

# COVID-19 Baseline Risk Score Analysis Report

## MockCOVE Study

USG COVID-19 Response Biostatistics Team

July 16, 2021



# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Disclaimers</b>   | <b>9</b>  |
| <b>2</b> | <b>Baseline Risk Score (Proxy for SARS-CoV-2 Exposure)</b> | <b>11</b> |
| <b>3</b> | <b>Appendix</b>  | <b>21</b> |



# List of Tables

|     |   |    |
|-----|---|----|
| 2.1 | Variables considered for risk score analysis. . . . .                                       | 12 |
| 2.2 | All learner-screen combinations (28 in total) used as input to the<br>Superlearner. . . . . | 13 |
| 2.3 | Weights assigned by Superlearner. . . . .   | 17 |
| 2.4 | Predictors in learners assigned weight $> 0.0$ by Superlearner. . .                         | 18 |



# List of Figures

|     |  |    |
|-----|--|----|
| 2.1 | Cross-validated AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57. . . . .  | 14 |
| 2.2 | 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. . . . . | 15 |
| 2.3 | ROC curves based off CV-estimated predicted probabilities for the top 2 learners, Superlearner and Discrete SL. . . . .  | 16 |
| 2.4 | Superlearner predicted probabilities of COVID-19 disease in vaccinees 7 days after Day 57 by case/control status. . . . .                                      | 19 |
| 2.5 | ROC curve based off Superlearner predicted probabilities in vaccinees. . . . .   | 20 |

MOCK



# 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.

MOCK

## 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) | 0/13336 (0.0%)       | NA       |
| EthnicityHispanic    | Indicator ethnicity = Hispanic (0 = Non-Hispanic)                                | 0/13336 (0.0%)       | NA       |
| EthnicityNotreported | Indicator ethnicity = Not reported (0 = Non-Hispanic)                            | 0/13336 (0.0%)       | NA       |
| EthnicityUnknown     | Indicator ethnicity = Unknown (0 = Non-Hispanic)                                 | 0/13336 (0.0%)       | NA       |
| Black                | Indicator race = Black (0 = White)   | 0/13336 (0.0%)       | NA       |
| Asian                | Indicator race = Asian (0 = White)   | 0/13336 (0.0%)       | NA       |
| NatAmer              | Indicator race = American Indian or Alaska Native (0 = White)                    | 0/13336 (0.0%)       | NA       |
| PacIsl               | Indicator race = Native Hawaiian or Other Pacific Islander (0 = White)           | 0/13336 (0.0%)       | NA       |
| Multiracial          | Indicator race = Multiracial (0 = White)   | 0/13336 (0.0%)       | NA       |
| Other                | Indicator race = Other (0 = White)   | 0/13336 (0.0%)       | NA       |
| Notreported          | Indicator race = Not reported (0 = White)  | 0/13336 (0.0%)       | NA       |
| Unknown              | Indicator race = unknown (0 = White)   | 0/13336 (0.0%)       | NA       |
| HighRiskInd          | Baseline covariate high risk pre-existing condition (1=yes, 0=no)                | 0/13336 (0.0%)       | NA       |
| Sex                  | Sex assigned at birth (1=female, 0=male)   | 0/13336 (0.0%)       | NA       |
| Age                  | Age at enrollment in years, between 18 and 85                                    | 0/13336 (0.0%)       | NA       |
| BMI                  | BMI at enrollment ( $\text{kg}/\text{m}^2$ )                                     | 0/13336 (0.0%)       | NA       |

Table 2.2: All learner-screen combinations (28 in total) used as input to the Superlearner.

| Learner | Screen*   |
|---------|---|
| SL.mean | all   |
| SL.glm  | all<br>glmnet<br>univar_logistic_pval<br>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 2.1: Cross-validated AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.



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.

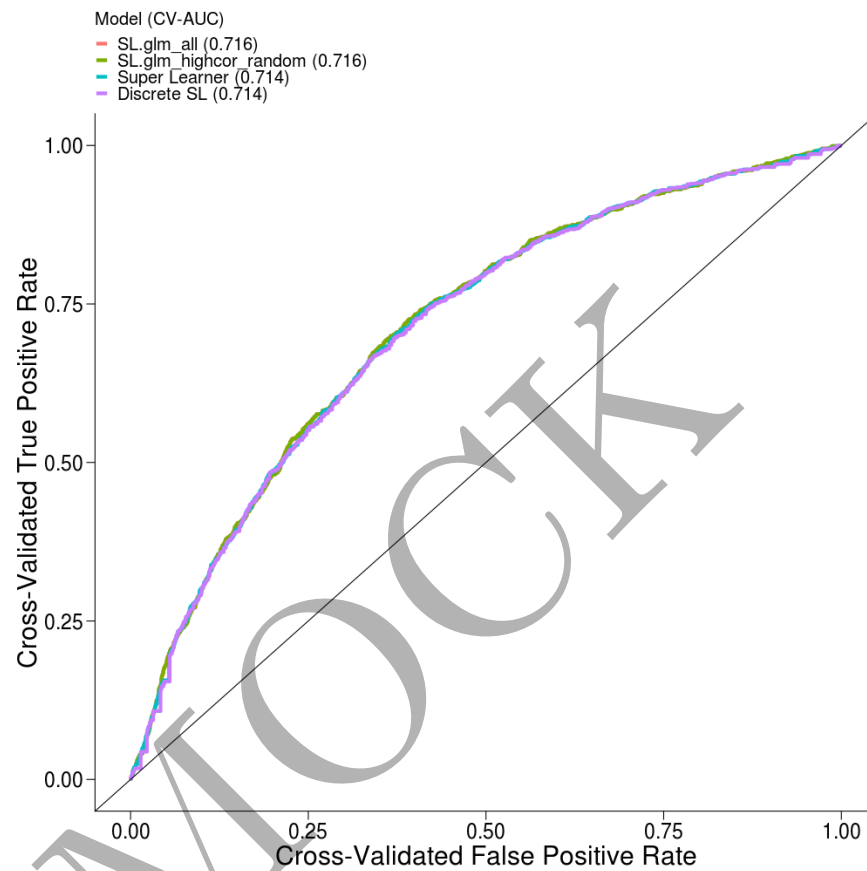


Figure 2.3: ROC curves based off CV-estimated predicted probabilities for the top 2 learners, Superlearner and Discrete SL.



Table 2.3: Weights assigned by Superlearner.

| <b>Learner</b> | <b>Screen</b>                   | <b>Weight</b> |
|----------------|---------------------------------|---------------|
| SL.glm         | screen_all                      | 0.590         |
| SL.glm         | screen_glmnet                   | 0.376         |
| SL.mean        | screen_all                      | 0.034         |
| SL.glm         | screen_univariate_logistic_pval | 0.000         |
| SL.glm         | screen_highcor_random           | 0.000         |

Table 2.4: Predictors in learners assigned weight  $> 0.0$  by Superlearner.

| Learner | Screen        | Weight | Predictors           | Coefficient | Odds.Ratio |
|---------|---------------|--------|----------------------|-------------|------------|
| SL.glm  | screen_all    | 0.59   | (Intercept)          | -2.767      | 0.063      |
| SL.glm  | screen_all    | 0.59   | MinorityInd          | -0.015      | 0.985      |
| SL.glm  | screen_all    | 0.59   | EthnicityHispanic    | -0.013      | 0.987      |
| SL.glm  | screen_all    | 0.59   | EthnicityNotreported | 0.046       | 1.048      |
| SL.glm  | screen_all    | 0.59   | EthnicityUnknown     | -0.072      | 0.931      |
| SL.glm  | screen_all    | 0.59   | Black                | -0.024      | 0.977      |
| SL.glm  | screen_all    | 0.59   | Asian                | 0.057       | 1.058      |
| SL.glm  | screen_all    | 0.59   | NatAmer              | 0.007       | 1.007      |
| SL.glm  | screen_all    | 0.59   | PacIsl               | -0.025      | 0.975      |
| SL.glm  | screen_all    | 0.59   | Multiracial          | -0.052      | 0.95       |
| SL.glm  | screen_all    | 0.59   | Other                | 0.055       | 1.056      |
| SL.glm  | screen_all    | 0.59   | Notreported          | 0.005       | 1.005      |
| SL.glm  | screen_all    | 0.59   | Unknown              | 0.01        | 1.01       |
| SL.glm  | screen_all    | 0.59   | HighRiskInd          | -0.019      | 0.981      |
| SL.glm  | screen_all    | 0.59   | Sex                  | -0.006      | 0.994      |
| SL.glm  | screen_all    | 0.59   | Age                  | 0.824       | 2.279      |
| SL.glm  | screen_all    | 0.59   | BMI                  | -0.045      | 0.956      |
| SL.glm  | screen_glmnet | 0.376  | (Intercept)          | -2.767      | 0.063      |
| SL.glm  | screen_glmnet | 0.376  | EthnicityNotreported | 0.048       | 1.049      |
| SL.glm  | screen_glmnet | 0.376  | EthnicityUnknown     | -0.071      | 0.932      |
| SL.glm  | screen_glmnet | 0.376  | Black                | -0.035      | 0.966      |
| SL.glm  | screen_glmnet | 0.376  | Asian                | 0.05        | 1.051      |
| SL.glm  | screen_glmnet | 0.376  | Multiracial          | -0.064      | 0.938      |
| SL.glm  | screen_glmnet | 0.376  | Other                | 0.048       | 1.049      |
| SL.glm  | screen_glmnet | 0.376  | Age                  | 0.824       | 2.279      |
| SL.glm  | screen_glmnet | 0.376  | BMI                  | -0.045      | 0.956      |
| SL.glm  | screen_glmnet | 0.376  | (Intercept)          | -2.767      | 0.063      |
| SL.glm  | screen_glmnet | 0.376  | EthnicityNotreported | 0.048       | 1.049      |
| SL.glm  | screen_glmnet | 0.376  | EthnicityUnknown     | -0.071      | 0.932      |
| SL.glm  | screen_glmnet | 0.376  | Black                | -0.035      | 0.966      |
| SL.glm  | screen_glmnet | 0.376  | Asian                | 0.05        | 1.051      |
| SL.glm  | screen_glmnet | 0.376  | Multiracial          | -0.064      | 0.938      |
| SL.glm  | screen_glmnet | 0.376  | Other                | 0.048       | 1.049      |
| SL.glm  | screen_glmnet | 0.376  | Age                  | 0.824       | 2.279      |
| SL.glm  | screen_glmnet | 0.376  | BMI                  | -0.045      | 0.956      |



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.

## Chapter 3

# Appendix

- This report was built from the [CoVPN/correlates\\_reporting](https://github.com/CoVPN/correlates_reporting) repository with commit hash 84d0fe7ee9a6464c7da1af24363efb62711c7a6e. A diff of the changes introduced by that commit may be viewed at [https://github.com/CoVPN/correlates\\_reporting/commit/84d0fe7ee9a6464c7da1af24363efb62711c7a6e](https://github.com/CoVPN/correlates_reporting/commit/84d0fe7ee9a6464c7da1af24363efb62711c7a6e)
- 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”:  
28964ce20cfd70a621aff9df412c42b1178d5e749e0f49ddd79fbff2dc7c90a