Final Project - BIOS 6643

Dominic Adducci

Introduction

Sarcoidosis is a potentially lethal disease with ~ 25,000 new cases diagnosed in the United States each year. The disease is characterized by granulomas, and the vast majority (90%) of cases primarily affect the lung. Currently, the standard in care for assessing sarcoidosis are lung images utilizing computed tomography (CT). While this method provides better characterizing of abnormalities compared to chest x-rays there are two significant problems: 1.) CT is primarily based on visual assessment, resulting in high inter- and intra-rater variation, and 2.) these visual assessments are not strongly predictive of disease course. Recently, a new method to better quantify measures of lung CT has been developed utilizing variograms, a standard geostatistical tool for measuring and modelling spatial covariance within an image and spatial process as a function of distance between points.

Variogram data is typically represented by three parameters: 1.) the sill, 2.) the range, and 3.) the nugget. The sill is the overall variation present in an image, the range represents the spatial scale of correlation, and the nugget is the variation due to measurement error. For sufficiently high resolution images the nugget is 0, which will be assumed for this analysis. The location of each individual scan is represented by the normalized slice variable spanning 0 to 1, corresponding to the bottom and top of the lung respectively.

The purpose of this analysis is to answer two primary questions of interest:

Should race be included as a demographic feature that is adjusted for?

How do clinical measures of disease associate with the lungs after adjusting for demographics?

There are six total clinical measures of disease. Three of these are PFT measurements: FEV1 (Forced Expiratory Volume in 1 second), FVC (Forced Vital Capacity), and FEV1/FVC ratio (provided as a fraction). The other three are VAS (Visual Assessment Scores): fibrosis, mediastinal lymphadenopathy, and traction bronchiectasis. Demographics other than race are BMI, age, sex, and height. Analysis involved comparing

models with and without race to determine the impact of including this as a demographic feature, and plots will be presented to illustrated trends in clinical features for both sill and range. Additionally, further research on the ethical and clinical/scientific justification for inclusion of race as a predictor in sarcoidosis was performed.

Methods

Data

The data for this analysis was compiled from two main sources, clinical measurements and variogram measurements. The clinical measurements provided the demographic data and the PFT and VAS measurements, and included 368 subjects. Cleaning the clinical data involved several steps. For simplicity in modeling VAS measurements were considered a binary indicating presence of the condition or no presence of the condition. Both traction bronchiectasis and mediastinal lymphadenopathy had four potential levels, which were then collapsed into binary indicators. Fibrosis was already a binary variable, meaning this step was not necessary for this variable. The construction of the race variable involved combining the provided race and ethnicity variables. The race categories initially had 6 possible options, Black or African American, American Indian or Alaskan Native, white, multi-racial or no primary race, Asian, and unknown. The ethnicity variable denoted whether a subject was Hispanic or not Hispanic. The race variable used in this analysis first involved separating white subjects into Hispanic white and non-Hispanic white. Next, race was collapsed into three categories: 1.) non-Hispanic white, 2.) Black or African American, 3.) and other, which incorporated American Indian or Alaskan Native, multi-racial or no primary race, Asian, unknown, and Hispanic white. Two important features are that subjects who were missing race/ethnicity data were considered "NA", which is distinct from the "unknown" category, and the majority of subjects were non-Hispanic white, followed by Black or African American. The composite category of "other" had the smallest amount of subjects.

Variogram data included information regarding the subject ID for each measurement, as well as which lung each measurement came from. Variogram measurements had three components, the range, the psill, and the slice normalized variable, which indicated location in the lung. Values closer to 0 correspond to the bottom of the lung, and values closer to 1 correspond to the top of the lung. For this analysis normalized slice values between 0.1 and 0.9 were used to avoid measurement variations of the top and bottom-most portion of scans. In general each subject had 50 slices per lung, meaning a total of 100 scans per subject. Two subjects were dropped during initial data cleaning, one due to an outlier in the range measurement (> 300), and another

due to missing data on the left lung. Out of the original 368 subject 366 were kept after dropping these two. Lastly, psill and range were log transformed due to prominent right-skew in distribution.

After initial cleaning both data sets were combined. As there were more subject with clinical data than variogram data the subject IDs for variogram data were used for the combining, resulting in a composite data frame of 342 subjects. A complete case analysis was performed, meaning subjects missing any of the relevant demographic data and clinical measurements were dropped from the analysis. After dropping there were 320 subjects.

Table 1: Descriptive Statistics

	Black or African American	Other Races	Non-Hispanic White	Overall
	(N=7800)	(N=2300)	(N=21900)	(N=32000)
Age				
Mean (SD)	52.5 (8.87)	51.7 (8.37)	53.3 (9.98)	53.0 (9.63)
Median [Min, Max]	52.8 [30.8, 70.1]	52.9 [34.9, 66.9]	53.9 [26.3, 77.3]	53.4 [26.3, 77.3]
BMI				
Mean (SD)	31.8 (6.85)	31.5 (6.38)	30.1 (6.22)	30.6 (6.44)
Median [Min, Max]	31.3 [19.8, 51.2]	29.6 [24.1, 48.7]	29.1 [18.2, 51.8]	29.7 [18.2, 51.8]
Height				
Mean (SD)	66.6 (3.84)	67.0 (3.96)	67.2 (4.04)	67.0 (4.00)
Median [Min, Max]	66.0 [60.0, 77.0]	67.0 [61.0, 76.0]	67.0 [56.0, 77.0]	67.0 [56.0, 77.0]
Gender				
Female	5400 (69.2%)	900 (39.1%)	10600 (48.4%)	16900 (52.8%)
Male	2400 (30.8%)	1400 (60.9%)	11300 (51.6%)	15100 (47.2%)
Fibrosis				
No	7200 (92.3%)	2000 (87.0%)	19300 (88.1%)	28500 (89.1%)
Yes	600 (7.7%)	300 (13.0%)	2600 (11.9%)	3500 (10.9%)
Traction Bronchiectasis				
No	3800 (48.7%)	1600 (69.6%)	13700 (62.6%)	19100 (59.7%)
Yes	4000 (51.3%)	700 (30.4%)	8200 (37.4%)	12900 (40.3%)
Mediastinal Lymphadenopathy				
No	4100 (52.6%)	900 (39.1%)	9300 (42.5%)	14300 (44.7%)
Yes	3700 (47.4%)	1400 (60.9%)	12600 (57.5%)	17700 (55.3%)
POST FEV1				
Mean (SD)	2.21 (0.749)	2.74 (0.733)	2.88 (0.975)	2.71 (0.953)
Median [Min, Max]	2.15 [0.830, 4.26]	2.87 [1.14, 4.28]	2.77 [0.740, 5.78]	2.58 [0.740, 5.78]
POST FVC				
Mean (SD)	2.96 (0.924)	3.78 (0.983)	3.81 (1.13)	3.60 (1.14)
Median [Min, Max]	2.83 [1.12, 5.57]	3.74 [2.02, 5.92]	3.65 [1.62, 7.27]	3.45 [1.12, 7.27]
POST FEV/FVC				
Mean (SD)	0.751 (0.118)	0.738 (0.129)	0.753 (0.108)	0.752 (0.112)
Median [Min, Max]	0.775 [0.322, 0.964]	0.757 [0.313, 0.883]	0.775 [0.258, 0.976]	0.775 [0.258, 0.976]

The table above shows descriptive statistics of the study cohort. The majority of subjects are non-Hispanic white, followed by Black or African American. Other makes up the smallest group, being around a tenth of the number of non-Hispanic white subjects. The average age, BMI, and height are similar between the three race groups. Rates of fibrosis and mediastinal lymphadenopathy hare highest in the other race category,

and rate of traction bronchiectasis are highest in the Black and African American category. The white race group has the highest average PFT measurements.

Analysis

Analysis utilized cubic penalized splines with 4 degrees of freedom to model the relationship between location in the lung and each clinical predictor, allowing for non-linear associations between these two variables. The logged outcomes for both psill and range were considered Gaussian-distributed. Models for each clinical predictor were made separately, with each outcome being having its own model, and each lung having its own model. Additionally, models including race and excluding race were compared. In total, each clinical predictor had 4 models for each outcome (right lung psill and range, left lung psill and range) including race and 4 models for each outcome excluding race. Between the six predictors this results in a total of 48 models. Additional models parameters include demographic data beyond race (BMI, age, sex, height), and a subject-specific random effect.

Models including and excluding race were compared with AIC, where lower scores indicate better model fitting. Plots of trend within the lung for each model were included to understand how each clinical predictor affects psill and range measurements, with point estimates and 95% confidence intervals included in the plots. The VAS predictors are stratified into yes/no, and the PFT measurements are stratified into the 25th percentile, 50th percentile, and 75th percentile for subjects in the cohort. For plots predicted values were based on a subject who is an African American/Black female who is the average age (53 years), average BMI (30.6), and average height(67 inches) for all subjects in the cohort.

Additional literature analysis regarding the inclusion/exclusion of race as a predictor was included when discussing significance of the results. For this analysis the category of “Black or African American” will be the reference group against which “non-Hispanic white” and “other” will be compared. All analysis were performed in R version 4.3.1.

Results

Direct Model Comparison

Table 2: AIC Across Models

	Psill		Range	
Model Type	Left	Right	Left	Right
Fibrosis				
Including Race	-470.2398	839.6175	-4143.537	-4302.651
Excluding Race	-470.0661	839.7928	-4143.586	-4302.715
Mediastinal Lymphadenopathy				
Including Race	436.4790	1531.6211	-3393.730	-3680.141
Excluding Race	436.6231	1531.7868	-3393.806	-3680.243
Traction Bronchiectasis				
Including Race	-368.3259	800.9055	-4224.133	-4452.619
Excluding Race	-368.2448	800.9907	-4224.248	-4452.822
FEV1				
Including Race	279.5477	1421.0749	-3591.071	-3877.065
Excluding Race	279.4999	1421.0288	-3591.168	-3877.219
FVC				
Including Race	368.8143	1457.8577	-3431.659	-3782.729
Excluding Race	368.5150	1457.7810	-3431.766	-3782.876
FEV1/FVC				
Including Race	146.8869	1311.4856	-3788.683	-3838.928
Excluding Race	147.0078	1311.6224	-3788.805	-3839.067

The table above compares AIC values between different models including and excluding race as a predictor. Psill models including race have a lower AIC for fibrosis, mediastinal lymphadenopahty, traction bronchiectasis, and FEV1/FVC. For FEV1 and FVC, psill models excluding race have a lower AIC. However, the delta in AIC between models including and excluding race are less than 1, suggesting that both models perform similarly.

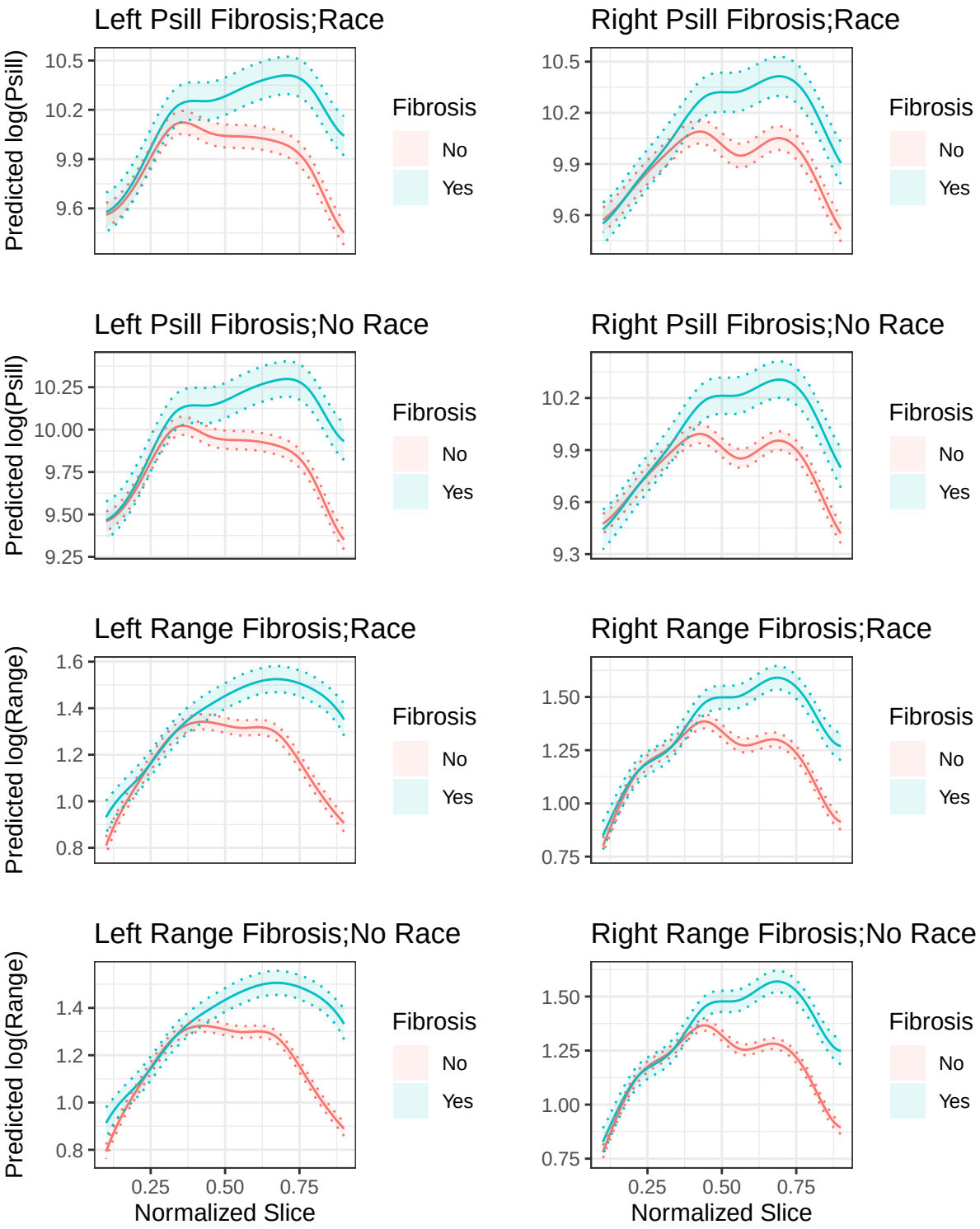
Range models excluding race have lower AIC values compared to models including race. Again, the differences are less than 1 suggesting that these models perform similarly. Between both lungs and both outcomes models excluding race have a higher proportion of lower AIC values compared to models including race.

Fibrosis

Table 3: Summary of Models for Fibrosis

	Left Lung					Right Lung				
Term	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**
Outcome: Psill | With Race										
Intercept	10.2597	0.3977	<2e-16	NA	NA	10.0292	0.3965	<2e-16	NA	NA
BMI	0.0031	0.0025	0.2	NA	NA	0.0034	0.0025	0.18	NA	NA
Age	0.0017	0.0017	0.3	NA	NA	0.0031	0.0017	0.06	NA	NA
Male	0.0271	0.0434	0.5	NA	NA	0.0480	0.0433	0.27	NA	NA
Height	-0.0081	0.0054	0.1	NA	NA	-0.0060	0.0054	0.27	NA	NA
Other Races	-0.1084	0.0675	0.1	NA	NA	-0.1027	0.0673	0.13	NA	NA
Non-Hispanic White	-0.1616	0.0380	2e-05	NA	NA	-0.1574	0.0379	3e-05	NA	NA
s(Slice)	NA	NA	<2e-16	8.5137	8.8983	NA	NA	<2e-16	8.4116	8.8505
s(ID)	NA	NA	<2e-16	307.5549	312.0000	NA	NA	<2e-16	307.1460	312.0000
s(Slice):Fibrosis	NA	NA	<2e-16	4.2561	4.9289	NA	NA	<2e-16	6.7540	7.5551
Outcome: Psill | Without Race										
Intercept	9.9869	0.4022	<2e-16	NA	NA	9.7647	0.4005	<2e-16	NA	NA
BMI	0.0044	0.0026	0.09	NA	NA	0.0046	0.0026	0.07	NA	NA
Age	0.0015	0.0017	0.39	NA	NA	0.0029	0.0017	0.09	NA	NA
Male	-0.0094	0.0435	0.83	NA	NA	0.0126	0.0433	0.77	NA	NA
Height	-0.0059	0.0055	0.28	NA	NA	-0.0039	0.0055	0.48	NA	NA
s(Slice)	NA	NA	<2e-16	8.5136	8.8983	NA	NA	<2e-16	8.4117	8.8505
s(ID)	NA	NA	<2e-16	309.7666	314.0000	NA	NA	<2e-16	309.3663	314.0000
s(Slice):Fibrosis	NA	NA	<2e-16	4.2565	4.9294	NA	NA	<2e-16	6.7551	7.5561
Outcome: Range | With Race										
Intercept	1.2840	0.1805	1e-12	NA	NA	1.3706	0.1803	3e-14	NA	NA
BMI	-0.0011	0.0011	0.349	NA	NA	-0.0016	0.0011	0.17	NA	NA
Age	0.0008	0.0008	0.279	NA	NA	0.0016	0.0008	0.03	NA	NA
Male	0.0515	0.0197	0.009	NA	NA	0.0406	0.0197	0.04	NA	NA
Height	-0.0016	0.0025	0.504	NA	NA	-0.0032	0.0024	0.20	NA	NA
Other Races	0.0009	0.0306	0.976	NA	NA	-0.0245	0.0306	0.42	NA	NA
Non-Hispanic White	-0.0293	0.0172	0.090	NA	NA	-0.0299	0.0172	0.08	NA	NA
s(Slice)	NA	NA	<2e-16	7.7680	8.3997	NA	NA	<2e-16	8.5513	8.9123
s(ID)	NA	NA	<2e-16	294.8383	312.0000	NA	NA	<2e-16	294.9661	312.0000
s(Slice):Fibrosis	NA	NA	<2e-16	7.2006	7.9544	NA	NA	<2e-16	5.9516	6.7765
Outcome: Range | Without Race										
Intercept	1.2441	0.1786	3e-12	NA	NA	1.3181	0.1782	1e-13	NA	NA
BMI	-0.0008	0.0011	0.48	NA	NA	-0.0013	0.0011	0.25	NA	NA
Age	0.0008	0.0008	0.32	NA	NA	0.0016	0.0008	0.04	NA	NA
Male	0.0463	0.0193	0.02	NA	NA	0.0336	0.0193	0.08	NA	NA
Height	-0.0014	0.0024	0.58	NA	NA	-0.0027	0.0024	0.26	NA	NA
s(Slice)	NA	NA	<2e-16	7.7681	8.3998	NA	NA	<2e-16	8.5513	8.9123
s(ID)	NA	NA	<2e-16	296.9197	314.0000	NA	NA	<2e-16	297.0172	314.0000
s(Slice):Fibrosis	NA	NA	<2e-16	7.2024	7.9561	NA	NA	<2e-16	5.9519	6.7767

The table above illustrates differences between models including and excluding race when fibrosis is the main clinical predictor. In all models the interaction between normalized slice and fibrosis is a significant predictor for range and psill values regardless of whether or not race was included as a predictor. For psill non-Hispanic white is a significant predictor for both lungs while the other race group is not. For range neither non-Hispanic white nor the other race group are significant predictors in either lung.



The plots above show the trend in range and psill when moving up the lung (normalized slice) when fibrosis is the main clinical predictor. Between the race and no race plots the trends are very similar, and subsequent analysis of fibrosis will consider the models excluding race. For psill the left and right lung both

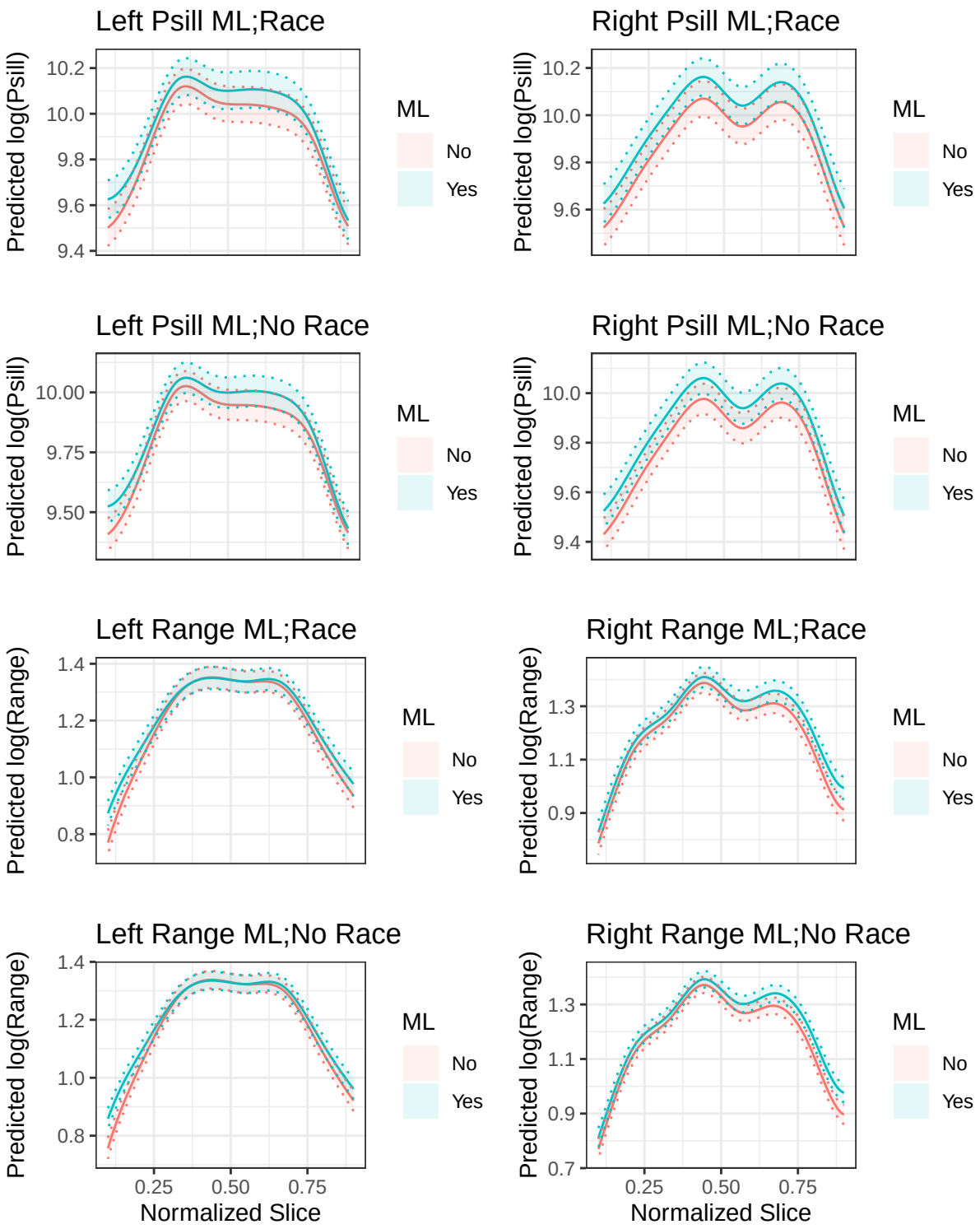
show similar trends where those without fibrosis have noticeably lower psill when moving up the lung. In the lower portion of the lung there is almost no difference between those with and without fibrosis, with the confidence intervals overlapping. For range the left and right plots again show similar trends, where those without fibrosis have lower values compared to subjects who have fibrosis when moving up the lung. The bottom portion of both lungs has significant overlap in range measurements between the two groups.

Mediastinal Lymphadenopathy (ML)

Table 4: Summary of Models for Mediastinal Lymphadenopathy (ML)

	Left Lung					Right Lung				
Term	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**
Outcome: Psill | With Race										
Intercept	10.1695	0.4134	<2e-16	NA	NA	9.9049	0.4055	<2e-16	NA	NA
BMI	0.0012	0.0026	0.6	NA	NA	0.0018	0.0025	0.487	NA	NA
Age	0.0014	0.0017	0.4	NA	NA	0.0026	0.0017	0.121	NA	NA
Male	0.0011	0.0450	1.0	NA	NA	0.0201	0.0442	0.648	NA	NA
Height	-0.0056	0.0056	0.3	NA	NA	-0.0032	0.0055	0.562	NA	NA
Other Races	-0.0965	0.0698	0.2	NA	NA	-0.0963	0.0685	0.160	NA	NA
Non-Hispanic White	-0.1558	0.0393	7e-05	NA	NA	-0.1551	0.0386	6e-05	NA	NA
s(Slice)	NA	NA	<2e-16	8.4725	8.8765	NA	NA	<2e-16	8.3674	8.8304
s(ID)	NA	NA	<2e-16	307.6041	312.0000	NA	NA	<2e-16	307.1081	312.0000
s(Slice):ML	NA	NA	2e-05	4.9908	5.7524	NA	NA	0.005	2.0011	2.0022
Outcome: Psill | Without Race										
Intercept	9.9204	0.4172	<2e-16	NA	NA	9.6568	0.4095	<2e-16	NA	NA
BMI	0.0025	0.0026	0.3	NA	NA	0.0031	0.0026	0.23	NA	NA
Age	0.0012	0.0018	0.5	NA	NA	0.0025	0.0017	0.16	NA	NA
Male	-0.0317	0.0450	0.5	NA	NA	-0.0126	0.0442	0.78	NA	NA
Height	-0.0037	0.0057	0.5	NA	NA	-0.0013	0.0056	0.81	NA	NA
s(Slice)	NA	NA	<2e-16	8.4724	8.8765	NA	NA	<2e-16	8.3675	8.8304
s(ID)	NA	NA	<2e-16	309.7843	314.0000	NA	NA	<2e-16	309.3152	314.0000
s(Slice):ML	NA	NA	2e-05	4.9914	5.7530	NA	NA	0.01	2.0003	2.0007
Outcome: Range | With Race										
Intercept	1.2443	0.1948	2e-10	NA	NA	1.3121	0.1923	9e-12	NA	NA
BMI	-0.0024	0.0012	0.05	NA	NA	-0.0028	0.0012	0.02	NA	NA
Age	0.0007	0.0008	0.38	NA	NA	0.0014	0.0008	0.08	NA	NA
Male	0.0369	0.0212	0.08	NA	NA	0.0242	0.0209	0.25	NA	NA
Height	-0.0003	0.0026	0.92	NA	NA	-0.0016	0.0026	0.55	NA	NA
Other Races	0.0101	0.0329	0.76	NA	NA	-0.0174	0.0325	0.59	NA	NA
Non-Hispanic White	-0.0244	0.0185	0.19	NA	NA	-0.0265	0.0183	0.15	NA	NA
s(Slice)	NA	NA	<2e-16	7.6260	8.2708	NA	NA	<2e-16	8.5329	8.9055
s(ID)	NA	NA	<2e-16	296.4072	312.0000	NA	NA	<2e-16	296.2826	312.0000
s(Slice):ML	NA	NA	<2e-16	4.7276	5.4583	NA	NA	4e-08	3.6236	4.1963
Outcome: Range | Without Race										
Intercept	1.2165	0.1927	3e-10	NA	NA	1.2693	0.1901	2e-11	NA	NA
BMI	-0.0021	0.0012	0.08	NA	NA	-0.0025	0.0012	0.03	NA	NA
Age	0.0007	0.0008	0.42	NA	NA	0.0014	0.0008	0.08	NA	NA
Male	0.0334	0.0208	0.11	NA	NA	0.0186	0.0205	0.36	NA	NA
Height	-0.0001	0.0026	0.97	NA	NA	-0.0012	0.0026	0.63	NA	NA
s(Slice)	NA	NA	<2e-16	7.6260	8.2709	NA	NA	<2e-16	8.5329	8.9055
s(ID)	NA	NA	<2e-16	298.4374	314.0000	NA	NA	<2e-16	298.2852	314.0000
s(Slice):ML	NA	NA	<2e-16	4.7278	5.4586	NA	NA	<2e-16	3.6236	4.1964

The table above illustrates differences between models including and excluding race when mediastinal lymphadenopathy is the main clinical predictor. In all models the interaction between normalized slice and mediastinal lymphadenopathy is a significant predictor for range and psill values regardless of whether or not race was included as a predictor. For psill the non-Hispanic white category is a significant predictor for both lungs while the other race category is not. For range neither non-Hispanic white nor the other race category were significant predictors for either lung.



The plots above show the trend of range and psill when moving up the lung (normalized slice) when mediastinal lymphadenopathy is the main clinical predictor. Similar to fibrosis, the trends in plots including and excluding race are nearly the same, and subsequent analysis will consider plots excluding race. For psill

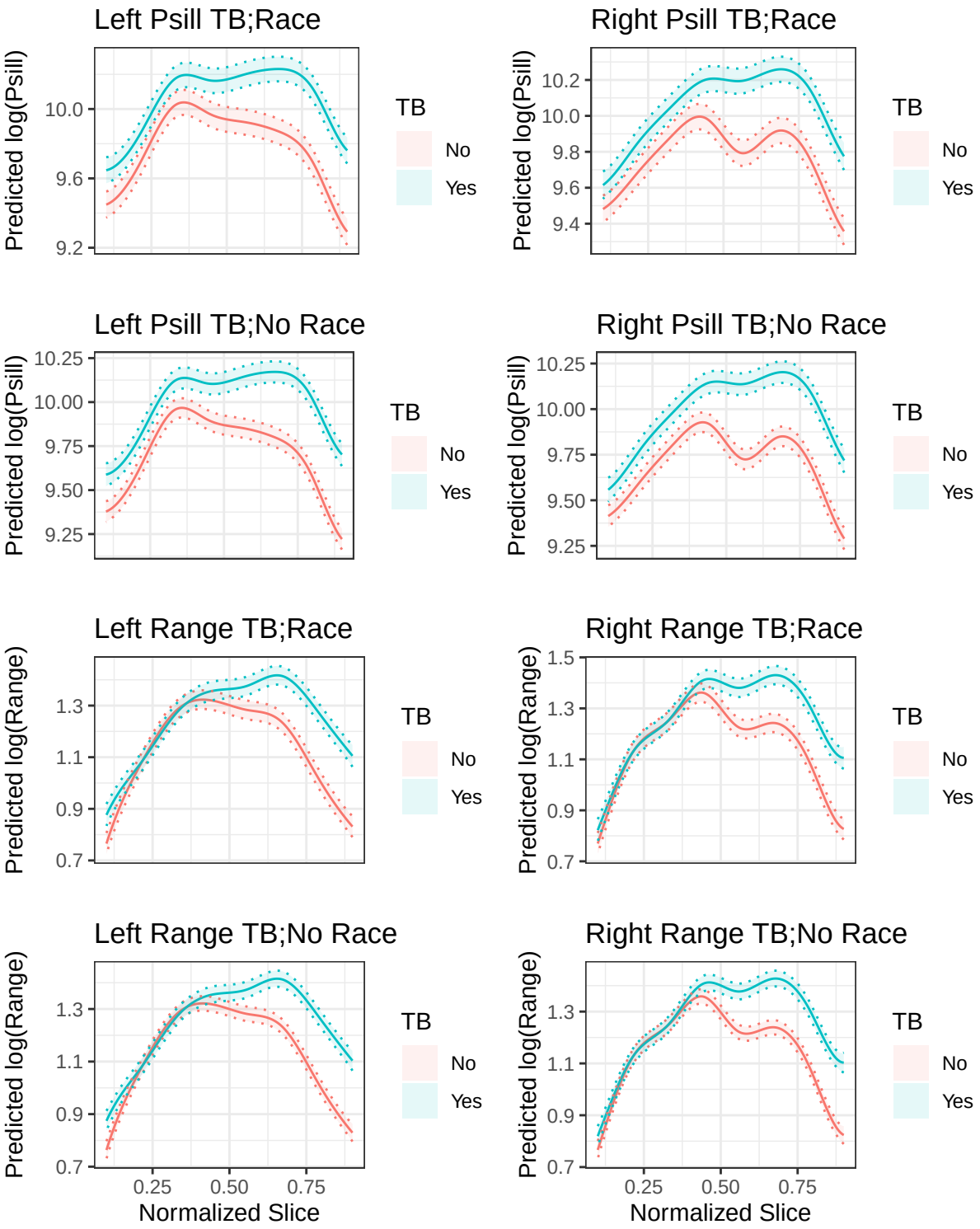
the left lung has more overlap for those with and without mediastinal lymphadenopathy compared to the right lung, however both models have overlap in their 95% confidence intervals. The right lung also has a more noticeable dip in the middle portion of the lung and then a subsequent rise in psill before quickly declining compared to the left lung. For range both left and right lung have significant overlap in 95% confidence intervals between those with and without mediastinal lymphadenopathy, with those having the condition showing slightly higher range values. The left lung plateaus in the middle, while the right lung shows a dip in range values.

Traction Bronchiectasis (TB)

Table 5: Summary of Models for Traction Bronchiectasis (TB)

	Left Lung					Right Lung				
Term	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**
Outcome: Psill | With Race										
Intercept	10.0063	0.3718	<2e-16	NA	NA	9.7719	0.3639	<2e-16	NA	NA
BMI	0.0044	0.0024	0.061	NA	NA	0.0050	0.0023	0.029	NA	NA
Age	-0.0003	0.0016	0.858	NA	NA	0.0011	0.0015	0.473	NA	NA
Male	0.0096	0.0403	0.812	NA	NA	0.0328	0.0395	0.406	NA	NA
Height	-0.0050	0.0050	0.316	NA	NA	-0.0031	0.0049	0.528	NA	NA
Other Races	-0.0350	0.0631	0.580	NA	NA	-0.0301	0.0618	0.626	NA	NA
Non-Hispanic White	-0.1072	0.0357	0.003	NA	NA	-0.1032	0.0350	0.003	NA	NA
s(Slice)	NA	NA	<2e-16	8.5134	8.8950	NA	NA	<2e-16	8.4902	8.8563
s(ID)	NA	NA	<2e-16	306.8533	312.0000	NA	NA	<2e-16	306.2211	312.0000
s(Slice):TB	NA	NA	<2e-16	5.3706	6.1663	NA	NA	<2e-16	8.0633	8.7170
Outcome: Psill | Without Race										
Intercept	9.8358	0.3706	<2e-16	NA	NA	9.6096	0.3627	<2e-16	NA	NA
BMI	0.0055	0.0024	0.02	NA	NA	0.0061	0.0023	0.009	NA	NA
Age	-0.0006	0.0016	0.72	NA	NA	0.0008	0.0015	0.593	NA	NA
Male	-0.0111	0.0400	0.78	NA	NA	0.0131	0.0391	0.738	NA	NA
Height	-0.0038	0.0051	0.45	NA	NA	-0.0020	0.0050	0.692	NA	NA
s(Slice)	NA	NA	<2e-16	8.5134	8.8950	NA	NA	<2e-16	8.4903	8.8564
s(ID)	NA	NA	<2e-16	308.9728	314.0000	NA	NA	<2e-16	308.3517	314.0000
s(Slice):TB	NA	NA	<2e-16	5.3709	6.1666	NA	NA	<2e-16	8.0639	8.7176
Outcome: Range | With Race										
Intercept	1.1797	0.1812	8e-11	NA	NA	1.2639	0.1790	2e-12	NA	NA
BMI	-0.0011	0.0011	0.34	NA	NA	-0.0015	0.0011	0.2	NA	NA
Age	0.0001	0.0008	0.94	NA	NA	0.0008	0.0008	0.3	NA	NA
Male	0.0402	0.0197	0.04	NA	NA	0.0298	0.0194	0.1	NA	NA
Height	0.0000	0.0025	0.98	NA	NA	-0.0016	0.0024	0.5	NA	NA
Other Races	0.0343	0.0308	0.26	NA	NA	0.0092	0.0304	0.8	NA	NA
Non-Hispanic White	-0.0053	0.0174	0.76	NA	NA	-0.0057	0.0172	0.7	NA	NA
s(Slice)	NA	NA	<2e-16	7.6879	8.3277	NA	NA	<2e-16	8.5729	8.9120
s(ID)	NA	NA	<2e-16	294.9688	312.0000	NA	NA	<2e-16	294.7919	312.0000
s(Slice):TB	NA	NA	<2e-16	6.4385	7.2560	NA	NA	<2e-16	6.6566	7.4628
Outcome: Range | Without Race										
Intercept	1.1883	0.1786	3e-11	NA	NA	1.2601	0.1759	8e-13	NA	NA
BMI	-0.0010	0.0011	0.38	NA	NA	-0.0014	0.0011	0.2	NA	NA
Age	0.0000	0.0008	0.99	NA	NA	0.0008	0.0008	0.3	NA	NA
Male	0.0416	0.0192	0.03	NA	NA	0.0294	0.0190	0.1	NA	NA
Height	-0.0002	0.0024	0.93	NA	NA	-0.0016	0.0024	0.5	NA	NA
s(Slice)	NA	NA	<2e-16	7.6879	8.3277	NA	NA	<2e-16	8.5729	8.9120
s(ID)	NA	NA	<2e-16	296.9672	314.0000	NA	NA	<2e-16	296.7007	314.0000
s(Slice):TB	NA	NA	<2e-16	6.4386	7.2562	NA	NA	<2e-16	6.6567	7.4628

The table above illustrates differences between models including and excluding race when traction bronchiectasis is the main clinical predictor. In all models the interaction between normalized slice and traction bronchiectasis is a significant predictor for range and psill regardless of including or excluding race. For psill non-Hispanic race is a significant predictor while the other race category is insignificant in both lungs. For range models neither race category was a significant predictor for either lung.



The plots above show the trend of range and psill when moving up the lung (normalized slice) when traction bronchiectasis is the main clinical predictor. The trends between models including and excluding race are similar, with the psill models including race having slightly more overlap. Subsequent analysis will focus on

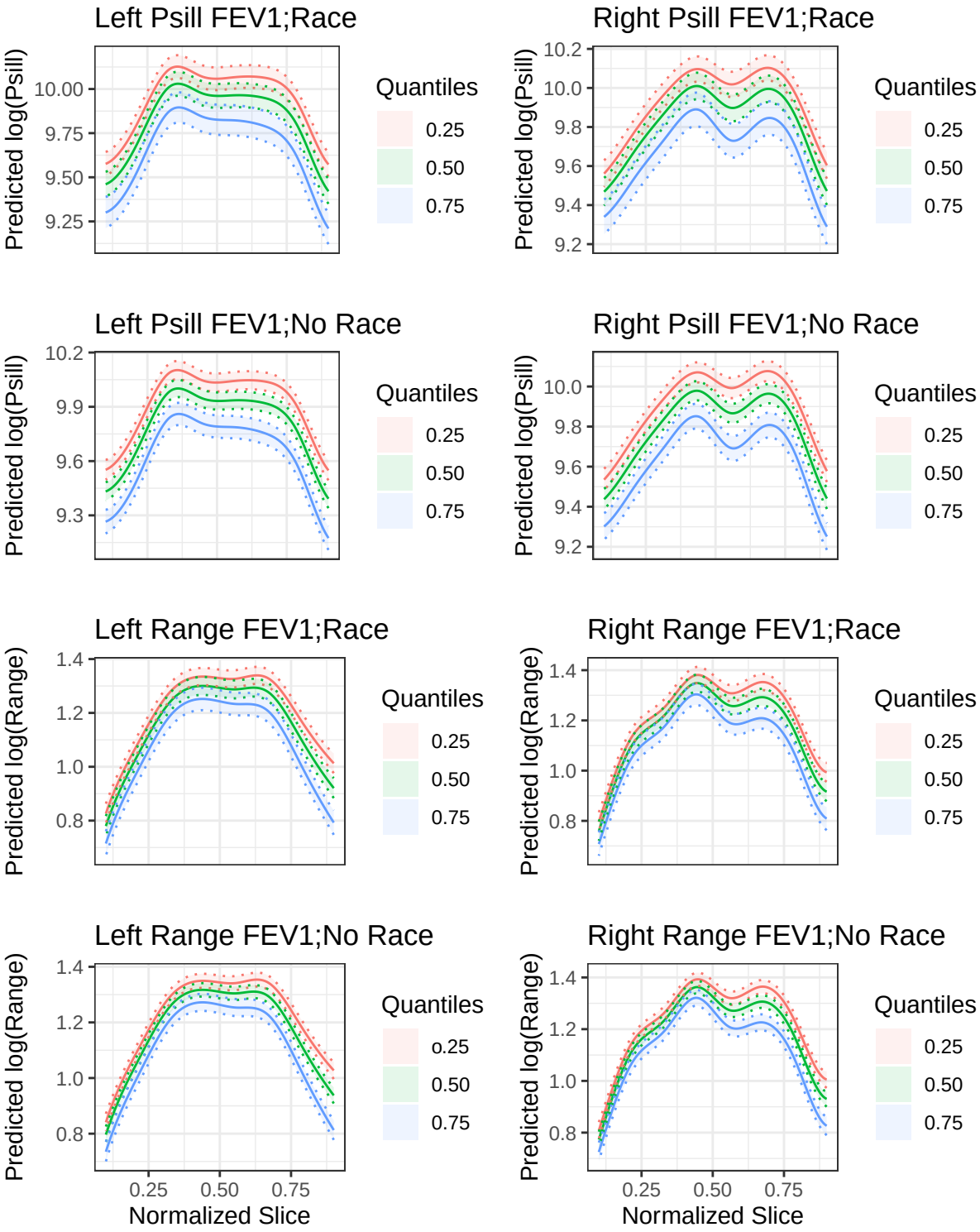
the plots excluding race. For psill the subjects with traction bronchiectasis have higher values in both lungs compared to subjects without traction bronchiectasis, and in both lungs subjects with the condition increase substantially more when moving up the lung. For range there is significant overlap in the lower portion of normalized slice for both lungs, and then a divergence, with subjects not having traction bronchiectasis showing lower range values. Similar to the psill measurements subjects with the condition have increasing range values further up the lung compared to subjects without the condition.

FEV1

Table 6: Summary of Models for FEV1

	Left Lung					Right Lung				
Term	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**
Outcome: Psill | With Race										
Intercept	10.0424	0.3699	<2e-16	NA	NA	9.8267	0.3701	<2e-16	NA	NA
BMI	0.0003	0.0023	0.903	NA	NA	0.0009	0.0023	0.713	NA	NA
Age	-0.0045	0.0017	0.008	NA	NA	-0.0027	0.0017	0.114	NA	NA
Male	0.1160	0.0419	0.006	NA	NA	0.1326	0.0419	0.002	NA	NA
Height	0.0079	0.0053	0.133	NA	NA	0.0089	0.0053	0.091	NA	NA
Other Races	-0.0271	0.0630	0.667	NA	NA	-0.0276	0.0630	0.661	NA	NA
Non-Hispanic White	-0.0435	0.0372	0.242	NA	NA	-0.0474	0.0372	0.203	NA	NA
s(Slice)	NA	NA	<2e-16	8.3005	8.7450	NA	NA	<2e-16	5.9098	6.6774
s(ID)	NA	NA	<2e-16	306.5939	312.0000	NA	NA	<2e-16	306.2032	312.0000
s(Slice):FEV1	NA	NA	<2e-16	5.6961	6.4936	NA	NA	<2e-16	8.9336	9.4575
Outcome: Psill | Without Race										
Intercept	9.9690	0.3640	<2e-16	NA	NA	9.7476	0.3644	<2e-16	NA	NA
BMI	0.0006	0.0023	0.801	NA	NA	0.0012	0.0023	0.608	NA	NA
Age	-0.0048	0.0017	0.004	NA	NA	-0.0030	0.0017	0.070	NA	NA
Male	0.1120	0.0416	0.007	NA	NA	0.1283	0.0416	0.002	NA	NA
Height	0.0090	0.0052	0.082	NA	NA	0.0101	0.0052	0.051	NA	NA
s(Slice)	NA	NA	<2e-16	8.3004	8.7449	NA	NA	<2e-16	5.9098	6.6775
s(ID)	NA	NA	<2e-16	308.5826	314.0000	NA	NA	<2e-16	308.1958	314.0000
s(Slice):FEV1	NA	NA	<2e-16	5.6966	6.4941	NA	NA	<2e-16	8.9337	9.4576
Outcome: Range | With Race										
Intercept	1.1858	0.1772	2e-11	NA	NA	1.2762	0.1774	7e-13	NA	NA
BMI	-0.0027	0.0011	0.01	NA	NA	-0.0032	0.0011	0.004	NA	NA
Age	-0.0019	0.0008	0.02	NA	NA	-0.0009	0.0008	0.242	NA	NA
Male	0.0865	0.0201	2e-05	NA	NA	0.0742	0.0201	2e-04	NA	NA
Height	0.0056	0.0025	0.03	NA	NA	0.0038	0.0025	0.129	NA	NA
Other Races	0.0401	0.0302	0.18	NA	NA	0.0131	0.0302	0.664	NA	NA
Non-Hispanic White	0.0242	0.0178	0.17	NA	NA	0.0215	0.0178	0.229	NA	NA
s(Slice)	NA	NA	<2e-16	7.4098	8.0619	NA	NA	<2e-16	8.3263	8.7445
s(ID)	NA	NA	<2e-16	293.5011	312.0000	NA	NA	<2e-16	293.8604	312.0000
s(Slice):FEV1	NA	NA	<2e-16	4.8409	5.5638	NA	NA	<2e-16	6.1030	6.9085
Outcome: Range | Without Race										
Intercept	1.2381	0.1747	1e-12	NA	NA	1.3123	0.1746	6e-14	NA	NA
BMI	-0.0029	0.0011	0.009	NA	NA	-0.0033	0.0011	0.003	NA	NA
Age	-0.0017	0.0008	0.032	NA	NA	-0.0008	0.0008	0.324	NA	NA
Male	0.0906	0.0199	6e-06	NA	NA	0.0762	0.0199	1e-04	NA	NA
Height	0.0049	0.0025	0.049	NA	NA	0.0033	0.0025	0.185	NA	NA
s(Slice)	NA	NA	<2e-16	7.4097	8.0618	NA	NA	<2e-16	8.3262	8.7444
s(ID)	NA	NA	<2e-16	295.5273	314.0000	NA	NA	<2e-16	295.8271	314.0000
s(Slice):FEV1	NA	NA	<2e-16	4.8412	5.5641	NA	NA	<2e-16	6.1034	6.9089

The table above illustrates the differences between models including and excluding race when FEV1 is the main clinical predictor. In all models the interaction between normalized slice and FEV1 is a significant predictor for range and psill regardless of including or excluding race. For both outcomes neither race category is a significant predictor.



The plots above show the trend of range and psill when moving up the lung (normalized slice) when FEV1 is the main clinical predictor. Models including race have slightly more overlap compared to models excluding race. Subsequent analysis will consider plots excluding race. For both range and psill higher FEV1 values

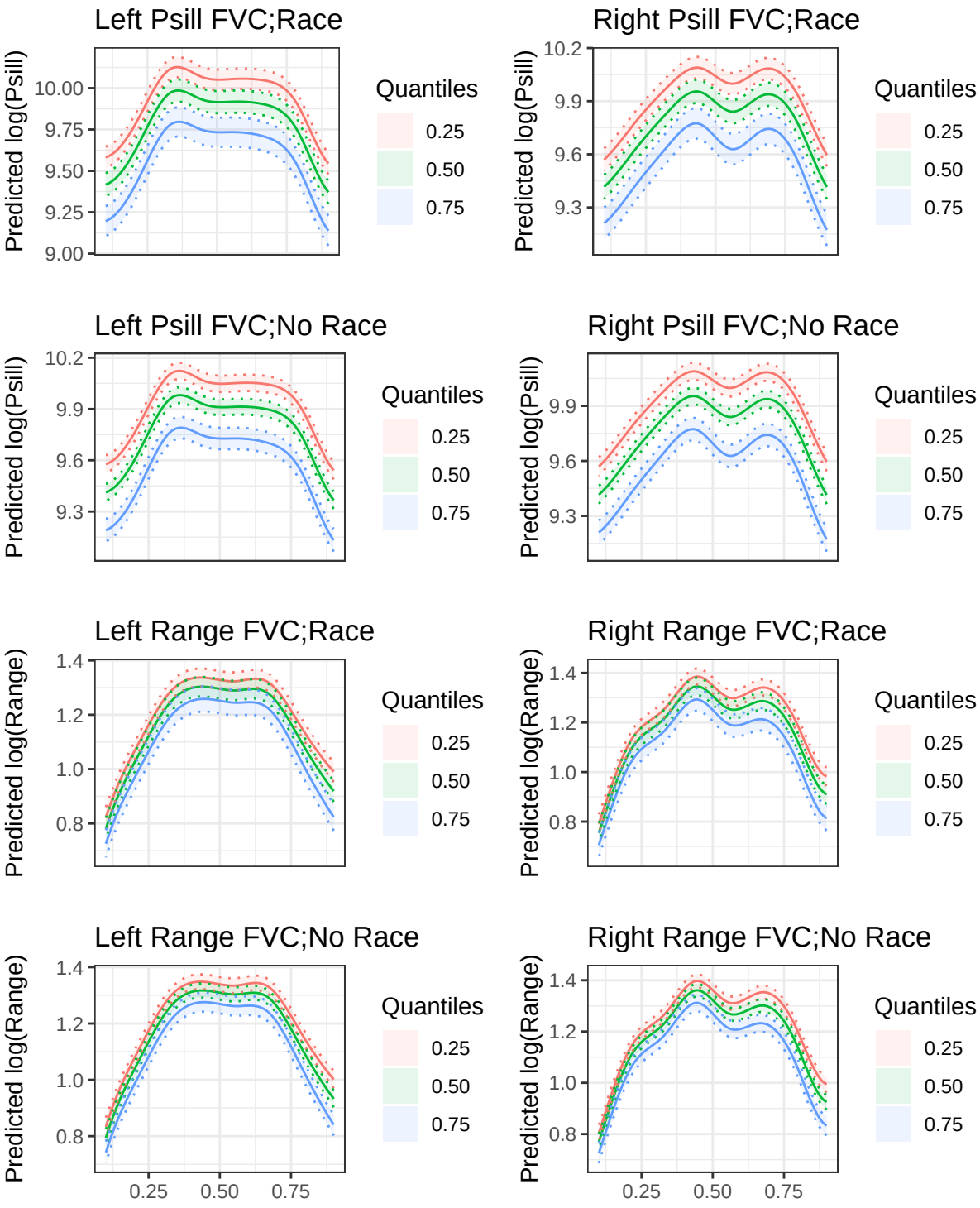
are associated with lower predicted outcomes. For left lung psill there is a peak in the first half of the lung, and a general decreasing trend when moving up the lung. For right lung psill there is a noticeable dip around the center of the lung. For range the left lung has a plateau in the middle of the plot before decreasing when moving up the lung, and the right lung has a dip near the center before rising slightly and then dropping off sharply when moving up the lung.

FVC

Table 7: Summary of Models for FVC

	Left Lung					Right Lung				
Term	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**
Outcome: Psill | With Race										
Intercept	9.6609	0.3573	<2e-16	NA	NA	9.4240	0.3495	<2e-16	NA	NA
BMI	-0.0011	0.0022	0.607	NA	NA	-0.0007	0.0022	0.76	NA	NA
Age	-0.0049	0.0016	0.002	NA	NA	-0.0035	0.0016	0.02	NA	NA
Male	0.1695	0.0412	4e-05	NA	NA	0.1951	0.0403	1e-06	NA	NA
Height	0.0175	0.0053	9e-04	NA	NA	0.0197	0.0052	1e-04	NA	NA
Other Races	0.0259	0.0608	0.671	NA	NA	0.0310	0.0595	0.60	NA	NA
Non-Hispanic White	-0.0087	0.0362	0.809	NA	NA	-0.0037	0.0354	0.92	NA	NA
s(Slice)	NA	NA	<2e-16	8.4089	8.8247	NA	NA	<2e-16	5.8687	6.6351
s(ID)	NA	NA	<2e-16	306.0638	312.0000	NA	NA	<2e-16	305.3588	312.0000
s(Slice):FVC	NA	NA	<2e-16	4.7560	5.4832	NA	NA	<2e-16	8.6809	9.2596
Outcome: Psill | Without Race										
Intercept	9.6583	0.3490	<2e-16	NA	NA	9.4323	0.3413	<2e-16	NA	NA
BMI	-0.0011	0.0022	0.631	NA	NA	-0.0006	0.0022	0.78	NA	NA
Age	-0.0050	0.0016	0.002	NA	NA	-0.0036	0.0015	0.02	NA	NA
Male	0.1712	0.0410	3e-05	NA	NA	0.1971	0.0401	9e-07	NA	NA
Height	0.0176	0.0051	6e-04	NA	NA	0.0197	0.0050	9e-05	NA	NA
s(Slice)	NA	NA	<2e-16	8.3199	8.7388	NA	NA	<2e-16	5.8687	6.6351
s(ID)	NA	NA	<2e-16	308.0350	314.0000	NA	NA	<2e-16	307.3250	314.0000
s(Slice):FVC	NA	NA	<2e-16	5.1125	5.9817	NA	NA	<2e-16	8.6810	9.2596
Outcome: Range | With Race										
Intercept	1.1058	0.1867	3e-09	NA	NA	1.1742	0.1829	1e-10	NA	NA
BMI	-0.0030	0.0012	0.009	NA	NA	-0.0035	0.0011	0.002	NA	NA
Age	-0.0010	0.0008	0.212	NA	NA	-0.0004	0.0008	0.583	NA	NA
Male	0.0860	0.0215	6e-05	NA	NA	0.0808	0.0211	1e-04	NA	NA
Height	0.0063	0.0028	0.023	NA	NA	0.0055	0.0027	0.043	NA	NA
Other Races	0.0459	0.0318	0.149	NA	NA	0.0238	0.0311	0.445	NA	NA
Non-Hispanic White	0.0182	0.0189	0.336	NA	NA	0.0219	0.0185	0.237	NA	NA
s(Slice)	NA	NA	<2e-16	7.4768	8.1224	NA	NA	<2e-16	8.5150	8.8885
s(ID)	NA	NA	<2e-16	294.8459	312.0000	NA	NA	<2e-16	294.5047	312.0000
s(Slice):FVC	NA	NA	<2e-16	4.9969	5.7311	NA	NA	<2e-16	3.9926	4.6227
Outcome: Range | Without Race										
Intercept	1.1578	0.1829	2e-10	NA	NA	1.2221	0.1789	9e-12	NA	NA
BMI	-0.0031	0.0012	0.008	NA	NA	-0.0036	0.0011	0.001	NA	NA
Age	-0.0009	0.0008	0.255	NA	NA	-0.0003	0.0008	0.714	NA	NA
Male	0.0890	0.0215	3e-05	NA	NA	0.0823	0.0210	9e-05	NA	NA
Height	0.0054	0.0027	0.043	NA	NA	0.0047	0.0026	0.075	NA	NA
s(Slice)	NA	NA	<2e-16	7.4767	8.1223	NA	NA	<2e-16	8.5150	8.8884
s(ID)	NA	NA	<2e-16	296.8539	314.0000	NA	NA	<2e-16	296.4726	314.0000
s(Slice):FVC	NA	NA	<2e-16	4.9973	5.7315	NA	NA	<2e-16	3.9929	4.6231

The table above illustrates the differences between models including and excluding race when FVC is the main clinical predictor. In all models the interaction between normalized slice and FVC is a significant predictor for range and psill regardless of whether race is included or excluded. Race is an insignificant predictor in both lungs for both outcomes.



The plots above show the trend of range and psill when moving up the lung (normalized slice) when FVC is the main clinical predictor. The models excluding race have narrower 95% confidence intervals compared to models including race. Subsequent analysis will only consider models excluding race. For both psill and

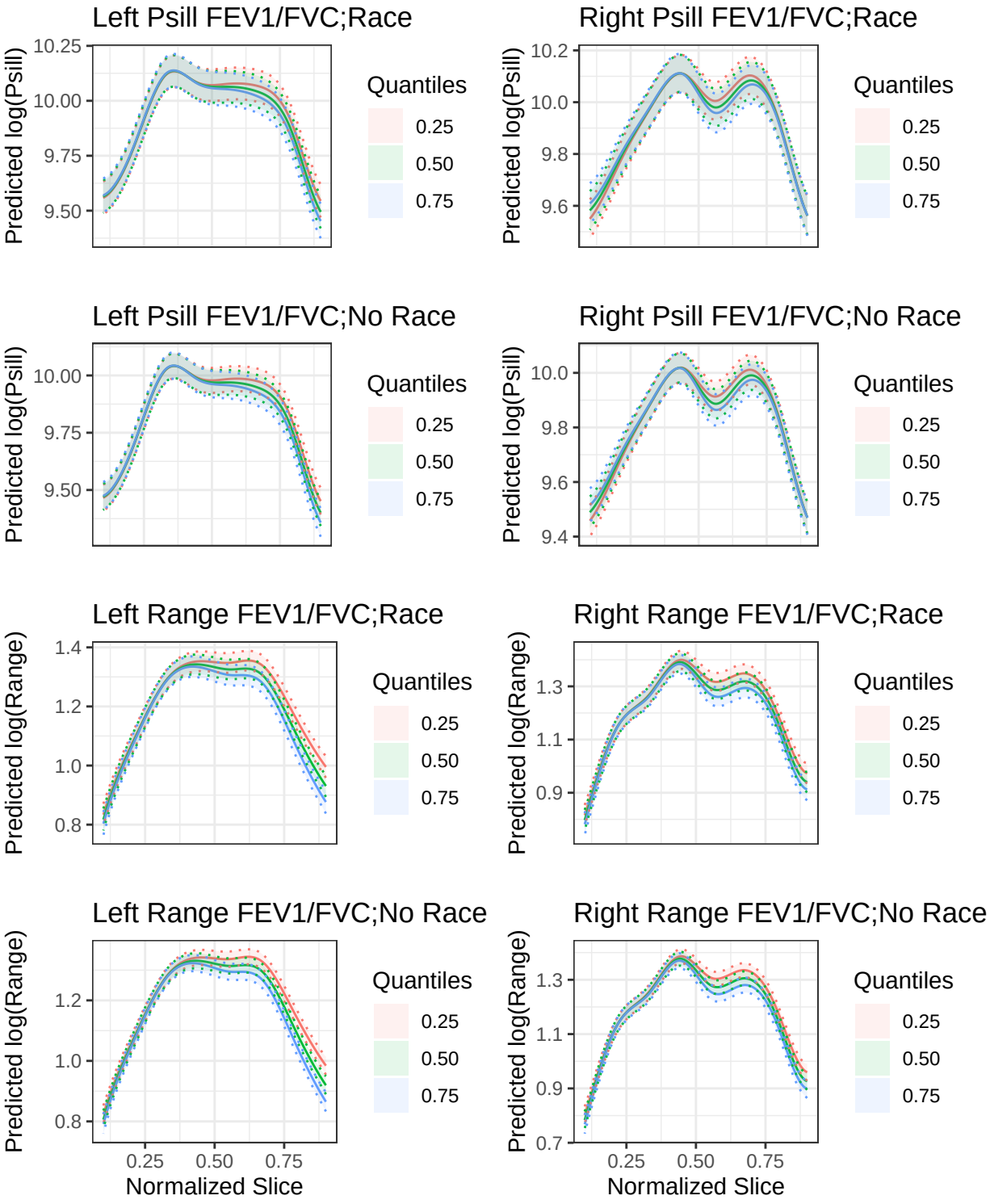
range higher FVC values are associated with lower predicted outcomes. For left lung psill there is a peak before the middle of the lung and then a decrease when moving up the lung. For the right lung there is a noticeable decrease in the middle of the lung before increasing and then subsequently decreasing sharply. For range the left lung has a plateau in the middle, while the right lung has a slight increase after the middle portion of the lung, and then a sharp decrease.

FEV1/FVC

Table 8: Summary of Models for FEV1/FVC

	Left Lung					Right Lung				
Term	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**	**Estimate**	**SE**	**P-Value**	**EDF**	**Ref. DF**
Outcome: Psill | With Race										
Intercept	10.4616	0.4370	<2e-16	NA	NA	10.0622	0.4333	<2e-16	NA	NA
BMI	0.0013	0.0026	0.6	NA	NA	0.0016	0.0026	0.52	NA	NA
Age	0.0013	0.0017	0.5	NA	NA	0.0031	0.0017	0.07	NA	NA
Male	0.0085	0.0449	0.8	NA	NA	0.0324	0.0445	0.47	NA	NA
Height	-0.0070	0.0056	0.2	NA	NA	-0.0047	0.0056	0.40	NA	NA
Other Races	-0.0931	0.0699	0.2	NA	NA	-0.0875	0.0693	0.21	NA	NA
Non-Hispanic White	-0.1496	0.0394	1e-04	NA	NA	-0.1482	0.0390	1e-04	NA	NA
s(Slice)	NA	NA	<2e-16	8.2830	8.7110	NA	NA	<2e-16	6.0184	6.7044
s(ID)	NA	NA	<2e-16	307.7012	312.0000	NA	NA	<2e-16	307.2993	312.0000
s(Slice):FEV1/FVC	NA	NA	<2e-16	4.4894	5.1625	NA	NA	<2e-16	8.4197	8.9320
Outcome: Psill | Without Race										
Intercept	10.2322	0.4406	<2e-16	NA	NA	9.8375	0.4369	<2e-16	NA	NA
BMI	0.0027	0.0026	0.3	NA	NA	0.0029	0.0026	0.3	NA	NA
Age	0.0011	0.0018	0.6	NA	NA	0.0028	0.0018	0.1	NA	NA
Male	-0.0243	0.0447	0.6	NA	NA	0.0002	0.0444	1.0	NA	NA
Height	-0.0052	0.0057	0.4	NA	NA	-0.0029	0.0056	0.6	NA	NA
s(Slice)	NA	NA	<2e-16	8.2821	8.7101	NA	NA	<2e-16	6.0188	6.7047
s(ID)	NA	NA	<2e-16	309.8619	314.0000	NA	NA	<2e-16	309.4756	314.0000
s(Slice):FEV1/FVC	NA	NA	<2e-16	4.4937	5.1673	NA	NA	<2e-16	8.4200	8.9322
Outcome: Range | With Race										
Intercept	1.6507	0.1962	<2e-16	NA	NA	1.6505	0.1992	<2e-16	NA	NA
BMI	-0.0019	0.0012	0.10	NA	NA	-0.0024	0.0012	0.04	NA	NA
Age	0.0001	0.0008	0.91	NA	NA	0.0011	0.0008	0.18	NA	NA
Male	0.0386	0.0202	0.06	NA	NA	0.0286	0.0205	0.16	NA	NA
Height	-0.0016	0.0025	0.54	NA	NA	-0.0029	0.0026	0.25	NA	NA
Other Races	0.0078	0.0314	0.80	NA	NA	-0.0170	0.0319	0.59	NA	NA
Non-Hispanic White	-0.0194	0.0177	0.27	NA	NA	-0.0211	0.0179	0.24	NA	NA
s(Slice)	NA	NA	<2e-16	3.8070	4.4648	NA	NA	<2e-16	4.8910	5.6389
s(ID)	NA	NA	<2e-16	295.3200	312.0000	NA	NA	<2e-16	295.8590	312.0000
s(Slice):FEV1/FVC	NA	NA	<2e-16	7.8989	8.5510	NA	NA	<2e-16	9.1032	9.5000
Outcome: Range | Without Race										
Intercept	1.6321	0.1941	<2e-16	NA	NA	1.6159	0.1968	2e-16	NA	NA
BMI	-0.0017	0.0012	0.14	NA	NA	-0.0023	0.0012	0.05	NA	NA
Age	0.0000	0.0008	0.97	NA	NA	0.0010	0.0008	0.19	NA	NA
Male	0.0357	0.0197	0.07	NA	NA	0.0237	0.0200	0.24	NA	NA
Height	-0.0014	0.0025	0.57	NA	NA	-0.0026	0.0025	0.30	NA	NA
s(Slice)	NA	NA	<2e-16	3.8070	4.4648	NA	NA	<2e-16	4.8911	5.6389
s(ID)	NA	NA	<2e-16	297.3092	314.0000	NA	NA	<2e-16	297.8259	314.0000
s(Slice):FEV1/FVC	NA	NA	<2e-16	7.8990	8.5511	NA	NA	<2e-16	9.1032	9.4999

The table above illustrates the differences between models including and excluding race when FEV1/FVC is the main clinical predictor. In all models the interaction between normalized slice and FEV1/FVC are significant regardless of whether race was included or excluded. For psill non-Hispanic white is significant for both lungs, while the other races category is not significant. For range neither race category is a significant predictor in either lung.



The plots above show the trend of range and psill when moving up the lung (normalized slice) when FEV1/FVC is the main clinical predictor. Models including race have slightly wider confidence intervals, but trends in both sets of plots including and excluding race are similar. Subsequent analysis will only

consider models excluding race. For both psill and range there is significant overlap before the middle portion of the lung. When moving towards the middle portion of the lung there is a slight spread of lines, with higher FEV1/FVC values corresponding to lower psill and range values. As the scans move further up the lung the expected range and psill move closer together.

Discussion

Race as a Predictor

Direct model comparison using AIC suggested that there were minimal differences between models including race and models excluding race. For example, the left lung psill model including race had an AIC of -470.2398, while the model excluding race had a psill of -470.0661, meaning that the model including race had an AIC only 0.1737 units lower than the model excluding race. Between all models only the psill models where fibrosis, mediastinal lymphadenopathy, traction bronchiectasis, and FEV1/FVC were the clinical predictor had AIC values which were smaller in the models including race. Psill models where FEV1 and FVC were the clinical predictors and all models for range had lower AIC values when excluding race. Again, the differences were all below 1. In deciding on whether to include race or not for based on AIC values models that excluded race had a higher proportion of lower AIC values compared to models which included race.

Based on model outputs from the tables provided in the results neither race category was a significant predictor for the range outcome. For psill models the non-Hispanic white category was generally significant while the other races category was never a significant predictor. Additionally, the interaction between the clinical measurement and normalized slice was significant in all models regardless of whether or not race was included as a predictor. From the plots the relationship between models including and excluding race had the same general trend, with the main difference being that models without race tended to have slightly narrower confidence intervals. These narrower confidence intervals gave better discrepancy in predicted outcome between the binary VAS conditions and the three levels of PFT measurements (25th percentile, 50th percentile, and 75 percentile).

Beyond data analyzed in this study, past studies have noted difficulty in assessing the prevalence of sarcoidosis among different racial groups. For example, a study in the American Journal of Epidemiology noted that while African-Americans were found to have threefold higher risk of sarcoidosis compared to Caucasians, it is difficult to determine the degree that environmental exposure versus genetic susceptibility accounts for this difference[1]. Additionally, socioeconomic status (SES) is a known predictor of health outcomes, and patients of low SES tend to have worse pulmonary function, lower baseline physical health, and higher likelihood

of multi-organ disease after adjusting for comorbidities[2]. In the context of this study, simply adjusting for race will likely overlook more important predictors of sarcoidosis, potentially impeding care for those who are most at risk. In another study an age-adjusted mortality rate increase of 50% between 1988 and 2007 in the United States was observed for Caucasians and not African Americans, with the authors suggesting that access to more advanced medical technology by Caucasian patients could explain higher rates of detection and bias in diagnoses[3]. As more advanced imaging options become available past assumptions regarding sarcoidosis susceptibility need careful scrutiny to avoid perpetuating bias in treatment and diagnosis. After running comparisons between models including race and models excluding race there is no tangible benefit for including race as a predictor. Additional covariates which may influence sarcoidosis should be investigated for better modeling.

Relationship Between Clinical Predictors and Outcomes

Differences within clinical measurements were most obvious in plots showing the relationship between the log of psill/range and normalized slice. Analysis for these plots focused on the models which did not control for race. For fibrosis, subjects with no disease had lower measurements in both psill and range when moving up the lung. The lower in the lung measurements were taken the closer the two groups were.

For models where mediastinal lymphadenopathy was the primary clinical measurement psill was lower among subjects who did not have the condition compared to subjects who had the condition. In the left lung the largest difference was in the middle portion of the lung, with more similarity moving when moving towards the top and bottom most portion of the lungs. The right lung showed more consistant difference between subjects with and without mediastinal lymphadenopathy throughout the lung. For models of range both lungs had small differences between those with and without mediastinal lymphadenopathy, with subjects who did not have the condition generally having slightly lower values.

For models where traction bronchiectasis was the primary clinical measurement psill was lower among subjects who did not have the condition compared to subjects who had the condition. Both the left and right lung showed similar trends, with a steep dip in the middle of the right lung. For models of range the lower portion of the lung showed similar measurements between subjects who did and did not have traction bronchiectasis. As scans moved up both lungs there were larger differences in measurements.

For models where FEV1 was the primary clinical variable psill and range was lower in both lungs for subjects who had higher FEV1 measurements. Both outcomes followed similar trends between the left and right lung, with higher psill and range measurements in the middle portion of the lung.

For models where FVC was the primary variable psill and range was lower in both lungs for subjects who had higher FVC measurements. Differences were smaller for the range outcome compared to psill, but the same general trend was present in both outcomes where the upper and lower portions of the lung had lower measurements.

For models where FEV1/FVC was the primary variable psill and range were in general similar regardless of FEV1/FVC measurements. The largest differences were in the middle portion of the lung where higher FEV1/FVC values corresponded to lower psill and range measurements.

In summary subjects with better PFT lung function, no fibrosis, no mediastinal lymphadenopathy, and no traction bronchiectasis had smaller psill and range measurements. Differences tended to be largest in the middle portion of the lung, suggesting that scans focused on this area show the largest discrepancy among subjects with worse health.

Limitations of this analysis include lack of diversity among subjects. The majority of subjects were non-Hispanic whites, followed by African Americans/Black. The other race category incorporated multiple different races, which strongly limits the inference beyond the two primary race categories mentioned previously. Additionally, the diagnosis of sarcoidosis was not explicitly incorporated into this analysis, but rather potential indicators of disease represented by the clinical measurements. Additional study will need to be performed to better understand the relationship between variogram scans and sarcoidosis. Lastly, a complete case analysis was performed for better comparison of models, meaning that the sample size was smaller than originally expected.

References

1. Benjamin A. Rybicki, Marcie Major, John Popovich Jr., Mary J. Maliarik, Lannuzzi MC. Racial differences in sarcoidosis incidence: A 5-year study in a health maintenance organization. American Journal of Epidemiology 1997;145:234–41.
2. Obi ON. Diversity and disparity in the treatment and care of sarcoidosis. Current Opinion in Pulmonary Medicine Sep. 2023;29:512–24.
3. Mehdi Mirsaeidi, Roberto F. Machado, Dean Schraufnagel, Nadera J. Sweiss, Baughman RP. Racial difference in sarcoidosis mortality in the united states. Original Research Diffuse Lung Disease May 2014;147:438–49.