

**Effects of a Physical Exercise Program on the Responses Measured by HIF-1 Related
to Ventilatory and Hematological Function in Patients With COPD Resident at 2600
m.s.n.m.**

Statistical Analysis Plan (SAP)

December 2020

1. Data analysis

1.1 Descriptive analysis

For the descriptive analysis of the characteristics of the population studied, the variables on the qualitative measurement scale will be described as proportional frequencies, the continuous variables will be described with measures of central tendency (mean and median) and of dispersion (standard deviation, interquartile range). according to whether or not they present a normal distribution according to graphic (histograms, dispersion, among others) and numerical (Shapiro-Wilk) methods.

1.2 Baseline analysis

To compare the three groups (G1: COPD 1-2, G2: COPD 3-4, G3: without COPD). Parametric and non-parametric tests will be carried out according to the verification of a priori criteria. For comparisons between proportions, Chi2 or Fischer's exact will be carried out, for the comparison of continuous variables between the three groups an ANOVA or kruskal wallis test will be used.

1.3 Multivariate analysis

As a longitudinal method, a repeated measures analysis will be carried out for correlated data. For each of the continuous outcome variables, the analysis will be performed with fixed effects (marginal) models. The evaluation of the assumption of normal distribution of the variables will be carried out using graphic methods (histograms, box plots and QQ graphs).

In the model, the main exposure will be included as covariates, referring to the group it belongs to (G1: COPD 1-2, G2: COPD 3-4, G3: without COPD). The time will be taken as a continuous variable according to the minutes between each of the measurements, for which its linearity and its interactions with the covariates of interest will be evaluated, in case of not complying with this assumption it will be categorized based on the

measurements carried out, considering the baseline measurement as a reference and the evaluation of possible interaction terms will be carried out.

For the variables with normal distribution, the restrictive model of maximum likelihood (REML) will be used to select the best correlation matrix; later, to select the covariates to keep in the model, the maximum likelihood estimate (MLE) will be used. The selection of the best model will be based on the lower value of the Akaike information criterion (AIC) and the Bayesian evaluation criterion (BIC) after having evaluated the interactions and the non-linear terms. For conditional models, models with intercept and slopes will be used. If it does not present a normal distribution, a generalized equation estimation model (GEE) with an unstructured matrix will be implemented. A residual analysis will be performed on all final models. With the above, it will be possible to evaluate the effects both within and between groups with a more robust model to analyze the outcome measures and control for possible confounding variables.