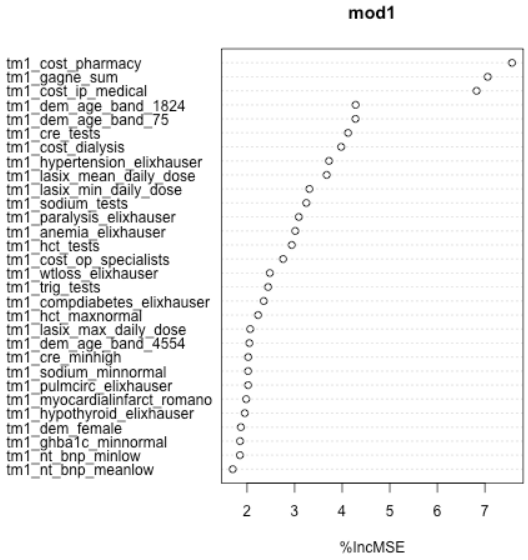
Lab 8 Report

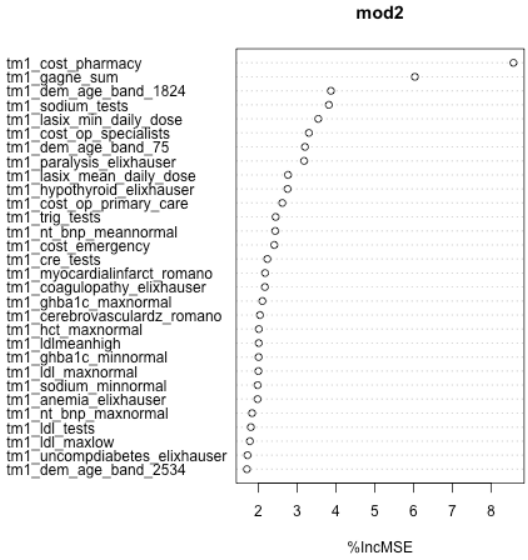
Evangelos Kassos

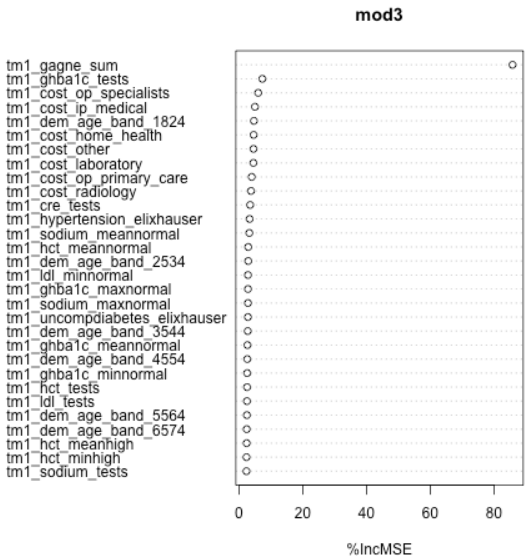
4/14/2022

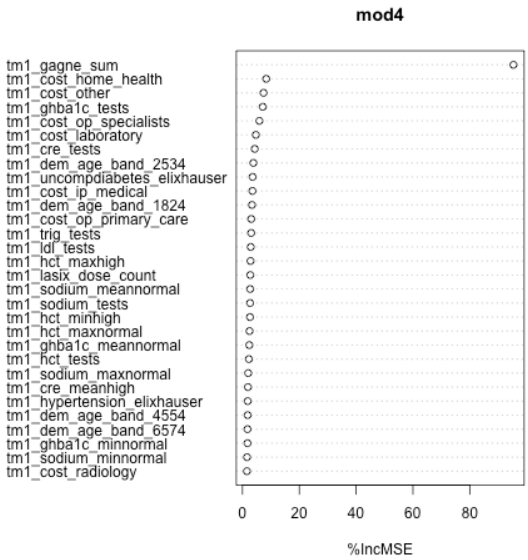
Question 1. See code.

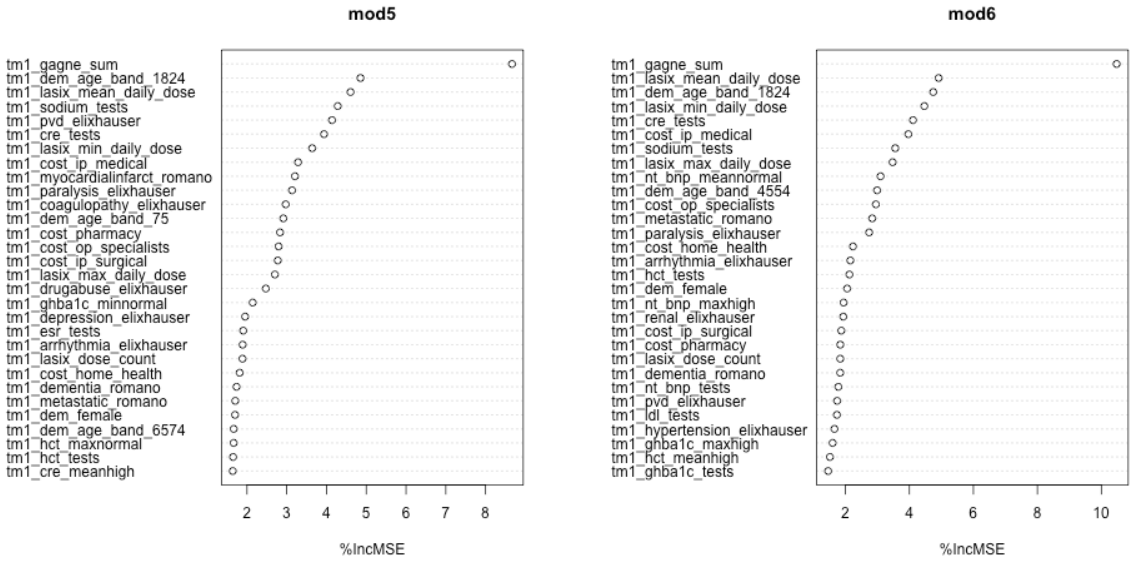
Question 2. See below for statistical model results.











Question 3. Based on the results presented below for the training sample, the root mean squared prediction error for the models that include patient race is higher compared to the respective models that do not include patient race for models concerning costs and health. The root mean squared prediction error for the model that includes patient race is lower compared to the respective model that does not include patient race for the models concerning avoidable costs.

Models	RMSPE of model excluding race	RMSPE of model including race
1 & 2 - Costs	7718.129	8038.495
3 & 4 - Health	0.461	0.465
5 & 6 - Avoidable costs	6234.884	6069.311

Question 4. Based on the results presented below for the test sample, the root mean squared prediction error for the models that include patient race is lower compared to the respective models that do not include patient race for models concerning avoidable costs and health. The root mean squared prediction error for the model that includes patient race is higher compared to the respective model that does not include patient race for the models concerning costs.

Models	RMSPE of model excluding race	RMSPE of model including race
1 & 2 - Costs	16 275.504	16 297.990
3 & 4 - Health	1.061	1.059
5 & 6 - Avoidable costs	12 136.502	12 010.001

Question 5. See code.

Question 6. See code.

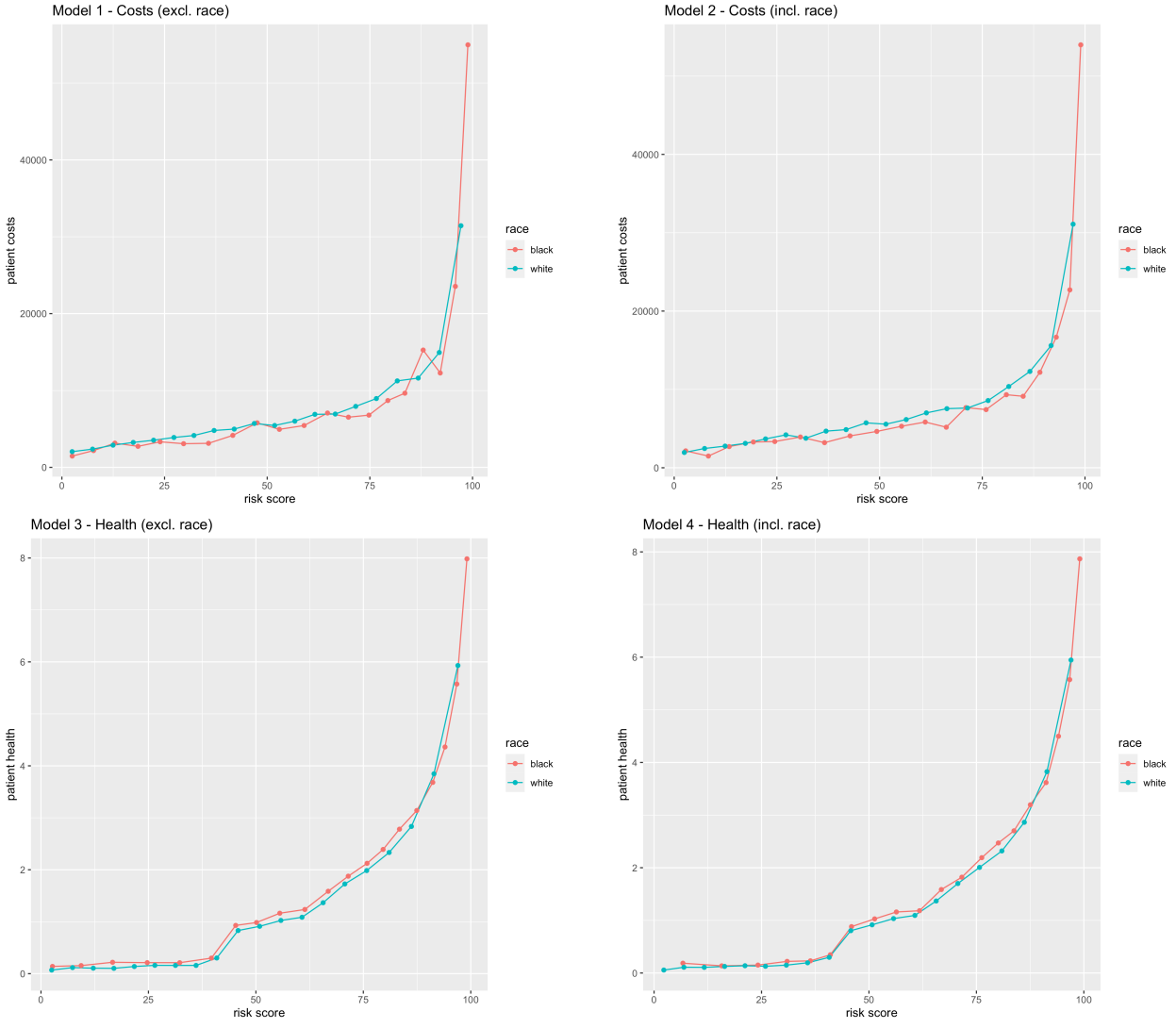
Question 7a.

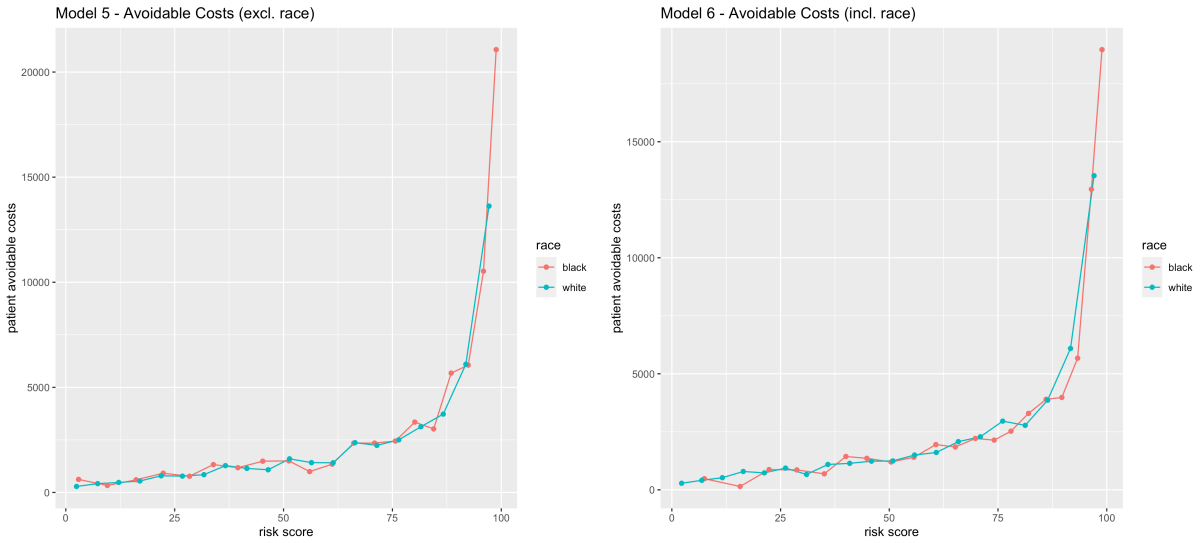
Model	Fraction
Model 1 - Costs (excl. race)	0.5064806
Model 2 - Costs (incl. race)	0.5264207
Model 3 - Health (excl. race)	0.5804586
Model 4 - Health (incl. race)	0.5884347
Model 5 - Avoidable costs (excl. race)	0.5339980
Model 6 - Avoidable costs (incl. race)	0.5804586

Question 7b.

Model	Fraction
Model 1 - Costs (excl. race)	0.1282958
Model 2 - Costs (incl. race)	0.1333468
Model 3 - Health (excl. race)	0.1470425
Model 4 - Health (incl. race)	0.1490630
Model 5 - Avoidable costs (excl. race)	0.1352662
Model 6 - Avoidable costs (incl. race)	0.1470351

Question 8.





Question 9. Often times, algorithms are aimed at the wrong target to begin with. The result is a "label choice bias," arising from a mismatch between the ideal target the algorithm should be predicting, and a biased proxy variable the algorithm is actually predicting. Here we see some evidence of this bias: compare models 3 & 4 with models 1 & 2/5 & 6. We can see that the shape of the curves in risk score intervals (40, 75) is different, meaning that if we chose models 1 & 2/5 & 6 over models 3 & 4 we would get completely different results, categorizing a significantly different fraction of patients as with increased risk. Comparing the duos in each model category, along with our tables in question 7, we can also see that this is likely not due to the inclusion or exclusion of the race variable.