Universidade de Aveiro
2021

JORGE VAZ RAMOS RODRIGUES DE CABRAL

COMPORTAMENTO TEMPORAL DA DPOC E INFLUÊNCIA DO CONFINAMENTO IMPOSTO PELA COVID-19: COMPARAÇÃO DE MÉTODOS DE SELEÇÃO DE VARIÁVEIS

BEHAVIOUR OF COPD OUTCOME MEASURES OVER TIME AND INFLUENCE OF THE COVID-19 LOCKDOWN: COMPARISON OF FIXED-EFFECTS SELECTION METHODS

Universidade de Aveiro
2021

JORGE VAZ RAMOS RODRIGUES DE CABRAL

COMPORTAMENTO TEMPORAL DA DPOC E INFLUÊNCIA DO CONFINAMENTO IMPOSTO PELA COVID-19: COMPARAÇÃO DE MÉTODOS DE SELEÇÃO DE VARIÁVEIS

BEHAVIOUR OF COPD OUTCOME MEASURES OVER TIME AND INFLUENCE OF THE COVID-19 LOCKDOWN: COMPARISON OF FIXED-EFFECTS SELECTION METHODS

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Estatística Médica, realizada sob a orientação científica da Doutora Vera Mónica Almeida Afreixo, Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro e da Doutora Alda Sofia Pires de Dias Marques, Professora Coordenadora da Universidade de Aveiro

o júri

presidente	Prof. Dr. Bruno Miguel Alves Fernandes do Gago professor auxiliar convidado do Departamento de Ciências Médicas da Universidade de Aveiro
arguente	Prof. Dr. Tiago José Andrade Gregório professor auxiliar convidado da Faculdade de Medicina da Universidade do Porto
orientador	Profª. Drª. Vera Mónica Almeida Afreixo professora auxiliar do Departamento de Matemática da Universidade de Aveiro

agradecimentos

Em primeiro lugar, e sempre nesse lugar cimeiro de onde nunca saíram e jamais sairão, quero manifestar a minha profunda e sentida gratidão à minha família, em especial à Daniela, ao Eduardo e à Teresa.
Não estaria hoje a escrever estes agradecimentos se não fosse todo o apoio dado e a confiança depositada em mim pela Profª. Dra. Vera Afreixo ao longo dos dois últimos anos. Por isso e por tudo o mais, obrigado.
À Profª. Dra. Alda Marques enalteço e agradeço a calma manifestada ao me ouvir e as sábias sugestões transmitidas nos últimos seis meses.
Não posso deixar de agradecer os valiosos contributos prestados pela Dra. Vânia Rocha e Sara Souto-Miranda na interpretação clínica dos resultados obtidos.

palavras-chave

COVID-19, dados longitudinais, Doença Pulmonar Obstrutiva Crónica, modelos lineares de efeitos mistos, seleção de variáveis.

resumo

Modelar um determinado resultado é desafiante e recorre-se habitualmente à recolha de diversas variáveis. Contudo, desconhecem-se ainda os métodos estatísticos apropriados para a seleção de variáveis importantes e com significado, nomeadamente em dados longitudinais.
Dados longitudinais podem ser agrupados e definem trajetórias alteráveis por inúmeros fatores, alguns deles inesperados.
Identificar as trajetórias individuais de determinados resultados em fases iniciais de uma doença, bem como os potenciais fatores de risco, deveria ser prioritário uma vez que esse conhecimento pode conduzir ao desenvolvimento de tratamentos individualizados e resultar em intervenções efetivas.
A doença pulmonar obstrutiva crónica é uma doença prevenível e progressiva e indivíduos com esta doença poderiam beneficiar com a identificação desses fatores de risco e do comportamento da doença ao longo do tempo.
Esta dissertação teve como objetivos comparar diferentes métodos de seleção de variáveis, em dados longitudinais, baseados em algoritmos de regressão, nomeadamente, random forest, Boruta, extreme gradient boosting, estimação com penalização L-1 e eliminação automática. Também pretendemos descrever o efeito provocado pelo confinamento decorrente da pandemia de COVID-19 no teste de sentar e levantar em 1 minuto, na força de preensão manual e no teste de avaliação do impacto da doença pulmonar obstrutiva crónica. Finalmente, explorámos os fatores que influenciam o comportamento do teste de sentar e levantar em 1 minuto ao longo de seis meses em indivíduos com doença pulmonar obstrutiva crónica.
O método de eliminação automática foi consistente na seleção de variáveis que produziram modelos lineares de efeitos mistos com menores valores de critério de informação de Akaike. O período de confinamento não teve efeito estatisticamente significativo no teste de sentar e levantar em 1 minuto nem na força de preensão manual. No entanto, foi observado um efeito negativo no impacto da doença. Foi também observada uma pior evolução dos resultados do teste de sentar e levantar em 1 minuto, ao longo do tempo, em indivíduos com doença pulmonar obstrutiva crónica mais velhos e com maior carga tabágica.

keywords

Chronic obstructive pulmonary disease, COVID-19, feature selection, linear mixed-effects models, longitudinal data.

abstract

Modelling a certain outcome is challenging and it is common practice to collect several features in that attempt. Nevertheless, the appropriate statistical methods to select important and meaningful features are still unknown, namely under repeated measurements

Longitudinal data can be grouped in forming trajectories that can be altered by countless factors, some of them unexpected. Identifying individuals' outcome trajectories at early stage of illness, as well as potential risk factors should be of high priority since this knowledge can guide to the development of individually tailored treatment and result in effective interventions. Chronic obstructive pulmonary disease is a progressive and preventable disease and people with this disease could benefit from the identification of such risk factors and over time behaviour.

In this dissertation we aimed to compare different feature selection methods based on regression algorithms, namely, random forest, Boruta, extreme gradient boosting, L-1 penalized estimation and automatic backward selection, adapted to longitudinal data. We also aimed to describe the effect of the Coronavirus disease 2019 lockdown on the one-minute sit-to-stand test, handgrip muscle strength and chronic obstructive pulmonary disease assessment test behaviour. We finally aimed to explore the factors influencing the behaviour of the one-minute sit-to-stand test over a six-month period in people with chronic obstructive pulmonary disease.

We showed that the automatic backward elimination of features was consistent when it came to select statistically relevant features to be included in linear mixed-effects models with the lowest values of Akaike information criterion. The COVID-19 lockdown period seemed to have had no effect in the one-minute sit-to-stand test and handgrip muscle strength behaviour but a negative effect in the impact of the disease was observed. Also, an increase of the smoking load or age seems to lead to a worse evolution in the one-minute sit-to-stand test results over time in people with chronic obstructive pulmonary disease.

Contents

List of Abbreviations

1minSTS	One-Minute Sit-To-Stand Test
A1	Assessment 1
A5	Assessment 5
AECOPD	Acute Exacerbation of COPD
AIC	Akaike information criterion
BMI	Body Mass Index
BPAAT	Brief Physical Activity Assessment Tool
CAT	COPD Assessment Test
CCI	Charlson Comorbidity Index
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
COVID-19	Coronavirus disease 2019
DALYs	Disability-adjusted life-years
EMM	Estimated marginal mean
ESSUA	School of Health Sciences
FEV_1	Forced Expiratory Volume in 1 Second
FVC	Forced Vital Capacity
GENIAL	GENetic and clinIcAL markers in COPD trajectory
GOLD	Global Initiative for Chronic Obstructive Lung Disease
HMS	Handgrip muscle strength
IQR	Interquartile range
Lab 3R	Respiratory Research and Rehabilitation Laboratory
LMM	Linear mixed-effects model
LTOT	Long-term oxygen therapy
MCID	Minimal Clinical Important Difference
mMRC	Modified Medical Council Dyspnoea Scale
MSE	Mean-square error
NIV	Non-invasive ventilation
OOB	Out-of-bag
PR	Pulmonary Rehabilitation
PRIME	Pulmonary Rehabilitation Innovation and Microbiota in Exacerbations of COPD
R^2	Coefficient of determination
RF	Random Forest
RMSE	Root-mean-square error
SD	Standard deviation
SE	Standard Error
SGQR	St. George's Respiratory Questionnaire
XGB	Extreme Gradient Boosting

List of Figures

List of Tables

1. INTRODUCTION

It is not uncommon to collect many predictors to model a certain outcome, but it is of great interest to find important and significant features among that set of candidate variables. When that outcome is measured repeatedly through different time points, we are in the presence of longitudinal data. Longitudinal data require special statistical methods because observations from the same individual are not independent, and variables tend to be correlated.

Despite many previous studies investigating how predictors affect an outcome, the question of choosing appropriate statistical methods to select important and meaningful predictors under repeated measurements is still not clearly answered (E. I. George, 2000). Some of the more common selection methods are time consuming and sometimes unreliable for marking inferences.

One of the characteristics of longitudinal data is that it can be grouped in forming trajectories. Identifying individuals' outcome trajectories at early stages of an illness and potential risk factors associated with a poor outcome trajectory are of high priority because this knowledge can guide the development of individually tailored treatments and effective interventions that potentially alter the course of the disease. (Hall et al., 2019)

Chronic obstructive pulmonary disease (COPD) is a progressive and preventable disease (Global Initiative for Chronic Obstructive Lung Disease, 2021) associated with enormous burden not just for individuals but also for their families, society and economy. People with this disease could benefit from the identifications of such risk factors and over time behaviour to manage the disease early and optimise outcomes.

In this dissertation we briefly introduce COPD, characterize longitudinal data and the methodology to analyse it, namely the linear mixed-effects models, and describe algorithms that enables us to select the variables to include in those models. We then present the dataset used to perform two studies, the main focus of this dissertation. The first study aimed to compare different feature selection methods and describe the effect of the COVID-19 lockdown on the one-minute sit-to-stand test, handgrip muscle strength and COPD Assessment test behaviour in people with COPD. The second study aimed to describe the one-minute sit-to-stand behaviour over a six-month period and explore the factors influencing this behaviour in people with COPD.

2. CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Chronic obstructive pulmonary disease is a progressive, treatable and highly preventable disease. It was estimated that more than 12% of the general population of the world suffers from COPD. It kills more than 3 million people worldwide every year (Rabe & Watz, 2017; Varmaghani et al., 2019) and accounts for a substantial and increasing individual, economic and societal burden (World Health Organization, 2008). COPD was projected to be the seventh leading cause of disability-adjusted life-years (DALYs) by 2030 (Mathers & Loncar, 2006) 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 (GBD 2019 Diseases and Injuries Collaborators et al., 2020).

COPD 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” (Global Initiative for Chronic Obstructive Lung Disease, 2021). The chronic airflow limitation that is characteristic of COPD is caused by a mixture of small airways disease and parenchymal destruction that may evolve at different rates over time (Global Initiative for Chronic Obstructive Lung Disease, 2021).

Dyspnoea, also known as “shortness of breath”, is the subjective sensation of running out of air and of not being able to breathe fast enough (Hashmi et al., 2021). This sense of uncomfortable breathing, comprised of various sensations of varying intensity, is the most characteristic symptom of COPD (Global Initiative for Chronic Obstructive Lung Disease, 2021). COPD diagnosis should be considered in any patient who has dyspnoea, chronic cough or sputum production, and/or a history of exposure to risk factors for the disease. Nevertheless, spirometry, a widely available and reproducible physiological test that measures the maximal volume of air that an individual can inspire and expire with maximal effort, is required to make the diagnosis in this clinical context.

To date studies indicate that morbidity and prevalence due to COPD increases at steady rates with age (Jarad, 2011; Varmaghani et al., 2019).

Although cigarette smoking is the leading and most well studied COPD environmental risk factor for COPD (GBD 2019 Diseases and Injuries Collaborators et al., 2020) it is not the only risk factor. For instance, sex, genetics and comorbidities playing an important role on the disease progression (Rennard & Drummond, 2015). Also, body mass index (BMI) is associated with the rate of lung function decline in COPD. Compared to normal BMI, low BMI is associated with faster, and high BMI is associated with slower, forced expiratory

volume in 1 second (FEV_1) decline. Overweight or obese have a protective effect against mortality, sometimes called the “obesity paradox” (Cao et al., 2012; Sun et al., 2019).

Physicians should refer COPD diagnosed patients to pulmonary rehabilitation (PR), an evidence-based, non-pharmacological multidisciplinary intervention, as patients with COPD remain symptomatic despite optimisation of pharmacological treatment. PR is defined as “a comprehensive intervention based on thorough patient assessment followed by patient-tailored therapies that include, but are 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 behaviours” (Spruit et al., 2013).

Patient-centred outcomes have historically been used for patient assessment and measurement of change or impact of PR in chronic respiratory disease. The strongest evidence of impact from PR has been for improvement in symptoms, exercise performance/ functional activities and quality of life since they are highly meaningful to individuals with COPD (Souto-Miranda & Marques, 2018; Spruit et al., 2013). Also, studies have focused on accurately defining and describing relevant outcomes and their measurement and interpretation. Analyses of outcomes have included descriptions of relevant change, such as the minimal clinically important difference (MCID). The MCID 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 (Kiley et al., 2005).

3. LONGITUDINAL DATA

Measuring change over time is not possible with cross-sectional data (Willett, 1989). To be able to describe and interpret change, longitudinal data, also referred as panel data is mandatory. Longitudinal data can be defined as “measurements or observations taken from multiple subjects repeatedly over time” (Funatogawa & Funatogawa, 2018).

To characterize changes in an outcome of interest over time, as well as the potential relationships between risk factors and the outcome, is of major importance that is the reason why longitudinal studies are a frequent practice. Although several benefits of longitudinal studies can be pointed out, such as the alleviation of the recall bias or the opportunity to observe patterns of change, they are not without a cost. There is, for instance, the risk of bias due to incomplete follow-up, or drop-out of study participants and the need to perform statistical analysis using methods that can account for the intra-subject correlation of measurements. From those methods, we will briefly describe the linear mixed-effects models (LMMs) (Laird & Ware, 1982).

The term mixed model refers to the use of both fixed and random effects in the same analysis. Fixed effects have levels that are of primary interest and would be used again if the experiment was repeated. Random effects have levels that are not of primary interest but rather are a random selection from a larger set of levels (Mallinckrodt & Lipkovich, 2017).

Let $Y_i = \left(Y_{i1}, Y_{i2}, \dots, Y_{in_i}\right)^T$ be the vector of the response corresponding to the *i*th $(i = 1, \dots, N)$ subject measured from 1 to n_i occasions. Y_{ij} is the *j*th measurement. Linear mixed-effects models are expressed by

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i\,, \quad \varepsilon_i = \left(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in_i}\right)^T \tag{3.1}$$

where β is a $p \times 1$ vector of unknown fixed effects parameters, X_i is a known $n_i \times p$ design matrix for fixed effects, b_i is a $q \times 1$ vector of unknown random effects parameters, Z_i is a known $n_i \times q$ design matrix for random effects, and ε_i is a $n_i \times 1$ vector of random errors.

It is assumed that b_i and ε_i are both independent across subjects and independently follow a multivariate normal distribution with the mean zero vector and variance covariance matrices *G* ($q \times q$ square matrix) and R_i ($n_i \times n_i$ square matrix), respectively. Responses from different subjects are assumed to be independent.

Different LMMs can be defined. We can, for instance, consider those who include means at each time point with a random intercept or time trend models with a random intercept and a random slope (Funatogawa & Funatogawa, 2018).

4. FEATURE SELECTION

4.1. MACHINE LEARNING

The problem of automatically searching patterns in data using computer algorithms and, with the use of these regularities, to take actions is a fundamental one and has a long successful history (Bishop, 2006). Machine learning is the scientific discipline that focuses on how computers learn from data. It seeks to learn relationships from data, with emphasis on efficient computing algorithms. The types of learning methods used by computers are conveniently subclassified into categories such as supervised learning and unsupervised learning. Applications in which the training data comprises input characteristics (also designated as features, independent variables or predictors) along with their corresponding outcome (also designated as response, result or dependent variable) are known as supervised learning problems (Bishop, 2006). Cases in which the aim is to assign input features to one of a finite number of discrete categories, are called classification problems. If the desired outcome consists of one or more continuous variables, then the task is called regression. On the other hand, with unsupervised learning, there are input features, but no supervising outcome. Nevertheless, we can learn relationships and structure from such data (Gareth et al., 2013).

4.2. REGRESSION

The goal of regression is build models that can predict the value of one or more continuous outcomes given the value of a n-dimensional vector *x* of input features (Bishop, 2006).
One of the most used supervised technique for relating a set of variables is the multiple linear regression model (Jobson, 1991) and we will use to establish some important definitions used throughout this thesis.
It takes the form

$$Y_{\mathrm{i}} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i \ , \ \epsilon_i \sim N\left(0, \sigma_i^2\right) \tag{4.1}$$

where Y_i represents a quantitative outcome, x_i represents the i^{th} predictor from the selected p, β_i quantifies the association between that feature and the outcome and ϵ_i is a normally

distributed, mean-zero and constant variance (σ_i^2), random error term, independent of Y_i. β_0 is called the intercept term and β_i is the slope associated to the *i*th predictor.

The regression constant parameters or coefficients $\beta_0, \beta_1, \dots, \beta_p$ are unknown. Data must be used to produce estimates of them, $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$.

Let

$$(x_{11}, x_{12}, \dots, x_{1p}, y_1), (x_{21}, x_{22}, \dots, x_{2p}, y_2), \dots, (x_{n1}, x_{n2}, \dots, x_{np}, y_n) \tag{4.2}$$

represent n independent observation (*p+1)*-tuples.

Let

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_p x_{ip} \tag{4.3}$$

be the prediction for y_i. Then

$$e_i = y_i - \hat{y}_i \tag{4.4}$$

represents the *i*th residual and we define the residual sum of squares (RSS) as

$$RSS = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip}\right)^2 \tag{4.5}$$

Parameters are estimated to minimize the residual sum of squares.

One of the most important assumptions of multiple linear regression state that the relationship between the predictors and response must be additive. This means that the effect of changes in a predictor x_i on the response Y_i is independent of the values of the other predictors. In practice, this assumption is often violated, and alternative approaches are demanded.

One common approach is the inclusion of more predictors, called interaction terms. These terms are constructed by computing the product of j independent variables, with $j > 1$. For a simpler understanding of what was stated, and without losing generality, let us consider model (4.1) with $p = 2$, which will include x_1 and x_2 as main effects, and the interaction between x_1 and x_2. The multiple linear regression model would be defined by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i \tag{4.6}$$

with $x_1 x_2$ representing the interaction term (further also represented as $x_1 * x_2$).

After defining a regression model, it is important to quantify the extent to which the model fits the data and, generally, access the quality of the model. That is, we need to quantify the extent to which the predicted response value for a given observation is close to the true response value for that observation (Gareth et al., 2013). Three of the commonly used measures are the root mean squared error (RMSE), the coefficient of multiple determination (R^2) and the Akaike's information criterion (AIC) (Akaike, 1973).

Let us first define the mean squared error (MSE), which is given by the following expression

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}. \tag{4.7}$$

The RMSE can now be defined as

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}. \tag{4.8}$$

For the same sample, a lower value for RMSE represents a better fit of the model.

R^2 takes the form of a proportion,

$$R^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{4.9}$$

where

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}. \tag{4.10}$$

R^2 assumes values from 0 to 1 with values closer to 1 indicating that a large proportion of the variability in the outcome was explained by the regression model.

The AIC is defined by

$$AIC = -2\log(L) + 2p \tag{4.11}$$

where $\log(L)$ is the log-likelihood function of the parameters in the model evaluated at maximum likelihood estimator and the term $2p$ is a penalty term for additional parameters in the model.

As for the RMSE, a lower value indicates a better fit.

Multiple linear regression models are challenging to build specially when massive datasets are used (Abu-Mostafa et al., 2012; Hastie et al., 2009), assuming massive datasets as the ones with a great number of features. It is possible that all of the features are associated with the response to be studied, but it is more often the case that the response is only related to a subset of the collected features (Gareth et al., 2013). Moreover, the criteria for considering one feature in a model may vary from one study to the other (Hosmer et al., 2013) which is corroborated by Jobson (Jobson, 1991) when he affirms that there are three different purposes of a regression model, i.e. description, control or prediction, and the variables to include may not be the same according to each end. So, among the many decisions to be taken, the selection of the variables to include in the model, from now on called the feature selection, is one of the most important aspects of model building since it can help in obtaining predictive models with less correlated variables, biases, and unwanted noise. Yet, studies have shown that when using feature selection algorithms one can obtain models which use only non-informative features and are still 100% accurate on the training set (Ambroise & McLachlan, 2002). Clinically, such models are useless and are returning random answers on the test set.

4.3. FEATURE SELECTION PROCEDURES

In the following sub-chapters, we will briefly define some of the feature selection procedures used in our study.

4.3.1. AUTOMATIC BACKWARD ELIMINATION

The feature selection process assumes greater importance as the number of features in a dataset increases and, therefore, the ratio between them and the size of the dataset also increases. If we have *p* features, we can compute at least 2^p models that contain subsets of those features. As *p* increases, trying out every possible subset of the features becomes not practical and, at some point, infeasible (Gareth et al., 2013; E. I. George, 2000). One of the classical approaches for this task is the automatic backward selection algorithm. This

algorithm starts with all *p* features in the regression model and remove the feature that is the least statistically significant one, for instance. The new *(p − 1)*-feature model is fit, and, once again, the least statistically significant feature is removed. This procedure continues until a stopping rule is reached. We may stop when all remaining variables have a p-value below a defined threshold, no significant improvement in a quality or performance parameter is achieved (tested for instance with a likelihood ratio test) or some other stopping rule. Some authors state that this process cannot be used if $p > n$, with *n* representing the size of a random sample (Gareth et al., 2013).

4.3.2. RANDOM FOREST

Aside from the previously described multiple linear regression model, there are many other regression methods. For instance, the tree-based methods involve stratifying or segmenting the predictor space into a number of cuboid disjoint regions, whose edges are aligned with the axes, and then assign, for instance, a constant to each region (Bishop, 2006; Gareth et al., 2013). The less complex of these tree-based methods is the regression tree. It consists of a series of splitting rules, starting at the top of the tree. The predictor space regions are formed recursively using binary partitions according to whether $x_j \leq \theta_k$ or $x_j > \theta_k$, were θ_k is parameter of the model. The regions are recursively subdivided. This is a greedy process since a particular split or node of the predictor's space is made to optimize that precise split, therefore not considering future splits. In fact, the objective is to define rectangular regions, R_k, that minimize the RSS, given by

$$\sum_{k=1}^{K} \sum_{i \in R_k} \left(y_i - \hat{y}_{R_k}\right)^2 . \tag{4.12}$$

After all the partitioning has been done, the model predicts the output based on the average response values for all observations that fall in that subgroup. Sometimes it is not practical to obtain many partitions of the space and the tree can be grown to a certain point only. It is possible to define a minimum node size (*nodesize*) which is the minimum cardinality of a region.

This binary partitioning process, that can be put in terms of answering questions with a "yes" or a "no" is very popular in the health-related sciences since it is readily interpretable by humans but regression trees are very sensitive to details in the dataset, prone to high variance and may overfit the data, leading to poor test set performance (Bishop, 2006;

Gareth et al., 2013). One procedure to avoid this is to use an ensemble method called the Bootstrap aggregation, also known as bagging (Breiman, 1996).

In the bagging procedure, B regression trees are grown using B bootstrapped training sets (Efron, 1982), and the resulting predictions are averaged. As it was said, each individual regression tree has high variance, but low bias, and by averaging these B regression trees the variance is reduced. At least one improvement can be made to bagged regression trees to reduce the correlation between the trees obtained by this method. One can define how many predictors are available to create the splitting rule, here named *mtry*. Since a new sample of predictors is randomly chosen at each split, with a typical default value of *mtry* = *p/3*, randomness is added into the tree growing and gives an opportunity to moderately strong predictors. This process is called the Random forest method (RF) (Breiman, 2001).

In each bootstrap training set, approximately one-third of the observations are left out and are called the out-of-bag (OOB) observations (Breiman, 2001; Gareth et al., 2013). Several authors (Breiman, 1996; Tibshirani, 1996a) proposed using OOB estimates as an estimate of generalization error. We can predict the outcome for an observation using each of the trees in which that observation was an OOB observation and average these predicted outcomes. An OOB prediction can be obtained in this way for each of the *n* observations, from which the overall OOB MSE (also referred as OOB error) can be computed.

The bagging and RF processes seem to enhance prediction accuracy at the expense of interpretability (Breiman, 2001; Gareth et al., 2013). This difficulty was well put by Breiman (Breiman, 2001): “a forest of trees is impenetrable as far as simple interpretations of its mechanism go”.

Despite this, it is possible to obtain information on the importance of each feature using that that purpose the total amount that the RSS (8.1) is decreased due to splits over a given feature, averaged over all B regression trees. One other related approach is based on how much the accuracy decreases when a certain feature is excluded. A larger value indicates a more important predictor but, as far as we know, there are no reference values.

4.3.3. BORUTA

Wrapper methods use a subset of features and train a model using them. Based on it, features are kept or removed from the subset. The already mentioned Automatic Backward Elimination algorithm is an example. Boruta is also a wrapper algorithm that uses RF (M. Kursa et al., 2010).

At a first step, the Boruta algorithm adds randomness to a dataset by creating shuffled copies of all features. These copies are called shadow features. It then trains a RF model on the extended dataset and uses the mean decrease accuracy to evaluate the importance of each feature. At every iteration, it checks whether a real feature has a higher importance than the best of its shadow features and constantly removes features which are considered highly unimportant. This is done using Z-scores. Z-scores of the shuffled copies of the features and the original features are compared to see if the latter performed better than the former. If it does, the algorithm will mark the feature as confirmed. Finally, the algorithm stops either when all features get confirmed or rejected or it reaches a defined number of RF iterations. Features that are not confirmed nor rejected are called tentative features, also referred as unconfirmed features. They have an importance that is close to their best shadow features but that the algorithm is not able to decide with a certain confidence, in the chosen number of RF iterations.

4.3.4. EXTREME GRADIENT BOOSTING

Gradient descent is an optimization algorithm based on a convex function used to adjust the coefficients of a model that minimize a cost function to its local minimum, by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. The goal is to move in the direction that decreases the cost of the coefficients. This process is taken iteratively until the bottom of the cost function is reached, where the values of the coefficients result in the minimum cost. An important parameter in Gradient descent is the learning rate *eta*, which specifies the size of the steps the algorithm takes toward the local minimum. A learning rate too big, means it is possible to change more the coefficients at each step, but there is the risk of overshooting the lowest point since the slope of the cost function is always changing. On the other hand, a learning rate too small is more precise, but it will take the algorithm a long time to converge.

Gradient tree boosting algorithm is a gradient descent algorithm that create an ensemble of trees that, unlike bagging or RF, do not involve bootstrap sampling and are grown sequentially using information from the previously grown trees (Gareth et al., 2013). A shallow and weak tree is first trained and then the next tree is trained based on the errors of the first tree. The process continues with a new tree being sequentially added to the ensemble and the new successive tree improves on the errors of the ensemble of preceding trees.

Extreme gradient boosting (XGB) is an improvement of the Gradient tree boosting model, designed for speed and performance (Chen & Guestrin, 2016).

The already mentioned learning rate is a hyperparameter for the XGB algorithm but there are others. We will only mention those we find the most important ones. For instance, the number of decision trees, also referred as iterations, used in the ensemble. Varying the maximum depth of each tree added to the ensemble is another important hyperparameter because it controls how specialized each tree is to the training dataset, that is to say, how general or overfit the model might be. The number of samples, also referred as *subsampling*, used to fit each tree can be adjusted. This means that each tree can be fitted on a randomly selected subset of the training dataset. Using fewer samples introduces more variance for each tree, although it can improve the overall performance of the model. Finally, the minimum number of instances needed in each node, also known as minimum child weight. The larger the weight, the more conservative the algorithm will be.

Features' importance can be compared using the gain. It implies the relative contribution of the corresponding feature to the model calculated by taking each feature's contribution for each tree in the model. A higher value of this metric when compared to another feature implies it is more important for generating a prediction but, as far as we know, there are no reference values.

4.3.5. L1-PENALIZED ESTIMATION

Embedded method are algorithms that have their own built-in feature selection procedures. The Least Absolute Shrinkage and Selection Operator (LASSO) method is an example of an embedded method because it simultaneously performs variables selection and shrinkage (Tibshirani, 1996b).

It was originally a regularized estimation approach for regression models. Regularization is an important concept that is used to prevent overfitting when modelling. It can be implemented by adding a penalty term λ that reduces the magnitude of coefficients. This reduction is called shrinkage and, in the case of LASSO, is also called L1 penalty. As $\lambda \to \infty$, the impact of the shrinkage penalty grows, and the regression coefficient estimates will approach zero.

The estimation of the LASSO coefficients is made to minimize

$$\sum_{i=1}^{n}\left(y_i\hat{\beta}_0 - \hat{\beta}_1 x_{\mathrm{i}1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip}\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right|. \tag{4.13}$$

The penalty term must be chosen. One of the possible procedures is to choose the one that produces a model with the lowest AIC.

Although Tibshirani (Tibshirani, 1996b) originally proposed quadratic programming to solve the optimization of the LASSO problem it were Groll and Tutz who developed a LASSO method applied to models that can predict repeated outcomes (Groll & Tutz, 2014). The method primarily reduces the dimensionality and then performs refitting by maximum likelihood estimation to get accurate parameter estimates.

5. DATASET

5.1. RESPIRATORY RESEARCH AND REHABILITATION LABORATORY

The Respiratory Research and Rehabilitation Laboratory (Lab-3R) was created in 2009 but moved to new infrastructures in September 2013. It is currently situated at School of Health Sciences (ESSUA) of the University of Aveiro and it works in close collaboration with health institutions (e.g., Centro Hospitalar do Baixo Vouga, Agrupamento dos Centros de Saúde do Baixo Vouga), city councils (e.g., Aveiro, Estarreja) and patients' associations (e.g., RESPIRA).

Lab3R aims to empower people with chronic respiratory diseases and their families to better adjust and manage the impact caused by the respiratory disease; and serve and empower the community on understanding these diseases. Lab-3R is also concerned in increasing scientific knowledge about the normal and pathological behaviour of clinical measures in the field of respiratory health and to do so it has built an extensive database.

Among the several Lab3R research projects, two contributed the most to the dataset analysed in this dissertation. "*GENetic and clinIcAL markers in COPD trajectory (GENIAL project)*" which aimed to "establish the role of genetic mutations on the development and trajectory of COPD and identify clinical markers (e.g., dyspnoea; number of exacerbations per year; lung function; exercise capacity) to detect AECOPD episodes" (*Lab 3R*, 2018) and "*Pulmonary Rehabilitation Innovation and Microbiota in Exacerbations of COPD (PRIME project)*" which aimed "to determine the role of microbiome and clinical data on predicting acute exacerbations of COPD, and to contribute for the knowledge of pulmonary rehabilitation during acute exacerbations of COPD" (*Lab 3R*, 2018).

5.2. STUDY DESIGN AND PARTICIPANTS

Data collected in GENIAL (PTDC/DTP-PIC/2284/2014) and PRIME (PTDC/SAU-SER/28806/2017) research projects were used in the following studies. Five independent Ethics Committees (Centro Hospitalar do Médio Ave ref. 09/2016 and 10/2018; Unidade Local de Saúde de Matosinhos ref. 10/CES/JAS 17/02/2017 and 73/CE/JAS 12/10/2018; Centro Hospitalar Baixo Vouga ref. 777638 and 086892; Hospital Distrital da Figueira da Foz ref. 1807/2017 and 27/05/2019; Administração Regional de Saúde do Centro ref.

64/2016 and 85/2018) approved the study. Written informed consent was obtained from all participants before data collection. Data protection was ensured by the National Committee for Data Protection (no. 7295/2016) and followed the General Data Protection Regulation.
Individuals were eligible if diagnosed with COPD, according to the GOLD criteria (Global Initiative for Chronic Obstructive Lung Disease, 2021) (i.e. FEV_1 / forced vital capacity (FVC) x 100 < 70).
Inclusion criteria also stated that individuals should have been clinically stable over the previous month, i.e., no hospital admissions or exacerbations and not have had a change in medication for the cardiorespiratory system. Individuals 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 collection and interpretation were excluded.

5.3. DATA COLLECTION

Sociodemographic (age, sex, educational level, marital status and current occupation), anthropometric (height and weight to calculate BMI (body weight in kilograms divided by height in meters squared)) and clinical data (smoking status, smoking number of years, number of cigarettes' packs smoked per day, use of long-term oxygen therapy (LTOT) and non-invasive ventilation (NIV), number of acute exacerbations in the previous year (AECOPD), number of respiratory related hospitalisations and emergencies in the previous year and comorbidities) were assessed with a structured questionnaire to characterise the sample. The pack-years were computed by multiplying the number of cigarettes' packs smoked per day by the number of years the person has smoked. The severity of comorbid diseases was scored according to Charlson comorbidity index (CCI): i) scores of 1–2; ii) scores of 3–4; and iii) scores ≥5 (Charlson et al., 1994). Lung function, specifically the predicted percentage of FEV_1 (FEV_1 % of predicted) and the FEV_1/FVC ratio, was then assessed with spirometry (MicroLab 3535, CareFusion, Kent, UK) as recommended by the American Thoracic Society and the European Respiratory Society (Graham et al., 2019). The severity of airway obstruction was classified using the FEV_1 % of predicted according to the GOLD (Global Initiative for Chronic Obstructive Lung Disease, 2021). The modified British medical research council questionnaire (mMRC) was used to assess activity-related dyspnoea (Crisafulli & Clini, 2010). Scores range from 0 (no trouble with breathlessness) to

4 (too breathless to leave the house) (Bestall et al., 1999). The modified Borg scale is a categorical vertical scale with scores ranging from 0 to 10 with corresponding verbal expressions of perceived sensation intensity, where 0 corresponds to the sensation of normal breathing and 10 corresponds to the patients' maximum possible sensation of dyspnoea (Borg, 1982; Mahler et al., 1987; Wilson & Jones, 1989). Self-reported physical activity was assessed with the brief physical activity assessment tool (BPAAT) (Marshall et al., 2005) which was found to be a valid tool to assess physical activity of people with COPD (Cruz et al., 2017). Scores range from 0 t0 8 being interpreted as 0-3 "insufficiently active" or ≥4 "sufficiently active".

The Saint George's respiratory questionnaire (SGRQ) was used to measure health-related quality of life (P W Jones et al., 1991). Scores (per domain or total) range from 0 (no impairment) to 100 (worst possible health-related quality of life) units.

Information on the effects of pulmonary rehabilitation throughout the follow-up period was also recorded as a dichotomous variable (i.e., yes - participants were under the effect of PR during the follow-up period or no – participants were not under this effect). The PR program consisted of supervised exercise training twice per week and education and psychosocial support every other week (World Health Organization, 2002). More detailed information on the PR programme is available elsewhere (Marques et al., 2019).

Functional status was assessed using the one-minute sit-to-stand test (1minSTS). This test consists of sitting and standing from a 46-48 cm height chair as many times as possible for one minute (Ozalevli et al., 2007; Vaidya et al., 2017). A change of 3 repetitions between assessments is commonly used as MCID for 1minSTS (Vaidya et al., 2016).

The COPD assessment test (CAT) is a disease-specific questionnaire consisting of eight items (i.e., cough, sputum, chest tightness, breathlessness going up hills/stairs, activity limitations at home, confidence leaving home, sleep, and energy) scored from 0 to 5 (P W Jones et al., 2009). Each item individual score is added to provide a total CAT score that can range from 0 to 40. Scores range from 0 to 40 and are interpreted as ≤10 low, 11-20 medium, 21-30 high and 31-40 very high impact (Paul W Jones et al., 2011). It was used to evaluate the disease impact (F. George, 2013; P W Jones et al., 2009) and to classify participants according to the ABCD assessment tool, following the Global Initiative for Chronic Obstructive Lung Disease (GOLD) (Global Initiative for Chronic Obstructive Lung Disease, 2021). The MCID of the CAT was found to be a two-point decrease at both the individual and population level (Kon et al., 2014).

Handgrip muscle strength (HMS) was used as a marker of overall upper-limb muscle strength (Clegg et al., 2013). A systematic review concluded that changes of 5.0 to 6.5 kg may be reasonable estimates of meaningful changes in HMS (Bohannon, 2019).

All data were collected cross-sectionally at baseline. Over the monthly follow-up assessments, 1minSTS, HMS and CAT were collected.

6. STUDY 1 – COVID-19 LOCKDOWN EFFECT IN COPD: A COMPARISON OF FIXED-EFFECTS SELECTION METHODS*

*** Some of the preliminary results from this study were submitted in the form of an abstract to the XXV Congress of the Portuguese Statistical Society, to be held at Évora between the 13th and the 16th of October 2021 (Appendix B).**

6.1. RATIONALE

The 2020 imposed lockdown due to the recent the Coronavirus Disease 2019 (COVID-19) pandemic is likely to have influenced the functional status and the wellbeing and daily life of patients with COPD, but this is still unknown. Few feature selection algorithms are available for longitudinal data. We aimed to compare different feature selection methods and describe the effect of the COVID-19 lockdown on the 1minSTS, HMS and CAT behaviour in people with COPD.

6.2. STATISTICAL METHODS

An observational, prospective cohort study was conducted with data described in chapters 5.2 and 5.3. Two groups were defined: participants with baseline date between the 1st of February 2019 and the 15th of March 2019 were classified as pre-lockdown; participants with baseline date between the 1st of February 2020 and the 15th of March 2020 were classified as lockdown. For the pre-lockdown group, the baseline assessment, and the assessments after one (A1) and five (A5) months were analysed, For the lockdown group, the baseline assessment, the assessment after one month (A1) and the assessment immediately after the period of lockdown (A5) were analysed,

Descriptive analysis was performed by group. Quantitative variables were summarized using mean and standard deviation values if normally distributed or median values and interquartile ranges, otherwise. Categorical variables were summarized through count values and percentages. Shapiro-Wilk test was used to assess the assumption of normality. Welch t-tests and Mann-Whitney-Wilcoxon tests were used to compare, respectively, means and medians of baseline characteristics and outcome measures between groups.

Fisher's exact test was used to compare proportions of baseline characteristics between groups.

Three repeated outcome measures (dependent variables) were considered: 1minSTS, HMS and CAT. For each of the dependent variables, five feature selection algorithms were performed, RF, Boruta, XGB, L1 penalized estimation and automatic backward elimination. Due to the nature of the algorithms and the longitudinal characteristics of data, the three first algorithms used 21 standardized independent variables and the last two also included evaluation. For the same reason, the difference of the dependent variable between two consecutive assessments was determined and, for RF, Boruta and XGB two sets of features were obtained, one for the difference between A1 and baseline and another for the difference between A5 and A1.

A preliminary tunning of RF parameters was performed with a grid of values for the number of features to consider at each split point (mtry) and the minimum number of observations in a terminal node (nodesize). The pair of values that produced the lowest OOB error was used in the algorithm with 1000 trees. The importance of each variable in the RF was based on how much the accuracy decreased when the variable was excluded. Two cut-off values were tested: 2% and 5%.

For the Boruta algorithm, variables were classified as confirmed, unconfirmed and unimportant according with shadow features (M. Kursa et al., 2010).

The XGB models were trained using a 4-fold cross-validation process with 750 iterations using the values of a grid containing all the possible combination of the following tuning parameters: learning rate (*eta*) = 0.010, 0.015, 0.020, 0.025; subsampling = 0.4, 0.5, 0.6; minimum child weight=1, 2, 3; maximum depth of a tree =5, 8, 10, 11, 12, 14, 17. The total grid size was 252. The remaining of the XGB parameters were set to the default values and we used a *gbtree* booster and an objective of *reg:squarederror*. Finally, we evaluated the prediction accuracy of the model based on the test RMSE at each of the 750 iterations. The best iteration was recorded, and the best combination of parameters was used. The importance of each variable was based on the fractional contribution of each feature to the model based on the total gain of this feature's splits, with higher proportion meaning a more predictive feature (Chen et al., 2021). Two cut-off values were tested: 0.075 and 0.15.

The penalty parameter λ used in the L1 penalized estimation was determined from a grid of 1000 $\log_{10}$ values ranging from -1 to 4, equally spaced. The λ value that produced the lowest AIC, here defined as the negative of twice the log-likelihood plus twice the corresponding degrees of freedom. These degrees of freedom are determined by the sum

of nonzero coefficients corresponding to fixed-effects plus the number of random effects covariance parameters that have to be estimated (Abugaber, 2020; Groll, 2017).

The automatic backward elimination consisted firstly, of a single random-effect terms deletion and secondly, a backward elimination of single main fixed-effect terms in order to obtain a model with the lowest AIC (Zuur et al., 2009).

Whenever possible, R^2 values were determined to evaluate the quality of the predictive models.

LMMs were applied to assess the mean change in dependent variables and to account for correlations in repeated measurements per participant using the features selected by the different algorithms. Evaluation, group and their interaction were always considered, independently of their previous selection/rejection. Random intercepts and slopes were used to incorporate individual response trajectories. The p-values were computed based on conditional F-tests with Kenward-Roger approximation for the degrees of freedom (Kenward & Roger, 1997). Marginal coefficient of determination (marginal R^2) and conditional coefficient of determination (conditional R^2) were determined for assessing the model's quality and level of adjustment according to Nakagawa (Nakagawa et al., 2017). The models with the lowest AICs were kept. LMM assumptions were assessed by visual inspection of residuals' boxplots, scatterplots, and Q-Q plots. Estimated marginal means (EMM) were averaged and weighted in proportion to the frequencies, in the original data, over the levels of categorical variables (Lenth, 2021) and the p-value was adjusted by the Tukey method.

Two-sided $P < 0.05$ was considered statistically significant for all analyses.

Statistical analyses were performed using R packages *randomForestSRC* (Ishwaran & Kogalur, 2021), *randomForest* (Liaw & Wiener, 2002), *Boruta* (M. B. Kursa & Rudnicki, 2010), *xgboost* (Chen et al., 2021), *glmmLasso* (Groll, 2017), *lmer* (Bates et al., 2015), *lmerTest* (Kuznetsova et al., 2017), *emmeans* (Lenth, 2021) and *ggeffects* (Lüdecke, 2018) in RStudio Version 1.4.1103 (RStudio Team, 2021) running R version 4.0.5 (R Core Team, 2021).

6.3. RESULTS

6.3.1. DESCRIPTIVE ANALYSIS

A total of 59 participants with COPD were included, 34 (57.6%) of whom were assessed for baseline characteristics between the 1st of February 2019 and the 15th of March 2019 (Figure 6.1).

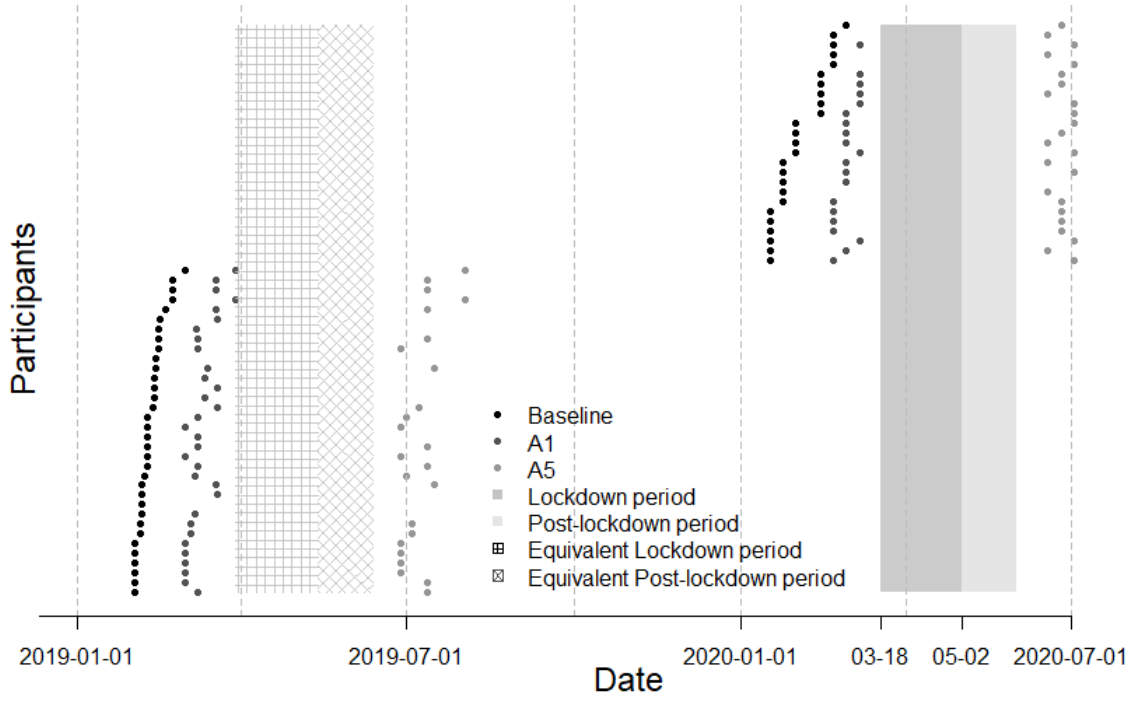

Figure 6.1 - *Distribution of the participants' assessments date.* ***Abbreviations:*** *A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment.*

Participants mean age was 66.9 (standard deviation 8.0) years, most were men (83.1%), with a mean BMI of 27.4 kg/m^2 (standard deviation 5.1) and presented 3 to 4 comorbidities (61.0%). Additionally, most participants were former smokers (83%), without non-invasive ventilation (86.4%) nor long-term oxygen therapy. Finally, the majority of participants did not have an AECOPD in the previous year (71.4%), had a median mMRC score of 1 (IQR [1;2]) and a SGRQ total score of 31.7 (IQR [23.0;48.3]) points. No statistically significant differences were found between the pre-lockdown and the lockdown groups. Detailed baseline characteristics are presented in Table 6.1.

Table 6.1 - *Descriptive statistics of selected variables at baseline (n=59).*

Characteristics		All (n=59)	Group		p-value
			Pre-Lockdown (n=34)	Lockdown (n=25)	
Age, *years*, mean (SD)		66.9 (8.0)	67.5 (7.9)	66 (8.3)	0.497[T]
Sex	Male	49 (83.1%)	28 (82.4%)	21 (84.0%)	1[F]
BMI, *kg/m²*, mean (SD)		27.4 (5.1)	26.6 (5.2)	28.4 (4.8)	0.902[T]
Smoking status	Never Former Current	5 (8.5%) 49 (83.0%) 5 (8.5%)	4 (11.8%) 26 (76.4%) 4 (11.8%)	1 (4.0%) 23 (92.0%) 1 (4.0%)	0.329[F]
Smoking no. of years, median [IQR]		38.0 [30.0;45.0]	36.5 [30.0;43.0]	40.0 [32.0;45.0]	0.271[M]
Pack-years, median [IQR]		56.0 [30.0;80.0]	52.0 [30.0;77.3]	60.0 [32.0;90.0]	0.509[M]
FEV_1/FVC, *%*, median [IQR]		56.7 [44.2;63.4]	54.8 [39.3;62.9]	58.2 [46.7;63.2]	0.495[M]
FEV_1, *% predicted*, mean (SD)		61.4 (23.3)	58.6 (25.5)	65.2 (19.9)	0.269[T]
mMRC, *points*, median [IQR]		1.0 [1.0;2.0]	1.0 [1.0;3.0]	1.0 [0.0;2.0]	0.312[T]
SGRQ Total, median [IQR]		31.7 [23.0;48.3]	35.3 [27.7;51.2]	26.6 [20.1;47.4]	0.103[M]
BPAAT Moderate, *score*	0 1 2 4	22 (37.3%) 12 (20.3%) 8 (13.6%) 17 (28.8%)	11 (32.4%) 6 (17.6%) 6 (17.6%) 11 (32.4%)	11 (44.0%) 6 (24.0%) 2 (8.0%) 6 (24.0%)	0.593[F]
BPAAT Vigorous, *score*	0 2 4	56 (94.9%) 1 (1.7%) 2 (3.4%)	32 (94.2%) 1 (2.9%) 1 (2.9%)	24 (96.0%) 0 (0.0%) 1 (4.0%)	1[F]
Modified BORG scale (Dyspnoea), median [IQR]		0.0 [0.0;1.0]	0.0 [0.0;0.9]	0.0 [0.0;2.0]	0.103[M]
Modified BORG scale (Fatigue), median [IQR]		0.0 [0.0;2.0]	0.0 [0.0;1.0]	0.0 [0.0;2.0]	0.879[M]
AECOPD*, in the previous year*	Yes	11 (18.6%)	6 (17.6%)	5 (20.0%)	1[F]
Respiratory related emergencies*, in the previous year*	Yes	8 (13.6%)	4 (11.8%)	4 (16.0%)	0.711[F]
Respiratory related hospitalizations*, in the previous year*	Yes	4 (8.5%)	3 (8.8%)	2 (8.0%)	1[F]
Long-term oxygen therapy	Yes	8 (13.6%)	4 (11.8%)	4 (16.0%)	0.711[F]

Characteristics		All (n=59)	Group		p-value
			Pre-Lockdown (n=34)	Lockdown (n=25)	
Non-invasive ventilation	Yes	12 (20.3%)	6 (17.6%)	6 (24.0%)	0.745[F]
CCI, *score*	1-2	11 (18.7%)	6 (17.6%)	5 (20.0%)	1[F]
	3-4	36 (61.0%)	21 (61.8%)	15 (60.0%)	
	>=5	12 (20.3%)	7 (20.6%)	5 (20.0%)	

Note: *Data presented as mean (standard deviation) or median [1st quartile;3rd quartile] for continuous variables and number (percentage) for categorical variables. T, Welch t-test; M, Mann-Whitney-Wilcoxon test; F, Fischer's exact test.*

Abbreviations: *1minSTS, one-minute sit-to-stand test; AECOPD, acute exacerbation of COPD; BMI, body mass index; BPAAT, brief physical activity assessment tool; CAT, COPD assessment test; CCI, Charlson comorbidity index; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; GOLD, global initiative for chronic obstructive lung disease; IQR, interquartile range; mMRC, modified medical council dyspnoea scale; SD, standard deviation; SGRQ, St. George's respiratory questionnaire.*

From the initial 59 individuals assessed at baseline, a decrease in number was registered in the first and last assessments to 52 and 46 participants, respectively. A total of 43 individuals attended all 2 assessments (24 from the pre-lockdown group) while 14 individuals (8 from the pre-lockdown group) attended only 1 assessment aside from the baseline one. The median time between baseline and last assessment was of 147 days (interquartile range [143.8;154.0]) and 146 days (interquartile range [132.0;154.8] in the pre-lockdown and lockdown groups, respectively. Figure 6.2 show the distribution of the time between assessments in both groups.

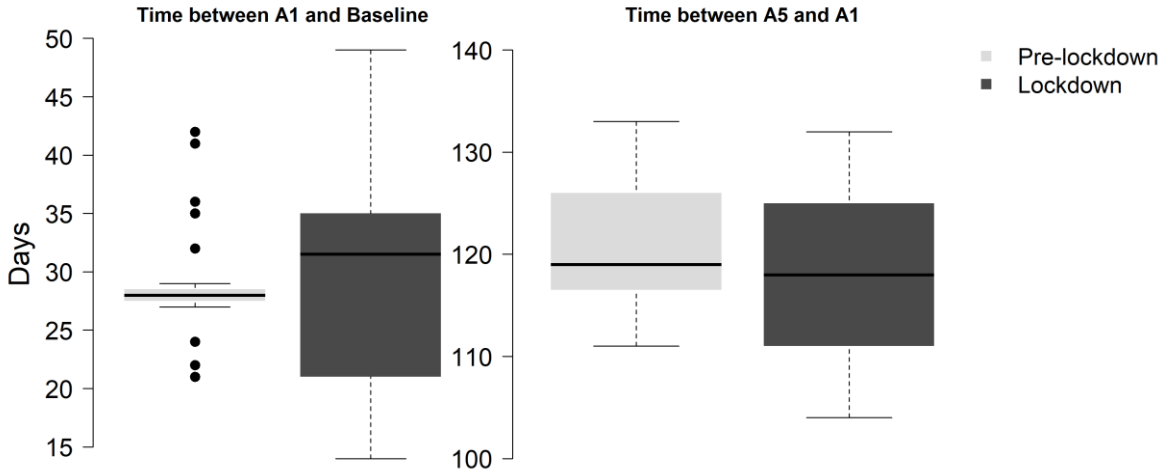

Figure 6.2 – *Distribution of the time, in days, between consecutive assessments by groups.* ***Abbreviations:*** *A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment.*

At baseline, the pre-lockdown group registered a median of 26 repetitions in the 1minSTS which is statistically different from the 23 repetitions achieved in the lockdown group (p=0.033). No other statistically significant difference was found. Table 6.2 presents the detailed characteristics of all variables considered at baseline and at the 2 different assessments.

Table 6.2 - *Descriptive statistics of collected variables at baseline, assessment 1 and assessment 5 (n=59).*

Variable	Baseline (n=59)		A1 (n=52)		A5 (n=48)	
	Pre-lockdown (n=34)	Lockdown (n=25)	Pre-lockdown (n=32)	Lockdown (n=20)	Pre-lockdown (n=24)	Lockdown (n=24)
1minSTS, median [IQR]	26.0 [23.0;31.0]	23.0 [18.0;27.0]	29.0 [24.5;32.0] **(b)**	(a) 25.0 [22.5;28.5]	28.0 [23.0;38.0]	27.5 [22.5;31.8] **c)**
	p = 0.033^{M}		p = 0.149^{M}		p = 0.253^{M}	
HMS, *kg,* mean (SD)	33.4 (10.5)	34.4 (8.4)	34.2 (10.8)	36.9 (7.5) **(a)**	36.4 (12.7) **(d)**	37.0 (8.8) **(e)**
	p = 0.710^{T}		p = 0.099^{T}		p = 0.839^{T}	
CAT, median [IQR]	11.5 [6.0;15.8]	9.0 [8.0;15.0]	9.5 [5.8;14.3]	9.0 [6.5;10.0] **(a)**	6.0 [2.8;13.3]	7.0 [4.0;12.8]
	p = 0.461^{M}		p = 0.660^{M}		p = 0.469^{M}	

(a) 19 participants; (b) 31 participants; (c) 20 participants; (d) 22 participants; (e) 21 participants.
Note: *Data presented as mean (standard deviation) or median [1st quartile;3rd quartile] for continuous variables and number (percentage) for categorical variables. T, Welch t-test; M, Mann-Whitney-Wilcoxon test.* ***Abbreviations:*** *1minSTS, one minute sit-to-stand Test; A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; CAT, COPD assessment test; COPD, chronic obstructive pulmonary disease; HMS, handgrip muscle strength; IQR, interquartile range; SD, standard deviation.*

Individual longitudinal trajectories and global linear trajectories between assessments of each of the repeated outcome measures is shown in Figure 6.3. From visual inspection, differences between groups in the 4-month period separating A1 and A5 seem to exist.

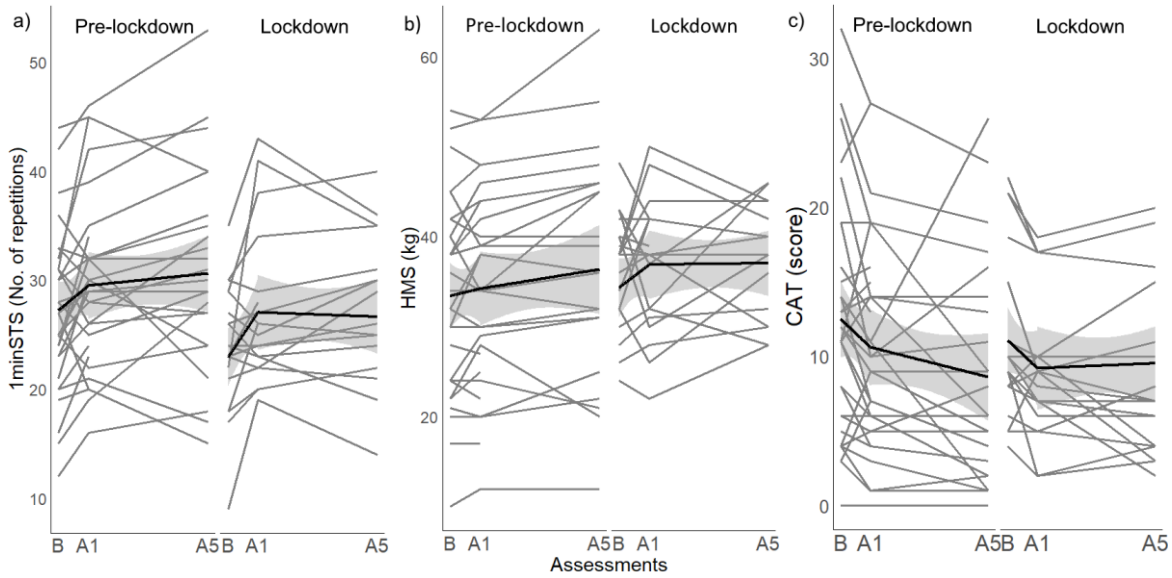

Figure 6.3 - *Individual values of the outcome measures over time.* a) *One-minute sit-to-stand test number of repetitions;* b) *Handgrip muscle strength;* c) *Chronic obstructive pulmonary disease assessment test score. The black line represents the global linear regression line between each assessment; The grey band represents the 95% confidence interval.* ***Abbreviations:*** *1minSTS, one-minute sit-to-stand test; B, baseline assessment; A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; CAT, chronic obstructive pulmonary disease assessment test score; HMS, handgrip muscle strength.*

6.3.2. FEATURE SELECTION PROCEDURES

In the following sub-chapters, we will be presenting the full results for all the feature selection algorithms considered for 1minSTS and only the summary tables for HMS and CAT.

6.3.2.1. ONE-MINUTE SIT-TO-STAND

Five feature selection algorithms were performed. For each of them, the main results are presented next.

6.3.2.1.1. RANDOM FOREST

RF was applied to a sample of 50 participants in order to predict the difference of the number of repetitions in the 1minSTS between A1 and Baseline assessment. The parameters' tunning process returned the lowest OOB error of 0.991 for an *mtry* of 8 and a *nodesize* of 6. The number of participants available for determining the difference in the number of

repetitions between A5 and A1 was only 39. For an *mtry* of 13 and a *nodesize* of 9 a value of 0.886 for the OOB error was obtained (Figure 6.4).

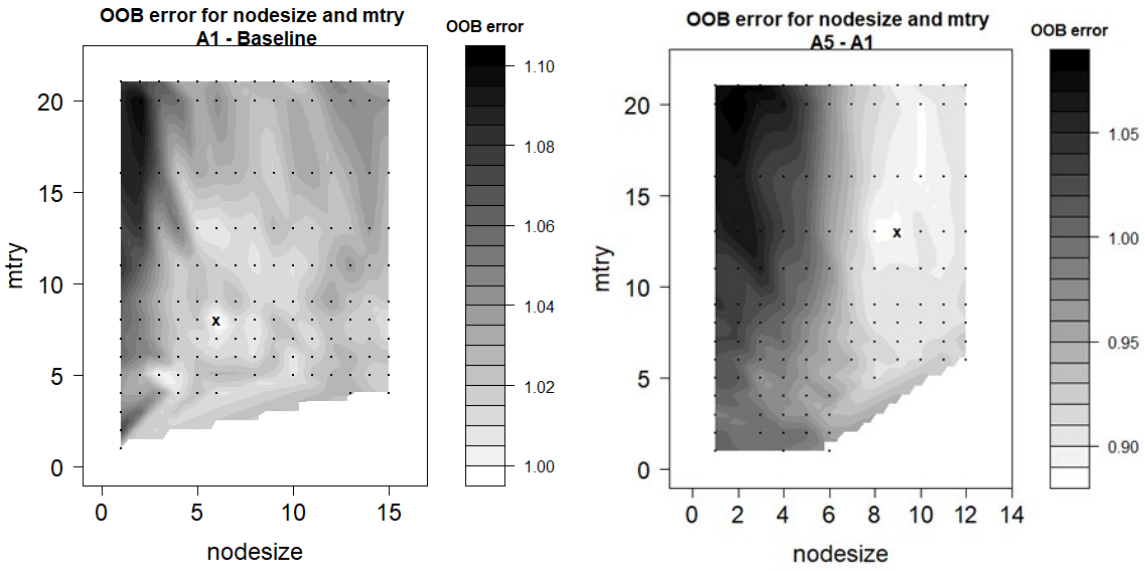

Figure 6.4 - *Random forest's out-of-bag error for different values of mtry and nodesize.* ***Abbreviations:*** *A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; OOB, out-of-bag.*

Pack-years and FEV_1 % of predicted features contributed with a percentage of increase of the RMSE greater than 5 in the RF model for the difference between A1 and baseline (9.4% and 9.1%, respectively). Four more features were retained when the cut-off value considered was 2%, namely, SGRQ (4.9%), smoking number of years (3.8%), respiratory hospitalizations in the previous year (3.0%) and smoking status (3.0%). In the RF model for the difference between the two last assessments only pack-years (12.7%) overcame the higher cut-off value and SGRQ (4.9%), sex (3.5%) and smoking status (2.5%) the lower one (Figure 6.5).

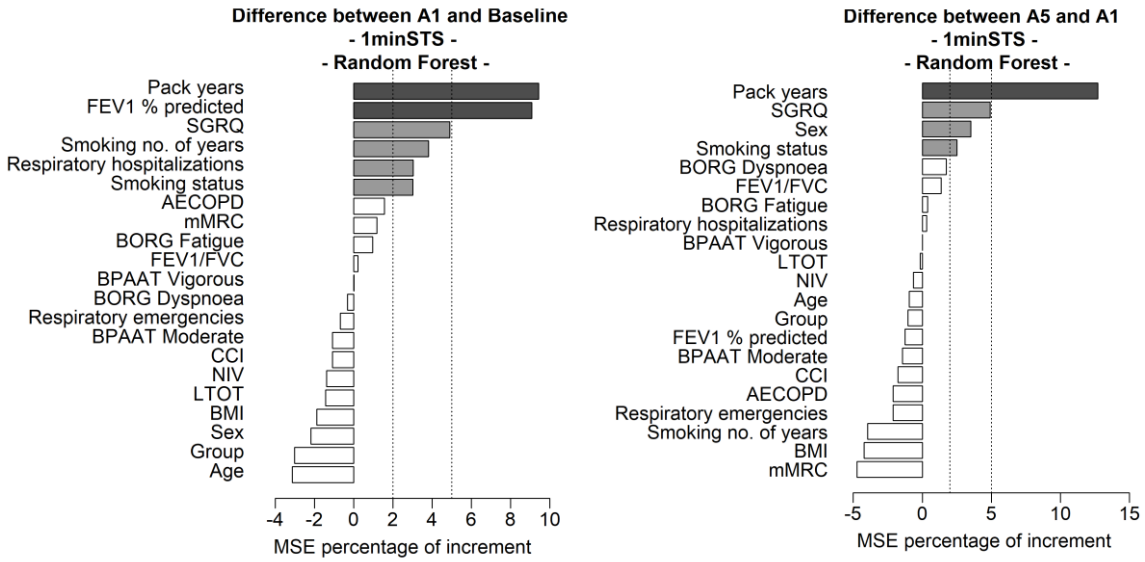

Figure 6.5 - *Feature importance given by the random forest model for the difference between consecutive assessments in the number of repetitions obtained by the one-minute sit-to-stand test. Light grey corresponds to the 2% cut-off and dark grey corresponds to the 5% cut-off.* ***Abbreviations:*** *1minSTS, one-minute sit-to-stand test; A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; AECOPD, acute exacerbation of COPD; BPAAT, brief physical activity assessment tool; BMI, body mass index; CCI, Charlson comorbidity index; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; LTOT, long-term oxygen therapy; MSE, mean squared error; mMRC, modified medical council dyspnoea scale; NIV, non-invasive ventilation; SGRQ, St. George's respiratory questionnaire.*

An R^2 of 0.77 was obtained for the RF model for the difference between the two first assessments while for the difference between the last two assessments the value was 0.74. Figure 6.6 shows the standardized predicted difference in the number of repetitions in the 1minSTS against the observed one.

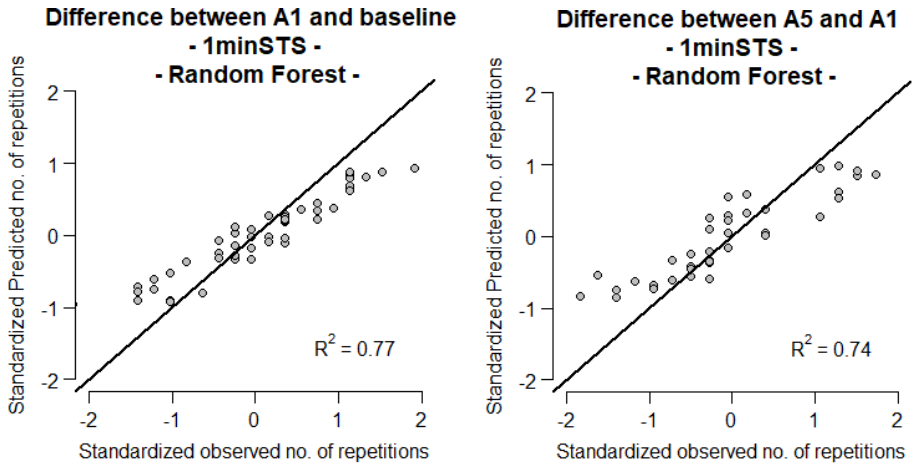

Figure 6.6 - *Standardized observed and predicted by the random forest algorithm difference of the number of repetitions of the one-minute sit-to-stand test between consecutive assessments.* ***Abbreviations:*** *1minSTS, one-minute sit-to-stand test; A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; R^2, R-squared.*

The LMMs with the features selected by the RF algorithm using the percentage of increase of MSE values of 2 and 5 as cut-off returned values of AIC of 942.052 and 960.919, respectively.

6.3.2.1.2. BORUTA

Boruta algorithm found two confirmed important features for the difference in the number of repetitions between A1 and baseline assessment. FEV_1 % of predicted and SGRQ had a mean importance of 5.2 and 3.2, respectively. Pack-years was classified as unconfirmed with a mean importance of 2.7. When the last difference of the number of repetitions was analysed, pack-years (7.2) and SGRQ (4.8) were found important features while sex (3.4) was classified as unconfirmed (Figure 6.7).

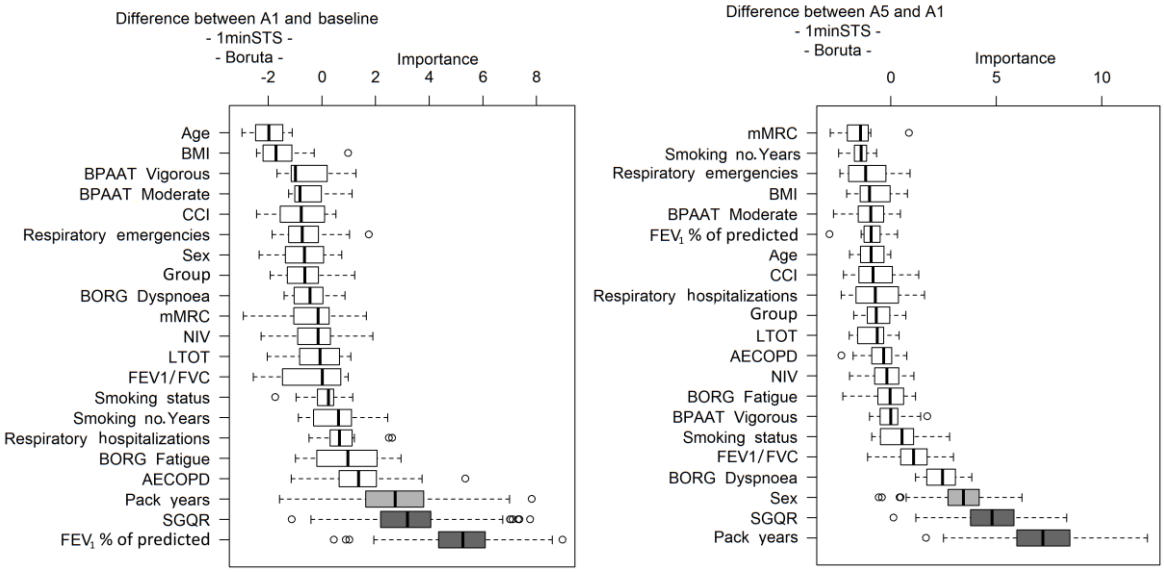

Figure 6.7 - *Feature importance given by the Boruta algorithm for the difference in the number of repetitions obtained by the one-minute sit-to-stand test between consecutive assessments. Light grey corresponds to the confirmed features and dark grey corresponds to the unconfirmed features.* ***Abbreviations:*** *1minSTS, one-minute sit-to-stand test; A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; AECOPD, acute exacerbation of COPD; BPAAT, brief physical activity assessment tool; BMI, body mass index; CCI, Charlson comorbidity index; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; LTOT, long-term oxygen therapy; mMRC, modified medical council dyspnoea scale; NIV, non-invasive ventilation; SGRQ, St. George's respiratory questionnaire.*

The LMM with the features selected by the Boruta algorithm using both unconfirmed and confirmed features returned an AIC of 960.542. When considering only the confirmed features the AIC increased to 962.294.

6.3.2.1.3. EXTREME GRADIENT BOOSTING

The hyperparameters' tunning for the XGB algorithm predicting the difference in the number of repetitions obtained by the 1minSTS between the A1 and the baseline assessment returned a minimum RMSE of 0.906 in the cross-validation test set. That value was obtained for a learning rate *eta* of 0.025 and was achieved at the maximum number of training iterations. It was considered a 40% of data to grow each tree, at a maximum depth of 5 and a minimum sum of instance weight needed in a child of 3. For the model predicting the difference between the last two assessments, the lowest RMSE in the cross-validation test set was also obtained at the maximum number of training iterations but considering 60% of data to grow each tree. A learning rate *eta* of 0.020, with a maximum depth of trees of 5 and a minimum sum of instance weight needed in a child of 3 were also registered (Table 6.3).

Table 6.3 – *Results from the hyperparameters tunning for the extreme gradient boosting model for the difference in the number of repetitions in the one-minute sit-to-stand test between consecutive assessments.*

	Eta	Maximum tree depth	Minimum child weight	Subsample ratio	Train set		Test set	
					Iteration number	Minimum RMSE	Iteration number	Minimum RMSE
A1 - Baseline	0.025	5	3	0.4	750	0.15907	197	0.90563
	0.025	8	3	0.4	750	0.15890	197	0.90563
	0.025	10	3	0.4	750	0.15890	197	0.90563
	0.025	11	3	0.4	750	0.15890	197	0.90563
	0.025	12	3	0.4	750	0.15890	197	0.90563
A5 – A1	0.020	5	3	0.6	750	0.06761	61	1.00457
	0.020	8	3	0.6	750	0.06763	61	1.00457
	0.020	10	3	0.6	750	0.06763	61	1.00457
	0.020	11	3	0.6	750	0.06763	61	1.00457
	0.020	12	3	0.6	750	0.06763	61	1.00457

Note: *Results are ordered by increasing minimum RMSE obtained in the cross-validation test set. Only the top 5 from the 252 combinations are presented for each difference.*

Abbreviations: *A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; Eta, learning rate; RMSE, root mean squared error*

Pack-years was the only feature with a gain superior to 0.15 in the XGB model for the difference between A1 and baseline. Five extra features were selected when the cut-off of 0.075 was considered, namely, SGRQ (0.14), FEV_1 % of predicted features (0.14), ratio

FEV/FVC (0.10), fatigue's BORG scale (0.08) and age (0.08). In the XGB predictive model of the difference in repetitions between the two last assessments, pack-years (0.24) and SGRQ (0.15) overcame the higher cut-off value and FEV_1 % of predicted (0.10) and BMI (0.09) the lower cut-off value (Figure 6.8).

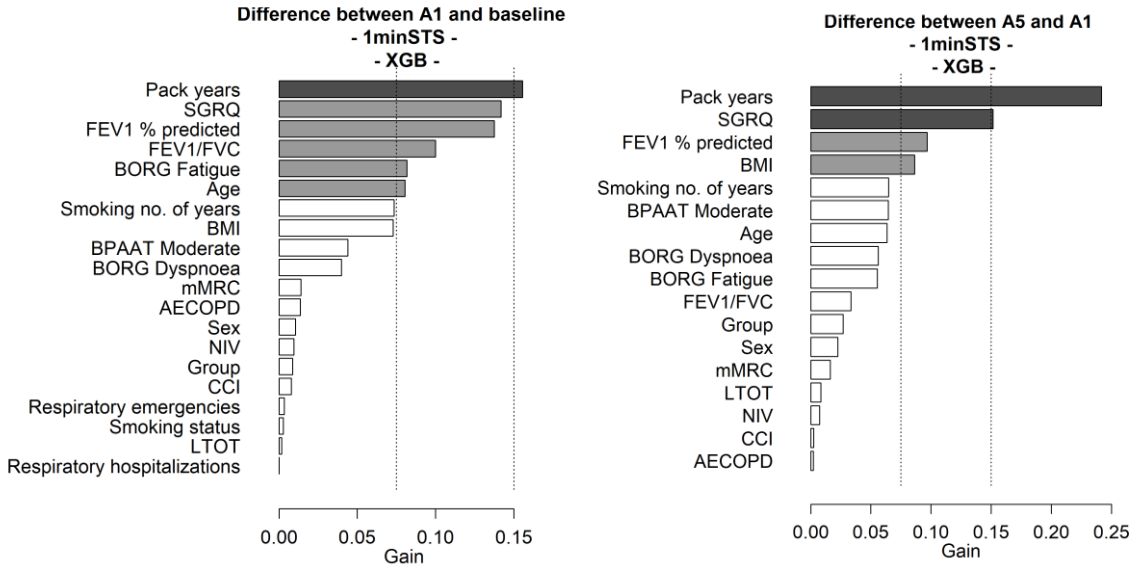

Figure 6.8 – *Feature importance given by the extreme gradient boosting model for the difference between consecutive assessments in the number of repetitions obtained by the one-minute sit-to-stand test. Light grey corresponds to the 0.075 cut-off and dark grey corresponds to the 0.15 cut-off.* ***Abbreviations:*** *1minSTS, one-minute sit-to-stand test; A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; AECOPD, acute exacerbation of COPD; BPAAT, brief physical activity assessment tool; BMI, body mass index; CCI, Charlson comorbidity index; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; LTOT, long-term oxygen therapy; mMRC, modified medical council dyspnoea scale; NIV, non-invasive ventilation; SGRQ, St. George's respiratory questionnaire; XGB; extreme gradient boosting.*

The R^2 obtained by the XGB model for the difference between the two first assessments was greater than the one obtained for the difference between the last two assessments (0.90 and 0.45, respectively) (Figure 6.9).

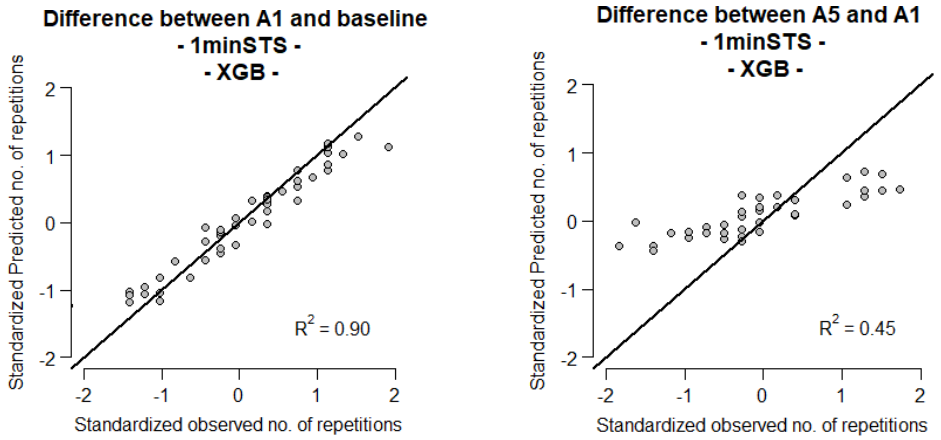

Figure 6.9 – *Standardized observed and predicted by the extreme gradient boosting algorithm difference of the number of repetitions of the one-minute sit-to-stand test between consecutive assessments.* ***Abbreviations:*** *1minSTS, one-minute sit-to-stand test; A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; R^2, R-squared; XGB, extreme gradient boosting.*

The LMM with the features selected by the XGB algorithm using the lowest gain value returned a value of AIC higher than the model obtain with the highest gain value (966.053 and 957.109, respectively).

6.3.2.1.4. L1-PENALIZED ESTIMATION

The optimal penalty parameter λ obtained was 378.9, which returned the lowest value of AIC of -123175 (Figure 6.10). Sex and respiratory emergencies in the previous year were selected as the features with an estimate coefficient different from 0 (Table 6.5).

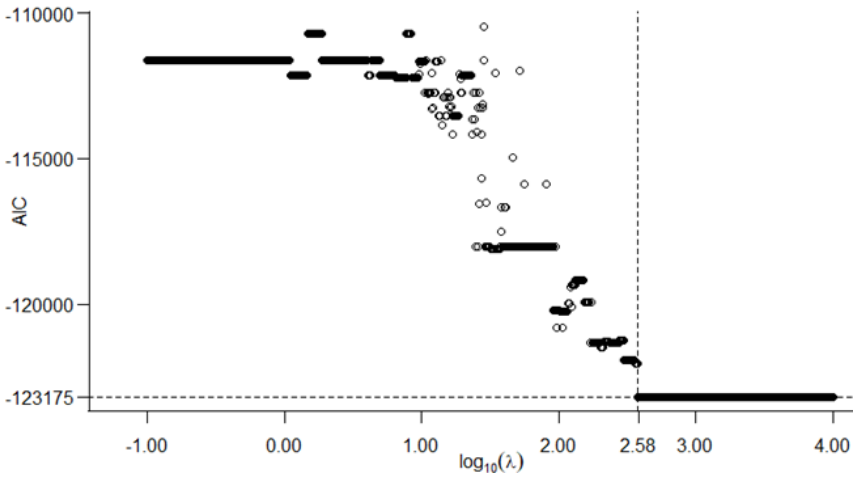

Figure 6.10. *Akaike information criterion (AIC) results for R glmmLasso as a function of the base 10 logarithm of the penalty parameter λ. The optimal value of $log_{10}(\lambda)$ is indicated by a vertical dashed line.*

The LMM with the features selected by the L1-penalized estimation algorithm returned an AIC of 947.842.

6.3.2.1.5. AUTOMATIC BACKWARD ELIMINATION

The automatic backward elimination algorithm removed sequentially 17 features starting with age and ending with the FEV_1/FVC ratio. Smoking status, pack-years, BPAAT moderate, group and evaluation were kept (Table 6.4).

Table 6.4 – *Order of the features' elimination by the Automatic Backward Elimination algorithm for the one-minute sit-to-stand test.*

Feature	Order of elimination	F value	p value
Age	1	0.0001	0.991
NIV	2	0.0007	0.979
LTOT	3	0.0163	0.899
BPAAT Vigorous	4	0.1322	0.718
BORG Fatigue	5	0.2155	0.645
mMRC	6	0.3241	0.573
CCI	7	0.5273	0.594
Respiratory hospitalizations	8	0.3551	0.554
FEV_1pp	9	0.3096	0.581
Smoking no. of years	10	0.3303	0.568
Sex	11	1.0757	0.265
Respiratory emergencies	12	1.6204	0.210
AECOPD	13	1.4279	0.250
SGRQ	14	1.9985	0.164
BMI	15	2.5788	0.115
BORG Dyspnoea	16	1.5328	0.221
FEV_1/FVC	17	3.9201	0.053
Smoking status	kept	11.4625	<0.001
Pack-years	kept	7.7154	0.008
BPAAT Moderate	kept	5.6814	0.021
Group	kept	5.4846	0.023
evaluation	kept	12.8615	<0.001

Abbreviations: *AECOPD, acute exacerbation of COPD; BPAAT, brief physical activity assessment Tool; BMI, body mass index; CCI, Charlson comorbidity index; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; LTOT, long-term oxygen therapy; mMRC, modified medical council dyspnoea scale; NIV, non-invasive ventilation; SGRQ, St. George's respiratory questionnaire.*

The LMM obtained with the selected features achieved an AIC of 931.793.

6.3.2.1.6. SUMMARY

The results obtained by the different feature selection algorithms used with the number of repetitions of the 1minSTS in patients with COPD, and described in the previous chapters, are summarized in Table 6.5. The LMM using as predictors the features selected by the

automatic backward elimination algorithm achieved the lowest AIC of 931.793 followed by the one which features were selected by the RF algorithm with a cut-off of 2%.

Table 6.5 – *Feature selection algorithms' results for the one-minute sit-to-stand test*

Features	Random Forest		Boruta		Extreme gradient boosting		L1 penalized estimation	Automatic backward elimination
	A1-B	A5-A1	A1-B	A5-A1	A1-B	A5-A1		
Sex								
Age								
BMI								
Smoking status								
Smoking no. of years								
Pack-years								
SGRQ								
BPAAT Moderate								
BORG Fatigue								
FEV_1 % predicted								
FEV_1/FVC								
Group	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ
Respiratory emergencies								
Respiratory hospitalizations								
Evaluation	#	#	#	#	#	#	Δ	Δ
Group*evaluation	#	#	#	#	#	#	#	#
Linear mixed-effects model AIC	942.052	960.919	960.545	962.294	966.053	957.109	947.842	931.793

Note: *Long-term oxygen therapy, non-invasive ventilation, CCI, mMRC, BPAAT vigorous, BORG dyspnoea and AECOPD were not included because no algorithm selected these features. Light grey corresponds to 2%, unconfirmed+confirmed and 0.075 cut-offs in the RF, Boruta and XGB algorithms, respectively. Dark grey corresponds to 5%, confirmed and 0.15 cut-offs in the RF, Boruta and XGB algorithms, respectively.*

Abbreviations: *A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; AECOPD, acute exacerbation of COPD; AIC, Akaike information criterion; B, baseline assessment; BPAAT, brief physical activity assessment tool; BMI, body mass index; CCI, Charlson comorbidity index; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; LTOT, long-term oxygen therapy; mMRC, modified medical council dyspnoea Scale; NIV, non-invasive ventilation; RF; random forest; SGRQ, St. George's respiratory questionnaire; XGB; extreme gradient boosting. "#" not included as a feature in the algorithm but included in all linear mixed-effects models; "Δ", included as a feature in the algorithm and in all linear mixed-effects models; "*", interaction.*

In all three algorithms where two cut-off values were considered, the lowest AICs were achieved when the features included in the LMM were selected using the less restrictive cut-off value.

6.3.2.2. HANDGRIP MUSCLE STRENGTH

The LMM using as dependent variable the HMS and as predictors the features selected by the automatic backward elimination algorithm achieved the lowest AIC of 931.793 followed by the one which features were selected by the Random Forest algorithm with a cut-off of 2% (Table 6.6).

Table 6.6 – *Feature selection algorithms' results for the handgrip muscle strength*

Features	Random Forest		Boruta		Extreme gradient boosting		L1 penalized estimation	Automatic backward elimination
	A1-B	A5-A1	A1-B	A5-A1	A1-B	A5-A1		
Sex								
Age								
BMI								
Smoking status								
Smoking no. of years								
Pack-years								
SGRQ								
BPAAT Moderate								
BORG Fatigue								
FEV_1 % predicted								
FEV_1/FVC								
AECOPD								
Group	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ
Respiratory emergencies								
Evaluation	#	#	#	#	#	#	Δ	Δ
Group*evaluation	#	#	#	#	#	#	#	#
Linear mixed-effects model AIC	956.629		976.651		1004.284 992.055		961.453	950.224

Note: *Long-term oxygen therapy, non-invasive ventilation, CCI, mMRC, respiratory hospitalizations, BPAAT vigorous and BORG dyspnoea were not included because no algorithm selected these features. Light grey corresponds to 2%, unconfirmed+confirmed and 0.075 cut-offs in the RF, Boruta and XGB algorithms, respectively. Dark grey corresponds to 5%, confirmed and 0.15 cut-offs in the RF, Boruta and XGB algorithms, respectively.*

Abbreviations: *A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; AECOPD, acute exacerbation of COPD; AIC, Akaike information criterion; B, baseline assessment; BPAAT, brief physical activity assessment tool; BMI, body mass index; CCI, Charlson comorbidity index; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; LTOT, long-term oxygen therapy; mMRC, modified medical council dyspnoea Scale; NIV, non-invasive ventilation; RF; random forest; SGRQ, St. George's respiratory questionnaire; XGB; extreme gradient boosting. "#" not included as a feature in the algorithm but included in all linear mixed-effects models; "Δ", included as a feature in the algorithm and in all linear mixed-effects models; "*", interaction.*

The RF algorithm returned no feature with a percentage of increase of mean square error greater than 5%. The same occurred for the Boruta algorithm, with no unconfirmed featured identified. The AIC obtained by the LMM that used the features generated by the XGB algorithm with the highest cut-off value was lesser than the one obtained with the other features.

6.3.2.3. COPD ASSESSMENT TEST

The feature selection algorithms' results for the CAT score are summarized in Table 6.7.

Table 6.7 - *Feature selection algorithms' results for the COPD assessment test.*

Features	Random Forest		Boruta		Extreme gradient boosting		L1 penalized estimation	Automatic backward elimination
	A1-B	A5-A1	A1-B	A5-A1	A1-B	A5-A1		
Sex								
Age								
BMI								
Smoking No years								
Pack-years								
LTOT								
CCI								
SGRQ								
mMRC								
BPAAT Moderate								
BORG Fatigue								
FEV_1 % predicted								
FEV_1/FVC								
Group	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ
Respiratory emergencies								
Respiratory hospitalizations								
Evaluation	#	#	#	#	#	#	Δ	Δ
Group*evaluation	#	#	#	#	#	#	#	#
Linear mixed-effects model AIC	921.643 945.063		918.436 945.063		910.155 957.898		948.016	814.885

Note: *Smoking status, non-invasive ventilation, BPAAT vigorous, AECOPD and BORG dyspnoea were not included because no algorithm selected these features. Light grey corresponds to 2%, unconfirmed+confirmed and 0.075 cut-offs in the RF, Boruta and XGB algorithms, respectively. Dark grey corresponds to 5%, confirmed and 0.15 cut-offs in the RF, Boruta and XGB algorithms, respectively.*

Abbreviations: *A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; AECOPD, acute exacerbation of COPD; AIC, Akaike information criterion; B, baseline assessment; BPAAT, brief physical activity assessment tool; BMI, body mass index; CCI, Charlson comorbidity index; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; LTOT, long-term oxygen therapy; mMRC, modified medical council dyspnoea Scale; NIV, non-invasive ventilation; RF; random forest; SGRQ, St. George's respiratory questionnaire; XGB; extreme gradient boosting. "#" not included as a feature in the algorithm but included in all linear mixed-effects models; "Δ", included as a feature in the algorithm and in all linear mixed-effects models; "*", interaction.*

The LMM using as predictors the features selected by the automatic backward elimination algorithm achieved the lowest AIC of 814.885 followed by the one which features were selected by the XGB algorithm with a cut-off gain of 0.075.

Similarly to the 1minSTS, for CAT, the lowest AICs were achieved when the features included in the LMM were selected using the less restrictive cut-off value.

6.3.3. SELECTED LINEAR MIXED-EFFECTS MODELS

We fitted three LMMs to predict each of the previously considered outcomes (Table 6.8).

Table 6.8 – *Features associated with the number of repetitions in the one-minute sit-to-stand test, the handgrip muscle strength and the COPD assessment test score in people with chronic obstructive pulmonary disease, over time.*

Features	One-minute sit-to-stand (n=59)			Handgrip muscle strength(n=59)			CAT (n=59)		
	$b_{1minSTS}$	CI 95%	p	b_{HMS}	CI 95%	p	b_{CAT}	CI 95%	p
(Intercept)	16.68	[10.63; 22.73]	**<0.001**	57.08	[41.28; 72.88]	**<0.001**	-13.22	[-20.84; -5.60]	**0.001**
Smoking status [former]	11.97	[5.84; 18.10]	**<0.001**	-	-	-	-	-	-
Smoking status [current]	18.55	[10.64; 26.46]	**<0.001**	-	-	-	-	-	-
Pack-years	-0.05	[-0.08; -0.01]	**0.007**	-	-	-	-	-	-
BPAAT Moderate	1.18	[0.19; 2.17]	**0.020**	-	-	-	-0.60	[-1.07; -0.13]	**0.013**
Sex [Male]	-	-	-	15.86	[10.93; 20.79]	**<0.001**	-	-	-
Age	-	-	-	-0.54	[-0.78; -0.31]	**<0.001**	0.19	[0.10; -0.29]	**<0.001**
SGRQ	-	-	-	-	-	-	0.25	[0.21; 0.30]	**<0.001**
BORG Fatigue	-	-	-	-	-	-	1.00	[0.44; 1.55]	**0.001**
FEV_1 % of predicted	-	-	-	-	-	-	0.05	[0.01; 0.08]	**0.012**
Respiratory hospitalizations [Yes]	-	-	-	-	-	-	4.58	[1.60; 7.56]	**0.003**
Group [Lockdown]	-4.06	[-7.56; -0.56]	**0.024**	-0.11	[-4.05; 3.84]	0.958	-0.04	[-1.91; 1.83]	0.966
Assessment [A1]	2.42	[0.70; 4.14]	**0.006**	0.89	[-0.82; 2.60]	0.306	-1.46	[-2.77; -0.16]	**0.029**
Assessment [A5]	3.11	[1.22; 5.01]	**0.002**	2.61	[0.64; 4.59]	**0.010**	-3.47	[-4.92; -2.02]	**<0.001**
Group [Lockdown] * Assessment [A1]	0.83	[-1.92; 3.58]	0.551	0.23	[-2.53; 2.98]	0.871	-0.11	[-2.20; 1.97]	0.915
Group [Lockdown] * Assessment [A5]	0.08	[-2.75; 2.91]	0.955	0.07	[-2.79; 2.93]	0.961	1.47	[-0.63; 3.57]	0.169
Random Effects σ^2	11.79			11.92			6.82		
τ_{00}	$30.49_{Participant}$			$44.06_{Participant}$			$4.87_{Participant}$		
Marginal R^2 / Conditional R^2	0.36 / 0.82			0.44 / 0.88			0.75 / 0.86		

Note: ** indicates "interaction with".*

Abbreviations: *1minSTS, one-minute sit-to-stand Test; A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; CI, confidence interval approximated by Kenward-Roger method; p, p value approximated by Kenward-Roger method; df, degrees of freedom approximated by Kenward-Roger method; BMI, body mass index; CCI, Charlson comorbidity index; mMRC, modified medical council dyspnoea scale; FEV_1 % of predicted, percentage of the predicted forced expiratory volume in 1 second; σ^2, residual variance; τ, random effect standard deviation; ρ, correlation between intercept and slope; R^2, coefficient of determination.*

The LMMs included the variable participant as random effects (*lmer* formula: ~1 | participant). The models' total explanatory power was substantial (conditional R^2 > 0.80) with the CAT's LMM achieving the highest value (conditional R^2 = 0.88). All the LMM assumptions were satisfied (Appendix A).

The predicted number of repetitions at baseline was 28.0 (CI95% = [25.1;30.8]) for the pre-lockdown group and 22.7 (CI95%=[19.4;25.9]) for the lockdown group (Table 6.9). A statistically significant increase of 2.3 repetitions was predicted after one month in the pre-lockdown group and a clinical and statistically significant increase of 3 repetitions was predicted in the lockdown group. No difference was detected between assessment 1 and assessment 5 in either groups suggesting that the lockdown period had no effect in the 1minSTS behaviour.

Table 6.9 – *Pairwise comparisons between assessments within each group and estimated marginal means of the number of repetitions in the one-minute sit-to-stand test for the different groups and assessments.*

	Pairwise comparisons				**Estimated marginal means**			
	contrast	**estimate**	**SE**	**p-value**	**evaluation**	**EMM**	**SE**	**95% CI**
Pre-lockdown	B-A1	-2.342	0.867	**0.022**	B	28.0	1.16	[25.1; 30.8]
	B-A5	-3.004	0.956	**0.006**	A1	30.3	1.18	[27.4; 33.2]
	A1-A5	-0.662	0.970	0.774	A5	31.0	1.25	[27.9; 34.0]
Lockdown	B-A1	-3.074	1.084	**0.015**	B	22.7	1.35	[19.4; 25.9]
	B-A5	-3.072	1.064	**0.013**	A1	25.7	1.43	[22.2; 29.2]
	A1-A5	0.002	1.148	1.000	A5	25.7	1.41	[22.3; 29.2]

Note: *Results are averaged and weighted in proportion to the frequencies (in the original data) over the levels of sex, Charlson comorbidity index and respiratory emergencies in the previous year. Degrees-of-freedom method: kenward-roger; p-value adjustment: tukey method for comparing a family of 3 estimates.*

Abbreviations: *A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; B, baseline assessment; EMM, estimated marginal mean; SE, standard error; CI, confidence Interval.*

The baseline and the final assessment predicted an HMS of approximately 34 kg and 36.6 kg respectively, in both groups (Table 6.10). The difference between the two assessments was statistically significant (p = 0.027 for the pre-lockdown and p = 0.031 for the lockdown group) but not clinically meaningful. No statistically significant difference was found between assessment 1 and assessment 5 in either groups suggesting that the lockdown period had no effect in the HMS behaviour.

Table 6.10 – *Pairwise comparisons between assessments within each group and estimated marginal means of the handgrip muscle strength for the different groups and assessments.*

	Pairwise comparisons				Estimated marginal means			
	contrast	**estimate**	**SE**	**p-value**	**evaluation**	**EMM**	**SE**	**95% CI**
Pre-lockdown	B-A1	-0.886	0.861	0.560	B	34.0	1.29	[30.9; 37.2]
	B-A5	-2.615	0.993	**0.027**	A1	34.9	1.30	[31.7; 38.1]
	A1-A5	-1.729	0.995	0.197	A5	36.6	1.39	[33.2; 40.0]
Lockdown	B-A1	-1.111	1.086	0.564	B	33.9	1.50	[30.2; 37.6]
	B-A5	-2.686	1.045	**0.031**	A1	35.0	1.57	[31.2; 38.8]
	A1-A5	-1.575	1.145	0.358	A5	36.6	1.54	[32.8; 40.4]

Note: *Results are averaged and weighted in proportion to the frequencies (in the original data) over the levels of sex. Degrees-of-freedom method: kenward-roger; p-value adjustment: tukey method for comparing a family of 3 estimates.*
Abbreviations: *A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; B, baseline assessment; EMM, estimated marginal mean; SE, standard error; CI, confidence Interval.*

In both groups, the score predicted at baseline for CAT was 11.6 and the difference between the first and last assessments was statistically and clinically significant ($p < 0.001$ for the pre-lockdown and $p = 0.028$ for the lockdown group) (Table 6.11). No statistically significant difference was found between A1 and A5 in the lockdown group but in the pre-lockdown group. This suggests that the lockdown period had a negative effect in the CAT behaviour.

Table 6.11 – *Pairwise comparisons between assessments within each group and estimated marginal means of the COPD assessment test score for the different groups and assessments.*

	Pairwise comparisons				Estimated marginal means			
	contrast	**estimate**	**SE**	**p-value**	**evaluation**	**EMM**	**SE**	**95% CI**
Pre-lockdown	B-A1	1.464	0.659	0.073	B	11.6	0.61	[10.1; 13.1]
	B-A5	3.470	0.732	**<0.001**	A1	10.1	0.62	[8.6; 11.6]
	A1-A5	2.006	0.736	**0.021**	A5	8.1	0.70	[6.5; 9.8]
Lockdown	B-A1	1.577	0.818	0.136	B	11.6	0.71	[9.9; 13.3]
	B-A5	2.003	0.765	**0.028**	A1	10.0	0.78	[8.1; 11.9]
	A1-A5	0.426	0.829	0.865	A5	9.6	0.72	[7.8; 11.3]

Note: *Results are averaged and weighted in proportion to the frequencies (in the original data) over the levels of respiratory hospitalizations in the previous year. Degrees-of-freedom method: kenward-roger; p-value adjustment: tukey method for comparing a family of 3 estimates.*
Abbreviations: *A1, assessment 1 month after baseline assessment; A5, assessment 5 months after baseline assessment; B, baseline assessment; EMM, estimated marginal mean; SE, standard error; CI, confidence Interval.*

6.4. DISCUSSION

Optimise the trade-off between the fit of a model and the model's complexity is one of the objectives of feature selection. Automatic backward elimination algorithms have come under criticism since they can overestimate the effect size of significant predictors (Whittingham et al., 2006). Despite the doubts concerning this method, in our study, it showed to be able to select features that produced the lowest AICs in LMMs.

The clinically significant differences found suggest that the lockdown period imposed by the COVID-19 pandemic had a negative impact in the wellbeing and daily life of patients with COPD, which was measured with CAT. The reason of this aggravation is beyond the scope of this study but should be given special attention in further studies so that, in future similar conditions, healthcare professionals can help improve the management of COPD and get the greatest benefit from treatment.

The strengths of our study include the consistence of the results since three different independent variables were considered with the same result and the high explanatory power of the LMMs computed. Limitations of this study include the low number of features considered and the fact that RF, Boruta and XGB algorithm considered the behaviour of the outcomes by consecutive steps and not as a whole.

Future studies should also explore these approaches but with a higher dimension of features to select from and include other pattern recognition and machine learning algorithms that take into account the longitudinal nature of data, for instance, the RF approach suggested by Capitaine (Capitaine et al., 2020) or the Gaussian Process Boosting (Sigrist, 2020).

6.5. CONCLUSION

Automatic backward elimination of features showed to be consistent when it came to select statistically relevant features that would compute LMMs with the lowest AIC. This study also showed that the COVID-19 lockdown period had no effect in the 1minSTS and HMS behaviour but a negative effect in the impact of the disease was observed in people with COPD.

7. STUDY 2 – BEHAVIOUR OF THE ONE-MINUTE SIT-TO-STAND TEST DURING SIX MONTHS IN PEOPLE WITH COPD*

*** The results from this study were submitted in the form of an extended abstract to the 3rd Statistics on Health Decision Making: Public Health, to be held at the University of Aveiro on the 22nd of July 2021 (Appendix C).**

7.1. RATIONALE

Functional status is highly meaningful to people with COPD although this outcome has been overlooked (Bui et al., 2017). Numerous field tests might be used to assess functional status (Vaidya et al., 2017), yet the 1minSTS has shown to be a simple test that mimics the common activity of sitting/standing from a chair which is essential to maintain independence among the elderly (Vaidya et al., 2016). Additionally, it is a valid and responsive measure that might be easily performed for follow-up assessment of people with COPD (Vaidya et al., 2017). In fact, regular assessment of people with COPD is essential (Global Initiative for Chronic Obstructive Lung Disease, 2021) and this study hypothesise that the 1minSTS might be an important indicator of functional status over time in people with COPD.
Thus, this study aimed to describe the 1minSTS behaviour over a six-month period and explore the factors influencing this behaviour in people with COPD.

7.2. STATISTICAL METHODS

An observational, prospective cohort study was conducted with data described in chapters 5.2 and 5.3. Quantitative variables were summarized using mean and standard deviation values if normally distributed or median values and interquartile ranges, otherwise. Categorical variables were summarized through count values and percentages. Shapiro-Wilk test was used to assess the assumption of normality.
LMMs were applied to assess the mean change in the dependent variable, 1minSTS, and to account for correlations in repeated measurements per participant. Random intercepts and slopes were used to incorporate individual response trajectories. Firstly, a backward elimination with single random-effect terms deletion was performed. Secondly, a model with the random effects kept was computed with backward elimination of single main fixed-effect

terms and their interaction with time (Zuur et al., 2009). Time, defined, in days, as a continuous variable starting at the date of the baseline assessment was kept in this process. The p-values were computed based on conditional F-tests with Kenward-Roger approximation for the degrees of freedom (Kenward & Roger, 1997). Marginal coefficient of determination (marginal R^2) and conditional coefficient of determination (conditional R^2) were determined for assessing the model's quality and level of adjustment according to Nakagawa (Nakagawa et al., 2017). LMM assumptions were assessed by visual inspection of residuals' boxplots, scatterplots, and Q-Q plots. EMM were determined for the reference levels of categorical variables and mean values of quantitative variables (Lüdecke, 2018). Two-sided $P < 0.05$ was considered statistically significant for all analyses.

Statistical analyses were performed using R *lmer* (Bates et al., 2015)*, lmerTest* (Kuznetsova et al., 2017) and *ggeffects* (Lüdecke, 2018) in RStudio Version 1.4.1103 (RStudio Team, 2021) running R version 4.0.5 (R Core Team, 2021).

7.3. RESULTS

A total of 149 participants with COPD were included. Participants mean age was 67.5 (Standard deviation 9.0) years, most were men (83.9%), slightly overweight (BMI=26.8 kg/m^2), former smokers (73.8%), presented severe airway obstruction (49.0 [38.0;70.0]), 3 to 4 comorbidities (53%) and were not under PR effect (72.5%). Additionally, most participants did not have an AECOPD in the previous year (71.1%), presented a GOLD grade 3 (39.6%) and a GOLD group B (48.3%) and had a median mMRC score of 2 [1;3], Further baseline characteristics are presented in Table 7.1.

Table 7.1 – *Baseline characteristics of participants (n=149).*

Characteristic	Categories	Measure
Age, *years*, mean (SD)		67.5 (9.0)
Sex	Female	24 (16.1)
	Male	125 (83.9)
Educational level	Without formal education	4 (2.7)
	4th degree	66 (44.3)
	6th degree	17 (11.4)
	9th degree	19 (12.8)
	12th degree	19 (12.8)
	College/Higher education	24 (16.1)
Marital status	Single	7 (4.7)
	Common-law marriage	115 (77.2)
	Married	16 (10.7)
	Divorced	4 (2.7)
	Widowed	7 (4.7)
Current occupation	Paid work	27 (18.1)

Characteristic	Categories	Measure
	Housework	1 (0.7)
	Unemployed (health reasons)	3 (2.0)
	Unemployed (other reasons)	8 (5.4)
	Retired	109 (73.2)
	Other	1 (0.7)
BMI, *kg/m²*, mean (SD)		26.8 (4.6)
Smoking status	Never	20 (13.4)
	Former	110 (73.8)
	Current	19 (12.8)
Smoking no. of years, median [IQR]		36.0 [20.0; 45.0]
Pack-years, median [IQR]		42.0 [15.0; 75.0]
Under PR effect during follow-up	No	108 (72.5)
	Yes	41 (27.5)
CCI*, score*	1-2	31 (20.8)
	3-4	79 (53.0)
	>=5	39 (26.2)
AECOPD*, in the previous year*	0	106 (71.1)
	1	19 (12.8)
	>1	24 (16.1)
FEV_1, *% predicted*, median [IQR]		49.0 [38.0; 70.0]
FEV_1/FVC, *%*, median [IQR]		53.0 [40.0; 63.0]
GOLD grades	1	25 (16.8)
	2	47 (31.5)
	3	59 (39.6)
	4	18 (12.1)
CAT, *points*, median [IQR]		12.0 [8.0; 18.0]
GOLD ABCD classification tool	A	49 (32.9)
	B	72 (48.3)
	C	2 (1.3)
	D	26 (17.5)
mMRC, *points*, median [IQR]		2 [1.0;3.0]
SGRQ Total, median [IQR]		40.5 [26.6; 56.9]
BPAAT Moderate, *score*	0	55 (39.9)
	1	27 (18.1)
	2	23 (15.4)
	4	44 (29.5)
BPAAT Vigorous, *score*	0	119 (79.9)
	2	24 (16.1)
	4	6 (4.0)
Modified BORG scale (Dyspnoea), score	0	98 (65.8)
	0.5	2 (1.3)
	1	15 (10.1)
	2	13 (8.7)
	3	13 (8.7)
	4	3 (2.0)
	5	2 (1.3)
	6	2 /1.3)
	7	1 (0.7)
Modified BORG scale (Fatigue), score	0	89 (59.7)
	0.5	3 (2.0)
	1	11 (7.4)
	2	13 (8.7)
	3	21 (14.1)
	4	7 (4.7)
	5	2 (1.3)
	6	2 (1.3)
	7	1 (0.7)

Characteristic	Categories	Measure
Long-term oxygen therapy	No	123 (82.6)
	Yes	26 (17.4)
Non-invasive ventilation	No	124 (83.2)
	Yes	25 (16.8)
Respiratory related hospitalizations, *in the previous year*	0	134 (89.9)
	1	13 (8.7)
	2	2 (1.4)
Respiratory related emergencies, *in the previous year*	0	122 (81.9)
	1	22 (14.7)
	2	3 (2.0)
	3	2 (1.4)
Years' quarter	1st	87 (58.4)
	2nd	0 (0.0)
	3rd	0 (0.0)
	4th	62 (41.6)
HMS, *kg* median [IQR]		35.0 [26.0; 41.0]
1minSTS, repetitions, median [IQR]		26.0 [21.0; 30.0]

Note: *Data presented as n (%), unless otherwise stated.*
Abbreviations: *1minSTS, one-minute sit-to-stand test; AECOPD, acute exacerbation of COPD; BMI, body mass index; BPAAT, brief physical activity assessment tool; CAT, COPD assessment test; CCI, Charlson comorbidity index; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; GOLD, global initiative for chronic obstructive lung disease; HMS, handgrip muscle strength; IQR, Interquartile range; mMRC, modified medical council dyspnoea scale; PR, pulmonary rehabilitation; SD, standard deviation; SGRQ, St. George's respiratory questionnaire.*

Participants' 1minSTS median number of repetitions at baseline was 26.0 [21.0;30.0] and a significant increase was observed reaching a median of 30.0 [24.0;37.5] repetitions at A5 (p<0.001) (Figure 7.1).

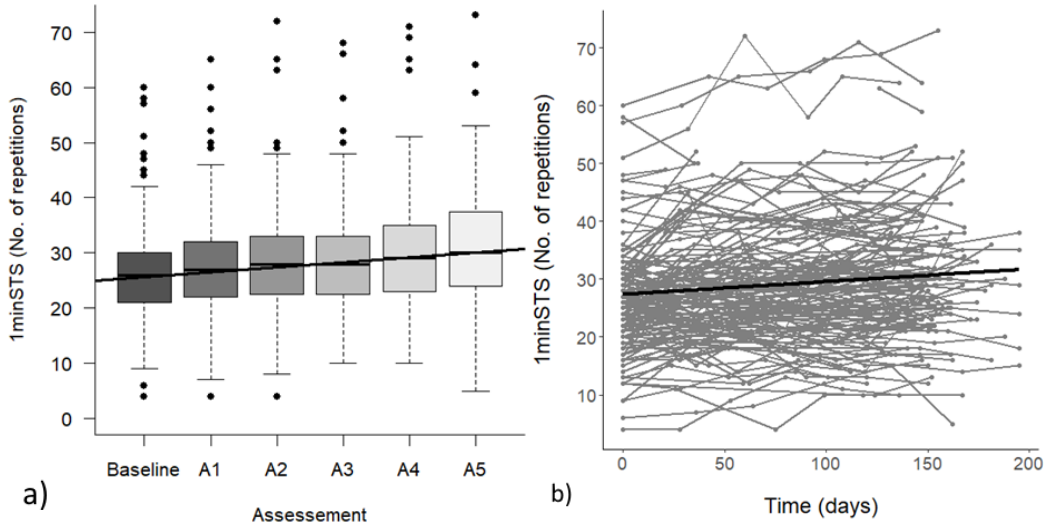

Figure 7.1 – *Description of the number of repetitions in the one-minute sit-to-stand (1minSTS) over time of participants with chronic obstructive pulmonary disease (n=149). (a) In the left panels, time was considered as a categorical variable, corresponding to the number of assessments; black line represents the linear regression line of the median values. (b) In the right panel, time was defined as number of days between baseline and follow-up assessments; the black line represents the global linear regression line.*

The 1minSTS model's total explanatory power was substantial (conditional R^2 0.92). The effect of time was positive and statistically significant (b = 0.08618, p < 0.001). The 1minSTS LMM predicted an increase of 3.8 repetitions after 195 days of follow up (Figure 7.2).

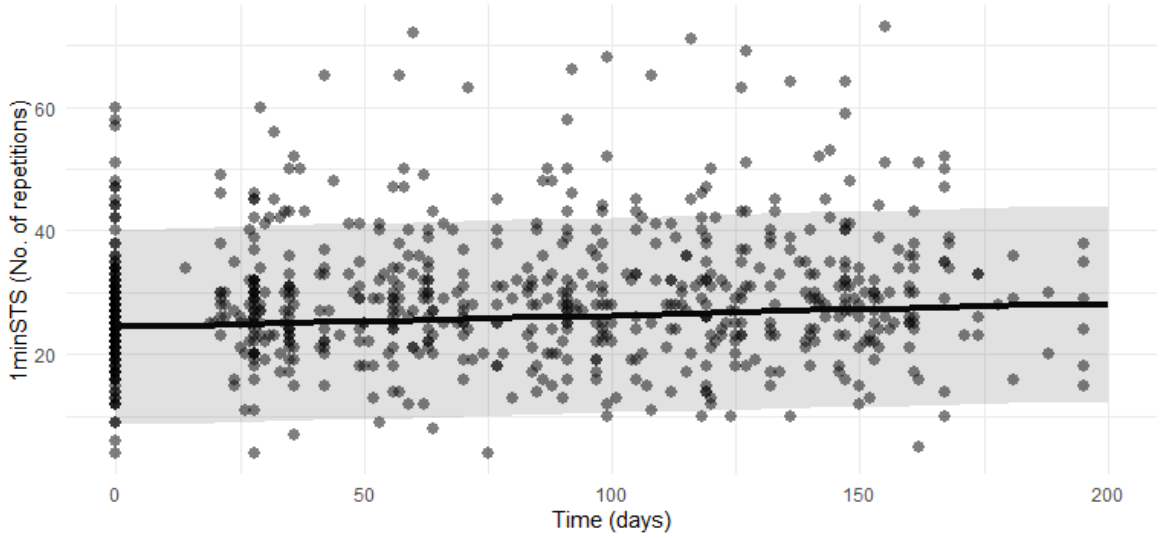

Figure 7.2 – *Predicted number of repetitions of the one-minute sit-to-stand test (1minSTS) of participants with chronic obstructive pulmonary disease over time. Adjusted for participants' mean characteristics values and reference categories. The linear mixed-effects model's predicted values are represented by a black line, 95% confidence prediction intervals by a grey band and observed values by dots.*

Table 7.2 shows the results of the LMM including time (continuous) and participant as random effects. Older participants (b = -0.56, p < 0.001) and higher scores of mMRC (b = -2.04, p = 0.001) were predicted with a statistically significant lower number of repetitions globally. Generally, participants with higher BMI and higher pack-years were expected to achieve a lower number of repetitions (b = -0.55, p < 0.001 and b = -0.03, p = 0.033, respectively). A difference of 1 repetition was predicted for a variation of 13.7% points in FEV_1 % of predicted (b = 0.07, p = 0.021). Females were expected to obtain approximately less 5 repetitions when compared with males (b = -4.69, p = 0.009).

Table 7.2 – *Factors associated with the number of repetitions of one-minute sit-to-stand test in people with chronic obstructive pulmonary disease, over time (n=149)*

	1minSTS (n=149)			
Factors	**Estimate (b)**	**CI 95%**	**p**	**df**
(Intercept)	77.56	[62.40; 92.71]	**<0.001**	140.70
Sex [Female]	-4.69	[-8.20; -1.18]	**0.009**	140.59
Age	-0.56	[-0.77; -0.34]	**<0.001**	141.39
BMI	-0.55	[-0.81; -0.28]	**<0.001**	139.05
Pack-years	-0.03	[-0.06; -0.01]	**0.033**	142.38
CCI [Moderate (3-4)]	4.80	[0.71; 8.89]	**0.022**	139.13
CCI [Severe (>=5)]	3.57	[-2.15; 9.28]	0.219	140.26
mMRC	-2.04	[-3.25; -0.83]	**0.001**	140.34
FEV_1 % of predicted	0.07	[0.07; 0.14]	**0.021**	139.75
Time	86.18 [a]	[44.05; 128.31] [a]	**<0.001**	132.41
Age*Time	-0.86[a]	[-1.48; -0.24] [a]	**0.007**	134.32
Pack-years*Time	-0.17 [a]	[-0.30; -0.05] [a]	**0.006**	133.37
Random Effects				
σ^2	0.08			
τ_{00}	$0.53_{Participant}$			
τ_{11}	$0.01_{Participant.Time}$			
ρ_{01}	$0.64_{Participant}$			
Observations	755			
Marginal R^2/Conditional R^2	0.41 / 0.92			

Abbreviations: *1minSTS, one-minute sit-to-stand test; CI, confidence interval approximated by Kenward-Roger method; p, p value approximated by Kenward-Roger method; df, degrees of freedom approximated by Kenward-Roger method; BMI, body mass index; CCI, Charlson comorbidity index; mMRC, modified medical council dyspnoea scale; FEV_1 % of predicted, percentage of the predicted forced expiratory volume in 1 second; σ^2, residual variance; τ, random effect standard deviation; ρ, correlation between intercept and slope; ICC, intraclass correlation coefficient; R^2, coefficient of determination. * indicates "interaction with". [a] multiplied by 10^{-3}.*

The interaction effect of time on age and time on pack-years in the 1minSTS LMM was negative and statistically significant (b = -0.00086, p = 0.007 and b = -0.00017, p = 0.006, respectively). Thus, older participants or/and participants that were heavy smokers were expected to increase less, or eventually even decrease their number of repetitions when compared with younger participants or/and participants with low smoking load (Figure 7.3).

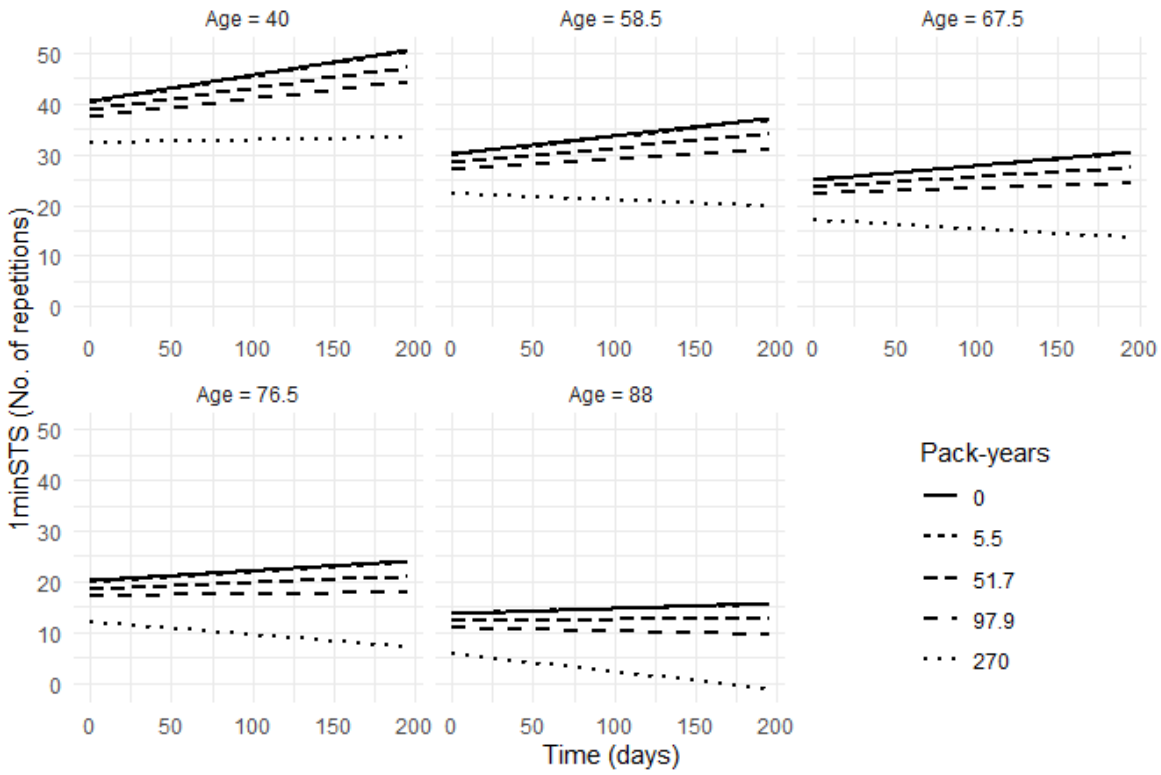

Figure 7.3 – *Predicted values of the number of repetitions of the one-minute sit-to-stand test (1minSTS) of participants with chronic obstructive pulmonary disease, at different ages (years) and pack-years' values over time, in days. Adjusted for male participants aged 67.5 years old, body mass index of 26.84, Charlson comorbidity index score of 1-2, forced expiratory volume in 1 second % of predicted of 53.31 and modified British medical research council questionnaire score of 2.*

For instance, a non-smoker male with 67.5 years is expected to increase of 5.5 repetitions whereas a 78.5-year-old male with a pack-years of 270 is expected to have an increase of only 0.67 repetitions over a period of 195 days (Table 7.3).

Table 7.3 – Predicted number of repetitions of the one-minute sit-to-stand test over time in people with chronic obstructive pulmonary disease according to the explanatory factors (age and pack-years).

Pack-years	Time	Age (years)				
		40 (minimum)	**58.5 (mean–1SD)**	**67.5 (mean)**	**78.5 (mean+1SD)**	**88 (maximum)**
		Estimate [95% CI]	Estimate [95% CI]	Estimate [95% CI]	Estimate [95% CI]	Estimate [95% CI]
0 (minimum)	0	40.46 [24.39; 56.53]	30.19 [14.52; 45.86]	25.19 [9.37; 41.01]	20.19 [4.00; 36.39]	13.81 [-3.17; 30.78]
	195	50.56 [34.02; 67.10]	37.19 [21.39; 53.00]	30.69 [14.77; 46.61]	24.18 [7.84; 40.52]	15.87 [-1.42; 33.16]
	Dif	**10.10**	**7.00**	**5.50**	**3.99**	2.06
5.5 (mean–1SD)	0	40.30 [24.25; 56.35]	30.02 [14.37; 45.68]	25.03 [9.21; 40.84]	20.03 [3.84; 36.22]	13.64 [-3.33; 30.61]
	195	50.21 [33.70; 66.73]	36.84 [21.06; 52.63]	30.34 [14.44; 46.24]	23.83 [7.51; 40.16]	15.52 [-1.76; 32.80]
	Dif	**9.92**	**6.82**	**5.31**	**3.80**	1.88
51.7 (mean)	0	38.92 [22.95; 54.89]	28.65 [13.03; 44.27]	23.65 [7.85; 39.45]	18.66 [2.45; 34.86]	12.27 [-4.74; 29.28]
	195	47.28 [30.86; 63.69]	33.90 [18.19; 49.61]	27.40 [11.55; 43.24]	20.89 [4.60; 37.18]	12.58 [-4.68; 29.84]
	Dif	**8.35**	**5.25**	**3.75**	2.24	0.31
97.9 (mean+1SD)	0	37.55 [21.56; 53.54]	27.28 [11.59; 42.96]	22.28 [6.39; 38.17]	17.28 [0.97; 33.59]	10.90 [-6.24; 28.04]
	195	44.34 [27.85; 60.82]	30.96 [15.14; 46.79]	24.46 [8.48; 40.44]	17.95 [1.52; 34.39]	9.64 [-7.78; 27.06]
	Dif	**6.79**	**3.69**	2.18	0.67	-1.25
270 (maximum)	0	32.43 [15.55; 49.32]	22.16 [5.39; 38.93]	17.16 [0.12; 34.20]	12.17 [-5.35; 29.68]	5.78 [-12.60; 24.16]
	195	33.39 [15.12; 51.66]	20.02 [2.22; 37.81]	13.51 [-4.48; 31.50]	7.01 [-11.45; 25.46]	-1.30 [-20.71; 18.10]
	Dif	0.96	-2.14	**-3.65**	**-5.16**	**-7.08**

Notes: *Adjusted for male participants aged 67.5, BMI of 26.84, Charlson comorbidity index score of 1-2, forced expiratory volume in 1 second % of predicted of 53.31 and mMRC of 2. Bold represent differences equal or greater than the absolute value of the minimal clinical important difference (3 repetitions).*

Abbreviations: *SD, standard deviation; Dif, difference between the number of repetitions of the one-minute sit-to-stand test at time equals to 195 days and the number of repetitions at baseline; 95% CI, 95% confidence prediction intervals.*

7.4. DISCUSSION

The clinically significant differences found in our study between the last assessment and the baseline suggest that monitoring patients with COPD monthly could benefit their functional status. Further studies with larger samples and control groups are needed to strengthen our findings.

Additionally, this study identified numerous explanatory factors of the 1minSTS behaviour. For instance, older participants or/and participants that were heavy smokers were expected to increase less, or eventually even decrease their number of repetitions when compared with younger participants or/and participants with low smoking load. This information is important to guide clinical decisions aiming to improve functional status of people with COPD. Future studies should explore the added benefit of monitoring the disease progression with meaningful outcomes.

The strengths of our study include the high explanatory power of the LMM computed.

Limitations of this study include the absence of a control group.

7.5. CONCLUSION

This study showed the potential of the 1minSTS to assess functional status over time in people with COPD and clarified the individual related factors of the 1minSTS behaviour.

8. CONCLUSION

From a methodological approach, this worked showed that the Automatic backward elimination of features was consistent when it came to select statistically relevant features to be included in linear mixed-effects models with the lowest values of AIC.

From a clinical perspective, this work seems to indicate that smoking load and age, two well-known environmental risk factors for COPD, are statistically influent in the diverse behaviour of the number of repetitions obtained in the 1minSTS by COPD patients monitored during a 6-months period. An increase in any of them might lead to a worse evolution in the 1minSTS results over time. Nevertheless, our study suggests that monitoring patients with COPD in a monthly base could generally improve their functional status. Even with an interruption in these assessments, as the one caused by the COVID-19 lockdown, the positive evolution seems to occur, which may suggest that the initial monitoring sessions are crucial, although further studies should be conducted to clarify this behaviour. On the contrary, the restrictions to circulation, the social distancing and isolation resulting from COVID-19 pandemic seem to have had a negative impact in the wellbeing and daily life of patients with COPD. Since situations like this one may be repeated in short or medium terms, it is important to identify strategies that healthcare professionals or patients can implement in order to improve the management of COPD and get the greatest benefit from treatment.

REFERENCES

Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H.-T. (2012). *Learning from data* (Vol. 4). AMLBook New York, NY, USA:

Abugaber, D. (2020). *Find the optimal mixed model for your data with glmmLasso*. https://davidabugaber.com/blog/f/find-the-optimal-mixed-model-for-your-data-with-glmmlasso

Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, *60*(2), 255–265. https://doi.org/10.1093/biomet/60.2.255

Ambroise, C., & McLachlan, G. J. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences*, *99*(10), 6562 LP – 6566. https://doi.org/10.1073/pnas.102102699

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software; Vol 1, Issue 1 (2015)*. https://www.jstatsoft.org/v067/i01

Bestall, J. C., Paul, E. A., Garrod, R., Garnham, R., Jones, P. W., & 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), 581–586. https://doi.org/10.1136/thx.54.7.581

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Bohannon, R. W. (2019). Minimal clinically important difference for grip strength: a systematic review. *Journal of Physical Therapy Science*, *31*(1), 75–78. https://doi.org/10.1589/jpts.31.75

Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Medicine and Science in Sports and Exercise*, *14*(5), 377–381.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bui, K.-L., Nyberg, A., Maltais, F., & Saey, D. (2017). Functional Tests in Chronic Obstructive Pulmonary Disease, Part 1: Clinical Relevance and Links to the International Classification of Functioning, Disability, and Health. *Annals of the American Thoracic Society*, *14*(5), 778–784. https://doi.org/10.1513/AnnalsATS.201609-733AS

Cao, C., Wang, R., Wang, J., Bunjhoo, H., Xu, Y., & Xiong, W. (2012). Body Mass Index and Mortality in Chronic Obstructive Pulmonary Disease: A Meta-Analysis. *PLOS ONE*, *7*(8), e43892-. https://doi.org/10.1371/journal.pone.0043892

Capitaine, L., Genuer, R., & Thiébaut, R. (2020). Random forests for high-dimensional longitudinal data. *Statistical Methods in Medical Research*, *30*(1), 166–184. https://doi.org/10.1177/0962280220946080

Charlson, M., Szatrowski, T. P., Peterson, J., & Gold, J. (1994). Validation of a combined comorbidity index. *Journal of Clinical Epidemiology*, *47*(11), 1245–1251. https://doi.org/10.1016/0895-4356(94)90129-5

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., & Li, Y. (2021). *xgboost: Extreme Gradient Boosting* (R package version 1.3.2.1). https://cran.r-project.org/package=xgboost

Clegg, A., Young, J., Iliffe, S., Rikkert, M. O., & Rockwood, K. (2013). Frailty in elderly people. *The Lancet*, *381*(9868), 752–762. https://doi.org/10.1016/S0140-6736(12)62167-9

Crisafulli, E., & Clini, E. M. (2010). Measures of dyspnea in pulmonary rehabilitation. *Multidisciplinary Respiratory Medicine*, *5*(3), 202. https://doi.org/10.1186/2049-6958-5-3-202

Cruz, J., Jácome, C., & Marques, A. (2017). Validity of the Brief physical activity assessment tool for clinical use in COPD. *European Respiratory Journal*, *50*(suppl 61), PA2565. https://doi.org/10.1183/1393003.congress-2017.PA2565

Efron, B. (1982). 5. The Bootstrap. In *The Jackknife, the Bootstrap and Other Resampling Plans* (pp. 27–36). Society for Industrial and Applied Mathematics. https://doi.org/doi:10.1137/1.9781611970319.ch5

Funatogawa, I., & Funatogawa, T. (2018). *Longitudinal Data and Linear Mixed Effects Models: Autoregressive Linear Mixed Effects Models* (pp. 1–26). https://doi.org/10.1007/978-981-10-0077-5_1

Gareth, J., Hastie, T., Tibshirani, R., & Witten, D. (2013). *An introduction to statistical learning : with applications in R*. Springer Science + Business Media, LLC.

GBD 2019 Diseases and Injuries Collaborators, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abolhassani, H., Aboyans, V., Abrams, E. M., Abreu, L. G., Abrigo, M. R. M., Abu-Raddad, L. J., Abushouk, A. I., … Murray, C. J. L. (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 (London, England)*, *396*(10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

George, E. I. (2000). The Variable Selection Problem. *Journal of the American Statistical Association*, *95*(452), 1304–1308. https://doi.org/10.1080/01621459.2000.10474336

George, F. (2013). *Diagnóstico e Tratamento da Doença Pulmonar Obstrutiva Crónica* (028/2011). Direção Geral da Saúde.

Global Initiative for Chronic Obstructive Lung Disease. (2021). GOLD Report 2020. *Global*

Initiative for Chronic Obstructive Lung Disease, 164.

Graham, B. L., Steenbruggen, I., Barjaktarevic, I. Z., Cooper, B. G., Hall, G. L., Hallstrand, T. S., Kaminsky, D. A., McCarthy, K., McCormack, M. C., Miller, M. R., Oropez, C. E., Rosenfeld, M., Stanojevic, S., Swanney, M. P., & Thompson, B. R. (2019). Standardization of spirometry 2019 update an official American Thoracic Society and European Respiratory Society technical statement. In *American Journal of Respiratory and Critical Care Medicine* (Vol. 200, Issue 8, pp. E70–E88). American Thoracic Society. https://doi.org/10.1164/rccm.201908-1590ST

Groll, A. (2017). *glmmLasso: Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation* (R package version 1.5.1). https://cran.r-project.org/package=glmmLasso

Groll, A., & Tutz, G. (2014). Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation. *Statistics and Computing*, *24*(2), 137–154. https://doi.org/10.1007/s11222-012-9359-z

Hall, M.-H., Holton, K. M., Öngu□r, D., Montrose, D., & Keshavan, M. S. (2019). Longitudinal Trajectory of Early Functional Recovery in Patients with First Episode Psychosis. *BioRxiv*, 525824. https://doi.org/10.1101/525824

Hashmi, M. F., Modi, P., & Sharma, S. (2021). *Dyspnea.*

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

Ishwaran, H., & Kogalur, U. B. (2021). *Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC)*. https://cran.r-project.org/package=randomForestSRC

Jarad, N. (2011). Chronic obstructive pulmonary disease (COPD) and old age? *Chronic Respiratory Disease*, *8*(2), 143–151. https://doi.org/10.1177/1479972311407218

Jobson, J. D. (1991). *Multiple Linear Regression BT - Applied Multivariate Data Analysis: Regression and Experimental Design* (J. D. Jobson (Ed.); pp. 219–398). Springer New York. https://doi.org/10.1007/978-1-4612-0955-3_4

Jones, P W, Harding, G., Berry, P., Wiklund, I., Chen, W.-H., & Kline Leidy, N. (2009). Development and first validation of the COPD Assessment Test. *European Respiratory Journal*, *34*(3), 648. https://doi.org/10.1183/09031936.00102509

Jones, P W, Quirk, F. H., & Baveystock, C. M. (1991). The St George's Respiratory Questionnaire. *Respiratory Medicine*, *85 Suppl B*, 25–27. https://doi.org/10.1016/s0954-6111(06)80166-6

Jones, Paul W, Tabberer, M., & Chen, W.-H. (2011). Creating scenarios of the impact of copd and their relationship to copd assessment test (CAT™) scores. *BMC Pulmonary Medicine*, *11*(1), 42. https://doi.org/10.1186/1471-2466-11-42

Kenward, M., & Roger, J. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 983–997. https://doi.org/10.2307/2533558

Kiley, J. P., Sri Ram, J., Croxton, T. L., & Weinmann, G. G. (2005). Challenges Associated with Estimating Minimal Clinically Important Differences in COPD—The NHLBI Perspective. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, *2*(1), 43–46. https://doi.org/10.1081/COPD-200050649

Kon, S. S. C., Canavan, J. L., Jones, S. E., Nolan, C. M., Clark, A. L., Dickson, M. J., Haselden, B. M., Polkey, M. I., & Man, W. D.-C. (2014). Minimum clinically important difference for the COPD Assessment Test: a prospective analysis. *The Lancet Respiratory Medicine*, *2*(3), 195–203. https://doi.org/10.1016/S2213-2600(14)70001-3

Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, *36*(11), 1–13. http://www.jstatsoft.org/v36/i11/

Kursa, M., Jankowski, A., & Rudnicki, W. (2010). Boruta - A System for Feature Selection. *Fundam. Informaticae*, 271–285. https://doi.org/10.3233/FI-2010-288

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software; Vol 1, Issue 13 (2017)*. https://www.jstatsoft.org/v082/i13

Lab 3R. (2018). https://www.ua.pt/en/lab3r

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963–974.

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (R package version 1.5.5-1). https://cran.r-project.org/package=emmeans

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22. https://cran.r-project.org/doc/Rnews/

Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Mahler, D. A., Rosiello, R. A., Harver, A., Lentine, T., McGovern, J. F., & Daubenspeck, J. A. (1987). Comparison of Clinical Dyspnea Ratings and Psychophysical Measurements of Respiratory Sensation in Obstructive Airway Disease [1–4]. *American Review of Respiratory Disease*, *135*(6), 1229–1233. https://doi.org/10.1164/arrd.1987.135.6.1229

Mallinckrodt, C., & Lipkovich, I. (2017). *Analysing Longitudinal Clinical Trial Data*. Taylor & Francis.

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., & 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. https://doi.org/10.1186/s12889-019-7045-1

Marshall, A. L., Smith, B. J., Bauman, A. E., & Kaur, S. (2005). Reliability and validity of a

brief physical activity assessment for use by family doctors. *British Journal of Sports Medicine*, *39*(5), 294–297. https://doi.org/10.1136/bjsm.2004.013771

Mathers, C. D., & Loncar, D. (2006). Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLOS Medicine*, *3*(11), e442-. https://doi.org/10.1371/journal.pmed.0030442

Nakagawa, S., Johnson, P., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisted and expanded. *J. R. Soc. Interface*, *14*. https://doi.org/10.1098/rsif.2017.0213

Ozalevli, S., Ozden, A., Itil, O., & 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), 286–293. https://doi.org/10.1016/j.rmed.2006.05.007

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Rabe, K. F., & Watz, H. (2017). Chronic obstructive pulmonary disease. *The Lancet*, *389*(10082), 1931–1940. https://doi.org/https://doi.org/10.1016/S0140-6736(17)31222-9

Rennard, S. I., & Drummond, M. B. (2015). Early chronic obstructive pulmonary disease: definition, assessment, and prevention. *Lancet (London, England)*, *385*(9979), 1778–1788. https://doi.org/10.1016/S0140-6736(15)60647-X

RStudio Team. (2021). *RStudio: Integrated Development Environment for R*. PBC. http://www.rstudio.com/

Sigrist, F. (2020). *Gaussian Process Boosting*. https://arxiv.org/abs/2004.02653

Souto-Miranda, S., & Marques, A. (2018). *"Outcomes of Pulmonary Rehabilitation valued by patients with COPD – patients' perspectives"*. https://doi.org/10.13140/RG.2.2.28146.96965

Spruit, M. A., Singh, S. J., Garvey, C., ZuWallack, R., Nici, L., Rochester, C., Hill, K.,

Holland, A. E., Lareau, S. C., Man, W. D.-C., Pitta, F., Sewell, L., Raskin, J., Bourbeau, J., Crouch, R., Franssen, F. M. E., Casaburi, R., Vercoulen, J. H., Vogiatzis, I., … Wouters, E. F. M. (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. https://doi.org/10.1164/rccm.201309-1634ST

Sun, Y., Milne, S., Jaw, J. E., Yang, C. X., Xu, F., Li, X., Obeidat, M., & Sin, D. D. (2019). BMI is associated with FEV1 decline in chronic obstructive pulmonary disease: a meta-analysis of clinical trials. *Respiratory Research*, *20*(1), 236. https://doi.org/10.1186/s12931-019-1209-5

Tibshirani, R. (1996a). *Bias, variance, and prediction error for classification rules*.

Tibshirani, R. (1996b). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. http://www.jstor.org/stable/2346178

Vaidya, T., Chambellan, A., & de Bisschop, C. (2017). Sit-to-stand tests for COPD: A literature review. *Respiratory Medicine*, *128*, 70–77. https://doi.org/10.1016/j.rmed.2017.05.003

Vaidya, T., de Bisschop, C., Beaumont, M., Ouksel, H., Jean, V., Dessables, F., & Chambellan, A. (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*, 2609–2616. https://doi.org/10.2147/COPD.S115439

Varmaghani, M., Dehghani, M., Heidari, E., Sharifi, F., Moghaddam, S. S., & Farzadfar, F. (2019). Global prevalence of chronic obstructive pulmonary disease: systematic review and meta-analysis. *Eastern Mediterranean Health Journal = La Revue de Sante de La Mediterranee Orientale = Al-Majallah Al-Sihhiyah Li-Sharq Al-Mutawassit*, *25*(1), 47–57. https://doi.org/10.26719/emhj.18.014

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, *75*(5), 1182–1189. https://doi.org/https://doi.org/10.1111/j.1365-2656.2006.01141.x

Willett, J. B. (1989). Some Results on Reliability for the Longitudinal Measurement of Change: Implications for the Design of Studies of Individual Growth. *Educational and Psychological Measurement*, *49*(3), 587–602. https://doi.org/10.1177/001316448904900309

Wilson, R. C., & Jones, P. W. (1989). A comparison of the visual analogue scale and modified Borg scale for the measurement of dyspnoea during exercise. *Clinical Science (London, England : 1979)*, *76*(3), 277–282. https://doi.org/10.1042/cs0760277

World Health Organization. (2002). *WHO Strategy for Prevention and Control of Chronic Respiratory Diseases*. World Health Organization.

World Health Organization. (2008). *The global burden of disease: 2004 update*. World Health Organization.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., & Smith, G. (2009). *Mixed Effects Models and Extensions in Ecology With R*. https://doi.org/10.1007/978-0-387-87458-6_1

APPENDICES

A. RESIDUAL ANALYSIS TO TEST LINEAR MIXED-EFFECTS MODELS' ASSUMPTIONS OF STUDY 1

A.1 ONE-MINUTE SIT-TO-STAND TEST

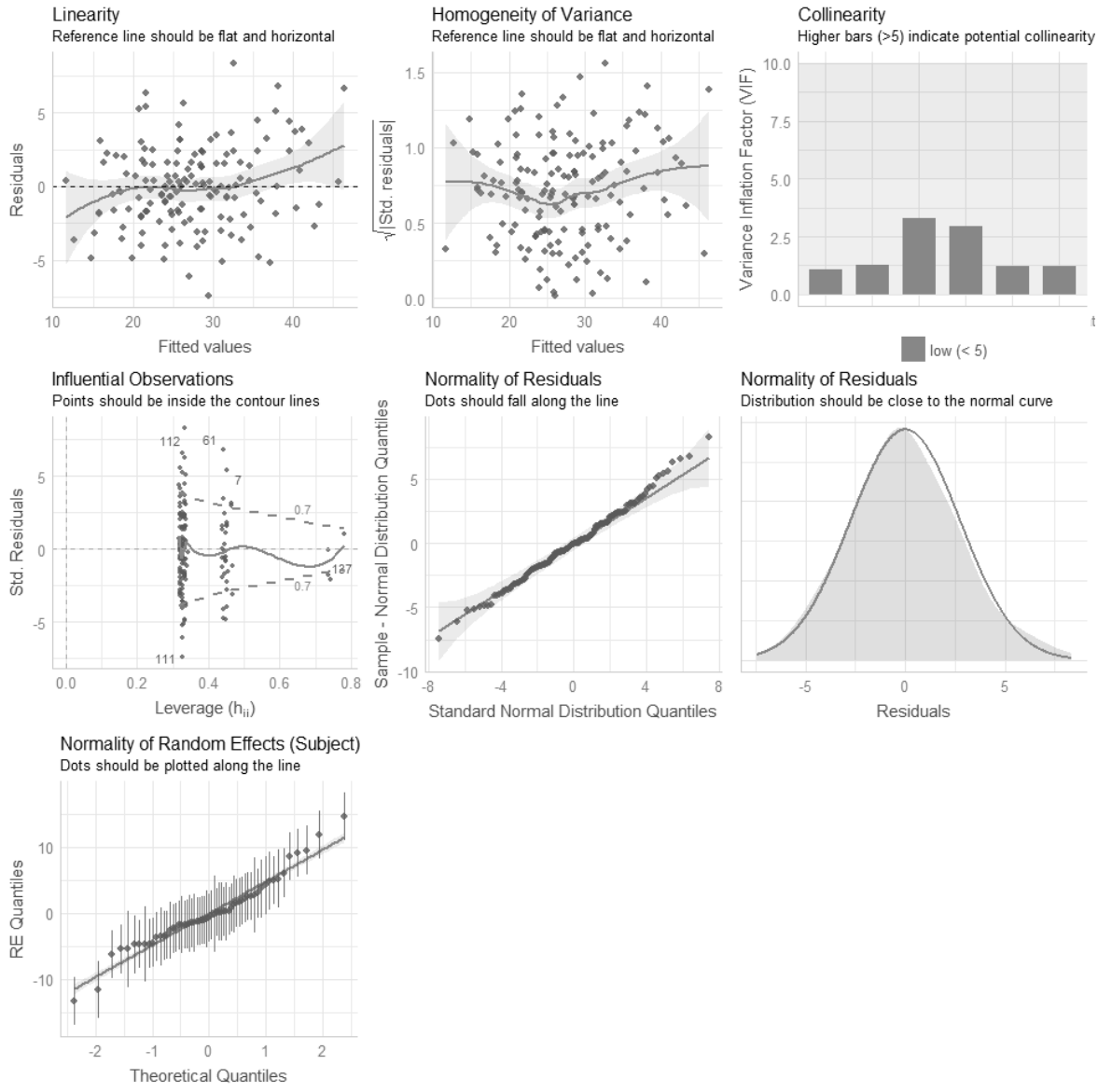

Figure A.1 – *Residual analysis for the linear mixed-effects model using as dependent variable the number of repetitions in the one-minute sit-to-stand test and as independent variables the ones obtained by automatic backward elimination of features*

A.2 HANDGRIP MUSCLE STRENGTH

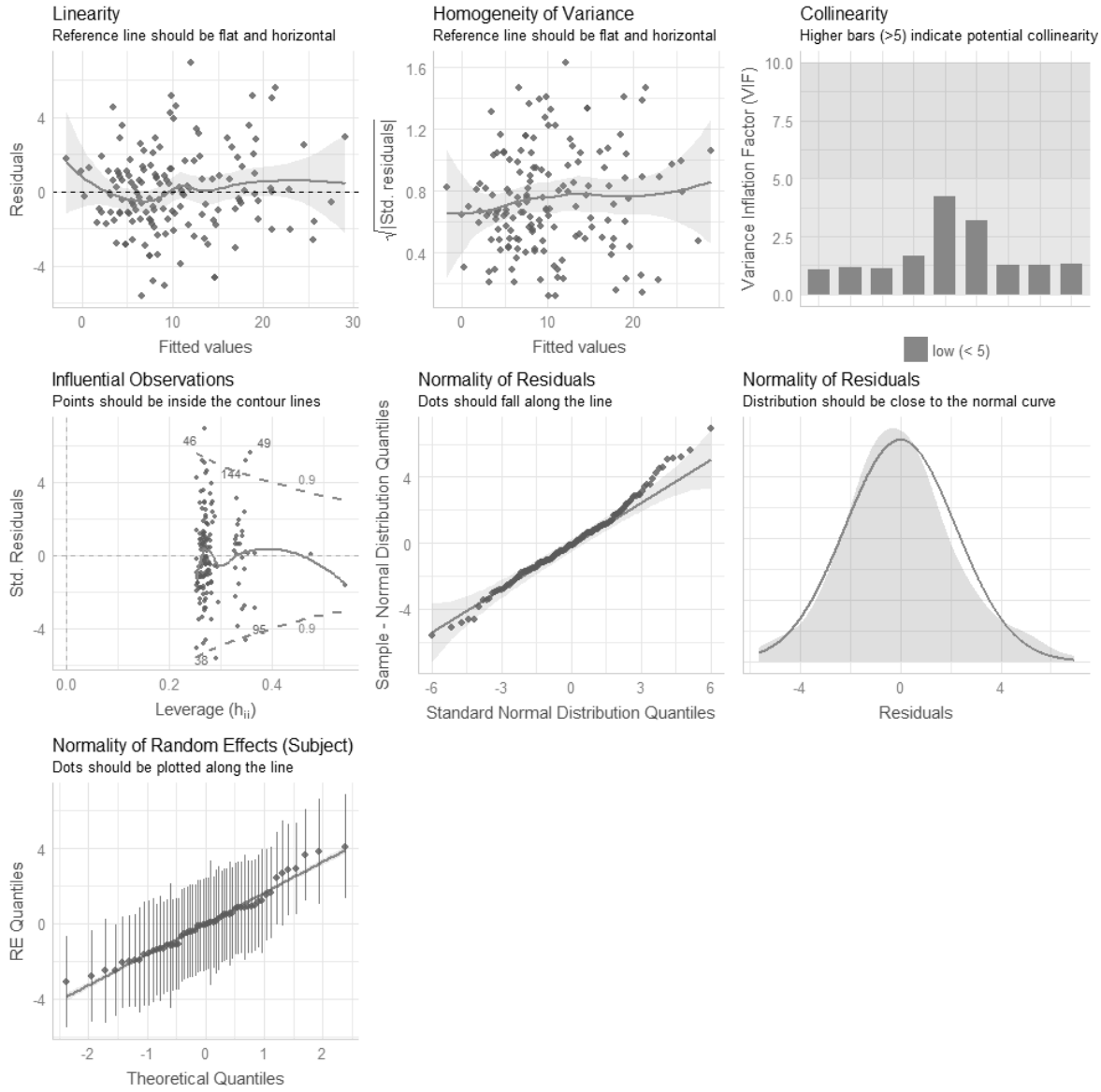

Figure A.2 – *Residual analysis for the linear mixed-effects model using as dependent variable the handgrip muscle strength and as independent variables the ones obtained by automatic backward elimination of features*

A.3 COPD ASSESSMENT TEST

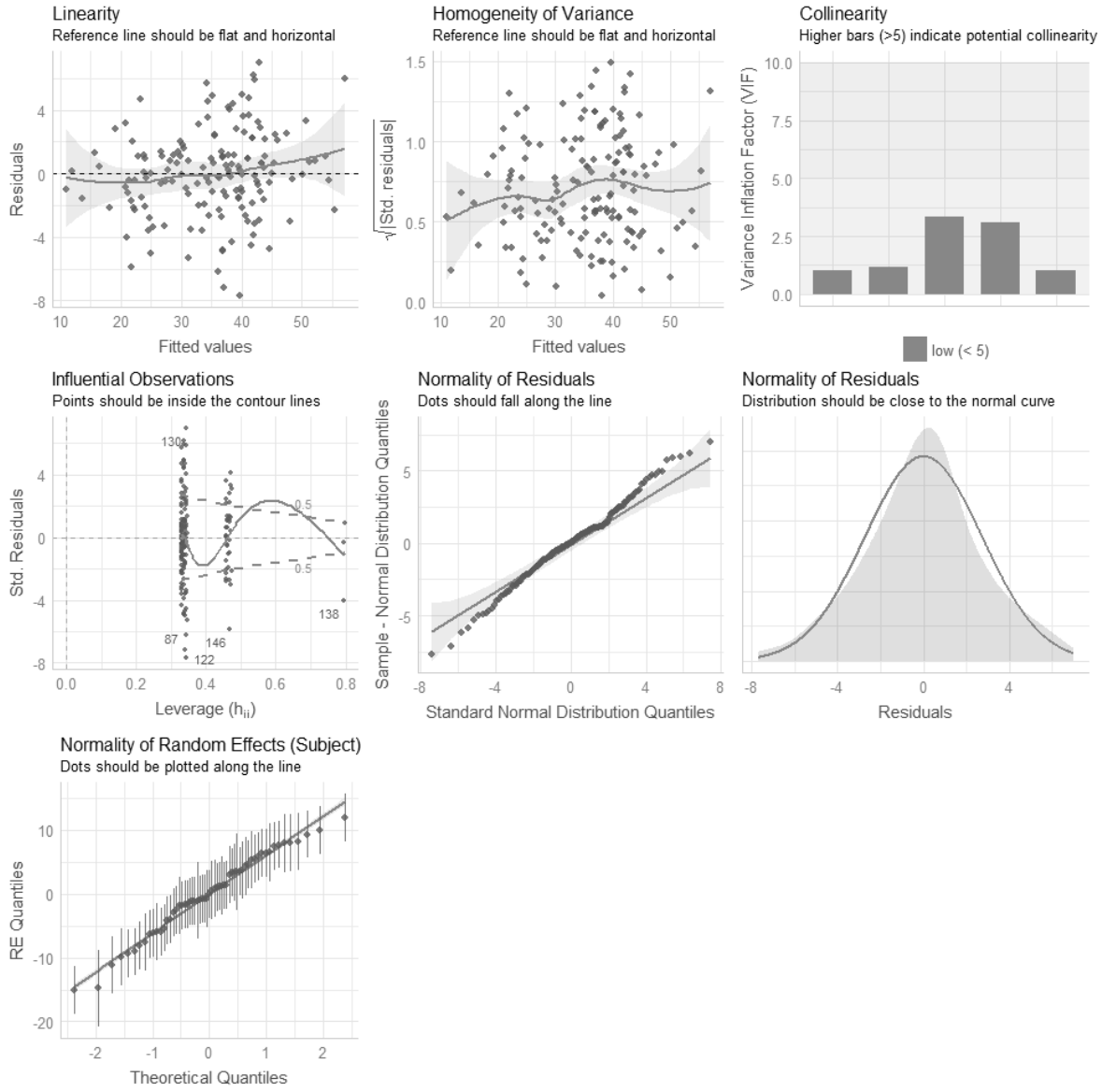

Figure A.3 – *Residual analysis for the linear mixed-effects model using as dependent variable the COPD Assessment Test score and as independent variables the ones obtained by automatic backward elimination of features*

B. EXTENDED ABSTRACT – 3RD STATISTICS ON HEALTH DECISION MAKING

Title:

Behaviour of the one-minute sit-to-stand test during six months in people with COPD

Short title (running head)

Behaviour of the one-minute sit-to-stand test in COPD

List of authors

Jorge Cabral, BSc[1,2], Vera Afreixo, PhD[1,3], Vânia Rocha, PhD[2,4], Sara Souto-Miranda, MSc[2,4] & Alda Marques, PhD[2,4]

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

Keywords: COPD; Functional Status; One-Minute Sit-To-Stand; Follow-up.

Introduction:

Chronic obstructive pulmonary disease (COPD) is a common, progressive and treatable disease that is characterized by persistent respiratory symptoms and airflow limitation [1]. One of the most frequent impacts of COPD on daily life is decreased functional status which includes struggling to perform basic, work and leisure activities [2]. Although, functional status is highly meaningful to people with COPD, this outcome has been overlooked [3]. Numerous field tests might be used to assess functional status [4], yet the one-minute sit-to-stand test (1minSTS) has shown to be a simple test that mimics the common activity of sitting/standing from a chair which is essential to maintain independence among the elderly [5]. Additionally, it is a valid and responsive measure that might be easily performed for follow-up assessment of people with COPD [4]. In fact, regular assessment of people with

COPD is essential [1] and this study hypothesise that the 1minSTS might be an important indicator of functional status over time in people with COPD.

Thus, this study aimed to describe the 1minSTS behaviour over a six-month period and explore the factors influencing this behaviour in people with COPD.

Methods:

Study design and participants

Data from an observational study including people with stable COPD was retrospectively analysed. Individuals were eligible if diagnosed with COPD [1] and clinically stable over the previous month (no acute exacerbations). Individuals with other respiratory diseases, signs of cognitive impairment or presence of a significant or unstable cardiovascular, neurological or musculoskeletal disease were excluded.

Data collection

Sociodemographic, anthropometric and clinical data were first collected with a structured questionnaire to characterise the sample. Spirometry was used to assess lung function [6]. Severity of comorbid diseases was scored according to the Charlson Comorbidity Index (CCI) [7]. Activity-related dyspnoea was assessed with the modified British medical research council dyspnoea questionnaire (mMRC) [8,9] and the impact of the disease with the COPD Assessment Test (CAT) [10,11].

Functional status was assessed with the 1minSTS which consists of sitting and standing from a 46-48 cm height chair as many times as possible for one minute [4,13]. A change of 3 repetitions was used as minimum clinically important difference (MCID) [13]. All data were collected at baseline and 1minSTS was repeated monthly up to six months.

Data analysis

Variables were summarized according to their nature. Linear-mixed effect models (LMM) with random intercepts and slopes were applied to assess the mean change in the number of repetitions of the 1minSTS [14,15]. A backward elimination with single terms deletion and keeping time was performed [16]. P-values were computed based on conditional F-tests with Kenward-Roger approximation [17]. Two-sided $P<0.05$ was considered statistically significant.

Results:

A total of 149 participants with COPD were included. Participants mean age was 67.5 (±9.0) years, most were men (83.9%), slightly overweight (BMI=26.8 kg/m^2), former smokers (73.8%), presented severe airflow obstruction (49.0 [38.0;70.0]), 3 to 4 comorbidities (53%),

were not under PR effect (72.5%) and the median of 1minSTS was 26 [21;30] repetitions. Further detailed baseline characteristics are presented in Table 1.

An increase of the number of repetitions performed over time was observed reaching a median of 30.0 [24.0;37.5] repetitions at assessment 5 (A5) (Figure 1a). Specifically, an increase of 3.8 repetitions after 195 days was predicted (Figure 1b).

Table 2 shows the results of the LMM including time and participant as random effects. The model's total explanatory power was substantial (conditional R^2 = 0.92). The effect of time was positive and statistically significant [0.09 (0.04; 0.13)]. Females [-4.68 (-8.20; -1.18)], older participants [-0.55 (-0.77; -0.34)], with higher BMI [-0.55 (-0.81; -0.28)], higher pack-years [-0.03 (-0.06; -0.00)], higher scores of mMRC [-2.04 (-3.25; -0.83)] and lower FEV_1%predicted [0.07 (0.07; 0.13)] showed a lower number of repetitions globally. The interaction effect of time on age and on pack-years was negative and statistically significant [-8.60E-4 (-1.48E-3; -2.40E-4) and -1.70E-4 (-3.00E-4; -5.00E-5), respectively].

Discussion:

The clinically significant differences found in our study between the last assessment and the baseline suggest that monitoring patients with COPD monthly could benefit their functional status. Further studies with larger samples and control groups are needed to strengthen our findings.

Additionally, this study identified numerous explanatory factors of the 1minSTS behaviour. For instance, older participants or/and participants that were heavy smokers were expected to increase less, or eventually even decrease their number of repetitions when compared with younger participants or/and participants with low smoking load. This information is important to guide clinical decisions aiming to improve functional status of people with COPD. Future studies should explore the added benefit of monitoring the disease progression with meaningful outcomes.

The strengths of our study include the high explanatory power of the LMM computed.

Limitations of this study include the absence of a control group.

In sum, this study showed the potential of the 1minSTS to assess functional status over time in people with COPD and clarified the individual related factors of the 1minSTS behaviour.

Ethics committee and informed consent:

Five independent Ethics Committees (Centro Hospitalar do Médio Ave ref. 09/2016 and 10/2018; Unidade Local de Saúde de Matosinhos ref. 10/CES/JAS 17/02/2017 and

73/CE/JAS 12/10/2018; Centro Hospitalar Baixo Vouga ref. 777638 and 086892; Hospital Distrital da Figueira da Foz ref. 1807/2017 and 27/05/2019; Administração Regional de Saúde do Centro ref. 64/2016 and 85/2018) approved the study. Written informed consent was obtained from all participants before data collection. Data protection was ensured by the National Committee for Data Protection (no. 7295/2016) and followed the General Data Protection Regulation.

Acknowledgements:

This work was funded by Programa Operacional de Competitividade e Internacionalização – POCI, through Fundo Europeu de Desenvolvimento Regional - FEDER (POCI-01-0145-FEDER-007628 and POCI-01-0145-FEDER-028806), Fundação para a Ciência e Tecnologia (PTDC/DTP-PIC/2284/2014 and PTDC/SAU-SER/28806/2017) and under projects UIDB/04501/2020 and UIDB/04106/2020.

References:

1. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: the GOLD science committee report 2021. 2021. https://goldcopd.org/wp-content/uploads/2020/11/GOLD-REPORT-2021-v1.1-25Nov20_WMV.pdf
2. Leidy NK. Functional status and the forward progress of merry-go-rounds: toward a coherent analytical framework. Nurs Res. 1994;43(4):196-202
3. Bui KL, Nyberg A, Maltais F, Saey D. Functional Tests in Chronic Obstructive Pulmonary Disease, Part 1: Clinical Relevance and Links to the International Classification of Functioning, Disability, and Health. Ann Am Thorac Soc. 2017 May;14(5):778-784. http://doi.org/10.1513/AnnalsATS.201609-733AS
4. Vaidya T, Chambellan A, de Bisschop C. Sit-to-stand tests for COPD: A literature review. Respiratory medicine. 2017;128:70-77. http://doi.org/10.1016/j.rmed.2017.05.003
5. Vaidya T, de Bisschop C, Beaumont M and others. 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. Int J Chron Obstruct Pulmon Dis. 2016;11:2609-2616. http://doi.org/10.2147/COPD.S115439
6. Graham BL, Steenbruggen I, Miller MR and others. Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. American journal of respiratory and critical care medicine. 2019;200(8):e70-e88. http://doi.org/10.1164/rccm.201908-1590ST

7. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. Journal of clinical epidemiology. 1994;47(11):1245-51. http://doi.org/10.1016/0895-4356(94)90129-5
8. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. Thorax. 1999;54(7):581-6. http://doi.org/10.1136/thx.54.7.581
9. Crisafulli E, Clini, EM. Measures of dyspnea in pulmonary rehabilitation. Multidisciplinary respiratory medicine. 2010;5(3):202-10. http://doi.org/10.1186/2049-6958-5-3-202
10. George F. Diagnóstico e Tratamento da Doença Pulmonar Obstrutiva Crónica. *Direção Geral da Saúde* 028/2011:1-15. 2013.
11. Jones PW, Harding G, Berry P, Wilklund I, Chen WH, Leidy NK. Development and first validation of the COPD Assessment Test. The European respiratory journal. 2009;34(3): 648-54. http://doi.org/10.1183/09031936.00102509
12. Ozalevli S, Ozden A, Itil O, Akkoclu A. Comparison of the Sit-to-Stand Test with 6 min walk test in patients with chronic obstructive pulmonary disease. Respiratory medicine. 2007;101(2):286-93. http://doi.org/10.1016/j.rmed.2006.05.007
13. Vaidya T, de Bisschop C, Beaumont M and others. 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. 2016;11:2609-16. http://doi.org/10.2147/COPD.S115439
14. Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015;67(1):1-48. http://doi.org/10.18637/jss.v067.i01
15. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software. 2017;82(13):1-26. http://doi.org/10.18637/jss.v082.i13
16. Zuur A, Ieno EN, Walker N, Saveliev A, Smith GM. Mixed Effects Models and Extensions in Ecology With R. 2009:121-22. http://doi.org/10.1007/978-0-387-87458-6_1
17. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997:983-97. http://doi.org/10.2307/2533558
18. Nakagawa S, Johnson P, Schielzeth H. *The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisted and expanded.* J. R. Soc. Interface 14:20170213. http://doi.org/10.1098/rsif.2017.0213

Tables:

Table 1. Baseline characteristics (n=149).

Age, *years*, mean (SD)		67.5 (9.0)
Sex	Female	24 (16.1)
	Male	125 (83.9)
BMI, *kg/m²*, mean (SD)		26.8 (4.6)
Smoking Status	Never	20 (13.4)
	Former	110 (73.8)
	Current	19 (12.8)
Pack-years, median [IQR]		42.0 [15.0;75.0]
Under PR Effect during follow-up	No	108 (72.5)
	Yes	41 (27.5)
CCI*, score*	1-2	31 (20.8)
	3-4	79 (53.0)
	>=5	39 (26.2)
AECOPD*, in the previous year*	0	106 (71.1)
	1	19 (12.8)
	>1	24 (16.1)
mMRC, *points*, median [IQR]		2 [1.0,3.0]
FEV_1, *% predicted*, median [IQR]		49.0 [38.0;70.0]
FEV_1/FVC, *%*, median [IQR]		53.0 [40.0;63.0]
GOLD grades	1	25 (16.8)
	2	47 (31.5)
	3	59 (39.6)
	4	18 (12.1)
CAT, *points*, median [IQR]		12.0 [8.0,18.0]
GOLD CAT, *stage*	A	49 (32.9)
	B	72 (48.3)
	C	2 (1.3)
	D	26 (17.5)
1minSTS, repetitions, median [IQR]		26.0 [21.0;30.0]

Note: Data presented as n (%), unless otherwise stated.

Abbreviations: COPD, Chronic Obstructive Pulmonary Disease; GOLD, Global Initiative for Chronic Obstructive Lung Disease; BMI, Body Mass Index; PR, Pulmonary Rehabilitation; CCI, Charlson Comorbidity Index; mMRC, Modified Medical Council Dyspnoea Scale; CAT, COPD Assessment Test; AECOPD, Acute Exacerbation of COPD; 1minSTS, One-Minute Sit-To-Stand Test; FEV_1, Forced Expiratory Volume in 1 Second; FVC, Forced Vital Capacity; SD, Standard deviation; IQR, Interquartile range.

Table 2. Factors associated with the number of repetitions of one-minute sit-to-stand test in people with COPD (n=149).

	1MinSTS (n=149)			
Factors	**Estimates**	**CI 95%**	**p**	**df**
(Intercept)	77.56	[62.40; 92.71]	**<0.001**	140.70
Sex [Female]	-4.69	[-8.20; -1.18]	**0.009**	140.59
Age	-0.56	[-0.77; -0.34]	**<0.001**	141.39
BMI	-0.55	[-0.81; -0.28]	**<0.001**	139.05
Pack-years	-0.03	[-0.06; -0.00]	**0.033**	142.38
CCI [Moderate (3-4)]	4.80	[0.71; 8.89]	**0.022**	139.13
CCI [Severe (>=5)]	3.57	[-2.15; 9.28]	0.219	140.26
mMRC	-2.04	[-3.25; -0.83]	**0.001**	140.34
FEV_1 % of predicted	0.07	[0.07; 0.14]	**0.021**	139.75
Time	0.09	[0.04; 0.13]	**<0.001**	132.41
Age*Time	-8.60E-4	[-1.48E-3; -2.40E-4]	**0.007**	134.32
Pack-years*Time	-1.70E-4	[-3.00E-4; -5.00E-5]	**0.006**	133.37
Random Effects				
σ^2	0.08			
τ_{00}	$0.53_{Participant}$			
τ_{11}	$0.01_{Participant.Time}$			
ρ_{01}	$0.64_{Participant}$			
ICC	0.87			
Observations	755			
Marginal R^2 / Conditional R^2	0.408 / 0.922			

*Abbreviations: 1minSTS, One-Minute Sit-To-Stand Test; CI, Confidence Interval approximated by Kenward-Roger method; p, p value approximated by Kenward-Roger method; df, degrees of freedom approximated by Kenward-Roger method; BMI, Body Mass Index; CCI, Charlson Comorbidity Index; mMRC, Modified Medical Council Dyspnoea Scale; FEV_1 % of predicted, percentage of the predicted Forced Expiratory Volume in 1 Second; σ^2, residual variance; τ, random effect standard deviation; ρ, correlation between intercept and slope; ICC, intraclass correlation coefficient; R^2, coefficient of determination [18]. * indicates "interaction with"*.

Figure's captions/legends:

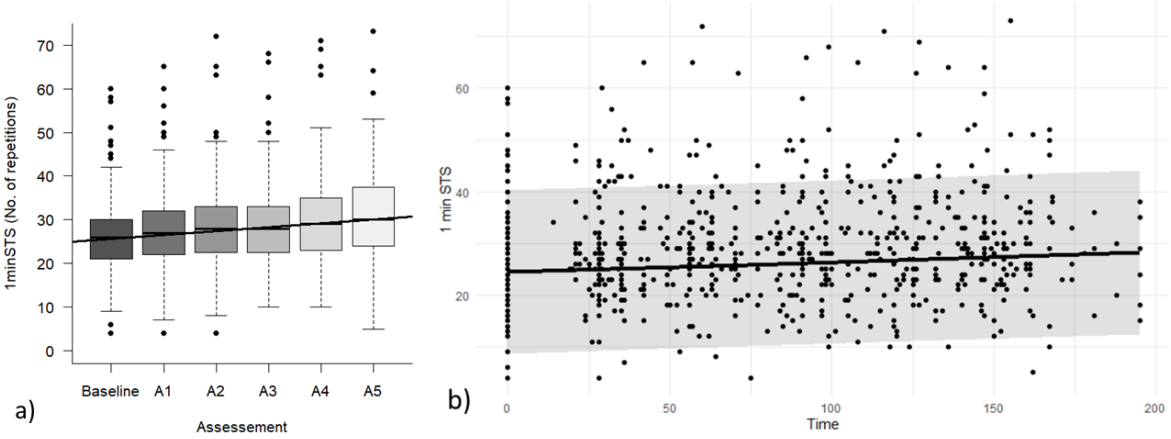

Figure 1. Description of the number of repetitions performed in the one-minute sit-to-stand (1minSTS) over time by participants with chronic obstructive pulmonary disease (n=149). (a) In the left panel, time was considered as a categorical variable, corresponding to the number of assessments; (b) in the right panel, time was defined as the number of days between baseline and follow-up assessments, predicted values are represented by a black line, 95% confidence intervals by a grey band and observed values by dots.

C. ABSTRACT - XXV CONGRESS OF THE PORTUGUESE STATISTICAL SOCIETY

COVID-19 lockdown effect in COPD: a comparison of fixed-effects selection methods

Jorge Cabral [a,b], Vera Afreixo [a,c], Alda Marques [b,d]

jorgecabral@ua.pt, vera@ua.pt, amarques@ua.pt

[a] Department of Mathematics, University of Aveiro, Aveiro, Portugal

[b] Respiratory Research and Rehabilitation Laboratory (Lab3R), School of Health Sciences (ESSUA), University of Aveiro, Aveiro, Portugal

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

[d] Institute of Biomedicine (iBiMED), School of Health Sciences, University of Aveiro, Aveiro, Portugal

Keywords: COPD, COVID-19, feature selection, linear mixed-effects models, longitudinal data

Abstract: Chronic obstructive pulmonary disease (COPD) is common and progressive. One of its major impacts on daily life is decreased functional status which can be assessed by the one-minute sit-to-stand test (1minSTS). The 2020 imposed lockdown due to the recent pandemic (COVID 19) is likely to have influenced the functional status of this population but this is still unknown.

Few feature selection algorithms are available for longitudinal data. We aimed to compare different feature selection methods and describe the effect of the COVID-19 lockdown on the 1minSTS behaviour in people with COPD.

Data from 59 people with COPD were collected at baseline (B), 34 of whom belonging to the no-lockdown group. 1minSTS was repeated after one (A1) and five months (A5), which corresponded to the assessments prior and after the lockdown in the lockdown group. Fixed-effects were included in different linear mixed-effects models (LMMs) according to the importance given by Random Forests, Boruta, Extreme Gradient Boosting, automatic backward elimination and L1-penalized estimation algorithms. The LMM with the lowest Akaike's information criterion (AIC) was chosen.

The LMM obtained by automatic backward elimination achieved the lowest AIC (919.7) and was followed by the one using L1-penalized estimation algorithm (923.5) although this one produced a higher conditional R-squared. Boruta algorithm returned the highest AIC (964.2). Difference between B and A1 number of repetitions in 1minSTS was statistically significant in both COVID-19 groups. No difference was found between A1 and A5 in either group suggesting that the lockdown had no effect in the 1minSTS behaviour.