COVID-19 Baseline Risk Score Analysis Report

MockCOVE Study

USG COVID-19 Response Biostatistics Team

September 20, 2021

Contents

List of Tables

List of Figures

Chapter 1

Disclaimers

- The data presented in the analysis originated from the Moderna Sponsored mRNA-1273-P301 clinical study and are provided to NIAID in accordance with Clinical Trial Agreement between the parties. The study was funded in part by BARDA under Government Contract No. 75A50120C00034

- The preliminary immunogenicity data presented here do not reflect the Sponsors statistical analysis plan and therefore should not be interpreted as a protocol defined read-out of the clinical study.

- These data are not to be disclosed without written permission of Moderna.

Chapter 2

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

Table 2.1: Variables considered for risk score analysis.

Variable.Name	Definition	Total.missing.values	Comments
MinorityInd	Baseline covariate underrepresented minority status (1=minority, 0=non-minority)	0/13294 (0.0%)	NA
EthnicityHispanic	Indicator ethnicity = Hispanic (0 = Non-Hispanic)	0/13294 (0.0%)	NA
EthnicityNotreported	Indicator ethnicity = Not reported (0 = Non-Hispanic)	0/13294 (0.0%)	NA
EthnicityUnknown	Indicator ethnicity = Unknown (0 = Non-Hispanic)	0/13294 (0.0%)	NA
Black	Indicator race = Black (0 = White)	0/13294 (0.0%)	NA
Asian	Indicator race = Asian (0 = White)	0/13294 (0.0%)	NA
NatAmer	Indicator race = American Indian or Alaska Native (0 = White)	0/13294 (0.0%)	NA
PacIsl	Indicator race = Native Hawaiian or Other Pacific Islander (0 = White)	0/13294 (0.0%)	NA
Multiracial	Indicator race = Multiracial (0 = White)	0/13294 (0.0%)	NA
Other	Indicator race = Other (0 = White)	0/13294 (0.0%)	NA
Notreported	Indicator race = Not reported (0 = White)	0/13294 (0.0%)	NA
Unknown	Indicator race = unknown (0 = White)	0/13294 (0.0%)	NA
HighRiskInd	Baseline covariate high risk pre-existing condition (1=yes, 0=no)	0/13294 (0.0%)	NA
Sex	Sex assigned at birth (1=female, 0=male)	0/13294 (0.0%)	NA
Age	Age at enrollment in years, between 18 and 85	0/13294 (0.0%)	NA
BMI	BMI at enrollment (kg/m^2)	0/13294 (0.0%)	NA

Table 2.2: 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

SL.glm	glmnet	0.717 [0.700, 0.733]
SL.glm	univar_logistic_pval	0.716 [0.700, 0.732]
SL.glm	all	0.716 [0.700, 0.732]
SL.glm	highcor_random	0.716 [0.700, 0.732]
SL	-	0.715 [0.699, 0.731]
Discrete SL	-	0.715 [0.699, 0.731]
SL.mean	all	0.500 [0.468, 0.532]

CV-AUC [95% CI]

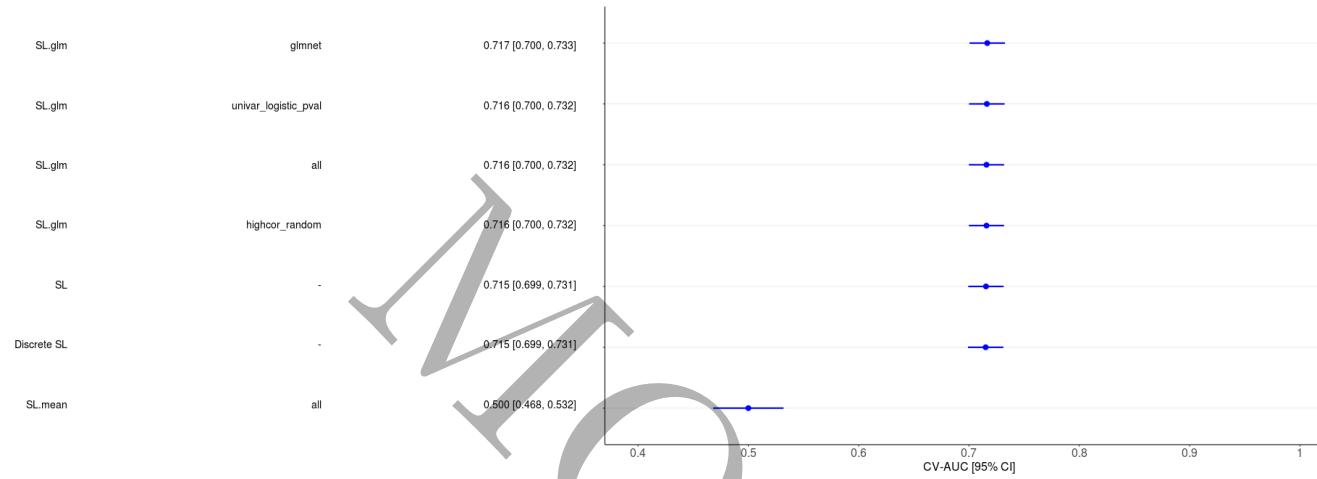Figure 2.1: Cross-validated AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57. CV-AUCs were computed using only data from the placebo arm.
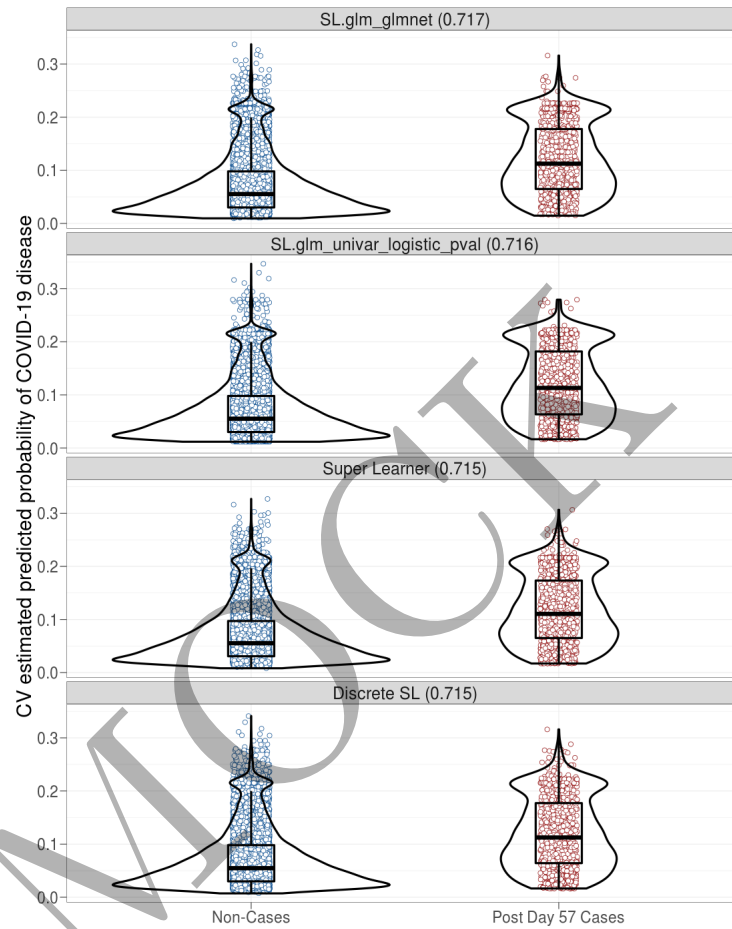
Figure 2.2: CV-estimated predicted probabilities of COVID-19 disease 7 days after Day 57 by case/control status for top 2 learners, SuperLearner and Discrete SL. CV-estimated predicted probabilities were computed using only data from the placebo arm.
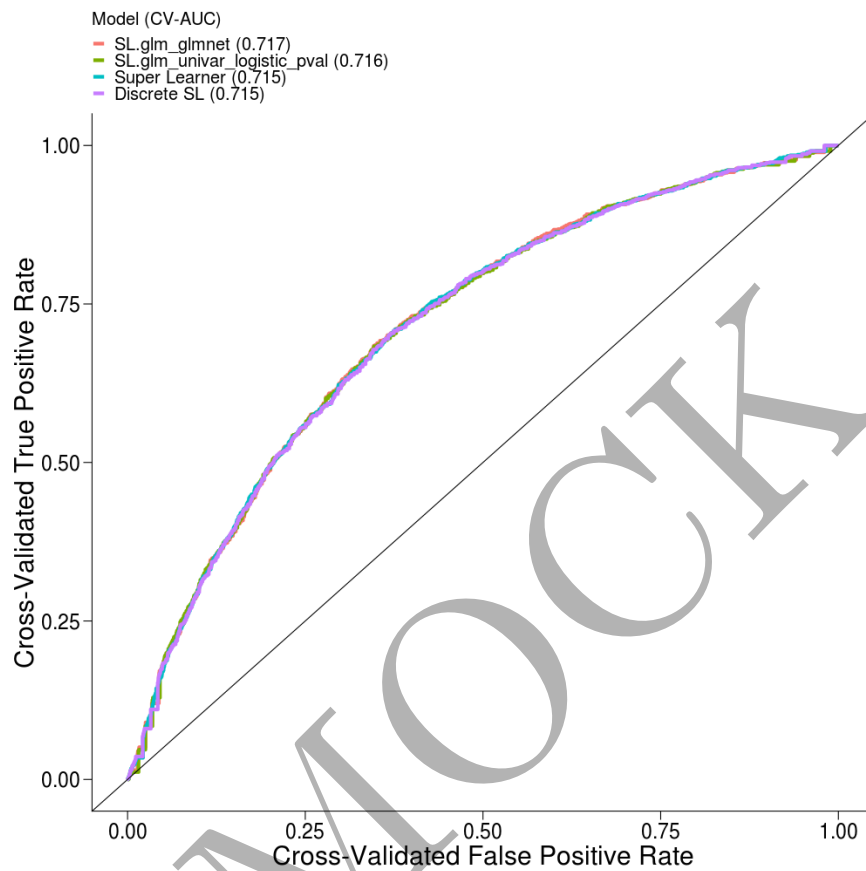
Figure 2.3: ROC curves based off CV-estimated predicted probabilities for the top 2 learners, Superlearner and Discrete SL. CV-estimated predicted probabilities were computed using only data from the placebo arm.

Table 2.3: Weights assigned by Superlearner.

Learner	Screen	Weight
SL.glm	screen_glmnet	0.397
SL.glm	screen_univariate_logistic_pval	0.366
SL.glm	screen_all	0.215
SL.mean	screen_all	0.022
SL.glm	screen_highcor_random	0.000

Table 2.4: Learners assigned weight > 0.0 by Superlearner sorted by weight. Predictors within each learner are sorted by variable importance which is the absolute value in Coefficient (in case of learners like SL.glm, SL.gam, SL.glm.interaction), or Gain (in case of SL.xgboost) or Importance (in case of SL.ranger.imp).

| Learner	Screen	Weight	Predictors	Coefficient	Odds.Ratio
SL.glm	screen__glmnet	0.397	(Intercept)	-2.771	0.063
SL.glm	screen__glmnet	0.397	Age	0.825	2.281
SL.glm	screen__glmnet	0.397	EthnicityUnknown	-0.071	0.931
SL.glm	screen__glmnet	0.397	Multiracial	-0.054	0.948
SL.glm	screen__glmnet	0.397	Asian	0.049	1.05
SL.glm	screen__glmnet	0.397	EthnicityNotreported	0.048	1.049
SL.glm	screen__glmnet	0.397	Other	0.048	1.05
SL.glm	screen__glmnet	0.397	BMI	-0.046	0.955
SL.glm	screen__glmnet	0.397	Black	-0.04	0.96
SL.glm	screen__glmnet	0.397	PacIsl	-0.027	0.973
SL.glm	screen__glmnet	0.397	EthnicityHispanic	-0.023	0.977
SL.glm	screen__univariate__logistic__pval	0.366	(Intercept)	-2.766	0.063
SL.glm	screen__univariate__logistic__pval	0.366	Age	0.823	2.278
SL.glm	screen__univariate__logistic__pval	0.366	EthnicityUnknown	-0.072	0.931
SL.glm	screen__univariate__logistic__pval	0.366	Asian	0.055	1.056
SL.glm	screen__all	0.215	(Intercept)	-2.771	0.063
SL.glm	screen__all	0.215	Age	0.825	2.281
SL.glm	screen__all	0.215	EthnicityUnknown	-0.071	0.931
SL.glm	screen__all	0.215	Asian	0.061	1.062
SL.glm	screen__all	0.215	Other	0.052	1.054
SL.glm	screen__all	0.215	EthnicityNotreported	0.049	1.05
SL.glm	screen__all	0.215	Multiracial	-0.048	0.954
SL.glm	screen__all	0.215	BMI	-0.046	0.955
SL.glm	screen__all	0.215	MinorityInd	-0.024	0.976
SL.glm	screen__all	0.215	Black	-0.023	0.978
SL.glm	screen__all	0.215	PacIsl	-0.022	0.979
SL.glm	screen__all	0.215	HighRiskInd	-0.019	0.981
SL.glm	screen__all	0.215	EthnicityHispanic	-0.014	0.987
SL.glm	screen__all	0.215	Unknown	0.01	1.01
SL.glm	screen__all	0.215	NatAmer	0.009	1.009
SL.glm	screen__all	0.215	Notreported	0.006	1.006
SL.glm	screen__all	0.215	Sex	-0.004	0.996
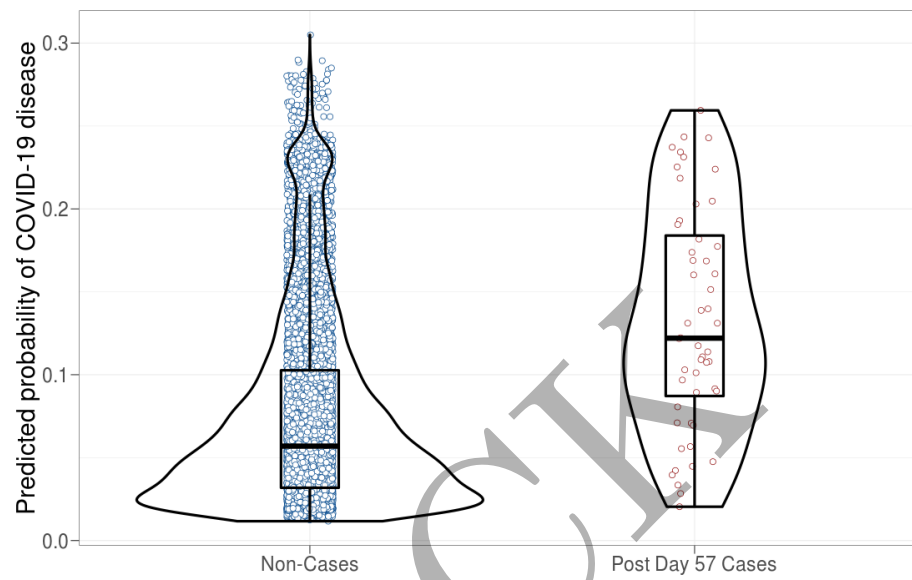
Figure 2.4: Superlearner predicted probabilities of COVID-19 disease in vaccinees 7 days after Day 57 by case/control status.
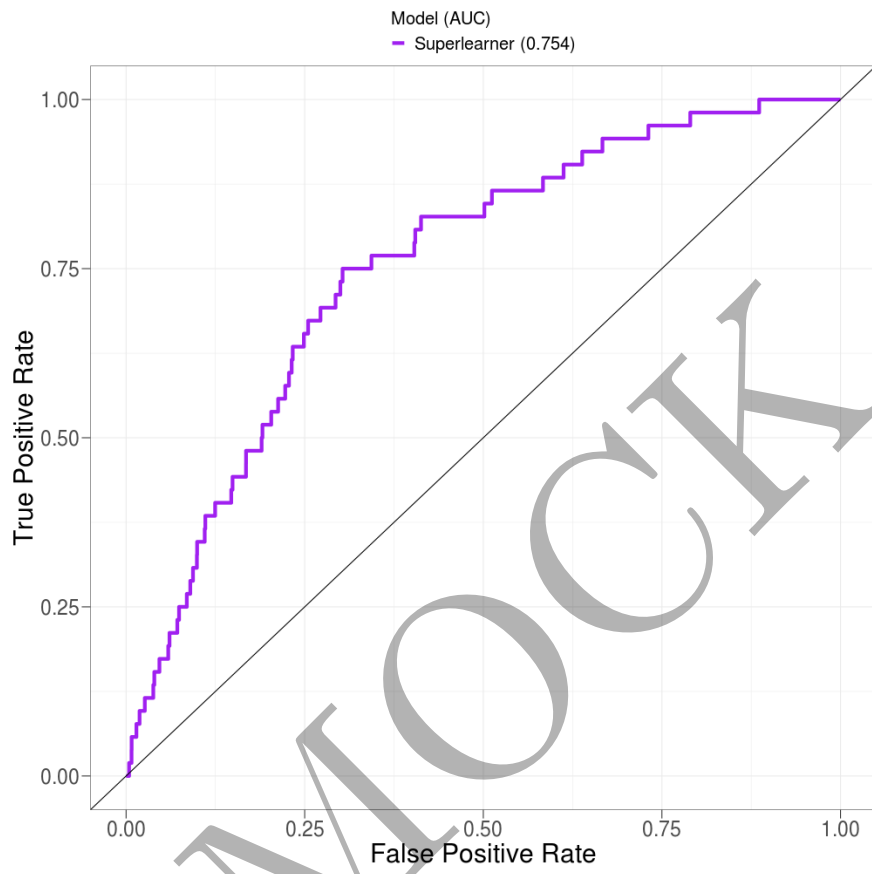
Figure 2.5: ROC curve based off Superlearner predicted probabilities in vaccinees.

Table 2.5: Cases per treatment arm prior to risk score analysis.

Study-Arm	Non-Cases	Post-Day 57-Cases
Placebo	12271	1023
Vaccine	13168	52

Table 2.6: Cases per treatment arm post risk score analysis.

Study-Arm	Non-Cases	Post-Day 57-Cases
Placebo	12271	1023
Vaccine	13168	52

```
#> [1] "running references ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~"
```

Chapter 3

Appendix

- This report was built from the `CoVPN/correlates_reporting` repository with commit hash d1bb29ddb0f88efadc38a0178632cdd2405f0005. A diff of the changes introduced by that commit may be viewed at https://github.com/CoVPN/correlates_reporting/commit/d1bb29ddb0f88efadc38a0178632cdd2405f0005

- The sha256 hash sum of the raw input file, "COVID_VEtrial_practicedata_primarystage1.csv": 83d0f55d1745ffd42be124d8f9ec9a9903abcc13cd22f95e537542a08b41300a

- The sha256 hash sum of the processed file, "moderna_mock_data_processed.csv": d806a7fb38690eff7ecb69f0bd6f74afc54fd2cab78067e383459bdcfa41cf32