



Weill Cornell Medicine

Machine learning to predict LCMM classes in COPD study

Merin Grace, Karina Camarena
Department of Population Health Sciences
Biostatistics and Data Science

Faculty Advisor: Dr. Paul Christos
Client: Yiyuan Wu

OBJECTIVE

This project aims to achieve two goals:

- Identify distinct quality of life (QOL) trajectories in patients diagnosed with Chronic Obstructive Pulmonary Disease (COPD) and major depression.
- Predict the identified life trajectory classes using machine learning techniques.

METHOD

QOL Trajectories:

- QOL scores were assessed using WHO Questionnaire at baseline, week 10, 14, and 26 respectively.
- Patients with at least 2 timepoint assessments were included in the analysis. Data from 51 patients was incorporated in the model.
- 5 separate latent mixed models were fit with varying number of classes, including time variable as continuous, quadratic and discrete.
- Models were also compared with fixed effects only vs fixed effects and patient specific random effects.

Trajectory Class Prediction:

- Patient specific baseline covariates and clinical measurements were used to build a predictive model for predicting the trajectory membership.
- Lasso regression, Support Vector Machine (SVM), Random Forest and Logistic Regression models were fit for trajectory membership prediction.

RESULT

QOL Trajectories:

- LCM Model with discrete time and patient specific random intercept with 2 classes produced reliable results (BIC:1412).
- Final Model identified 2 trajectories:
 - Stable:** 37 (72.5%) patients, mean posterior probability: 0.86
 - Improved:** 14 (27.45%) patients, mean posterior probability: 0.88

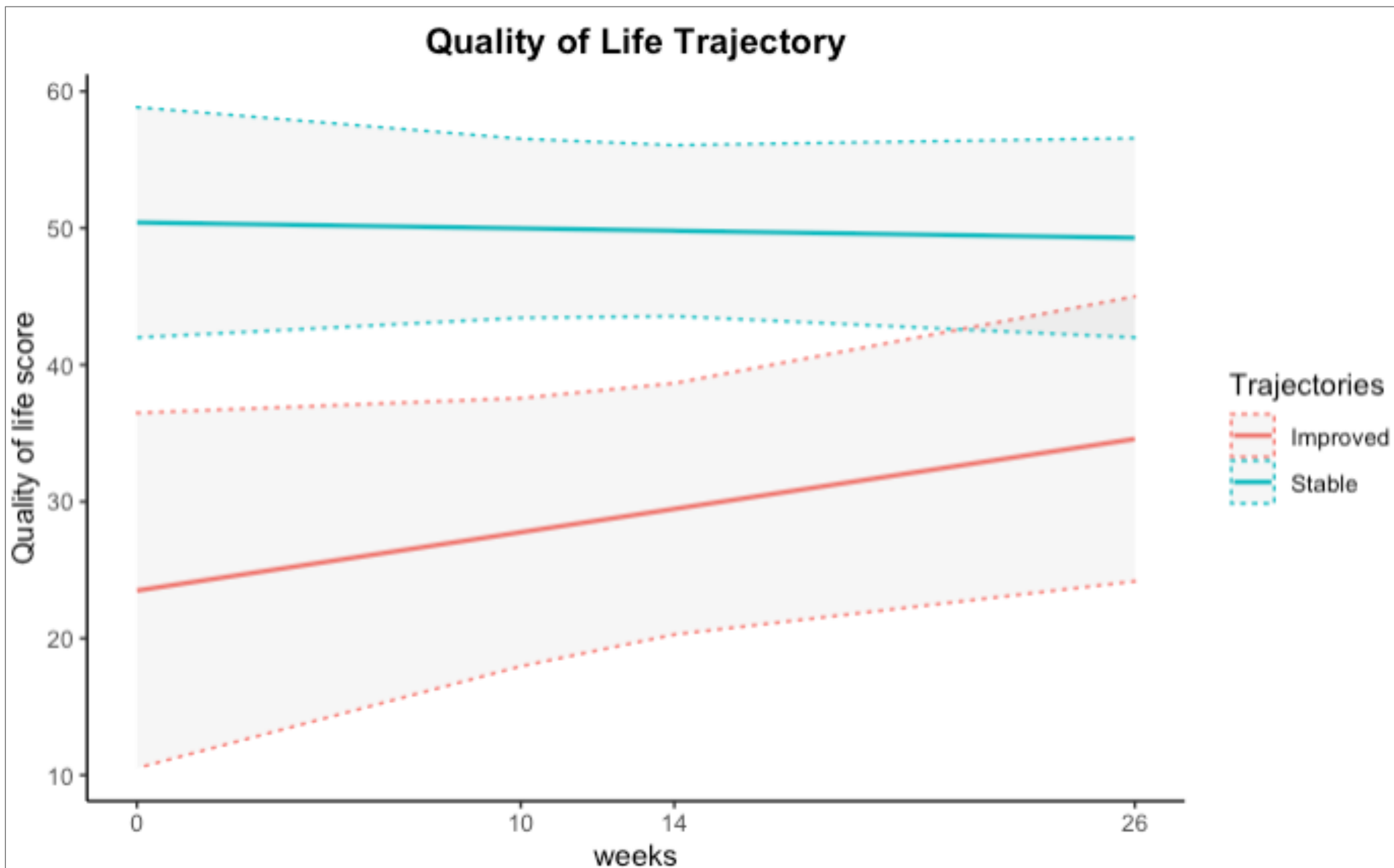
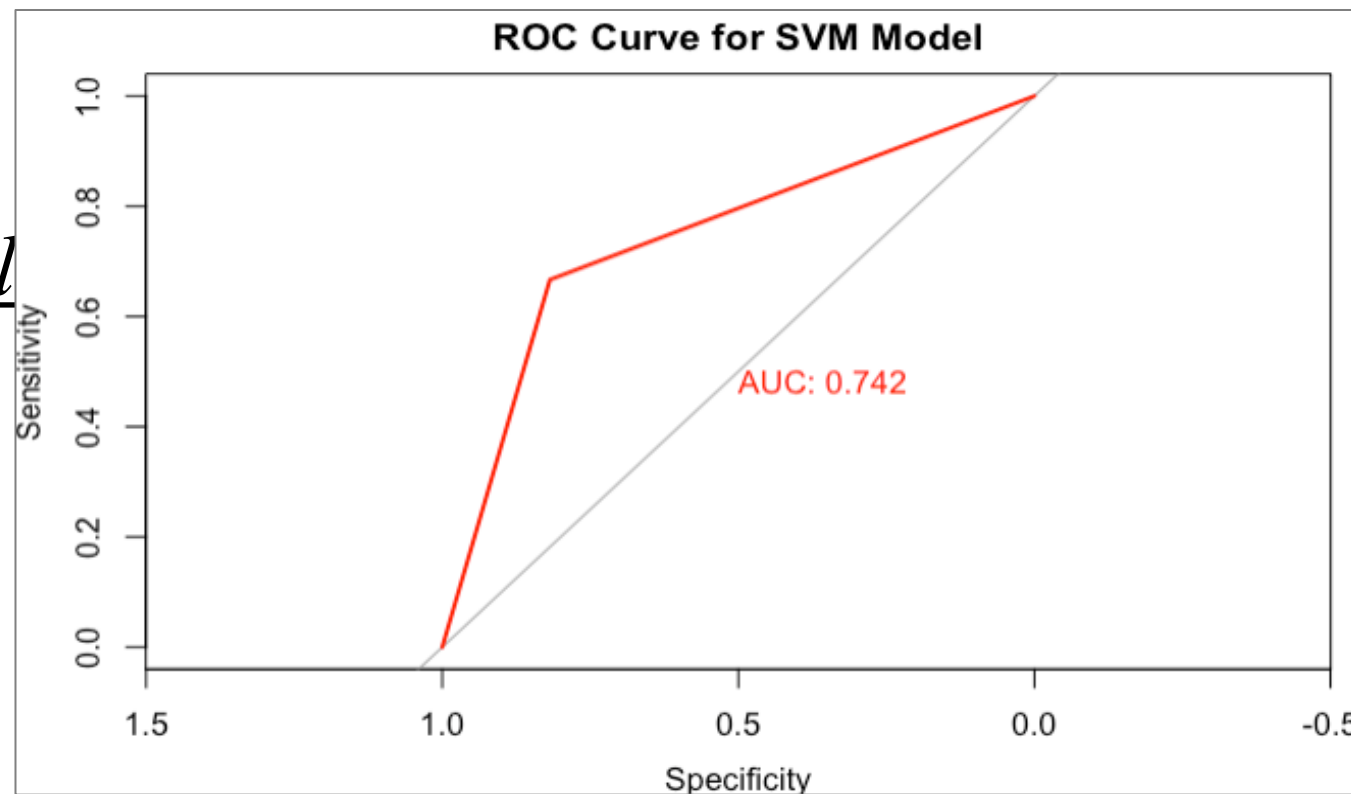


Figure 1: QOL trajectories identified by the latent class mixed model.

Classification:

- Classification algorithms were fit using the class membership from the LCM model as binary outcome and patient baseline and clinical characteristics as predictors.
- Lasso regression regression ($\lambda = 0.15$) lead to a complete shrinkage of 8 coefficients.
- Important variables from Random Forest were used to fit an SVM model.
- Prediction Accuracy and Recall
 - Lasso regression: 0.80, 0.45
 - Random Forest: 0.76, 0.45
 - SVM: 0.80, 0.67



CONCLUSION

- Latent class model indicates that there are 2 subgroups with distinct quality of life trajectory.
 - Stable group maintained an average score of 49.87 ± 4.6 throughout the study whereas the improved group started with average QOL score of 23.49 and improved to 34.58 over time.
 - Support Vector Machine indicated that the covariates such as years of education, WHO disability score, SF-12 Mental score, and Liverpool self-efficacy scales are significant predictors of trajectory classification.
- Limitations :
- Small sample size due to early dropouts
 - Aggregate of clinical scores utilized for classification

Table 1: Clinical and demographic characteristics of patients classified in the two groups

Characteristic	Improved Trajectory, N = 14 ¹	Stable Trajectory, N = 37 ¹	p-value ²
Treatment			0.6
PID-C	6 (43%)	21 (57%)	
PSA	8 (57%)	16 (43%)	
Gender			>0.9
Male	4 (29%)	12 (32%)	
Female	10 (71%)	25 (68%)	
Age	75(8)	72(10)	0.4
W.H.O. Disability Assessment	36(7)	28(7)	<0.001
Hamilton Depression Scale	19.3(4.1)	16.2(4.5)	0.021
LivControl.y	13.41(2.18)	15.11(2.51)	0.026
Self-Efficacy Scale-Personal	11.06(1.00)	12.56(1.75)	<0.001
SF-12 Mental Score	37(6)	47(9)	<0.001
SF-12 Physical Score	27(4)	32(9)	0.023
Dyspnea Score	50(20)	32(19)	0.004

¹ Statistics presented: n (%); mean(SD)

² Statistical tests performed: chi-square test of independence; Fisher's exact test; Wilcoxon rank-sum test

BIBLIOGRAPHY

- Wardenaar, Klaas J. "Latent Class Growth Analysis and Growth Mixture Modeling Using R: A Tutorial for Two R-packages and a Comparison with Mplus." PsyArXiv, 7 Apr. 2020. Web.
- Lennon H, Kelly S, Sperrin M, *et al.* Framework to construct and interpret latentT class trajectory modelling. *BMJ Open* 2018;**8**:e020683. doi: 10.1136/bmjopen-2017-020683