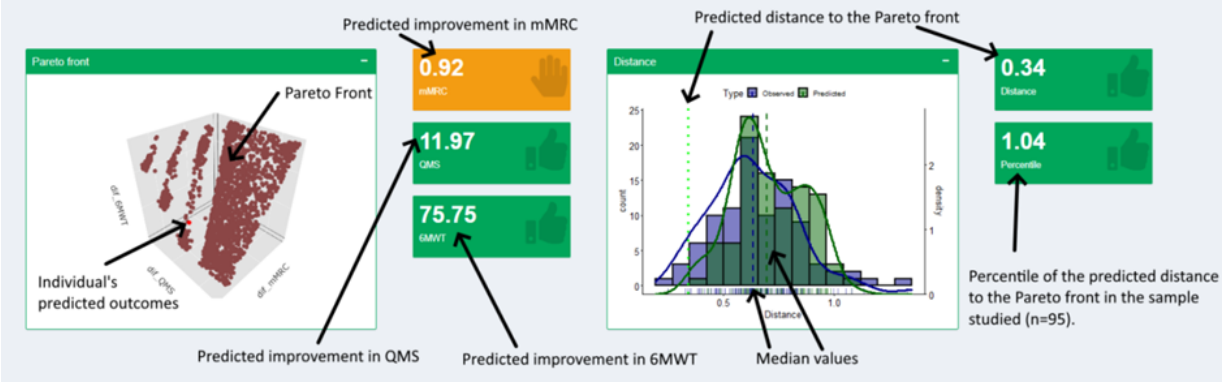

Figure 7. R Shiny web application. Help tab detail.

user to introduce the patients' predictor values required by each model using simple slider inputs. Each tab returns information about outcomes' estimates, confidence and prediction intervals, percentiles and distributions (Figure 8).

Figure 8. R Shiny web application. 6MWT difference predicted tab.

Distance predicted tab allows the user to introduce patients' predictor values required by all three models. It returns the three outcomes' estimates, the mean Euclidean distance to all points of the Pareto front and the correspondent percentile in the original data set. It is also possible to visually locate the predicted patient´s outcomes and the Pareto front in a 3-dimension plot or using 2-dimension projections.
Several configurations are possible including visual settings, MCID adjustments, changing chart types and information included (Figure 9).

Figure 9. R Shiny web application. Distance predicted tab.

4. Discussion

To the best of our knowledge, a study considering multiobjective optimization of PR outcomes in patients COPD has never been conducted. We proposed an optimized combination of three predictive models (exercise capacity-6MWT, dyspnoea-mMRC and isometric quadriceps muscle strength-QMS) of PR outcomes. We determined a measure based on the mean Euclidean distance to the Pareto front points that, along with the estimate given by the models are easily calculated with a built web-based, user-friendly interface which may help in PR prioritisation.

Our three predictive models confirmed that patients have statistically significant benefit when included in a PR programme compared to those who received standard pharmacological treatment. Furthermore, only 0.1% of patients would achieve an optimal outcome difference without attending PR. Corroborating our findings, patients included in a PR programme were statistically closer to the set of optimal solutions. Although the relationship between change in PR outcomes and baseline characteristics of patients has been extensively investigated, results are still conflicting [55-59]. The cross-validation measures used in our study indicate that the PR outcomes chosen could be partly explained by the socio-demographic and baseline clinical factors included in our models, corroborating the need for further investigation in this field to prioritise PR.

Environmental and genetic factors differences between male and female contribute to significant gender differences in COPD [60]. Our data set was mostly composed by male patients which is in accordance with the prevalence of COPD reported in the past [1], although recent findings suggest that this difference is narrowing due to the increased female smoking habits in developed, high-income countries [61]. Some studies suggest that the improvement in exercise capacity and quality of life after PR is independent of gender [62-63] but some found insufficient evidence to support or refute that conclusion [64]. Our 6MWD predictive model included female gender as a significant predictor. Because the other two models did not include this predictor, the Pareto front was only composed of female patients and the predicted female group was statistically closer to the set of optimal outcomes.

Smoking cessation is a key intervention for all COPD patients [1]. After three months, ex-smokers are expected to walk during six-minutes approximately more 35 metres than active smokers under the same baseline characteristics considered in the 6MWD predictive model. This difference is above the MCID established for the outcome [44]. As expected, all patients with COPD achieving optimal solutions were ex-smokers but the distance to the Pareto front was not statistically different between the three smoking

status groups. The unbalanced distribution of patients (1:4.7:1) and a low dimension of the non-smoker and actual smoker groups may justify this finding.

Despite smoking duration alone appeared to provide stronger risk estimates of COPD than the composite index of pack-years [65] we considered this last one as a significant predictor for the post-pre difference of mMRC and 6MWD. A higher baseline PY contributed to a greater improvement in the mMRC after PR whereas the opposite was observed for the QMS. This finding justifies the range in PY values obtained for the simulated patients in the Pareto front. Nonetheless, when considering the multiobjective optimization problem, a significant low positive correlation with the distance to the Pareto front was found, indicating that a lower value of baseline PY tends to approximate patients to an optimal solution.

A higher BMI has been associated with a slower decline in the rate of lung function in COPD [66] and has a protective effect against mortality [67]. A predicted increase of approximately 0.4 kgf in the post-pre difference of QMS for each unit increase in the baseline BMI was found in our study. Since BMI was only used as predictor in one model, the Pareto front simulated patients assumed all the values in our sample range but with a median value of approximately 37 kg/m^2. Both results are consistent with the "obesity paradox".

Previous studies with patients with COPD who attended PR programmes found that those with the lowest baseline 6MWD achieved the greatest benefit in exercise capacity [14,68]. Our findings support those since it was predicted a decrease of approximately 0.2 metres in the post-pre difference of 6MWD for each metre increase in the baseline distance covered in six-minutes. But the opposite was found when considering the other two outcomes. Models predicted a decrease of approximately 0.003 units in the post-pre difference of mMRC and an increase of 0.017 kgf in the post-pre difference of QMS for each metre increase in the baseline 6MWD. This conflicting predictor is reflected in the full range of baseline 6MWD values of the patients who achieved an optimal solution and by the non-significant correlation with the distance to those solutions.

Baseline QMS was considered a statistically significant predictor of its own value after PR as it was mMRC and 6MWD for its own models. Given the significant positive correlation of the distance with the QMS we can assume that patients with higher values at baseline will be further away from the optimal solution suggesting that a patient with greater baseline strength has less margin to improve.

Baseline mMRC and 6MWD were the only conflicting common predictors to all models. This resulted in simulated patients with optimal predicted outcomes that ranged all sample values for these variables except for the 0 score in mMRC. Although no correlation was found between distance and 6MWD, patients with COPD with higher baseline mMRC tended to be closer to the optimal solutions, suggesting that these patients not only have a greater margin to improve but also achieved this improvement.

When it comes to multiobjective optimization and from the decision maker's perspective, few solutions are desired for ease of decision [17]. We used R Shiny package [53] to create a web-based application that can be deployed centrally and can be easily shared as a URL. This kind of web-based application framework allows statistics, for instance, to put data insights into the hands of healthcare professionals, who will be responsible for the decision-making of PR prioritisation and commonly have no experience with using R nor the installation of R or RStudio.

Our study has several strengths. Study participants were followed prospectively; we used numerous validated questionnaires to assess baseline and change in function; the predictors included in our models moderately explain the variance in the post-pre difference of the considered outcomes and the R Shiny application is very user-friendly.

Some limitations also need to be acknowledged. We did not assessed changes in outcomes in those who did not complete the 12-week PR programme and hence our findings may not be generalizable to those who dropped out.

5. Conclusion

The mean Euclidean distance to all the Pareto front points along with the estimates given by our models are easily obtained using the R Shiny interface designed and may help healthcare professionals decide on the prioritisation of patients with COPD to PR programmes whenever the resources are limited. It is still up to the healthcare professional to carefully interpret the prediction and the distance to the optimal solutions and comprehensively evaluate each patient.

Software

Software in the form of R code is available in the GitHub repository https://github.com/jorgevazcabral/COPD_Optimization and in shinyapps.io platform https://jorgevazcabral.shinyapps.io/COPD/. Data sets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgement

All participants provided written informed consent to allow for their inclusion in the PR database for future research. Ethical approval was obtained from three different Ethics Committees for the studies: Administração Regional de Saúde do Centro (Ref. 73/2016; 85/2018), Centro Hospitalar do Baixo Vouga, EPE, Aveiro, Portugal (N/Ref 0863926) and from the Unidade de Investigação em Ciências da Saúde: Enfermagem (UICISA: E) of the Escola Superior de Enfermagem de Coimbra, Portugal (NºP517-08/2018). Data protection was ensured by the National Committee for Data Protection (no. 7295/2016) and followed the European regulation (https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN).
This work was supported by "Fundação para a Ciência e Tecnologia/MCTES", through the BioMath thematic line of the "Centro de Investigação e Desenvolvimento em Matemática e Aplicações (CIDMA), reference UIDB/04106/2020 and UIDP/04106/2020 and to the Institute of biomedicine (iBiMED, UIDB/04501/2020), using national funds (PIDDAC). We are also grateful to the funding agencies of the 3 research projects included in the study, i.e., LabexDRIIM (OHM-E/2018/ 1912), Fundo Europeu de Desenvolvimento Regional (FEDER) - Comissão Diretiva do Programa Operacional Regional do Centro and by Fundação para a Ciência e Tecnologia - FCT (SAICT-POL/23926/2016, PTDC/SAU-SER/28806/2017), and Programa Operacional Competitividade e Internacionalização (COMPETE), through COMPETE 2020 (POCI-01-0145-FEDER-016701 and POCI-01-0145-FEDER-007628). All authors critically revised the manuscript, ensured accuracy and integrity of the work, approved the final version to be published, and agree to be accountable for all aspects of the work. Declarations of interest: none.

REFERENCES

1. GOLD. (2021). Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: the GOLD science committee report 2021.

2. RABE, K. F. AND WATZ, H. (2017). Chronic obstructive pulmonary disease. Lancet 389(10082), 1931-1940. doi: 10.1016/S0140-6736(17)31222-9.

3. MATHERS, C. D. AND LONCAR, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. PLoS Medicine 3, 2011–2030. doi:10.1371/journal.pmed.0030442.

4. VOS, T. and others. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 396(10258), 1204-1222. doi:10.1016/S0140-6736(20)30925-9.

5. WORLD HEALTH ORGANIZATION. (2008). The global burden of disease: 2004 update. WHO Library Cataloguing-Publication Data.

6. SPRUIT, M. A., SINGH, S. J., GARVEY, C. and others. (2013). An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. American journal of respiratory and critical care medicine 188(8), e13-e64. doi:10.1164/rccm.201309-1634ST.

7. SPRUIT, M. A. AND WOUTERS, E. F. (2019). Organizational aspects of pulmonary rehabilitation in chronic respiratory diseases. Respirology 24(9), 838-843. doi:10.1111/resp.13512.

8. BOLTON, C.E., BEVAN-SMITH, E. F., BLAKEY, J. D. and others. (2013). British Thoracic Society guideline on pulmonary rehabilitation in adults: accredited by NICE. Thorax 68, ii1-ii30. doi:10.1136/thoraxjnl-2013-203808.

9. MACHADO, A., MARQUES, A. AND BURTIN, C. (2021). Extra-pulmonary manifestations of COPD and the role of pulmonary rehabilitation: a symptom-centered approach. Expert Review of Respiratory Medicine 15:1, 131-142. doi:10.1080/17476348.2021.1854737.

10. ROCHESTER, C. L., VOGIATZIS, I., HOLLAND, A.E. and others. (2015). An Official American Thoracic Society/European Respiratory Society Policy Statement: Enhancing Implementation, Use, and Delivery of Pulmonary Rehabilitation. Am J Respir Crit Care Med 192, 1373-1386. doi:10.1164/rccm.201510-1966ST.

11. WOUTERS, E. F. M., POSTHUMA, R., KOOPMAN, M., LIU, W. Y., SILLEN, M. J., HAJIAN, B., SASTRY, M., SPRUIT, M. A., AND FRANSSEN, F. M. (2020) An update on pulmonary rehabilitation techniques for patients with chronic obstructive pulmonary disease. Expert Review of Respiratory Medicine 14(2), 149-161. doi:10.1080/17476348.2020.1700796.

12. MARQUES, A., REBELO, P., PAIXÃO, C., MIRANDA, S., MACHADO, A., ALVES, A., SANTOS, L., PINHO, T., ALMEIDA, S., OLIVEIRA, A., CRUZ, J., JÁCOME, C., TAVARES, A., ANDRADE, L. AND VALENTE, C. (2019). Pulmonary rehabilitation closer to patients – feasibility and effectiveness study. European Respiratory Journal 54, PA565; doi:10.1183/13993003.congress-2019.PA565.

13. JOBSON J.D. (1991). Multiple Linear Regression. Applied Multivariate Data Analysis. Springer Texts in Statistics. Springer, New York, NY. doi:10.1007/978-1-4612-0955-3_4.

14. SCHROFF, P., HITCHCOCK, J., SCHUMANN, C., WELLS, J. M., DRANSFIELD, M. T. AND BHATT, S.P. (2017). Pulmonary Rehabilitation Improves Outcomes in Chronic Obstructive Pulmonary Disease Independent of Disease Burden. Ann Am Thorac Soc 14(1), 26-32. doi:10.1513/AnnalsATS.201607-551OC.

15. ROCHESTER, C.L. (2019). Patient assessment and selection for pulmonary rehabilitation. Respirology 24, 844– 853. doi:10.1111/resp.13616.

16. COELLO COELLO, C. A., GONZÁLEZ BRAMBILA, S., FIGUEROA GAMBOA, J., CASTILLO TAPIA, M. A. AND GÓMEZ, R. H. (2020). Evolutionary multiobjective optimization: open research areas and some challenges lying ahead. Complex Intell. Syst. 6, 221–236. doi: 10.1007/s40747-019-0113-4.

17. COELLO COELLO, C. A., VAN VELDHUIZEN, D. A. AND LAMONT, G. B. (2002). Evolutionary Algorithms for Solving Multi-Objective Problems. New York: Kluwer.

18. LIAO, X., LI, Q., YANG, X., YANG, X., ZHANG, W. AND LI, W. (2008). Multiobjective optimization for crash safety design of vehicles using stepwise regression model. Struct Multidisc Optim 35, 561–569. doi:10.1007/s00158-007-0163-x.

19. VARGAS-HÁKIM, G. A., MEZURA-MONTES, E. AND GALVÁN, E. (2020). Evolutionary Multi-Objective Energy Production Optimization: An Empirical Comparison. Math. Comput. Appl. 25, 32. doi: 10.3390/mca25020032.

20. DOUMPOS, M. AND ZOPOUNIDIS, C. (2020). Multi-objective optimization models in finance and investments. J Glob Optim 76, 243–244. doi:10.1007/s10898-019-00873-z.

21. ZHANG, W., CAO, K., LIU, S. AND HUANG, B. (2016). A multi-objective optimization approach for health-care facility location-allocation problems in highly developed cities such as Hong Kong. Computers, Environment and Urban Systems 59, 220-230. doi:10.1016/j.compenvurbsys.2016.07.001.

22. LU, Q., ZHU, X., WEI, D., BAI, K., GAO, J. AND ZHANG, R. (2019). Multi-Phase and Integrated Multi-Objective Cyclic Operating Room Scheduling Based on an Improved NSGA-II Approach. Symmetry 11, 599. doi:10.3390/sym11050599.

23. HERIS, S. M. K. AND KHALOOZADEH, H. (2011). Open- and Closed-Loop Multiobjective Optimal Strategies for HIV Therapy Using NSGA-II. IEEE Transactions on Biomedical Engineering 58(6), 1678–1685. doi:10.1109/tbme.2011.2110651.

24. COSTA-CARRAPIÇO, I., RASLAN, R. AND GONZÁLEZ, J.N. (2020). A systematic review of genetic algorithm-based multi-objective optimisation for building retrofitting strategies towards energy efficiency. Energy and Buildings, 210. doi: 10.1016/j.enbuild.2019.109690.

25. EMMERICH, M. T. M. AND DEUTZ, A. H. (2018). A tutorial on multiobjective optimization: fundamentals and evolutionary methods. Nat Comput 17, 585–609. doi: 10.1007/s11047-018-9685-y.

26. MARQUES, A., JÁCOME, C., REBELO, P., PAIXÃO, C., OLIVEIRA, A., CRUZ, J., FREITAS, C., RUA, M., LOUREIRO, H., PEGUINHO, C., MARQUES, F., SIMÕES, A., SANTOS, M., MARTINS, P., ANDRÉ, A., DE FRANCESCO, S., MARTINS, V., BROOKS, D. AND SIMÃO, P. (2019). Improving access to community-based pulmonary rehabilitation: 3R protocol for real-world settings with cost-benefit analysis. BMC Public Health 19(1), 676. doi:10.1186/s12889-019-7045-1.

27. CHARLSON, M., SZATROWSKI, T. P., PETERSON, J. AND GOLD, J. (1994) Validation of a combined comorbidity index. Journal of clinical epidemiology 47(11), 1245-51. doi:10.1016/0895-4356(94)90129-5.

28. GRAHAM, B. L., STEENBRUGGEN, I., MILLER, M. R. and others. (2019). Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. American journal of respiratory and critical care medicine 200(8), e70-e88. doi:10.1164/rccm.201908-1590ST.

29. BESTALL, J. C., PAUL, E. A., GARROD, R., GARNHAM, R., JONES, P. W. AND WEDZICHA, J. A. (1999). Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. Thorax 54(7), 581-6. doi:10.1136/thx.54.7.581.

30. CRISAFULLI, E. AND CLINI, E. M. (2010). Measures of dyspnea in pulmonary rehabilitation. Multidisciplinary respiratory medicine 2010;5(3):202-10. doi:10.1186/2049-6958-5-3-202.

31. GEORGE, F. (2013). Diagnóstico e Tratamento da Doença Pulmonar Obstrutiva Crónica. Direção Geral da Saúde 028/2011:1-15.

32. JONES, P. W., HARDING, G., BERRY, P., WILKLUND, I., CHEN, W-H. AND LEIDY, N. K. (2009). Development and first validation of the COPD Assessment Test. The European respiratory journal 34(3), 648-54. doi:10.1183/09031936.00102509.

33. JONES, P. W., TABBERER, M. AND CHEN, W.H. (2011). Creating scenarios of the impact of COPD and their relationship to COPD Assessment Test (CAT) scores. BMC pulmonary medicine 11, 42. doi:10.1186/1471-2466-11-42.

34. SMID, D. E., FRANSSEN, F. M., HOUBEN-WILKE, S. and others. (2017). Responsiveness and MCID Estimates for CAT, CCQ, and HADS in Patients With COPD Undergoing Pulmonary

Rehabilitation: A Prospective Analysis. Journal of the American Medical Directors Association 18(1), 53-58. doi:10.1016/j.jamda.2016.08.002.

35. ANDREWS, A. W., THOMAS, M. W. AND BOHANNON, R. W. (1996). Normative values for isometric muscle force measurements obtained with hand-held dynamometers. Physical therapy 76(3), 248-59.

36. BOHANNON, R. W. (1997). Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. Archives of physical medicine and rehabilitation 78(1), 26-32. doi:10.1016/s0003-9993(97)90005-8.

37. O'SHEA, S. D., TAYLOR, N. F. AND PARATZ, J. D. (2007). Measuring muscle strength for people with chronic obstructive pulmonary disease: retest reliability of hand-held dynamometry. Archives of physical medicine and rehabilitation 88(1), 32-6. doi:10.1016/j.apmr.2006.10.002.

38. SPRUIT, M.A., SILLEN, M. J., GROENEN, M. T., WOUTERS, E.F. AND FRANSSEN, F. M. (2013). New normative values for handgrip strength: results from the UK Biobank. Journal of the American Medical Directors Association 14(10), 775.e5-11. doi:10.1016/j.jamda.2013.06.013.

39. OLIVEIRA, A., REBELO, P., PAIXÃO, C., JÁCOME, C., CRUZ, J., MARTINS, V., SIMÃO, P., BROOKS, D. AND MARQUES, A. (2021). Minimal Clinically Important Difference for Quadriceps Muscle Strength in People with COPD following Pulmonary Rehabilitation. COPD: Journal of Chronic Obstructive Pulmonary Disease. doi:10.1080/15412555.2021.1874897.

40. OZALEVLI, S., OZDEN, A., ITIL, O. AND AKKOCLU, A. (2007). Comparison of the Sit-to-Stand Test with 6 min walk test in patients with chronic obstructive pulmonary disease. Respiratory medicine 101(2), 286-93. doi:10.1016/j.rmed.2006.05.007.

41. VAIDYA, T., CHAMBELLAN, A. AND DE BISSCHOP, C. (2017). Sit-to-stand tests for COPD: A literature review. Respiratory medicine 128, 70-77. doi:10.1016/j.rmed.2017.05.003.

42. VAIDYA, T., DE BISSCHOP, C., BEAUMONT, M. and others. (2016). Is the 1-minute sit-to-stand test a good tool for the evaluation of the impact of pulmonary rehabilitation? Determination of the minimal important difference in COPD. International journal of chronic obstructive pulmonary disease 11, 2609-16. doi:10.2147/COPD.S115439.

43. HOLLAND, A. E., SPRUIT, M.A., TROOSTERS, T. and others. (2014). An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease. The European respiratory journal 44(6), 1428-46. doi:10.1183/09031936.00150314.

44. SINGH, S.J., PUHAN, M.A., ANDRIANOPOULOS, V. and others. (2014). An official systematic review of the European Respiratory Society/American Thoracic Society: measurement properties of field walking tests in chronic respiratory disease. The European respiratory journal 44(6), 1447-78. doi:10.1183/09031936.00150414.

45. KILEY, J. P., SRI RAM, J., CROXTON T. L. AND WEINMANN, G. G. (2005). Challenges associated with estimating minimal clinically important differences in COPD — the NHLBI perspective. COPD 2, 43–46. doi:10.1081/copd-200050649.

46. CHANG, K. H. (2015). Chapter 19 - Multiobjective Optimization and Advanced Topics. Kuang-Hua Chang, e-Design,Academic Press, 1105-1173. doi: 10.1016/B978-0-12-382038-9.00019-3.

47. DEB, K., PRATAP, A., AGARWAL, S. AND MEYARIVAN, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2), 182-197. doi: 10.1109/4235.996017.

48. AHMADI, M., ARABI, M., HOAG, D. L., AND ENGEL, B. A. (2013). A mixed discrete-continuous variable multiobjective genetic algorithm for targeted implementation of nonpoint source pollution control practices. Water Resour. Res. 49, 8344– 8356. doi:10.1002/2013WR013656.

49. BROWNLEE, A. E. I. AND WRIGHT, J. A. (2015). Constrained, mixed-integer and multi-objective optimisation of building designs by NSGA-II with fitness approximation. Applied Soft Computing 33, 114-126. doi:10.1016/j.asoc.2015.04.010.

50. MERSMANN, O. (2020). mco: Multiple Criteria Optimization Algorithms and Related Functions. R package version 1.15.6.

51. RSTUDIO TEAM. (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA.

52. R CORE TEAM. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

53. CHANG, W., CHENG, J., ALLAIRE, J. J., SIEVERT, C., SCHLOERKE, B., XIE, Y., ALLEN, J., MCPHERSON, J., DIPERT, A., AND BORGES, B. (2021). shiny: Web Application Framework for R. R package version 1.6.0.

54. CILIONE, C., LORENZI, C., DELL'ORSO, D. and others. (2002). Predictors of change in exercise capacity after comprehensive COPD inpatient rehabilitation. Medical Science Monitor 8(11), CR740–CR745.

55. GARROD, R., MARSHALL, J., BARLEY, E. AND JONES, P. W. (2006). Predictors of success and failure in pulmonary rehabilitation. European Respiratory Journal 27(4), 788–794. doi:10.1183/09031936.06.00130605.

56. VAGAGGINI, B., COSTA, F., ANTONELLI, S. and others. (2009). Clinical predictors of the efficacy of a pulmonary rehabilitation programme in patients with COPD. Respiratory Medicine 103(8), 1224–1230. doi:10.1016/j.rmed.2009.01.023.

57. VAN RANST, D., OTTEN, H., MEIJER, J. W. AND VAN 'T HUL, A. J. (2011). Outcome of pulmonary rehabilitation in COPD patients with severely impaired health status. International Journal of Chronic Obstructive Pulmonary Disease 6, 647–657. doi:10.2147/COPD.S24579.

58. NIEDERMAN, M. S., CLEMENTE, P. H., FEIN, A. M. and others. (1991). Benefits of a multidisciplinary pulmonary rehabilitation program; improvements are independent of lung function. Chest 99(4), 798–804. doi:10.1378/chest.99.4.798.

59. ALTENBURG, W. A., DE GREEF, M. H., TEN HACKEN, N. H. AND WEMPE, J. B. (2012). A better response in exercise capacity after pulmonary rehabilitation in more severe COPD patients. Respiratory Medicine 106(5), 694–700. doi: 10.1016/j.rmed.2011.11.008.

60. ARYAL, S., DIAZ-GUZMAN, E., AND MANNINO, D. M. (2013). COPD and gender differences: an update. Translational Research 162(4), 208–218. doi:10.1016/j.trsl.2013.04.003.

61. NTRITSOS, G., FRANEK, J., BELBASIS, L., CHRISTOU, M. A., MARKOZANNES, G., ALTMAN, P., FOGEL, R., SAYRE, T., NTZANI, E. E. AND EVANGELOU, E. (2018). Gender-specific estimates of COPD prevalence: a systematic review and meta-analysis. Int J Chron Obstruct Pulmon Dis 13, 1507-1514. doi:10.2147/COPD.S146390.

62. DE TORRES, J. P., CASANOVA, C., COTE, C. G., AND CELLI, B. R. (2009). Women with chronic obstructive pulmonary disease: an emerging phenotype of the disease. Therapy 6(6), 821–830. doi:10.2217/thy.09.64.

63. MARCINIUK, D. D., BROOKS, D., BUTCHER, S., DEBIGARE, R. and others. (2010). Optimizing pulmonary rehabilitation in chronic obstructive pulmonary disease-practical issues: a Canadian Thoracic Society Clinical Practice Guideline. Can Respir J 17(4), 159-68. doi:10.1155/2010/425975.

64. ROBLES, P.G., BROOKS, D., GOLDSTEIN, R., SALBACH, N. AND MATHUR, S. (2014). Gender-associated differences in pulmonary rehabilitation outcomes in people with chronic obstructive pulmonary disease: a systematic review. J Cardiopulm Rehabil Prev 34(2), 87-97. doi: 10.1097/HCR.0000000000000018.

65. BHATT, S. P., KIM, Y. I., HARRINGTON, K. F. and others. (2018). Smoking duration alone provides stronger risk estimates of chronic obstructive pulmonary disease than pack-years. Thorax 73(5), 414-421. doi:10.1136/thoraxjnl-2017-210722.

66. SUN, Y., MILNE, S., JAW, J. E., YANG, C. X., XU, F., LI, X., OBEIDAT, M. AND SIN, D. D. (2019). BMI is associated with FEV1 decline in chronic obstructive pulmonary disease: a meta-analysis of clinical trials. Respir Res 20, 236. doi:10.1186/s12931-019-1209-5.

67. CAO, C., WANG, R., WANG, J., BUNJHOO, H., XU, Y. AND XIONG, W. (2012). Body Mass Index and Mortality in Chronic Obstructive Pulmonary Disease: A Meta-Analysis. PloS One 7(8), e43892. doi:10.1371/journal.pone.0043892.

68. ZUWALLACK, R. L., PATEL, K., REARDON, J.Z., CLARK, B.A. III AND NORMANDIN, E.A. (1991). Predictors of improvement in the 12-minute walking distance following a six-week outpatient pulmonary rehabilitation program. Chest 99, 805–808. doi:10.1378/chest.99.4.805.

Appendix

Section A Results from dif_mMRC linear regression.

Table I. Univariate linear regression results for the difference between Post- and Pre-pulmonary rehabilitation obtained in the modified British Medical Research Council Questionnaire of patients with chronic obstructive pulmonary disease.

Predictor	Coeff	p value	R^2	Adjusted R^2
Group (PR)	-0.78831	0.00003	0.170	0.162
Gender (female)	0.05263	0.82105	0.001	-0.010
Age	-0.00779	0.48468	0.005	-0.005
EdLevel (>=5th)	-0.10873	0.55990	0.004	-0.007
PY	-0.00297	0.16856	0.020	0.010
CCI	-0.12240	0.08260	0.032	0.022
LTOT (Yes)	0.13961	0.63110	0.002	-0.008
AECOPD	-0.07186	0.24521	0.014	0.004
BMI	0.00849	0.69955	0.002	-0.009
FEV1	0.23849	0.15718	0.021	0.011
FVC	0.11546	0.29006	0.012	0.001
mMRC	-0.44531	0.00000	0.273	0.265
1minSTS	0.00967	0.30170	0.011	0.001
6MWD	0.00016	0.84086	0.000	-0.010
CAT	-0.02794	0.01625	0.061	0.050
QMS	-0.00401	0.74622	0.001	-0.010
SmokSt (Ex-smoker)	-0.11250	0.66692	0.003	-0.018
SmokSt (Smoker)	-0.17500	0.59319	0.003	-0.018

Coeff - Regression coefficient; R^2 – Determination coefficient; COPD - Chronic Obstructive Cardiopulmonary Disease; PR – Pulmonary Rehabilitation; BMI – Body Mass Index; PY – Pack Years; CCI - Charlson Comorbidity Index; AECOPD - Acute Exacerbations of COPD; FEV_1 – Forced Expiratory Volume in one second; FVC - Forced Vital Capacity; CAT – COPD Assessment Test; QMS – Quadriceps Muscle Strength; mMRC - Modified Medical Research Council Questionnaire; 1minSTS - One-minute Sit To Stand test; 6MWD – Six-Minutes Walking Distance; EdLevel – Educational Level; SmokSt – Smoking Status; LTOT – Long-Term Oxygen Therapy; Dif – Difference between post- and pre-pulmonary rehabilitation

measures.

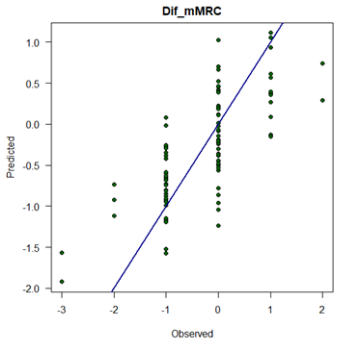

Figure I – Observed values versus the predicted values for the difference between Post- and Pre- pulmonary rehabilitation obtained in the modified British Medical Research Council Questionnaire of patients with chronic obstructive pulmonary disease.

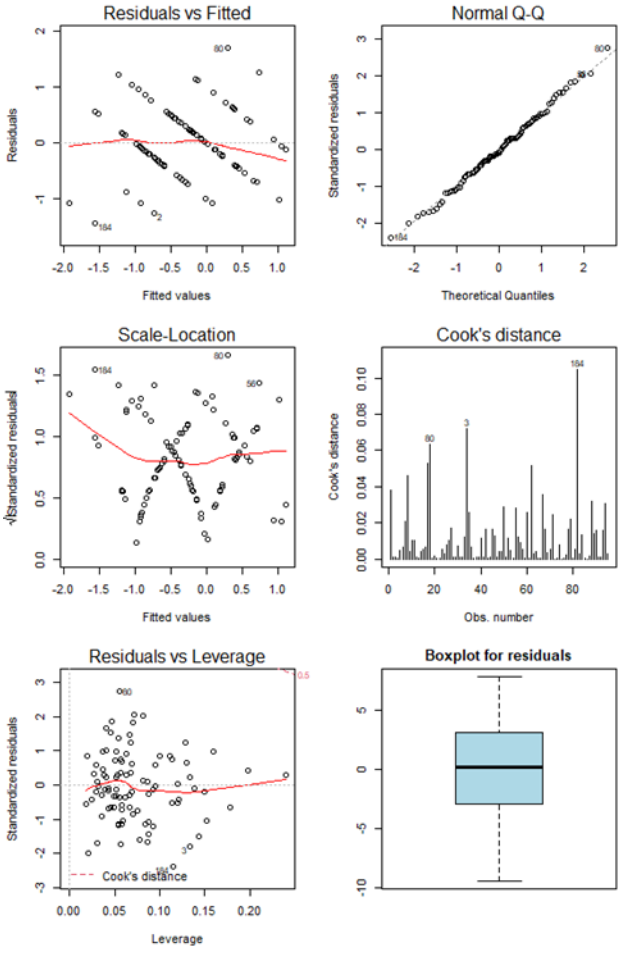

Figure II – Residual plots for the multiple linear regression of the difference between Post- and Pre- pulmonary rehabilitation

obtained in the modified British Medical Research Council Questionnaire of patients with chronic obstructive pulmonary disease.

Section B – Results from dif_QMS linear regression

Table II –Univariate linear regression results for the difference between Post- and Pre- pulmonary rehabilitation obtained in the Quadriceps muscle strength of patients with chronic obstructive pulmonary disease.

Predictor	Coeff	p value	R^2	Adjusted R^2
Group (PR)	1.75423	0.24901	0.014	0.004
Gender (female)	-1.35395	0.44876	0.006	-0.005
Age	0.01122	0.89599	0.000	-0.011
EdLevel (>=5th)	-0.77210	0.59053	0.003	-0.008
PY	0.01484	0.37198	0.009	-0.002
CCI	0.33646	0.53797	0.004	-0.007
LTOT (Yes)	-1.29643	0.56203	0.004	-0.007
AECOPD	-0.05610	0.90645	0.000	-0.011
BMI	0.03448	0.83852	0.000	-0.010
FEV1	1.69699	0.19104	0.018	0.008
FVC	0.46582	0.57982	0.003	-0.007
mMRC	-1.30767	0.05256	0.040	0.029
1minSTS	-0.02946	0.68322	0.002	-0.009
6MWD	0.00601	0.32982	0.010	0.000
CAT	-0.15603	0.08334	0.032	0.021
QMS	-0.31887	0.00055	0.121	0.112
SmokSt (Ex-smoker)	2.70969	0.17448	0.027	0.006
SmokSt (Smoker)	0.59875	0.80981	0.027	0.006

Coeff - Regression coefficient; R^2 – Determination coefficient; COPD - Chronic Obstructive Cardiopulmonary Disease; PR – Pulmonary Rehabilitation; BMI – Body Mass Index; PY – Pack Years; CCI - Charlson Comorbidity Index; AECOPD - Acute Exacerbations of COPD; FEV_1 – Forced Expiratory Volume in one second; FVC - Forced Vital Capacity; CAT – COPD Assessment Test; QMS – Quadriceps Muscle Strength; mMRC - Modified Medical Research Council Questionnaire; 1minSTS - One-minute Sit To Stand test; 6MWD – Six-Minutes Walking Distance; EdLevel – Educational Level; SmokSt – Smoking Status; LTOT – Long-Term Oxygen Therapy; Dif – Difference between post- and pre-pulmonary rehabilitation

measures.

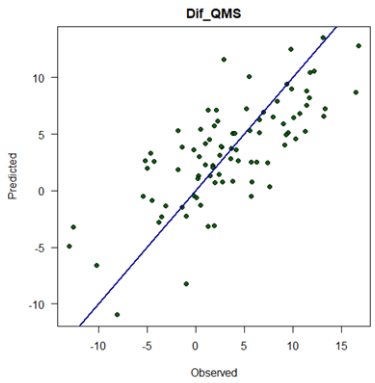

Figure III – Observed values versus the predicted values for the difference between Post- and Pre- pulmonary rehabilitation obtained in the quadriceps muscle strength of patients with chronic obstructive pulmonary disease.

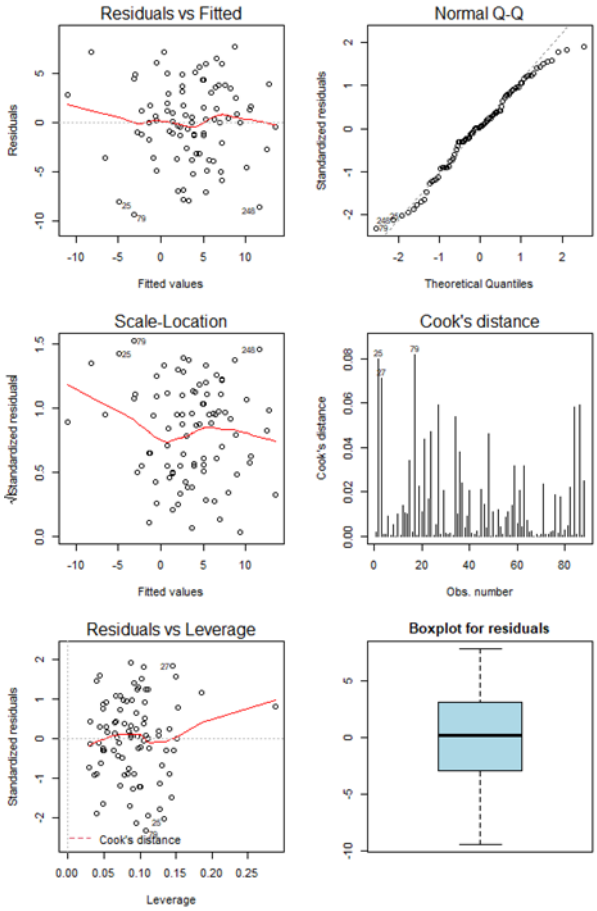

Figure IV – Residual plots for the multiple linear regression of the difference between Post- and Pre- pulmonary rehabilitation obtained in the quadriceps muscle strength of patients with chronic obstructive pulmonary disease.

Section C – Results from dif_6MWD linear regression

Table III – Univariate linear regression results for the difference between Post- and Pre- pulmonary rehabilitation obtained in the Six minutes walking distance of patients with chronic obstructive pulmonary disease.

Predictor	Coeff	p value	R^2	Adjusted R^2
Group (PR)	45.51072	0.00040	0.127	0.117
Gender (female)	16.99789	0.27400	0.013	0.002
Age	-0.69506	0.35140	0.009	-0.001
EdLevel (>=5th)	1.50858	0.90398	0.000	-0.011
PY	-0.20206	0.16152	0.021	0.010
CCI	-1.89079	0.69121	0.002	-0.009
LTOT (Yes)	-24.64617	0.20401	0.017	0.007
AECOPD	2.71570	0.51312	0.005	-0.006
BMI	1.80710	0.21819	0.016	0.006
FEV1	-5.65108	0.61846	0.003	-0.008
FVC	-4.19347	0.56706	0.004	-0.007
mMRC	3.74903	0.52708	0.004	-0.006
1minSTS	-0.37861	0.54674	0.004	-0.007
6MWD	-0.10022	0.06019	0.037	0.027
CAT	0.79166	0.31559	0.011	0.000
QMS	-0.80194	0.33238	0.010	-0.001
SmokSt (Ex-smoker)	18.37881	0.29208	0.016	-0.006
SmokSt (Smoker)	5.11475	0.81444	0.016	-0.006

Coeff - Regression coefficient; R^2 – Determination coefficient; COPD - Chronic Obstructive Cardiopulmonary Disease; PR – Pulmonary Rehabilitation; BMI – Body Mass Index; PY – Pack Years; CCI - Charlson Comorbidity Index; AECOPD - Acute Exacerbations of COPD; FEV_1 – Forced Expiratory Volume in one second; FVC - Forced Vital Capacity; CAT – COPD Assessment Test; QMS – Quadriceps Muscle Strength; mMRC - Modified Medical Research Council Questionnaire; 1minSTS - One-minute Sit To Stand test; 6MWD – Six-Minutes Walking Distance; EdLevel – Educational Level; SmokSt

– Smoking Status; LTOT – Long-Term Oxygen Therapy; Dif – Difference between post- and pre-pulmonary rehabilitation measures.

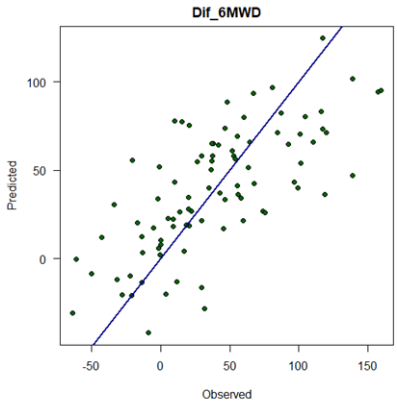

Figure V – Observed values versus the predicted values for the difference between Post- and Pre-pulmonary rehabilitation obtained in the Six minutes walking distance of patients with chronic obstructive pulmonary disease.

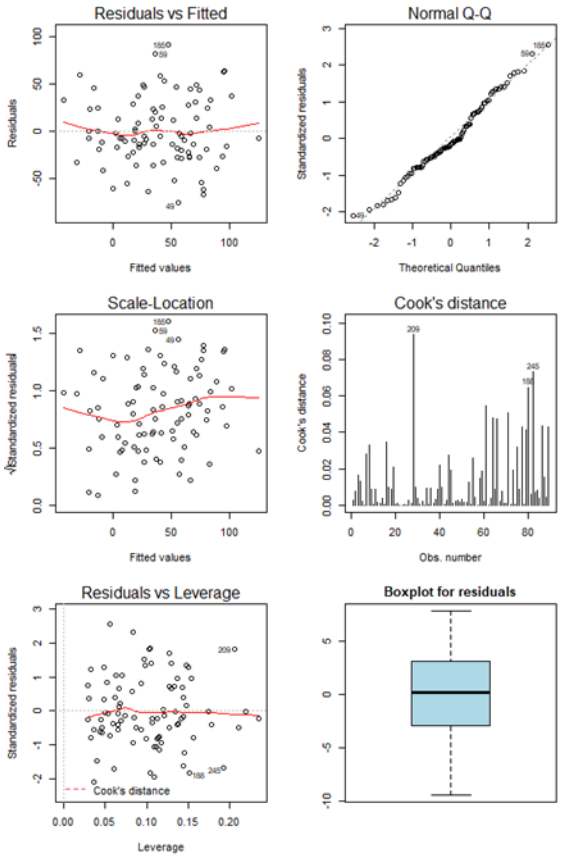

Figure VI – Residual plots for the multiple linear regression of the difference between Post- and Pre- pulmonary rehabilitation obtained in the Six minutes walking distance of patients with chronic obstructive pulmonary disease.

Section D – Characteristics of the individuals in the Pareto front coloured by outcome

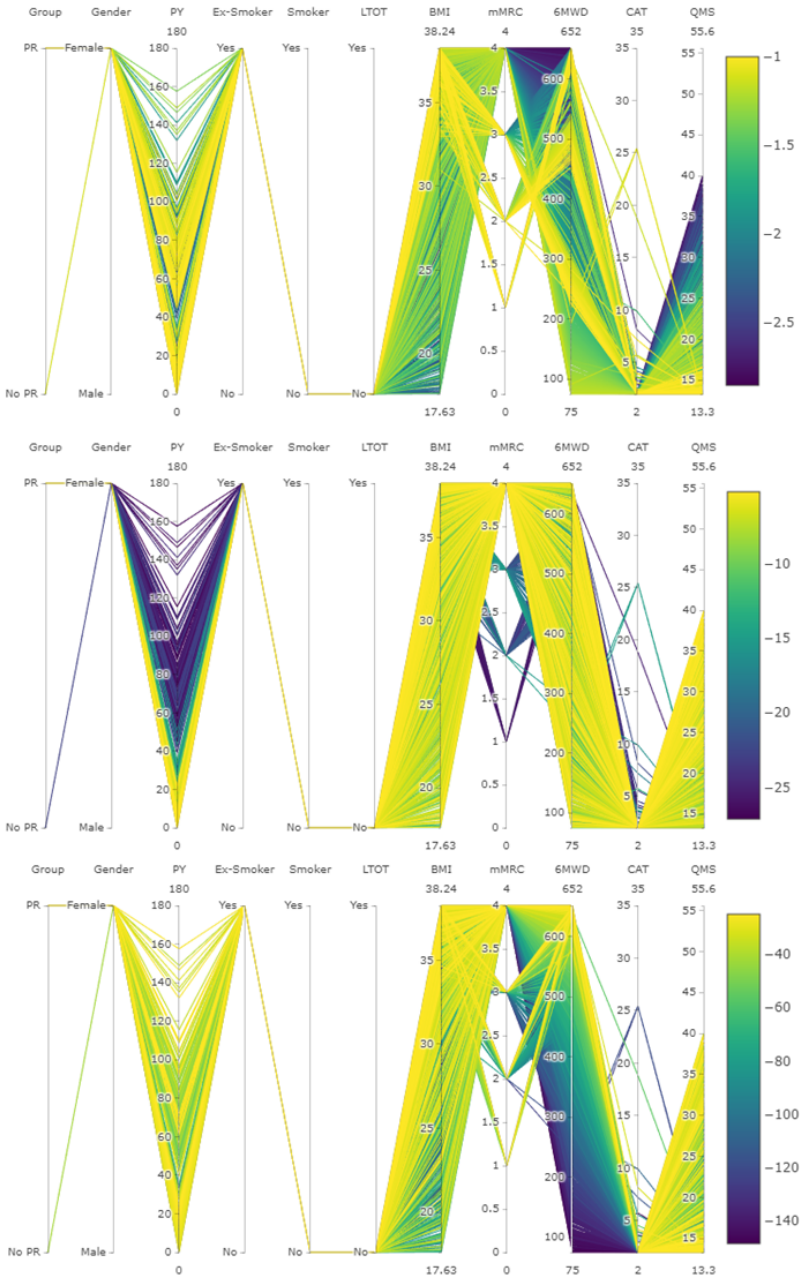

Figure VII – Characteristics of the patients with chronic obstructive pulmonary disease in the Pareto front. Lines are coloured according to the outcome (dif_mMRC – top, dif_QMS - middle; dif_6MWD – bottom). PR – Pulmonary Rehabilitation; BMI – Body Mass Index; PY – Pack Years; CAT – Chronic Obstructive Pulmonary Disease Assessment Test; QMS – Quadriceps Muscle Strength; mMRC - Modified Medical Research Council Questionnaire; 6MWD – Six-Minutes Walking Distance; LTOT – Long-Term Oxygen Therapy.

Section E – Pareto front projections

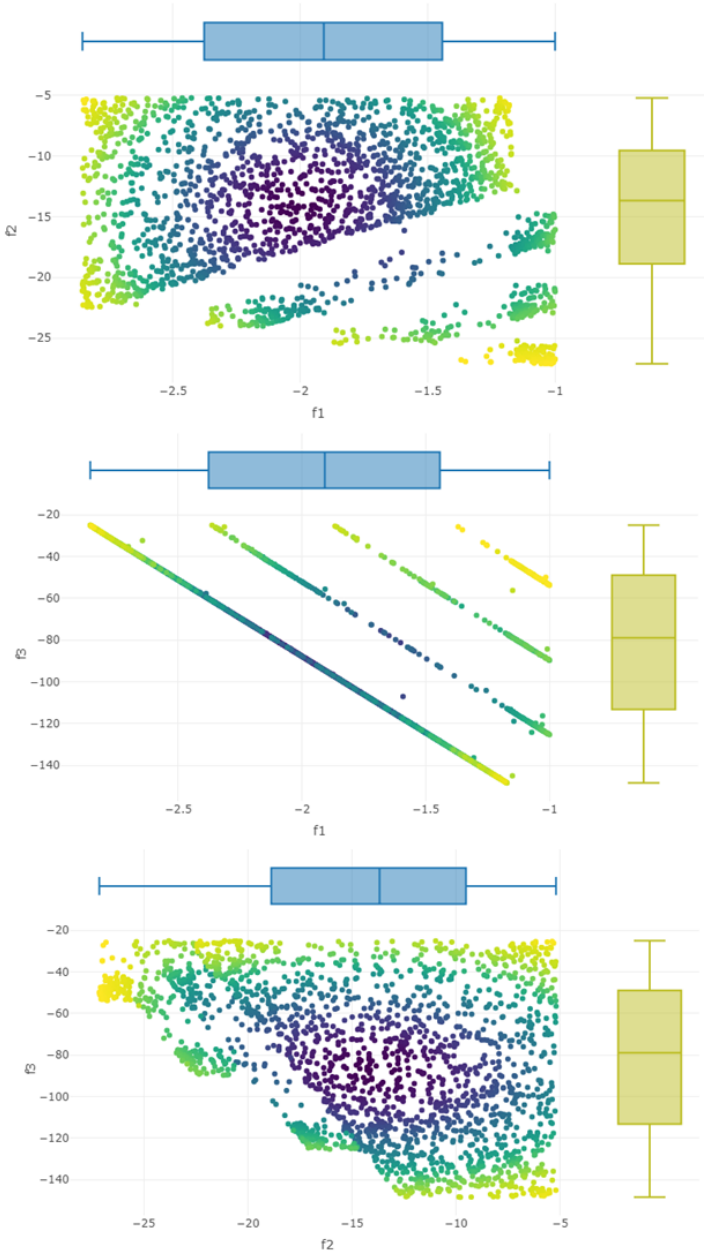

Figure VIII – Pareto front projections for pulmonary rehabilitation outcomes in patients with chronic obstructive pulmonary disease. Points are coloured according to the mean Euclidean distance to the remaining points. Colour scale ranges from Yellow, that represents a greater distance, to Purple, that represents a smaller distance. f1 (Dif_mMRC) – difference between the post- and the pre- Modified Medical Research Council Questionnaire; f2 (Dif_QMS) - difference between the pre- and the post-Quadriceps Muscle Strength; f3 (Dif_6MWD) - difference between the pre- and the post- Six-Minutes Walking Distance.

Section F – Pareto front, predicted and observed outcomes projections

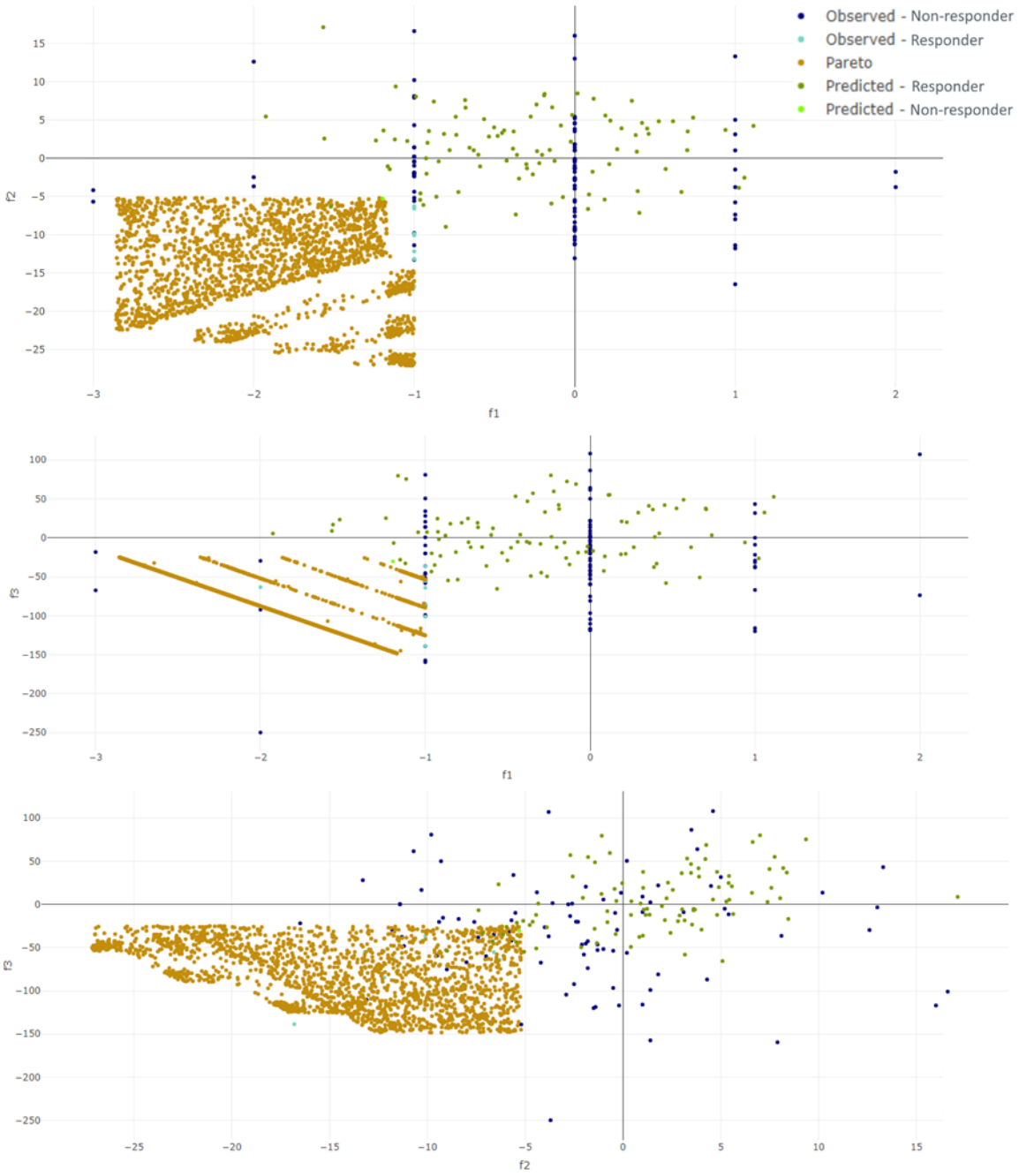

Figure IX – Pareto front, predicted and observed projections for pulmonary rehabilitation outcomes of patients with chronic obstructive pulmonary disease. f1 (Dif_mMRC) – difference between the post- and the pre- Modified Medical Research Council Questionnaire; f2 (Dif_QMS) - difference between the pre- and the post-Quadriceps Muscle Strength; f3 (Dif_6MWD) - difference between the pre- and the post- Six-Minutes Walking Distance.

Section G – Correlogram

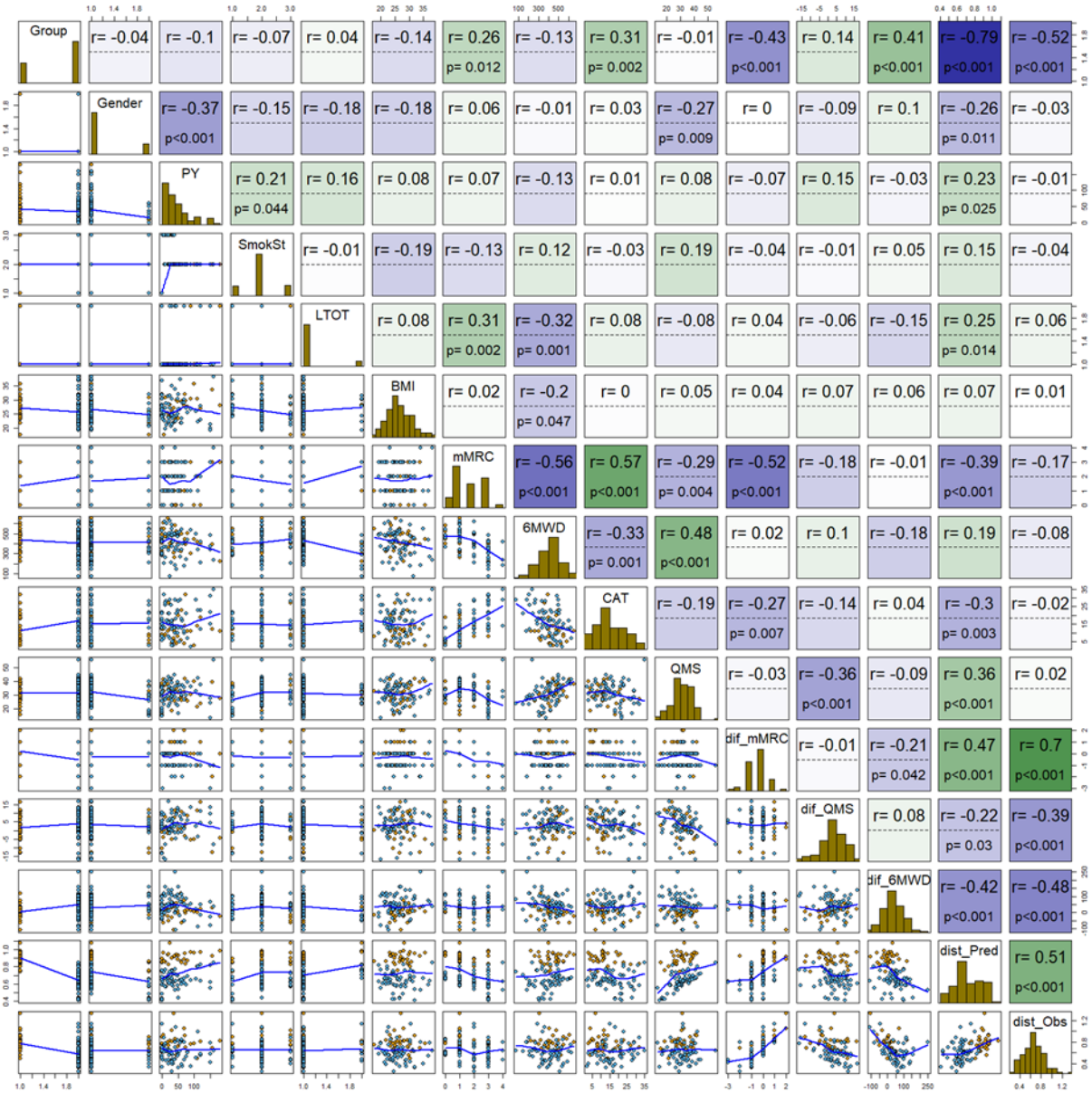

Figure X – Correlogram. r – Spearman's correlation coefficient. COPD - Chronic Obstructive Cardiopulmonary Disease; PR – Pulmonary Rehabilitation; BMI – Body Mass Index; PY – Pack Years; CAT – COPD Assessment Test; QMS – Quadriceps Muscle Strength; mMRC - Modified Medical Research Council Questionnaire; 6MWD – Six-Minutes Walking Distance; SmokSt – Smoking Status; LTOT – Long-Term Oxygen Therapy; Dif – Difference between post- and pre-pulmonary rehabilitation outcome measures; dist_Pred – Mean Euclidean distance for the predicted outcomes; Obs - Mean Euclidean distance for the observed outcomes.