COVID-19 Baseline Risk Score Analysis Report

MockENSEMBLE Study

USG COVID-19 Response Biostatistics Team

August 20, 2021

Contents

List of Tables

List of Figures

Chapter 1

Baseline Risk Score (Proxy for SARS-CoV-2 Exposure)

Table 1.1: Variables considered for risk score analysis.

Variable.Name	Definition	Total.missing.values	Comments
EthnicityHispanic	Indicator ethnicity = Hispanic (1 = Hispanic, 0 = complement)	0/19462 (0.0%)	NA
EthnicityNotreported	Indicator ethnicity = Not reported (1 = Not reported, 0 = complement)	0/19462 (0.0%)	NA
EthnicityUnknown	Indicator ethnicity = Unknown (1 = Unknown, 0 = complement)	0/19462 (0.0%)	NA
Black	Indicator race = Black (1=Black, 0=complement)	0/19462 (0.0%)	NA
Asian	Indicator race = Asian (1=Asian, 0=complement)	0/19462 (0.0%)	NA
NatAmer	Indicator race = American Indian or Alaska Native (1=NatAmer, 0=complement)	0/19462 (0.0%)	NA
Multiracial	Indicator race = Multiracial (1=Multiracial, 0=complement)	0/19462 (0.0%)	NA
Notreported	Indicator race = Not reported (1=Notreported, 0=complement)	0/19462 (0.0%)	NA
Unknown	Indicator race = unknown (1=Unknown, 0=complement)	0/19462 (0.0%)	NA
URMforsubcohortsampling	Indicator of under-represented minority (1=Yes, 0=No)	0/19462 (0.0%)	NA
HighRiskInd	Baseline covariate indicating >= 1 Co-existing conditions (1=yes, 0=no, NA=missing)	0/19462 (0.0%)	NA
HIVinfection	Indicator HIV infected at enrollment (1=infected, 0=not infected)	0/19462 (0.0%)	NA
Sex	Sex assigned at birth (1=female, 0=male/undifferentiated/unknown	0/19462 (0.0%)	NA
Country.X1	Dummy indicator country = Argentina (1 = Argentina, 0 = complement)	0/19462 (0.0%)	NA

Table 1.1: Variables considered for risk score analysis. *(continued)*

Variable.Name	Definition	Total.missing.values
Country.X2	Dummy indicator country = Brazil (1 = Brazil, 0 = complement)	0/19462 (0.0%)
Country.X3	Dummy indicator country = Chile (1 = Chile, 0 = complement)	0/19462 (0.0%)
Country.X4	Dummy indicator country = Columbia (1 = Columbia, 0 = complement)	0/19462 (0.0%)
Country.X5	Dummy indicator country = Mexico (1 = Mexico, 0 = complement)	0/19462 (0.0%)
Country.X6	Dummy indicator country = Peru (1 = Peru, 0 = complement)	0/19462 (0.0%)
Country.X7	Dummy indicator country = South Africa (1 = South Africa, 0 = complement)	0/19462 (0.0%)
Region.X1	Dummy indicator region = Latin America (1 = Latin America, 0 = complement)	0/19462 (0.0%)
Region.X2	Dummy indicator country = Southern Africa (1 = Southern Africa, 0 = complement)	0/19462 (0.0%)
CalDtEnrollIND.X1	Dummy indicator variable representing enrollment occurring between 4-8 weeks periods of first subject enrolled (1 = Enrollment between 4-8 weeks, 0 = complement).	0/19462 (0.0%)
Age	Age at enrollment in years (integer >= 18, NA=missing). Note that the randomization strata included Age 18-59 vs. Age >= 60.	0/19462 (0.0%)
BMI	BMI at enrollment (Ordered categorical 1,2, 3, 4, NA=missing); 1 = Underweight BMI < 18.5; 2 = Normal BMI 18.5 to < 25; 3 = Overweight BMI 25 to < 30; 4 = Obese BMI >= 30	0/19462 (0.0%)

Note:
Variables with more than 5% missing values were dropped from analysis; missing values for other variables w
Indicator variables not meeting the threshold, such that under the null of not a risk factor there were less th
subgroup with value 1 (or 0), were dropped from analysis.

Table 1.2: All learner-screen combinations (14 in total) used as input to the Superlearner.

Learner	**Screen***
SL.mean	all
SL.glm	all glmnet univar_logistic_pval highcor_random

Note:
*Screen details:
all: includes all variables
glmnet: includes variables with non-zero coefficients in the standard implementation of SL.glmnet that optimizes the lasso tuning parameter via cross-validation
univar_logistic_pval: Wald test 2-sided p-value in a logistic regression model < 0.10
highcor_random: if pairs of quantitative variables with Spearman rank correlation > 0.90, select one of the variables at random

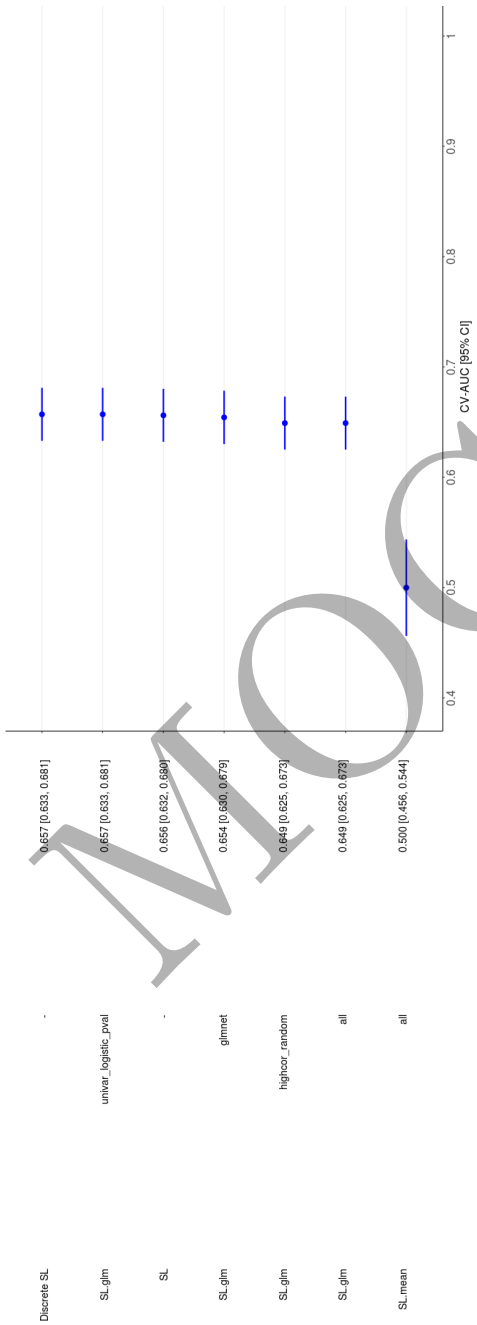

Figure 1.1: Cross-validated AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 29.

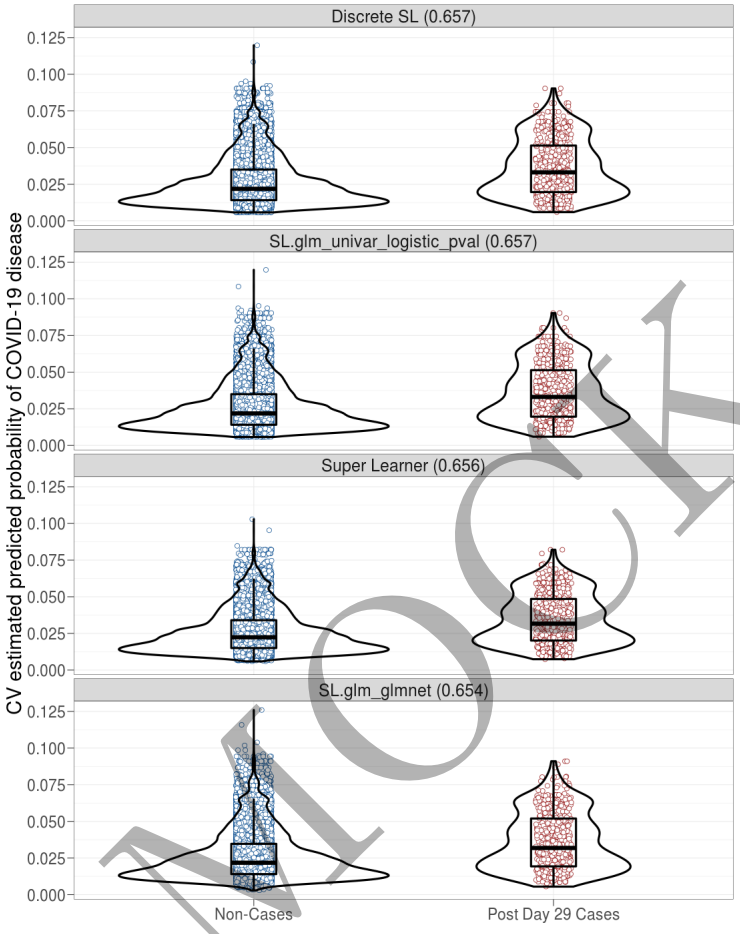

Figure 1.2: CV-estimated predicted probabilities of COVID-19 disease 7 days after Day 29 by case/control status for top 2 learners, SuperLearner and Discrete SL.

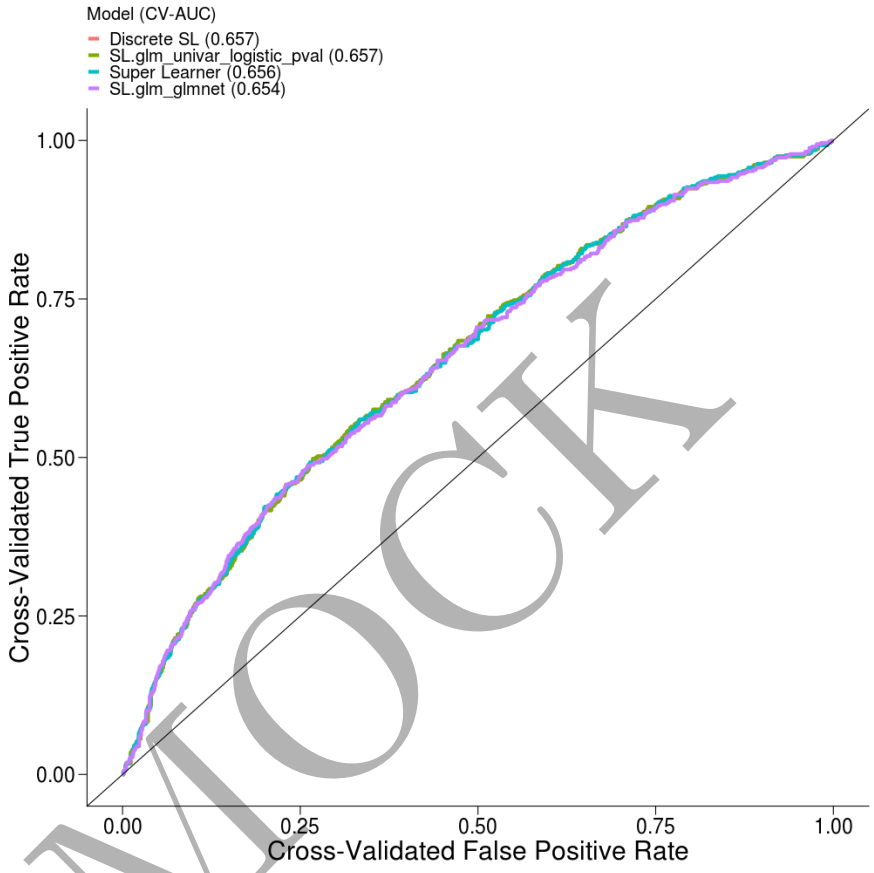

Figure 1.3: ROC curves based off CV-estimated predicted probabilities for the top 2 learners, Superlearner and Discrete SL.

Table 1.3: Weights assigned by Superlearner.

Learner	**Screen**	**Weight**
SL.glm	screen_univariate_logistic_pval	0.942
SL.mean	screen_all	0.058
SL.glm	screen_all	0.000
SL.glm	screen_glmnet	0.000
SL.glm	screen_highcor_random	0.000

Table 1.4: Predictors in learners assigned weight > 0.0 by Superlearner.

Learner	Screen	Weight	Predictors	Coefficient	Odds.Ratio
SL.glm	screen_univariate_logistic_pval	0.942	(Intercept)	-3.760	0.023
SL.glm	screen_univariate_logistic_pval	0.942	HighRiskInd	0.476	1.609
SL.glm	screen_univariate_logistic_pval	0.942	Country.X2	0.097	1.101
SL.glm	screen_univariate_logistic_pval	0.942	Age	0.343	1.409

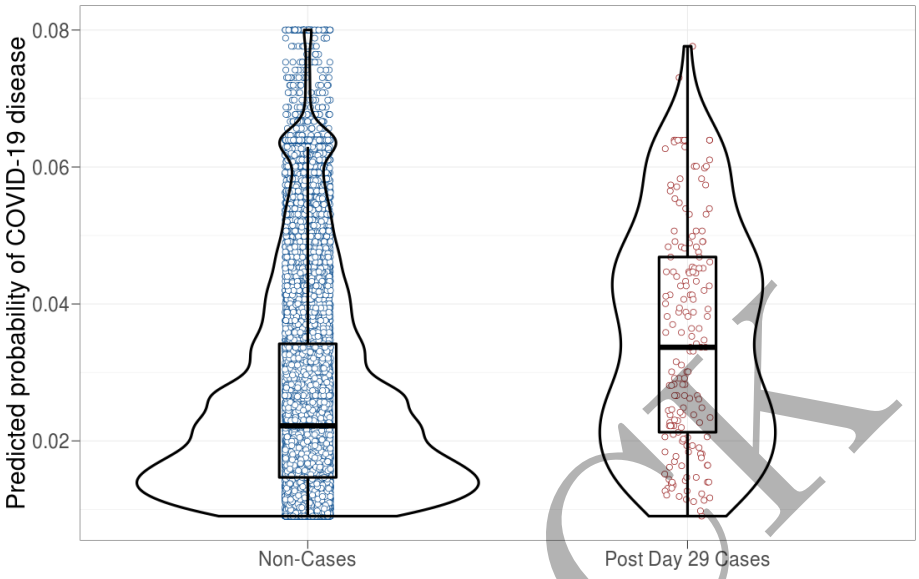

Figure 1.4: Superlearner predicted probabilities of COVID-19 disease in vaccinees 7 days after Day 29 by case/control status.

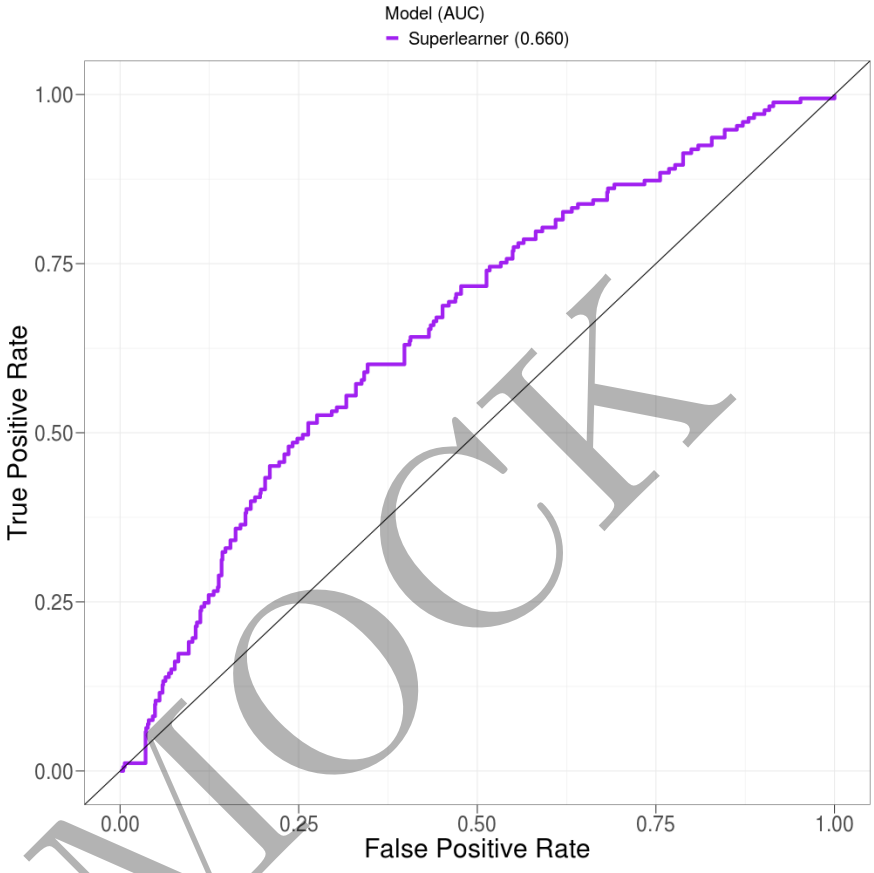

Figure 1.5: ROC curve based off Superlearner predicted probabilities in vaccinees.

Table 1.5: Cases per treatment arm prior to risk score analysis.

Study-Arm	Non-Cases	Post-Day 29-Cases
Placebo	18743	516
Vaccine	19183	173

Table 1.6: Cases per treatment arm post risk score analysis.

Study-Arm	Non-Cases	Post-Day 29-Cases
Placebo	18743	516
Vaccine	19183	173

Chapter 2

Appendix

- This report was built from the `CoVPN/correlates_reporting` repository with commit hash 323d462b7cac1904495f387773d15cbb8fe8b4af. A diff of the changes introduced by that commit may be viewed at https://github.com/CoVPN/correlates_reporting/commit/323d462b7cac1904495f387773d15cbb8fe8b4af
- The sha256 hash sum of the raw input file, "COVID_ENSEMBLE_practicedata.csv": 847161e464e2488f2d36717254de9e0d885d56cbe7205a3d174d747b6cb828d8
- The sha256 hash sum of the processed file, "janssen_pooled_mock_data_processed.csv": c05d0a8b66052e68358fc5fb76888a80b0437b151ae016a1b42f0072ac94d1eb