

# COVID-19 Baseline Risk Score Analysis Report

mock Study

USG COVID-19 Response Biostatistics Team

April 08, 2021



# Contents

|   |   |   |
|---|---|---|
| 1 | Baseline Risk Score (Proxy for SARS-CoV-2 Exposure) | 9 |
|---|---|---|



# List of Tables

- 1.1 Variables considered for risk score analysis. . . . . 10
- 1.2 All learner-screen combinations (28 in total) used as input to the  
Superlearner. . . . . 11
- 1.3 Weights assigned by Superlearner. . . . . 15
- 1.4 Predictors in learners assigned weight  $> 0.5$  by Superlearner. . . 16



# List of Figures

|     |  |    |
|-----|--|----|
| 1.1 | Cross-validated AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57. . . . .  | 12 |
| 1.2 | CV-estimated predicted probabilities of COVID-19 disease 7 days after Day 57 by case/control status for top 2 learners, Super-Learner and Discrete SL. . . . . | 13 |
| 1.3 | ROC curves based off CV-estimated predicted probabilities for the top 2 learners, Superlearner and Discrete SL. . . . .  | 14 |
| 1.4 | Superlearner predicted probabilities of COVID-19 disease in vaccinees 7 days after Day 57 by case/control status. . . . .                                      | 17 |
| 1.5 | ROC curve based off Superlearner predicted probabilities in vaccinees. . . . .   | 18 |

MOCK



## 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 |
|----------------------|--|----------------------|----------|
| MinorityInd          | Baseline covariate underrepresented minority status (1=minority, 0=non-minority) | 0/30000 (0.0%)       | NA       |
| EthnicityHispanic    | Indicator ethnicity = Hispanic (0 = Non-Hispanic)                                | 0/30000 (0.0%)       | NA       |
| EthnicityNotreported | Indicator ethnicity = Not reported (0 = Non-Hispanic)                            | 0/30000 (0.0%)       | NA       |
| EthnicityUnknown     | Indicator ethnicity = Unknown (0 = Non-Hispanic)                                 | 0/30000 (0.0%)       | NA       |
| Black                | Indicator race = Black (0 = White)   | 0/30000 (0.0%)       | NA       |
| Asian                | Indicator race = Asian (0 = White)   | 0/30000 (0.0%)       | NA       |
| NatAmer              | Indicator race = American Indian or Alaska Native (0 = White)                    | 0/30000 (0.0%)       | NA       |
| PacIsl               | Indicator race = Native Hawaiian or Other Pacific Islander (0 = White)           | 0/30000 (0.0%)       | NA       |
| Multiracial          | Indicator race = Multiracial (0 = White)   | 0/30000 (0.0%)       | NA       |
| Other                | Indicator race = Other (0 = White)   | 0/30000 (0.0%)       | NA       |
| Notreported          | Indicator race = Not reported (0 = White)  | 0/30000 (0.0%)       | NA       |
| Unknown              | Indicator race = unknown (0 = White)   | 0/30000 (0.0%)       | NA       |
| HighRiskInd          | Baseline covariate high risk pre-existing condition (1=yes, 0=no)                | 0/30000 (0.0%)       | NA       |
| Sex                  | Sex assigned at birth (1=female, 0=male)   | 0/30000 (0.0%)       | NA       |
| Age                  | Age at enrollment in years, between 18 and 85                                    | 0/30000 (0.0%)       | NA       |
| BMI                  | BMI at enrollment (kg/m <sup>2</sup> )   | 0/30000 (0.0%)       | NA       |

Table 1.2: All learner-screen combinations (28 in total) used as input to the Superlearner.

| <b>Learner</b>     | <b>Screen*</b>  |
|--------------------|---|
| SL.mean            | all   |
| SL.glm             | all<br>glmnet<br>univar_logistic_pval<br>highcor_random |
| SL.glm.interaction | glmnet<br>univar_logistic_pval<br>highcor_random        |
| SL.glmnet          | all   |
| SL.gam             | glmnet<br>univar_logistic_pval<br>highcor_random        |
| SL.xgboost         | all   |
| SL.ranger.imp      | all   |

*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

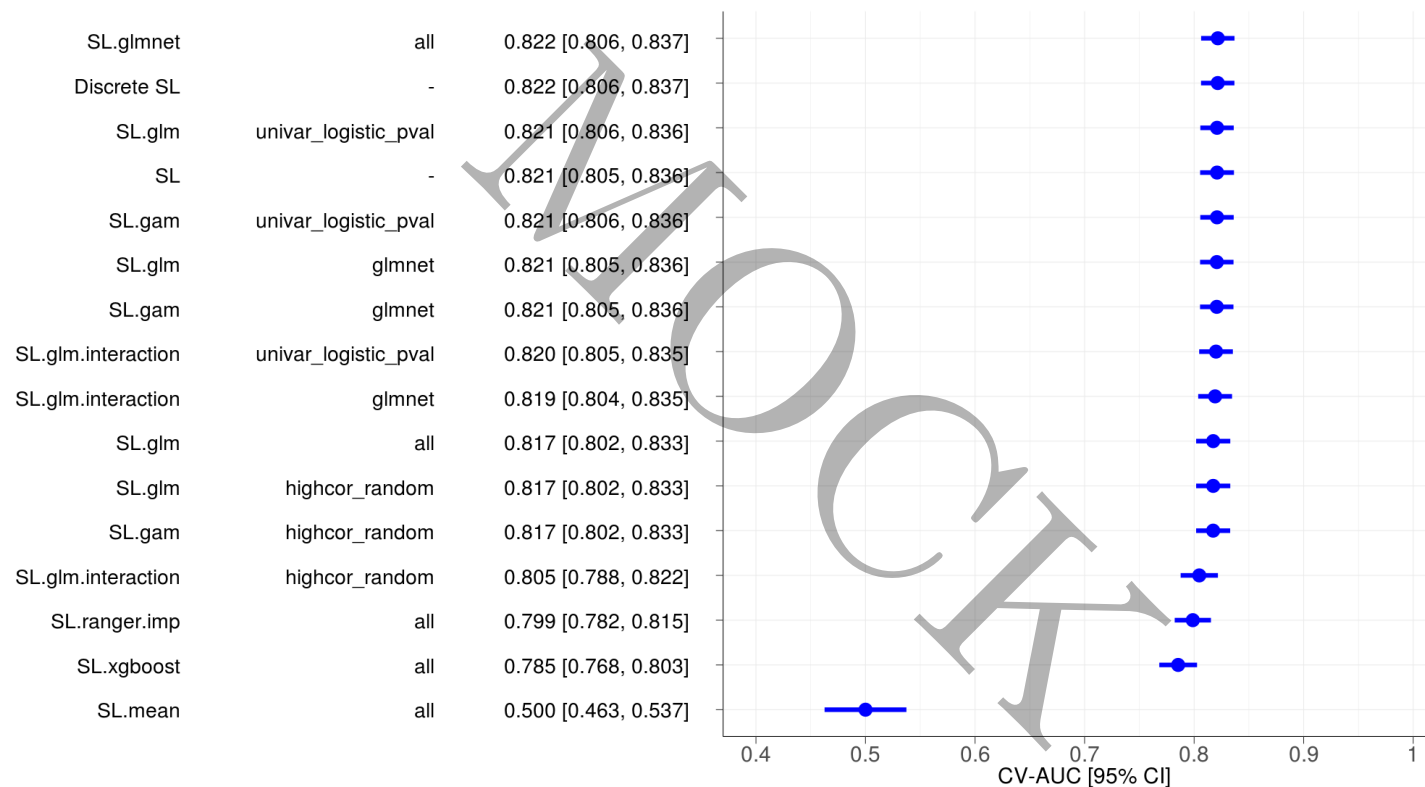


Figure 1.1: Cross-validated AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

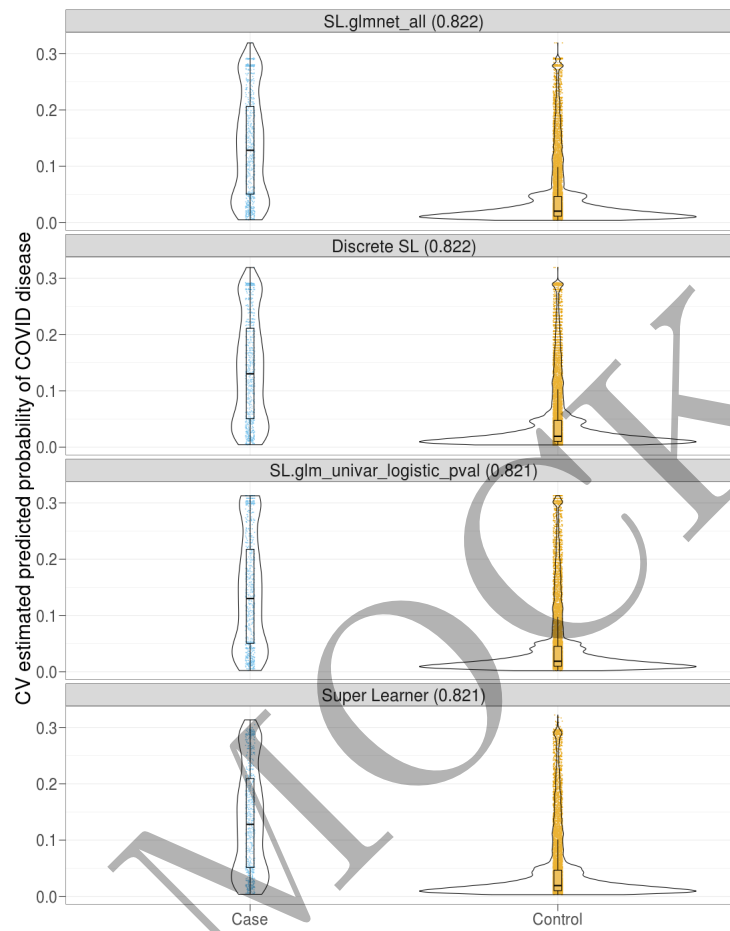


Figure 1.2: CV-estimated predicted probabilities of COVID-19 disease 7 days after Day 57 by case/control status for top 2 learners, SuperLearner and Discrete SL.

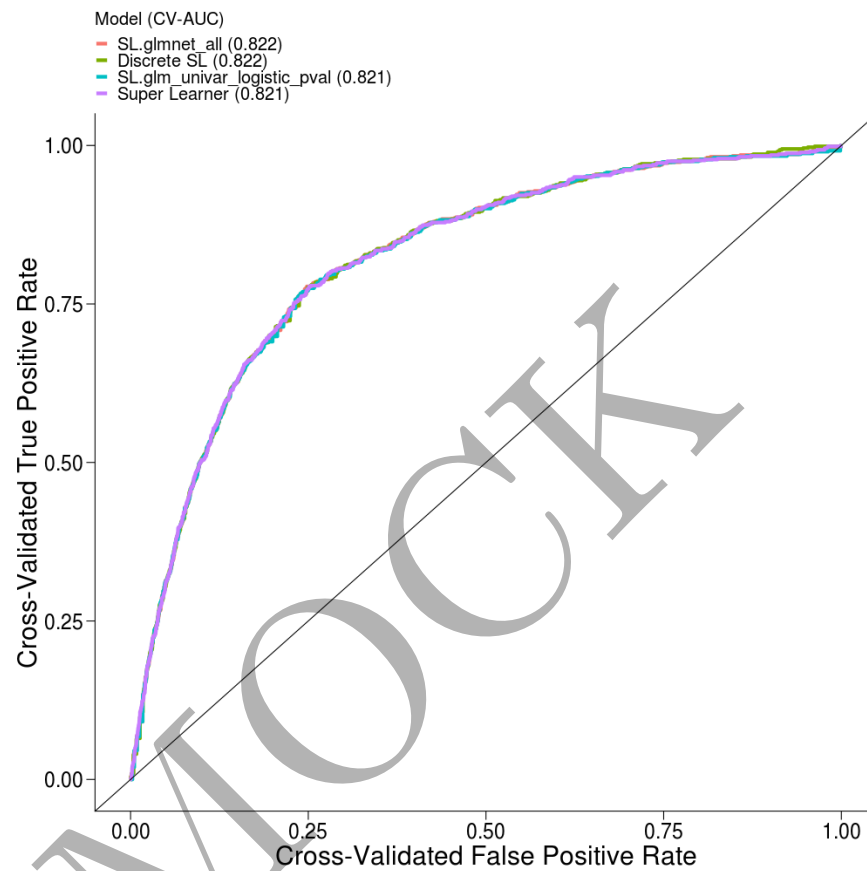


Figure 1.3: ROC curves based off CV-estimated predicted probabilities for the top 2 learners, Superlearner and Discrete SL.

Table 1.3: Weights assigned by Superlearner.

| Learner            | Screen                          | Weight |
|--------------------|---------------------------------|--------|
| SL.glmnet          | screen_all                      | 0.730  |
| SL.glm             | screen_univariate_logistic_pval | 0.245  |
| SL.glm.interaction | screen_highcor_random           | 0.024  |
| SL.mean            | screen_all                      | 0.000  |
| SL.glm             | screen_all                      | 0.000  |
| SL.xgboost         | screen_all                      | 0.000  |
| SL.ranger.imp      | screen_all                      | 0.000  |
| SL.glm             | screen_glmnet                   | 0.000  |
| SL.glm             | screen_highcor_random           | 0.000  |
| SL.glm.interaction | screen_glmnet                   | 0.000  |
| SL.glm.interaction | screen_univariate_logistic_pval | 0.000  |
| SL.gam             | screen_glmnet                   | 0.000  |
| SL.gam             | screen_univariate_logistic_pval | 0.000  |
| SL.gam             | screen_highcor_random           | 0.000  |

Table 1.4: Predictors in learners assigned weight  $> 0.5$  by Superlearner.

| Learner   | Screen     | Weight | Predictors           | Coefficient | Odds.Ratio |
|-----------|------------|--------|----------------------|-------------|------------|
| SL.glmnet | screen_all | 0.73   | (Intercept)          | -3.262      | 0.038      |
| SL.glmnet | screen_all | 0.73   | MinorityInd          | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | EthnicityHispanic    | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | EthnicityNotreported | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | EthnicityUnknown     | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | Black                | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | Asian                | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | NatAmer              | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | PacIsl               | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | Multiracial          | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | Other                | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | Notreported          | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | Unknown              | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | HighRiskInd          | 0.652       | 1.920      |
| SL.glmnet | screen_all | 0.73   | Sex                  | 0.000       | 1.000      |
| SL.glmnet | screen_all | 0.73   | Age                  | 0.358       | 1.430      |
| SL.glmnet | screen_all | 0.73   | BMI                  | 0.000       | 1.000      |



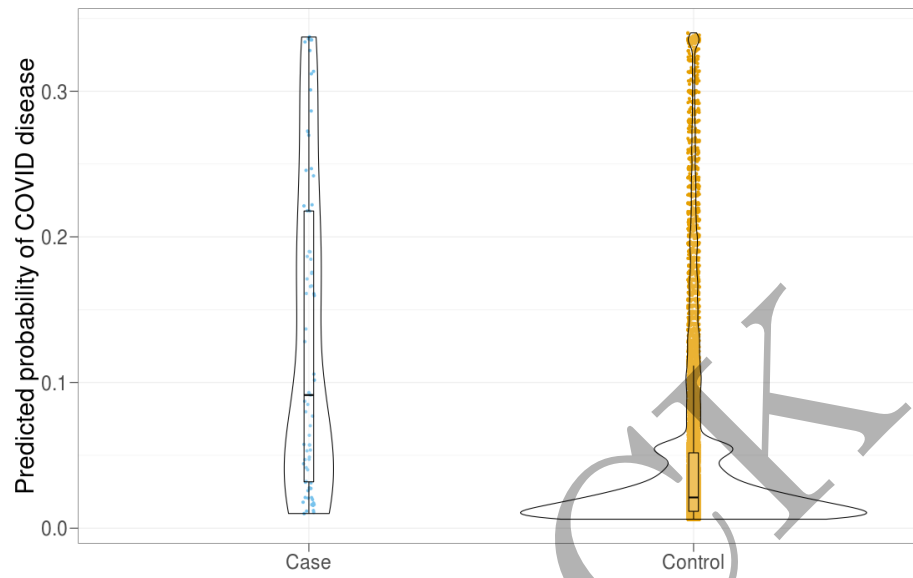


Figure 1.4: Superlearner predicted probabilities of COVID-19 disease in vaccinees 7 days after Day 57 by case/control status.

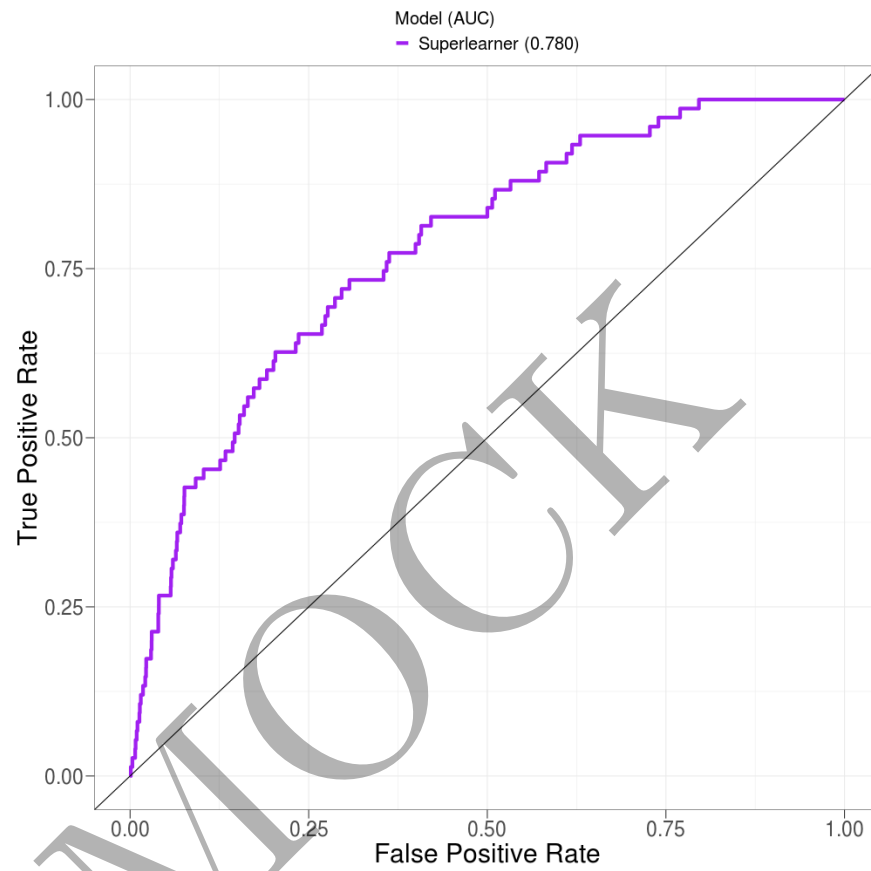


Figure 1.5: ROC curve based off Superlearner predicted probabilities in vaccinees.