* Non-tuberculous mycobacterial lung disease is caused by organisms belonging to the mycobacterium avium complex. The disease is associated with radiologic features that can be seen in the figure on the right. There are indications in the medical literature that the right middle lobe, lingula, and right upper lobe are more severely affected by the disease than other areas of the lung
* So our research questions for this project are “are there differences in the frequencies and severities for each of the specific CT features among the lung lobes in the disease? and are there differences in these frequencies/severities between the different species of mycobacteria. Our approach to this project is based on previous analyses performed by Matt. In his project, there were 71 patients and he used a consensus score achieved between raters to perform analysis. In our project, we will instead include each of the raters scores making rater a variable in our analysis. We will also compare different models fit through R and SAS
* The different features we will be studying are presented here. Features with ordinal outcomes are tree-in-bud, ground glass opacities, consolidation, bronchiectasis, and atelectasis, whose scores depend on the percent of the lobe affected by the disease, with 0 meaning no involvement, 1 meaning <25% of involvement, 2 meaning 25-50% involvement, and 3 meaning >50% involvement. Features with binary scores include large nodules and thin and thick wall cavities, with 0 indicating absence of this feature in the lobe and 1 indicating its presence
* Our study has 161 subjects, with 12 repeated measures per subject (6 scores for each of the lobes and 2 raters scoring each subject). The variables present in our preliminary analysis include the subject’s ID, lobe, score of the lobe, and rater. The data was provided to us in excel document with each rater as a separate sheet of scores. Minimal data cleaning was performed, mostly to get the data from wide to long format. I also put each feature in its own sheet for ease of extracting during modeling.
* Our preliminary modeling approaches include modeling both in SAS and R. In SAS we used glimmix to model both features with ordinal and binary outcomes, with lobe as a fixed effect, random intercepts, and rater as a random effect. A few of the binary models did not converge, so we used rater as fixed effect for these. In R, we modeled the ordinal outcome in two ways, one using simple fixed effect models in the polr() function, which allowed us to run the model through the brant() function to assess the proportional odds assumptions. We also modeled them using the clmm() function with lobe and rater as fixed effects and random intercepts. This function did not allow the use of rater as a fixed effect, as there were only two raters. We modeled the binary features in R through glmer() with lobe and rater as fixed effects and random intercepts. Similar to SAS, we had issues with convergence for thick wall cavity, so we modeled this with rater as fixed effect as well.
* Here is an example of the ordinal model set up in SAS, and on the right is how I computed pairwise comparisons between lobes using estimate statements, so that we can address the first research question
* Similarly, here is the set up for our binary models in SAS, and on the right again is how I computed pairwise comparisons, taking the difference of lsmeans between the lobes
* Here I present a few examples of significant differences between lobes for a given feature. We have here that RML and the LLS had greater severity of bronchiectasis, with RML having significantly higher grade vs all other lobes, and LLS having significantly higher grade than everything but RML. Our R models had similar results, but the RML-LLS and LLS-RUL comparisons were not significantly different in these models
* Similarly, RML and LLS had greater severity of atelectasis, with RML having significantly higher grade than all other lobes, and LLS having significantly higher grade than every lobe except RML. Again, our R models yielded similar results, but the RML-LLS comparison was not significantly different.
* The next steps in our analysis include thinking of ways to address the models that didn’t converge. Our analysis was also performed on incomplete data, and we should now use all of the data for our final analyses. We also need to incorporate the MAC species responsible for the disease to address the second research question, pertaining to differences in frequencies between the different species. It would also be interesting to calculate inter class correlation to determine raters’ consistency. We also want to think of ways to incorporate different correlation structures to our models to see how this affects our results.

Works Cited

Choi, Sangbong, et al. “Can physics principles help explain why non-tuberculous mycobacterial

lung disease is more severe in the right middle lobe and lingula?” Journal of Thoracic Disease, vol. 11, no. 11, Nov. 2019, pp. 4847–4854, https://doi.org/10.21037/jtd.2019.10.70.