Exploring Longitudinal Pulmonary Exacerbation Outcome Trajectories in Cystic Fibrosis Patients

Brenton Graham
12/08/2022

Introduction

Pulmonary exacerbations (PEx) are a leading cause of morbidity in cystic fibrosis (CF). Treatment response, which can be assessed through a variety of clinical markers, is often suboptimal despite seemingly appropriate antimicrobial therapy and can be quite variable among patients. In this study, CF patients were evaluated at three time points over the course of a PEx in order to assess treatment effectiveness. Time points include admission, hospital discharge and a follow-up clinic visit. PEx-related outcomes were measured at each time point, including lung function, PEx scores, and inflammation markers. Using neutrophil elastase, a marker of inflammation and lung disease, as a biological proxy for treatment response, we first aim to cluster patients into treatment response trajectory groups with a latent class mixed model (LCMM) approach. We then aim to identify potential baseline predictors of treatment response trajectory using the latent classes assigned by the model.

Methods

The Data

Data from 34 CF patients and 39 PEx are present in this data set (i.e., five subjects were evaluated over the course of two separate PEXs). For each PEx there are three repeated measures (T1 = Admission, T2 = Hospital Discharge, T3 = Follow-Up Visit). Time points are approximately separated by 10-14 days. All 39 PEx observations are treated independently throughout the study. The outcome of interest in this study is neutrophil elastase, a marker of inflammation and lung disease. Elastase measurements were log-transformed to handle skewness and more closely approximate a Gaussian distribution. Predictors of interest include baseline age, BMI, genotype (number of F508del mutations), number of hospitalizations in the past year, and CF pathogen measurements (as measured by 16S sequencing).

Exploratory Data Analysis

Summary statistics and data visualizations were used to explore clinical outcomes over time. Data visualizations of log-transformed elastase are shown in Figure S1 (Supplementary Section). Figure S1-A shows the distribution of the outcome stratified by time point within a PEx. Figure S1-B shows subject-level change in the outcome over the course of a PEx.

Statistical Analysis

The primary aim of this analysis was to characterize baseline risk factors associated with treatment response trajectory (as measured by neutrophil elastase). A two-step statistical analysis was conducted to perform this assessment. First, we used a latent class mixed model (LCMM) approach to cluster subjects into latent classes based on treatment response trajectories. Second, we performed a post-hoc analysis to determine if associations exist between baseline risk factors and treatment response trajectory group. Specifications for each analysis step are provided below.

Latent Class Mixed Modeling

As mentioned, latent class analysis (LCA) was used to assign latent classes to each subject based on elastase trajectory over time. The `lcmm` package in R was used to perform LCA. LCMMs require users to specify the number of latent classes that should be fit to the data. The optimal number of latent classes is not always trivial, however, so multiple models usually need to be fit and compared to determine the number of classes that best fit the data. We fit LCMMs with 2-5 latent classes for comparison. All LCMM models were fit using time as a fixed-effect, a random intercept for subject ID, and a random-effect for time. We include these random effects to account for the subject-to-subject variability at the intercept and additional between-subject correlation at each time point. Time was used as the only fixed effect so that latent classes were based solely on the outcome trajectory. Additionally, time was set as a class variable due to the limited number of time points and the apparent non-linear association between time and the outcome. We tested different variance-covariance matrix structures for the random-effects, including unstructured and diagonal matrices (similar to banded main diagonal, or UN(1) in SAS), for each number of latent classes. This resulted in a total of eight LCMM models for comparison. The final model was selected using a holistic approach; we considered a combination of AIC, latent class balance and visualizations to make the final decision (Table S1 and Figure S2). The selected model uses four latent classes and a diagonal variance-covariance matrix structures for the random-effects. Latent classes were assigned qualitative names based on mean group trajectory (Figure 1). First, latent classes are separated by baseline elastase levels, which can be considered "high" and "low" groups. Second, within each "high" and "low" group there are sub-groups of "responders" and "non-responders".

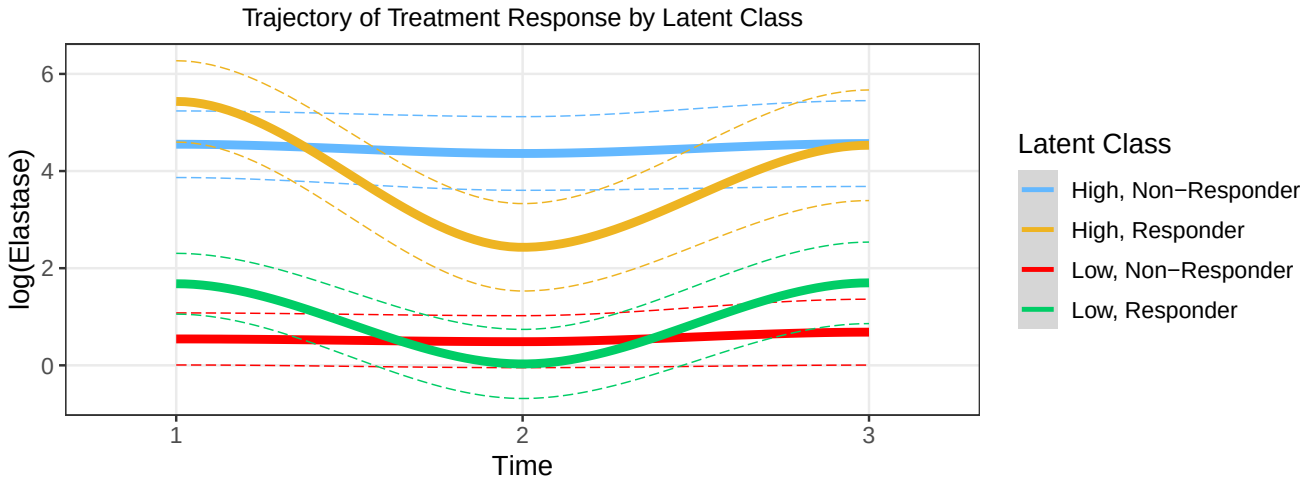

Figure 1: Mean structure of treatment response trajectory by Latent Class

Post-Hoc Analysis

Following LCMM selection and latent class assignment, we performed a post-hoc analysis that aimed to identify potential baseline risk factors associated with the different outcome trajectories. The post-hoc analysis was separated into two sections due to sample size considerations. In the first section we aim to identify differences between the high and low baseline elastase groups (i.e., responders and non-responders are grouped in together for each of the high and low groups). In the second section we aim to identify differences between responders and non-responders, regardless of baseline elastase group (i.e., all responders are grouped together and all non-responders are grouped together). A logistic regression framework is used in each section, using genotype (number of F508del mutations), presence of virus at admission, and baseline CF pathogen abundances as covariates. CF pathogens include Pseudomonas aeruginosa, Staphylococcus aureus, Achromobacter, Haemophilus, Stenotrophomonas, and Burkholderia, and were measured using 16S rRNA sequencing. A center-log ratio transformation was applied to the 16S compositional data.

Results

Table 1 describes the demographics and baseline risk factors of the study population, stratified by assigned latent class. Gender, age and BMI appear to be well-balanced across groups, aside from females not being represented in the "high responder" group with only 5 subjects, and are therefore not adjusted for within the post-hoc models.

Table 1: Demographics for study population by latent class

	High, Non-Responder	High, Responder	Low, Non-Responder	Low, Responder
n	8	5	16	10
Female (%)	4 (50.0)	0 (0.0)	10 (62.5)	5 (50.0)
Age	15.50 [13.00, 20.00]	16.00 [12.00, 18.00]	15.73 [11.93, 19.90]	17.77 [12.91, 22.05]
BMI	19.98 [16.97, 23.00]	18.51 [17.95, 20.48]	19.57 [16.42, 28.70]	19.89 [16.61, 25.14]
Mutations				
0 F508del (%)	2 (25.0)	0 (0.0)	1 (6.2)	1 (10.0)
1 F508del (%)	4 (50.0)	2 (40.0)	5 (31.2)	1 (10.0)
2 F508del (%)	2 (25.0)	3 (60.0)	10 (62.5)	8 (80.0)
Virus at Admission (%)	1 (12.5)	0 (0.0)	7 (46.7)	2 (28.6)
Pseudomonas aeruginosa	4.54 [-0.81, 11.51]	5.09 [4.54, 5.74]	3.48 [-0.99, 10.78]	3.48 [-0.82, 12.23]
Staphylococcus aureus	8.68 [-0.77, 12.47]	12.51 [6.33, 12.70]	5.31 [-0.96, 13.17]	6.55 [-1.04, 10.83]
Achromobacter	-0.71 [-1.07, 3.78]	-0.63 [-1.14, -0.28]	-0.50 [-0.95, 10.20]	-0.63 [-1.12, 12.07]
Haemophilus	1.07 [-0.86, 10.65]	4.45 [-0.63, 7.38]	3.31 [-0.79, 7.63]	5.54 [-0.82, 12.20]
Stenotrophomonas	-0.79 [-1.07, -0.56]	-0.63 [-1.14, 6.23]	1.94 [-0.99, 12.01]	1.51 [-0.82, 6.37]
Burkholderia	1.10 [-0.86, 3.67]	-0.63 [-1.14, -0.28]	-0.68 [-0.99, 3.22]	-0.56 [-1.12, 12.85]

Fixed-effects estimates from the selected latent class mixed model are summarized in Table 2. These estimates provide statistical insight into the differences in latent class trajectories over time. Responder trajectories, for both the high and low elastase groups, are characterized by significant decreases in elastase from admission to discharge. Average decreases of 3.00 and 1.65 units are observed for these groups, respectively, with corresponding p-values below 0.001. Non-responder trajectories, for both the high and low elastase groups, are characterized by no significant change from admission to discharge or from admission to follow-up. Interestingly, no significant change in elastase is observed between admission and follow-up for the low elastase responder group. Separately, the high elastase responder group is associated with a significant elastase decrease between admission and follow-up ($p = 0.028$). The variance-covariance matrix of the random effects is shown in Table S2.

Table 2: Fixed Effects from LCMM

Latent Class	Time Comparison	Estimate	p-value
High, Non-Responder	Discharge vs. Admission	-0.190	0.359
High, Responder	Discharge vs. Admission	-3.004	0.000
Low, Non-Responder	Discharge vs. Admission	-0.056	0.747
Low, Responder	Discharge vs. Admission	-1.652	0.000
High, Non-Responder	Follow-Up vs. Admission	0.014	0.966
High, Responder	Follow-Up vs. Admission	-0.903	0.029
Low, Non-Responder	Follow-Up vs. Admission	0.140	0.570
Low, Responder	Follow-Up vs. Admission	0.018	0.955

Results from the post-hoc logistic regression models are shown in Table 3. Estimates represent odds ratios. One-unit increases in Staphylococcus aureus and Stenotrophomonas (two CF pathogens) at baseline are associated with significantly increased odds of being in the high elastase group, as compared to being in the low elastase group. Separately, and unexpectedly, the odds of having 2 F508del mutations (compared to 0 F508del mutations) appear to be higher for the responder group as compared to the non-responder group. While not significant at an α-level of 0.05, a one unit-increase in Staphylococcus aureus is associated with increased odds of being a responder, as

compared to being a non-responder (OR = 1.37 [95% CI: 1.03, 2.02], p = 0.058). Box plots of CF pathogen distributions for each latent class are shown in Figure S3.

Table 3: Results from Logistic Regression Models Comparing Latent Groups

	High vs. Low Elastase		Responders vs. Non-Responders	
	Odds Ratio (95% CI)	p-value	Odds Ratio (95% CI)	p-value
1 F508del Mutation	5.30 (0.12, 491.16)	0.392	5.89 (0.28, 272.38)	0.291
2 F508del Mutations	3.93 (0.02, 809.40)	0.586	124.79 (3.02, 27742.79)	0.029*
Virus at Admission	0.76 (0.01, 54.80)	0.893	0.31 (0.01, 4.87)	0.409
Pseudomonas aeruginosa	1.05 (0.77, 1.50)	0.762	1.23 (0.96, 1.67)	0.135
Staphylococcus aureus	1.68 (1.14, 3.10)	0.030*	1.37 (1.03, 2.02)	0.058
Achromobacter	0.77 (0.43, 1.12)	0.247	1.08 (0.80, 1.45)	0.579
Haemophilus	1.50 (0.97, 2.88)	0.128	0.93 (0.64, 1.34)	0.690
Stenotrophomonas	0.52 (0.21, 0.82)	0.036*	0.92 (0.64, 1.22)	0.576
Burkholderia	0.69 (0.22, 1.12)	0.352	1.34 (0.94, 2.69)	0.229

Discussion

Fixed effects from the LCMM suggest that subjects in the low elastase responder group return to pre-treatment elastase levels 10-14 days after discharge. This could potentially be interesting biologically. Perhaps lung disease activity is only quenched for this group during and directly after treatment. Meanwhile, high elastase responders see a signficant decrease between admission and follow-up, suggesting a more sustained response. Nonetheless, this significant finding still might not be clinically significant; high levels of elastase are still measured at follow-up which indicates undesired activity. When comparing latent class differences, it was interesting to see that an increase in Staphylococcus aureus is associated with a borderline significant odds increase in responder versus non-responder. On the surface, this effect appears to go against intuition; we wouldn't necessarily expect an increase in baseline CF pathogen to be associated with better treatment response. One thought here is that this could be an effect of treatment type, rather than CF pathogen. It's possible that subjects with increased Staphylococcus aureus were treated with an anti-microbial that targets this pathogen, and that this anti-microbial treatment is the reason for improved response. This is very speculative, however, and clinical data sets should be further explored to determine if treatment information is available.

In light of the results here, limitations are prevalent and should be discussed. First, there should be skepticism when treating latent classes as biologically meaningful groups, especially since latent class assignments can be quite variable iteration-to-iteration (i.e., even for the same model). Correspondingly, post-hoc analysis results should be viewed as exploratory. Second, the sample size in this study is extremely limited, which is cause for concern when considering generalizability. Third, we treated all exacerbations as independent observations in this study even though some subjects enrolled more than once. Nonetheless, latent class analysis seems to be an interesting application of linear mixed models and can provide an interesting avenue of longitudinal analysis when meaningful classes aren't available.

Contributions

Thank you to Brandie Wagner, PhD, and Jonathan Kirk Harris, PhD.

Supplementary Figures & Tables

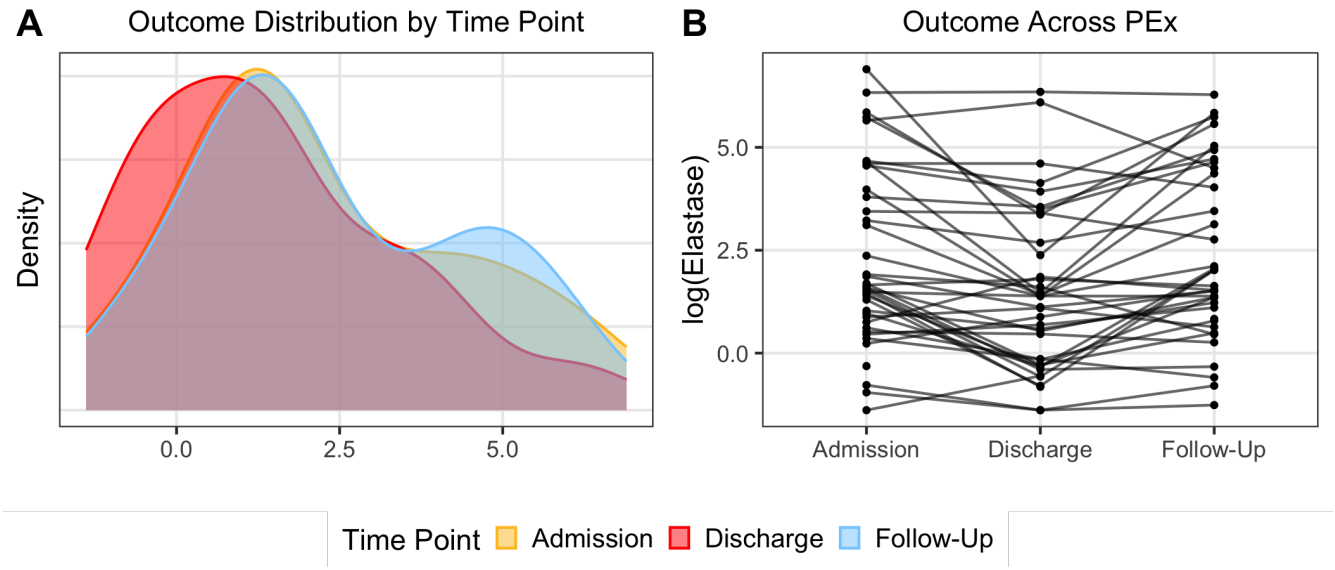

Figure S1: Data visualizations of log-transformed elastase

Table S1: Model comparison of latent class mixed-models fit with different numbers of latent classes

Latent Classes	Var-Cov Structure	AIC	LC1	LC2	LC3	LC4	LC5
2	Diagonal	390.5	66.7%	33.3%			
2	Unstructured	394.4	66.7%	33.3%			
3	Diagonal	391.7	69.2%	23.1%	7.7%		
3	Unstructured	395.4	48.7%	30.8%	20.5%		
4	Diagonal	392.3	41%	20.5%	12.8%	25.6%	
4	Unstructured	396.4	48.7%	20.5%	10.3%	20.5%	
5	Diagonal	392.8	43.6%	20.5%	10.3%	17.9%	7.7%
5	Unstructured	404.4	43.6%	10.3%	17.9%	17.9%	10.3%

Table S2: Variance-covariance matrix of the random effects in the LCMM

	intercept	timeDischarge	timeFollowUp
intercept	0.772		
timeDischarge	0.000	0.165	
timeFollowUp	0.000	0.000	0.632

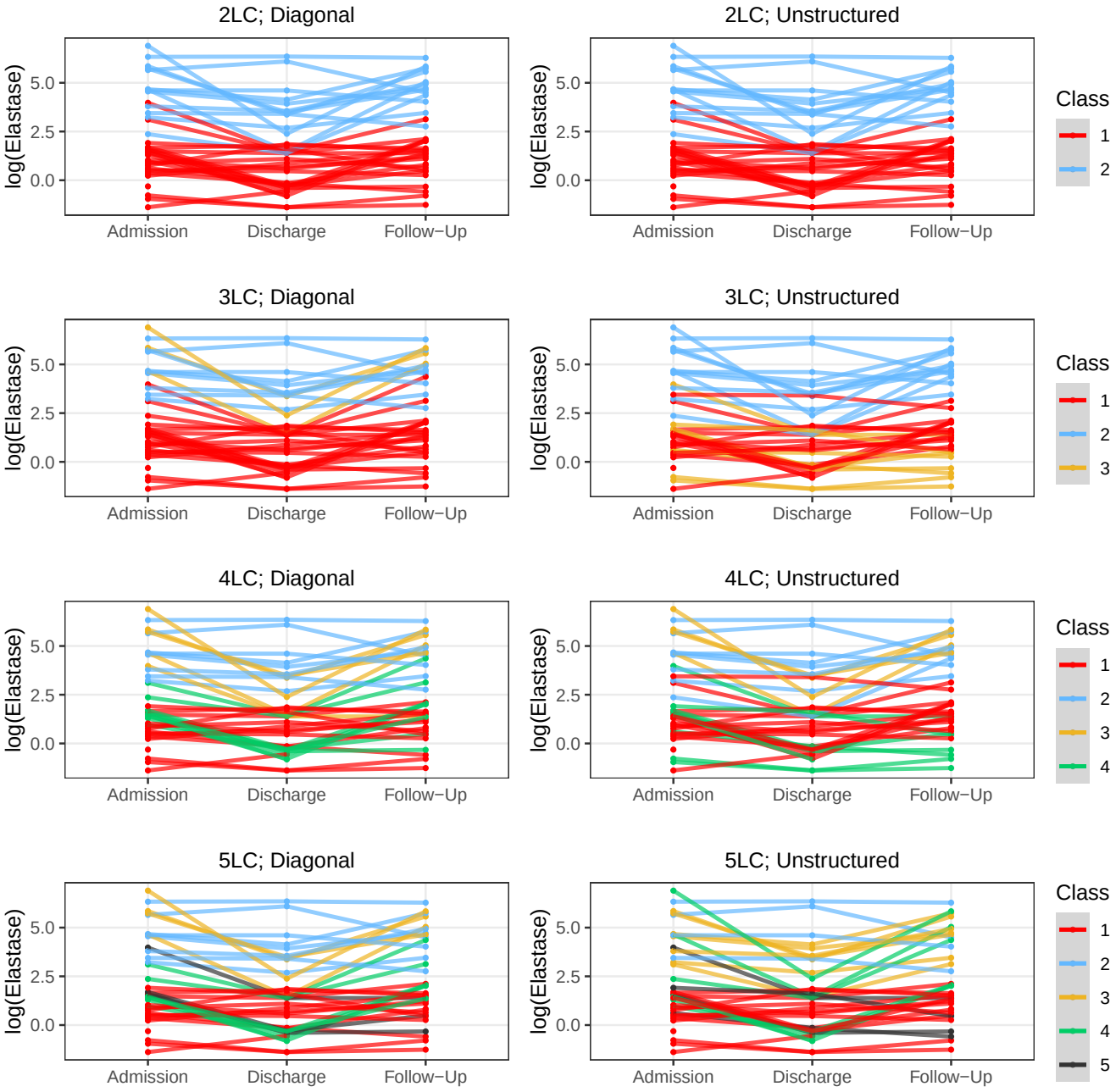

Figure S2: Spaghetti plots colored by latent class assignment for each model tested

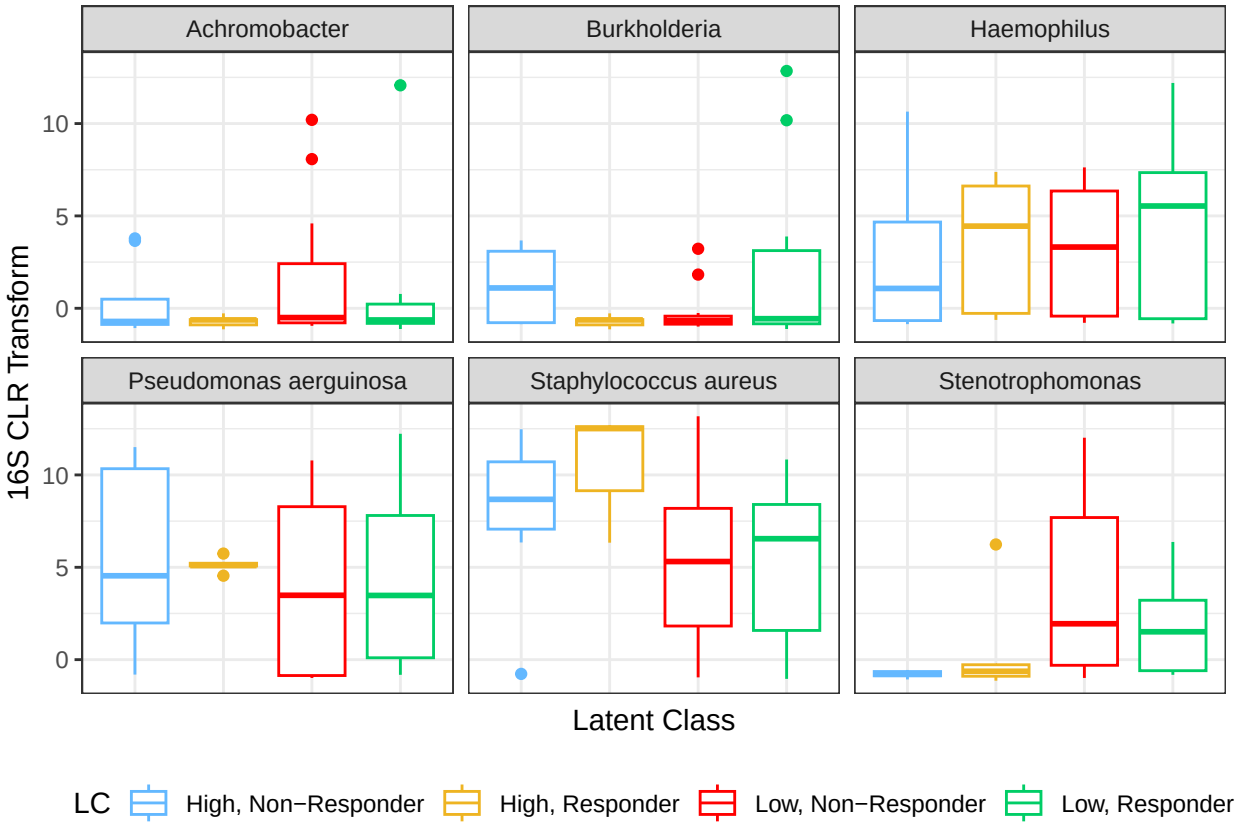

Figure S3: Distribution of CF pathogens at baseline for each latent class in final model