

COVID-19 Baseline Risk Score Analysis Report

MockCOVE Study

USG COVID-19 Response Biostatistics Team

June 23, 2021

Contents

| | | |
|----------|--|-----------|
| 1 | Disclaimers | 9 |
| 2 | Baseline Risk Score (Proxy for SARS-CoV-2 Exposure) | 11 |
| 3 | Appendix | 25 |

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 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. | 22 |
| 2.5 | ROC curve based off Superlearner predicted probabilities in vaccinees. | 23 |

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) | 0/30000 (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 ²) | 0/30000 (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 glmnet univar_logistic_pval highcor_random |
| SL.glm.interaction | glmnet univar_logistic_pval highcor_random |
| SL.glmnet | all |
| SL.gam | glmnet univar_logistic_pval 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 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.

| Learner | Screen | Weight |
|--------------------|---------------------------------|--------|
| SL.gam | screen_highcor_random | 0.488 |
| SL.glmnet | screen_all | 0.253 |
| SL.ranger.imp | screen_all | 0.139 |
| SL.glm.interaction | screen_glmnet | 0.120 |
| SL.mean | screen_all | 0.000 |
| SL.glm | screen_all | 0.000 |
| SL.xgboost | screen_all | 0.000 |
| SL.glm | screen_glmnet | 0.000 |
| SL.glm | screen_univariate_logistic_pval | 0.000 |
| SL.glm | screen_highcor_random | 0.000 |
| SL.glm.interaction | screen_univariate_logistic_pval | 0.000 |
| SL.glm.interaction | screen_highcor_random | 0.000 |
| SL.gam | screen_glmnet | 0.000 |
| SL.gam | screen_univariate_logistic_pval | 0.000 |

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

| Learner | Screen | Weight | Predictors | Coefficient | Odds.Ratio | Importance |
|---------------|-----------------------|--------|----------------------|-------------|------------|------------|
| SL.gam | screen_highcor_random | 0.488 | (Intercept) | -2.873 | 0.057 | NA |
| SL.gam | screen_highcor_random | 0.488 | s(Age,2) | 0.829 | 2.290 | NA |
| SL.gam | screen_highcor_random | 0.488 | s(BMI,2) | -0.046 | 0.955 | NA |
| SL.gam | screen_highcor_random | 0.488 | MinorityInd | -0.052 | 0.950 | NA |
| SL.gam | screen_highcor_random | 0.488 | EthnicityHispanic | 0.004 | 1.004 | NA |
| SL.gam | screen_highcor_random | 0.488 | EthnicityNotreported | 0.043 | 1.044 | NA |
| SL.gam | screen_highcor_random | 0.488 | EthnicityUnknown | -0.073 | 0.930 | NA |
| SL.gam | screen_highcor_random | 0.488 | Black | 0.004 | 1.004 | NA |
| SL.gam | screen_highcor_random | 0.488 | Asian | 0.069 | 1.071 | NA |
| SL.gam | screen_highcor_random | 0.488 | NatAmer | 0.015 | 1.016 | NA |
| SL.gam | screen_highcor_random | 0.488 | PacIsl | -0.013 | 0.987 | NA |
| SL.gam | screen_highcor_random | 0.488 | Multiracial | -0.046 | 0.955 | NA |
| SL.gam | screen_highcor_random | 0.488 | Other | 0.064 | 1.066 | NA |
| SL.gam | screen_highcor_random | 0.488 | Notreported | 0.003 | 1.003 | NA |
| SL.gam | screen_highcor_random | 0.488 | Unknown | 0.003 | 1.003 | NA |
| SL.gam | screen_highcor_random | 0.488 | HighRiskInd | -0.026 | 0.974 | NA |
| SL.gam | screen_highcor_random | 0.488 | Sex | -0.001 | 0.999 | NA |
| SL.glmnet | screen_all | 0.253 | (Intercept) | -2.665 | 0.070 | NA |
| SL.glmnet | screen_all | 0.253 | MinorityInd | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | EthnicityHispanic | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | EthnicityNotreported | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | EthnicityUnknown | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | Black | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | Asian | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | NatAmer | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | PacIsl | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | Multiracial | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | Other | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | Notreported | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | Unknown | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | HighRiskInd | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | Sex | 0.000 | 1.000 | NA |
| SL.glmnet | screen_all | 0.253 | Age | 0.444 | 1.559 | NA |
| SL.glmnet | screen_all | 0.253 | BMI | 0.000 | 1.000 | NA |
| SL.ranger.imp | screen_all | 0.139 | MinorityInd | NA | NA | 9.100 |
| SL.ranger.imp | screen_all | 0.139 | EthnicityHispanic | NA | NA | 9.685 |
| SL.ranger.imp | screen_all | 0.139 | EthnicityNotreported | NA | NA | 9.247 |
| SL.ranger.imp | screen_all | 0.139 | EthnicityUnknown | NA | NA | 6.908 |
| SL.ranger.imp | screen_all | 0.139 | Black | NA | NA | 9.156 |
| SL.ranger.imp | screen_all | 0.139 | Asian | NA | NA | 7.256 |
| SL.ranger.imp | screen_all | 0.139 | NatAmer | NA | NA | 5.381 |
| SL.ranger.imp | screen_all | 0.139 | PacIsl | NA | NA | 2.553 |
| SL.ranger.imp | screen_all | 0.139 | Multiracial | NA | NA | 5.054 |
| SL.ranger.imp | screen_all | 0.139 | Other | NA | NA | 6.375 |
| SL.ranger.imp | screen_all | 0.139 | Notreported | NA | NA | 6.779 |

Table 2.4: Predictors in learners assigned weight > 0.0 by Superlearner. (continued)

| Learner | Screen | Weight | Predictors | Coefficient | Odds.Ratio | Importance |
|--------------------|---------------|--------|---------------------------------------|-------------|------------|------------|
| SL.ranger.imp | screen_all | 0.139 | Unknown | NA | NA | 4.201 |
| SL.ranger.imp | screen_all | 0.139 | HighRiskInd | NA | NA | 17.116 |
| SL.ranger.imp | screen_all | 0.139 | Sex | NA | NA | 17.890 |
| SL.ranger.imp | screen_all | 0.139 | Age | NA | NA | 182.028 |
| SL.ranger.imp | screen_all | 0.139 | BMI | NA | NA | 231.544 |
| SL.glm.interaction | screen_glmnet | 0.12 | (Intercept) | -2.963 | 0.052 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported | -0.075 | 0.928 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown | -0.283 | 0.753 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black | -0.043 | 0.958 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian | 0.023 | 1.023 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Multiracial | -0.422 | 0.656 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Other | 0.074 | 1.077 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | HighRiskInd | -0.057 | 0.945 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Age | 0.849 | 2.337 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | BMI | -0.065 | 0.937 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:EthnicityUnknown | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:Black | 0.039 | 1.039 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:Asian | 0.002 | 1.002 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:Multiracial | -0.525 | 0.592 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:Other | 0.034 | 1.035 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:HighRiskInd | -0.069 | 0.934 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:Age | 0.001 | 1.001 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:BMI | 0.043 | 1.044 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:Black | 0.008 | 1.008 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:Asian | -0.067 | 0.936 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:Multiracial | -0.529 | 0.589 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:Other | 0.066 | 1.068 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:HighRiskInd | -0.009 | 0.991 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:Age | 0.114 | 1.121 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:BMI | -0.061 | 0.941 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:Asian | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:Multiracial | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:Other | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:HighRiskInd | -0.018 | 0.982 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:Age | 0.012 | 1.012 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:BMI | -0.028 | 0.972 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian:Multiracial | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian:Other | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian:HighRiskInd | 0.031 | 1.032 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian:Age | 0.018 | 1.018 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian:BMI | -0.025 | 0.976 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Multiracial:Other | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Multiracial:HighRiskInd | -0.082 | 0.921 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Multiracial:Age | 0.103 | 1.108 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Multiracial:BMI | 0.000 | 1.000 | NA |

Table 2.4: Predictors in learners assigned weight > 0.0 by Superlearner. (continued)

| Learner | Screen | Weight | Predictors | Coefficient | Odds.Ratio | Importance |
|--------------------|---------------|--------|---------------------------------------|-------------|------------|------------|
| SL.glm.interaction | screen_glmnet | 0.12 | Other:HighRiskInd | 0.036 | 1.037 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Other:Age | -0.019 | 0.981 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Other:BMI | 0.036 | 1.037 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | HighRiskInd:Age | 0.025 | 1.026 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | HighRiskInd:BMI | -0.090 | 0.914 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Age:BMI | 0.012 | 1.012 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | (Intercept) | -2.963 | 0.052 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported | -0.075 | 0.928 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown | -0.283 | 0.753 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black | -0.043 | 0.958 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian | 0.023 | 1.023 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Multiracial | -0.422 | 0.656 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Other | 0.074 | 1.077 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | HighRiskInd | -0.057 | 0.945 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Age | 0.849 | 2.337 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | BMI | -0.065 | 0.937 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:EthnicityUnknown | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:Black | 0.039 | 1.039 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:Asian | 0.002 | 1.002 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:Multiracial | -0.525 | 0.592 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:Other | 0.034 | 1.035 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:HighRiskInd | -0.069 | 0.934 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:Age | 0.001 | 1.001 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityNotreported:BMI | 0.043 | 1.044 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:Black | 0.008 | 1.008 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:Asian | -0.067 | 0.936 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:Multiracial | -0.529 | 0.589 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:Other | 0.066 | 1.068 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:HighRiskInd | -0.009 | 0.991 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:Age | 0.114 | 1.121 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | EthnicityUnknown:BMI | -0.061 | 0.941 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:Asian | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:Multiracial | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:Other | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:HighRiskInd | -0.018 | 0.982 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:Age | 0.012 | 1.012 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Black:BMI | -0.028 | 0.972 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian:Multiracial | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian:Other | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian:HighRiskInd | 0.031 | 1.032 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian:Age | 0.018 | 1.018 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Asian:BMI | -0.025 | 0.976 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Multiracial:Other | NA | NA | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Multiracial:HighRiskInd | -0.082 | 0.921 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Multiracial:Age | 0.103 | 1.108 | NA |

Table 2.4: Predictors in learners assigned weight > 0.0 by Superlearner. *(continued)*

| Learner | Screen | Weight | Predictors | Coefficient | Odds.Ratio | Importance |
|--------------------|---------------|--------|-------------------|-------------|------------|------------|
| SL.glm.interaction | screen_glmnet | 0.12 | Multiracial:BMI | 0.000 | 1.000 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Other:HighRiskInd | 0.036 | 1.037 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Other:Age | -0.019 | 0.981 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Other:BMI | 0.036 | 1.037 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | HighRiskInd:Age | 0.025 | 1.026 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | HighRiskInd:BMI | -0.090 | 0.914 | NA |
| SL.glm.interaction | screen_glmnet | 0.12 | Age:BMI | 0.012 | 1.012 | NA |



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.

MOCK

Chapter 3

Appendix

- This report was built from the [CoVPN/correlates_reporting](https://github.com/CoVPN/correlates_reporting/commit/c64a9538d4c443635a22f674b6725c4568d2e51a) repository with commit hash c64a9538d4c443635a22f674b6725c4568d2e51a. A diff of the changes introduced by that commit may be viewed at https://github.com/CoVPN/correlates_reporting/commit/c64a9538d4c443635a22f674b6725c4568d2e51a
- 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