Machine learning method to predict latent class mixed model (LCMM) classes in COPD study

Group Members: Merin Grace & Karina Camarena
Client: Yiyuan Wu
Faculty Advisor: Dr. Paul Christos

Objective: We aim to identify distinct quality of life trajectories in patients diagnosed with Chronic Obstructive Pulmonary Disease (COPD) and major depression. Also, predict the identified life trajectory classes using machine learning techniques.

Methods: Quality of life (QOL) were assessed using the World Health Organization Quality of Life Questionnaire at baseline, week 10, 14 and 26. Missing QOL scores were imputed using the mean. Patients with only 1 timepoint assessment were removed from the analysis. 51 patients were analyzed. Latent class modeling with random effects was used to identify distinct quality of life trajectories. Upon identification of trajectories, machine learning algorithms were utilized to predict the class membership of patients to these trajectories. Feature selection was performed using lasso. Classification algorithms such as Lasso, Support Vector machine, and Regression trees were utilized for the predictive model of patient baseline characteristics on class membership. Logistic regression was also fit for comparison.

Results: 5 separate models were fitted to identify the number of distinct trajectories. Models were fit with time as continuous, quadratic, and discrete variable. Models were compared with fixed and added random effects. The model with the lowest Bayesian Information Criterion (BIC=1412) was selected with discrete time and random effects. From the analysis, 2 distinct trajectories were identified. 37(72.5%) patients were identified in the stable QOL trajectory, while 14 (27.45%) patients showed an improved QOL trajectory. The mean posterior probability of group membership was 0.88 and 0.86 for improved and stable group respectively. Lasso regression predicted the classes with 80% prediction accuracy. Significant predictors included years of education, Liverpool personal agency, SF-12 Mental health score, and pulmonary functional status.

Conclusion: Latent class model shows 2 distinct quality of life trajectories. Lasso regression predicted the patients' class trajectory the best with prediction accuracy of 80% compared to the other algorithms.