


# Derivation With Internal Validation of a Multivariable Predictive Model to Predict COVID-19 Test Results in Emergency Department Patients

Samuel A. McDonald, MD, MS<sup>1,2</sup> , Richard J. Medford, MD<sup>2,3</sup>, Mujeeb A. Basit, MD, MMSc<sup>2,4</sup>, Deborah B. Diercks, MD, MSc<sup>1</sup>, and D. Mark Courtney, MD, MSc<sup>1</sup>

## ABSTRACT

**Objectives:** The COVID-19 pandemic has placed acute care providers in demanding situations in predicting disease given the clinical variability, desire to cohort patients, and high variance in testing availability. An approach to stratifying patients by likelihood of disease based on rapidly available emergency department (ED) clinical data would offer significant operational and clinical value. The purpose of this study was to develop and internally validate a predictive model to aid in the discrimination of patients undergoing investigation for COVID-19.

**Methods:** All patients greater than 18 years presenting to a single academic ED who were tested for COVID-19 during this index ED evaluation were included. Outcome was defined as the result of COVID-19 polymerase chain reaction (PCR) testing during the index visit or any positive result within the following 7 days. Variables included chest radiograph interpretation, disease-specific screening questions, and laboratory data. Three models were developed with a split-sample approach to predict outcome of the PCR test utilizing logistic regression, random forest, and gradient-boosted decision tree methods. Model discrimination was evaluated comparing area under the receiver operator curve (AUC) and point statistics at a predefined threshold.

**Results:** A total of 1,026 patients were included in the study collected between March and April 2020. Overall, there was disease prevalence of 9.6% in the population under study during this time frame. The logistic regression model was found to have an AUC of 0.89 (95% confidence interval [CI] = 0.84 to 0.94) when including four features: exposure history, temperature, white blood cell count (WBC), and chest radiograph result. Random forest method resulted in AUC of 0.86 (95% CI = 0.79 to 0.92) and gradient boosting had an AUC of 0.85 (95% CI = 0.79 to 0.91). With a consistently held negative predictive value, the logistic regression model had a positive predictive value of 0.29 (0.2–0.39) compared to 0.2 (0.14–0.28) for random forest and 0.22 (0.15–0.3) for the gradient-boosted method.

**Conclusion:** The derived predictive models offer good discriminating capacity for COVID-19 disease and provide interpretable and usable methods for those providers caring for these patients at the important crossroads of the community and the health system. We found utilization of the logistic regression model utilizing exposure history, temperature, WBC, and chest X-ray result had the greatest discriminatory capacity with the most interpretable model. Integrating a predictive model-based approach to COVID-19 testing decisions and patient care pathways and locations could add efficiency and accuracy to decrease uncertainty.

From <sup>1</sup>the Department of Emergency Medicine, University of Texas Southwestern Medical Center, Dallas, TX; <sup>2</sup>the Clinical Informatics Center, University of Texas Southwestern Medical Center, Dallas, TX; <sup>3</sup>the Department of Internal Medicine/Infectious Disease, University of Texas Southwestern Medical Center, Dallas, TX; <sup>4</sup>and the Department of Internal Medicine/Cardiology, University of Texas Southwestern Medical Center, Dallas, TX.

Received September 14, 2020; revision received November 20, 2020; accepted November 24, 2020.

The authors have no relevant financial information or potential conflicts to disclose.

Supervising Editor: Alan E. Jones, MD.

Address for correspondence: Samuel A. McDonald, MD, MS; e-mail: samuel.mcdonald@utsouthwestern.edu.

ACADEMIC EMERGENCY MEDICINE 2021;28:206–214.

The narrative of COVID-19 for the acute care provider in the United States has been dominated by testing and the question of likelihood of disease for individual patients.<sup>1</sup> In a matter of months providers have gone from relying on reported classic symptoms based on anecdotal experience to developing their own gestalt based on their own experience to having SARS-CoV-2 tests available. Despite very rapid development of viral RNA polymerase chain reaction (PCR) testing in many locations, the question of pretest probability (the probability of disease prior to a test) is still critical for the following reasons: testing throughout the United States is still not fully developed or immediately available,<sup>2</sup> risk of disease may be significantly overestimated resulting in clinician fear and impact on personal protective equipment (PPE) use, expansion of COVID-19 to at-risk countries and settings with no testing,<sup>3,4</sup> and its role in establishing a posttest probability after result of the test.<sup>5</sup>

This latter point has become critical for emergency clinicians, hospital administrators, and epidemiologists trying to limit both community- and health care-associated spread and follow disease trajectory. The posttest probability is the probability of disease after a test and is a function of two things: 1) the pretest probability and 2) the diagnostic performance and result of the test. Despite varying tests being used and the uncertainty of their diagnostic performance, the posttest probability is driven to a very large part by pretest probability. Efforts to more scientifically quantify pretest probability would aid individual and health system decision making.<sup>6,7</sup>

The greatest concern to limiting spread both within the community and importantly within health systems and care facilities is the reported nontrivial rate of false-negative testing.<sup>8</sup> In the context of high-risk patients, many are recommending repeat testing, recognizing that quality of nasopharyngeal swab technique, early-stage illness, and handling of media may contribute to false-negative tests.<sup>9,10</sup> For these reasons, many systems are adopting efforts to risk stratify patients based on gestalt, early internal data, or home-grown scoring systems. These approaches attempt to meet a significant operational need in the emergency department (ED) in particular as the decision of disposition and early management of suspected COVID-19 patients typically falls to the emergency physician. However, minimal work has been done to provide a data-driven method for COVID-19 risk stratification. Once developed, such a predictive approach would

further our ability to provide appropriate recommendations for mitigating disease spread in and out of the hospital, identify patients who warrant repeat testing if the initial test is negative, and inform future research. The purpose of this study was to develop and internally validate a predictive model that could aid in the discrimination and management of patients undergoing investigation for COVID-19 when presenting to the ED. Specifically, our objective was to use a robust electronic health record (EHR) that captures laboratory, clinical, vital sign, comorbidity, and radiographic data with traditional and novel modeling techniques to predict SARS-CoV-2 PCR test results in ED patients.

## METHODS

---

This is a retrospective cohort study of consecutive patients presenting to the ED who were tested for COVID-19 between March 6 and April 24, 2020, at a single academic quaternary care facility with a typical yearly volume of 50,000 patients. Given the pandemic environment and reduced ED presentations, the volume experienced during the study period was approximately 4,000 unique patient encounters. This quality improvement initiative was prioritized in the early evolution of our response to COVID-19 to best evaluate how to deploy testing strategies in a resource supply conscious environment and was formally approved by the Department of Emergency Medicine Quality Improvement Committee. Reporting of this derivation has followed guidelines established by TRIPOD statement.<sup>11</sup>

We included all patients greater than 18 years presenting to our ED who were tested for COVID-19 during this index ED visit and had not previously received testing in our health system. Testing decisions during the study period were protocolized by the health system based on a CDC guideline driven testing strategy requiring either a high-risk exposure or a high-prevalence location in addition to associated symptoms. Testing was also performed if there was high suspicion from the practicing clinician that fell outside of the protocolized guidelines. No asymptomatic testing was being performed during this period. Patients with repeat visits after the index visit were not included in the analysis because the available prior encounter data within the EHR could differentially affect pretest probability estimation as well as workup of disease relative to an encounter without

previous ED testing. The primary outcome of interest was the presence of COVID-19 infection evaluated by the performance of PCR testing via nasopharyngeal swab obtained at the index or any resulting positive test within the following week. Positive PCR testing was defined by any positive PCR test performed by laboratory technicians without awareness of other clinical data. If initial testing was negative but repeat testing within 1 week was positive, a positive result was attributed as the final outcome. This was done to account for the concern for possible initial false-negative testing in setting of early presenters. If no additional testing was performed the initial result was considered the final outcome.

Initially 44 variable candidates were collected from the clinical data warehouse that are routinely available during an ED encounter when evaluating influenza-like illnesses. The data set consisted of patient specific information, triage screening questions, initial vital signs, specific laboratory results, and chest x-ray interpretation. Patient-specific information included age at encounter; sex, race; and specific comorbid conditions consisting of hypertension, diabetes, end-stage renal disease, asthma, and COPD. The comorbidities are extracted from SNOMED concept-based patient registries, defined by a standardized vocabulary of clinical concepts, that are automatically populated within the EHR anytime a clinician adds a diagnosis or problem list item that maps to the specific SNOMED concept.<sup>12</sup> Given the number of patients available for our analysis limits our capacity to use unique comorbid conditions in a predictive model, we quantified the number of comorbid conditions per patient as a unique ordinal variable from 0 to 5. The screening questions performed at triage are a set of standardized general risk stratification questions for developing pathogens with increased specificity to COVID-19. These questions assessed presence of recent fever, travel to location of high disease prevalence, exposure to someone with COVID-19, and associated viral symptoms and are treated as dichotomous variables.

Continuous variables collected for evaluation consisted of laboratory values and vital signs. Labs selected for evaluation were total white blood cell count (WBC), absolute lymphocyte count and percentage, alanine aminotransferase, aspartate aminotransferase, C-reactive protein, ferritin, lactate dehydrogenase, high-sensitivity troponin T (hs-TnT), and NT-pro-BNP. These were selected based on emerging literature reporting value in disease specificity

and severity.<sup>13,14</sup> Vitals obtained for the visit were collected as additional covariates consisting of blood pressure, pulse, respiratory rate, oxygen saturation, and temperature. To have a uniform collection method the first vitals obtained during the ED visit were used for analysis.

Values of each of these covariates consisted of the first value obtained in the encounter for consistency and availability. Finally chest X-ray (CXR) interpretation was included as a categorical variable consisting of options of consistent with viral pneumonia, low likelihood findings related to viral pneumonia, and negative. This was stored as a structured result within the EHR, classified by the interpreting radiologist. This was a new process implemented specifically for COVID-19. There was a subset of images that were classified by the study group prior to implementation of that process. These were tagged in a similar fashion by the interpreting radiologist.

Listwise elimination was performed for any observation without an associated COVID-19 order result. Listwise elimination was also performed if no vital sign data or screening questionnaire data were available. This was done given the concern that these data were missing not at random as these points should be present in almost all ED visits. Variables with very low variance, more than 70% missing values, or high correlation with other features were removed prior to modeling. After initial variable selection and division of data, imputation was performed separately on the testing and validation data sets utilizing a random forest-based nonparametric method that can impute both numeric and categorical variables using R package *missRanger*.<sup>15,16</sup> It is likely that many encounters had missing data for CXR or laboratory values given they were not ill-appearing and unlikely to be admitted. Clinically, we did not mandate laboratory or CXR in all patients tested for COVID and the resulting data are likely what is to be found in real-world setting.

Data were randomly divided with a split-sample approach allocating 75% of data for training each model and 25% for internal validation. Utilizing the training data set, three distinct models were developed to predict COVID-19 test results to maximize our ability to arrive at an optimal method for prediction. A logistic regression model was developed using a backward stepwise variable selection method utilizing Wald chi-square statistics to evaluate the reduced model. This was performed using R RMS package.<sup>17</sup> Two

ensemble decision tree–based approaches were used, random forest using R Ranger package<sup>18</sup> and a gradient-boosted decision tree method using R XGBoost package<sup>19</sup> utilizing the default recommendations for tuning parameters. These models use the composite prediction of a group of base models to optimize bias and variance of the prediction, improving the prediction of a standard decision tree–based methodology. While there are many options for model types to use for prediction, our goal was to have a model that could be easily implemented within our EHR to provide automated alerting to providers. At this time, our capacity to implement models more complex than generalized linear models or tree-based models is limited, and we felt that the operational value for testing other model types was low.

We present central tendency with mean and associated 95% confidence intervals (CIs). Categorical data are presented as total number and percentages. Model performance is evaluated by examining the area under the receiver operator curve (AUC) for each model with associated 95% CIs in the validation data. To examine the specific threshold that would be used for clinical decision making, we compared the positive predictive values (PPV) at a set negative predictive value of 99%. This cut point was chosen to preferentially minimize false negatives. All analysis was done using R version 3.6.3

## RESULTS

Over the study period, 1,060 encounters met the inclusion criteria. There were 11 observations removed due to no associated outcome value and 23 observations were removed due to lack of triage screening data or vitals. Our final study sample consisted of 1,026 unique patients. The prevalence of COVID-19 in this population was found to be 9.6% (95% CI = 9.9% to 11.6%) with 99 patients testing positive. The cohort had a mean ( $\pm$ SD) age of 52 ( $\pm$ 19) years with 56% being female. Other demographic details are found in Table 1.

Alanine aminotransferase was found to have most frequent missing data with 29% in training and 31% in validation data subsets, followed by CXR (16 and 18%), lymphocytes (16 and 14%), and WBC (15 and 13%), visualized in Figure 1. The rest of the predictors had less than 3% missingness in both training and validation data. This data set reflects all subjects who were tested. The final selected covariates are presented in Table 2.

**Table 1**  
Patient Characteristics and Demographics

	Overall (N = 1,026)
Age (years)	52 ( $\pm$ 19)
Sex, Female	573 (55.8)
Race/ethnicity	
Asian	37 (3.6)
Black or African American	328 (32.0)
Other	71 (6.9)
Unavailable/unknown	55 (5.4)
White, Hispanic	120 (11.7)
White, non-Hispanic	415 (40.4)
Comorbidities	
HTN	601 (58.6)
DM	361 (35.2)
COPD	188 (18.3)
Asthma	329 (32.1)
ESRD	89 (8.7)

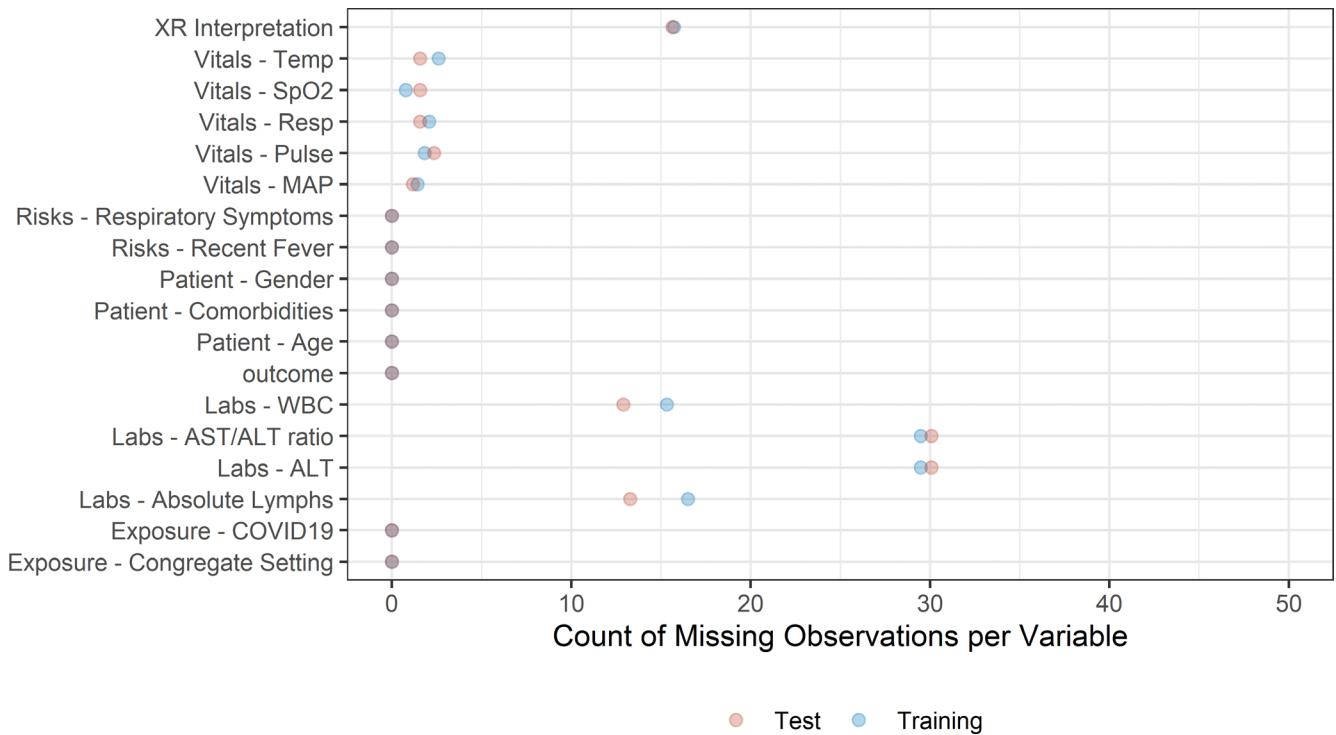
Data are reported as mean ( $\pm$ SD) or *n* (%).

COPD = chronic obstructive pulmonary disease; DM = diabetes mellitus; HTN = hypertension; ESRD = end-stage renal disease.

Three models were developed for comparison. The logistic regression model was found to have the largest AUC of 0.89 (95% CI = 0.84 to 0.94) and simplistically contained only known exposure to COVID-19, CXR consistent with viral pneumonia, elevated temperature, and reduced WBC. The two ensemble approaches used did not result in a superior AUC. Random forest method resulted in AUC of 0.86 (95% CI = 0.79 to 0.92) and gradient boosting method demonstrated an AUC of 0.85 (95% CI = 0.79 to 0.91). Associated receiver operator curves are presented in Figure 2. Discrimination of each model can be appreciated in Figure 3. To contextualize performance with respect to the necessary clinical decision making, sensitivity, specificity, and resulting PPV were reported in Table 3 at the pre-specified threshold where NPV was 99%. The logistic regression model offered a higher PPV of 0.29 (0.2–0.39) compared to 0.20 (0.14–0.28) for the random forest and 0.22 (0.15–0.3) for the gradient-boosted method. The selected logistic regression model is presented in Table 4 for review.

## DISCUSSION

Using data collected in the ED encounter of 1,026 patients presenting for evaluation of COVID-19, we developed three statistical learning models to aid in discrimination of patients when COVID-19 testing is not available, not timely, or of questionable result.



**Figure 1.** Visualization of missing data in training and validation data. AST/ALT = alanine aminotransferase/aspartate aminotransferase; MAP = mean arterial pressure; WBC = white blood cell count.

When comparing the derived models, all had similar discriminative capacity when predicting positive COVID-19 PCR result, but the logistic regression model offered identical performance at the predefined test threshold (NPV > 99%), the primary aim of this study, with a greater capacity to identify patients positive for COVID-19 as well (PPV = 29%).

We feel the results of this model provide great value to the bedside clinician in offering a relatively simplified and easy to interpret model with enough strength in discrimination to provide risk stratification for patients presenting to the ED for COVID-19 (Figure 3). The value in this derived model is its accessibility and relevance to an acute care clinician with key features that were identified by the feature selection process—abnormal x-ray, positive contact with disease, elevated temperature and WBC—clinically meaningful and regularly obtained during an ED encounter. While the ensemble methods had slightly decreased predictive capacity, they also attributed significant importance to these same variables further justifying their use in our final model.

As frontline clinicians continue to care for all patients presenting to the ED, the threat of COVID-19 creates a cognitive burden to ensure accurate diagnosis needed and expected of us to facilitate appropriate recommendations to prevent further spread of this

disease. Clinicians in the ED and point of care are being inundated daily with the question of “does this patient have COVID?” while at the same time they are charged with preserving PPE, assigning patients to the right team, and conserving test reagent resources. Models that offer some discrimination among patients provide an easy method for assigning pretest risk stratification. In settings where testing is not easily available or scarce this may offer enough discrimination to reduce the need for testing in low-risk patients. Even when testing may still be performed it could offer a data-driven mechanism to reduce PPE consumption. Some of the more accessible tests to emergency providers have reportedly poorer sensitivity causing concern for false-negative rates.<sup>20,21</sup> While not studied here specifically within this cohort, this model could be used to further lower the risk of COVID-19 after initially receiving a negative point-of-care test. This could offer reassurance to care teams, minimize PPE consumption for patients who are admitted, and reduce the need for patients being discharged to perform the self-isolation that would likely be unnecessary and unwarranted.

While none of these models offer near perfect discrimination there is a clear separation in the two distributions in our results (Figures 2 and 3). This

**Table 2**  
Distribution of Variables Included Within Split by Training and Validation Data Sets

	Total (N = 1026) Mean (SD)/n (%)	Training (N = 770) Mean (SD)/n (%)	Validation (N = 256) Mean (SD)/n (%)	p Value
<b>Patient specific</b>				
Age (years)	52 ( $\pm$ 19)	52 ( $\pm$ 19)	52 ( $\pm$ 19)	0.626
Sex, female	573 (55.8)	429 (55.7)	144 (56.2)	0.881
<b>Exposure</b>				
Congregate setting	24 (2.3)	20 (2.6)	4 (1.6)	0.343
Reported COVID-19 exposure	114(14.5)	77 (10)	37 (14.5)	0.050
<b>Labs</b>				
Absolute lymphocytes	1.59 (1.78)	1.63 (2.00)	1.47 (0.88)	0.226
ALT	38 (49)	37 (47)	40 (55)	0.409
AST/ALT ratio	1.317 (0.784)	1.333 (0.812)	1.267 (0.693)	0.330
WBC	8.8 (5.8)	8.7 (5.4)	9.2 (6.8)	0.191
Comorbidities	1.5 (1.3)	1.5 (1.3)	1.5 (1.3)	0.476
<b>Symptoms</b>				
Respiratory	680 (66.3)	500 (64.9)	180 (70.3)	0.115
Fever	374 (36.5)	281 (36.5)	93 (36.3)	0.962
<b>Vitals</b>				
MAP	97 (17)	97 (17)	97 (17)	0.922
Pulse	90 (19)	90 (18)	92 (19)	0.098
Respirations	18 (4)	18 (4)	19 (4)	0.654
SpO <sub>2</sub>	98 (2)	98 (2)	98 (3)	0.396
Temperature	98.5 (1.4)	98.5 (1.3)	98.5 (1.4)	0.830
<b>CXR interpretation</b>				
Negative	680 (78.6)	516 (79.5)	164 (75.9)	0.491
Typical for COVID	113 (13.1)	80 (12.3)	33 (15.3)	
Atypical for COVID	72 (8.3)	53 (8.2)	19 (8.8)	
COVID +	99 (9.6)	69 (9.0)	30 (11.7)	0.195

Data are reported as mean ( $\pm$ SD) or *n* (%).

ALT = alanine aminotransferase; AST = aspartate aminotransferase; MAP = mean arterial pressure; WBC = white blood cell count.

offers some flexibility in determination of an appropriate test threshold to divide the model prediction into two classes without significant loss in discrimination, increasing the capacity for utilizing the model for more than its capacity to rule out disease. Increasing the classification threshold could allow one to focus on PPV. While not typically a focus for emergency medicine, utilizing the model in this capacity could provide a method for identifying a high-risk cohort early in course of care to aid in better allocation of scarce hospital resources such as restricted treatment regimens, negative pressure rooms, and even personal protective equipment. A model of this type could become even more useful in the future when concern for COVID-19 is not the forefront of the acute care provider thoughts. If implemented appropriately, a provider caring for a patient with a high pretest probability for disease could receive an alert within the EHR that recommends testing and

appropriate cohorting to minimize unnecessary staff exposure. With a posttest probability of 30% found utilizing available ED data, with our model would be useful toward these efforts.

This is one of the first models to provide a context on the likelihood of COVID-19 with a population specific to emergency providers in the United States. There has been limited work that remains in preprint from China that attempted to predict COVID-19 with a goal of minimizing unnecessary testing, specifically CT scans because they were used as a primary decision point in China early in the pandemic.<sup>22,23</sup> Unfortunately, many of the features included in some of these early predictive models utilized laboratory tests that are not easily available or would not result during the normal course of an ED visit in the United States.

Health care researchers are more frequently utilizing novel techniques for prediction because consensus seems to be that novel “AI” machine learning

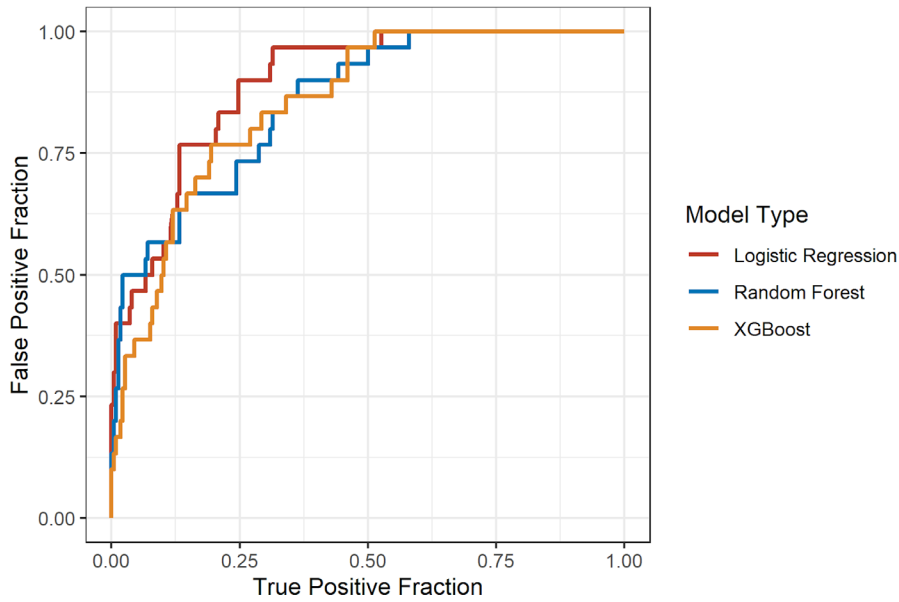


Figure 2. Area under the receiver operator curve of the three derived models.

Table 3

Test Statistics for the Three Derived Models With Threshold Established at Location Where the Model Had 99% NPV

Model	AUC (95% CI)	N	Predicted + /True +	Predicted - /True -	Sensitivity (95% CI)	Specificity (95% CI)	NPV (95% CI)	PPV (95% CI)	PLR (95% CI)	NLR (95% CI)
Logistic regression	0.89 (0.84–0.94)	256	29	155	0.97 (0.83–1)	0.69 (0.62–0.75)	0.99 (0.96–1)	0.29 (0.2–0.39)	3.1 (2.5–3.8)	0.05 (0.01–0.3)
Random forest	0.86 (0.79–0.92)	256	29	113	0.97 (0.83–1)	0.5 (0.43–0.57)	0.99 (0.95–1)	0.2 (0.14–0.28)	1.9 (1.7–2.2)	0.07 (0.01–0.47)
XGBoost	0.85 (0.79–0.91)	256	29	122	0.97 (0.83–1)	0.54 (0.47–0.61)	0.99 (0.96–1)	0.22 (0.15–0.3)	2.1 (1.8–2.5)	0.06 (0.01–0.4)

AUC = area under the receiver operator curve; NLR = negative likelihood ratio; NPV = negative predictive value; PLR = positive likelihood ratio; PPV = positive predictive value.

techniques offer improved prediction and discrimination.<sup>24,25</sup> Although allure exists for using some of these newer methods, the benefit of using a logistic regression model, in addition to its improved predictive performance for this use case, is the added interpretability offered.<sup>26</sup> It becomes operationally more lucrative than these black box machine learning methods as it provides the basis for development of a simple decision tool for clinicians to risk stratify patients when a health system lacks the capacity for implementation of a regression model within an EHR. Unlike standard decision trees, the nodes that enable decision making are not easily extrapolated into a single decision rule due to the nature of the composite prediction. Without implementation within an EHR for automated scoring, ensemble methods provide little value to the average clinician.

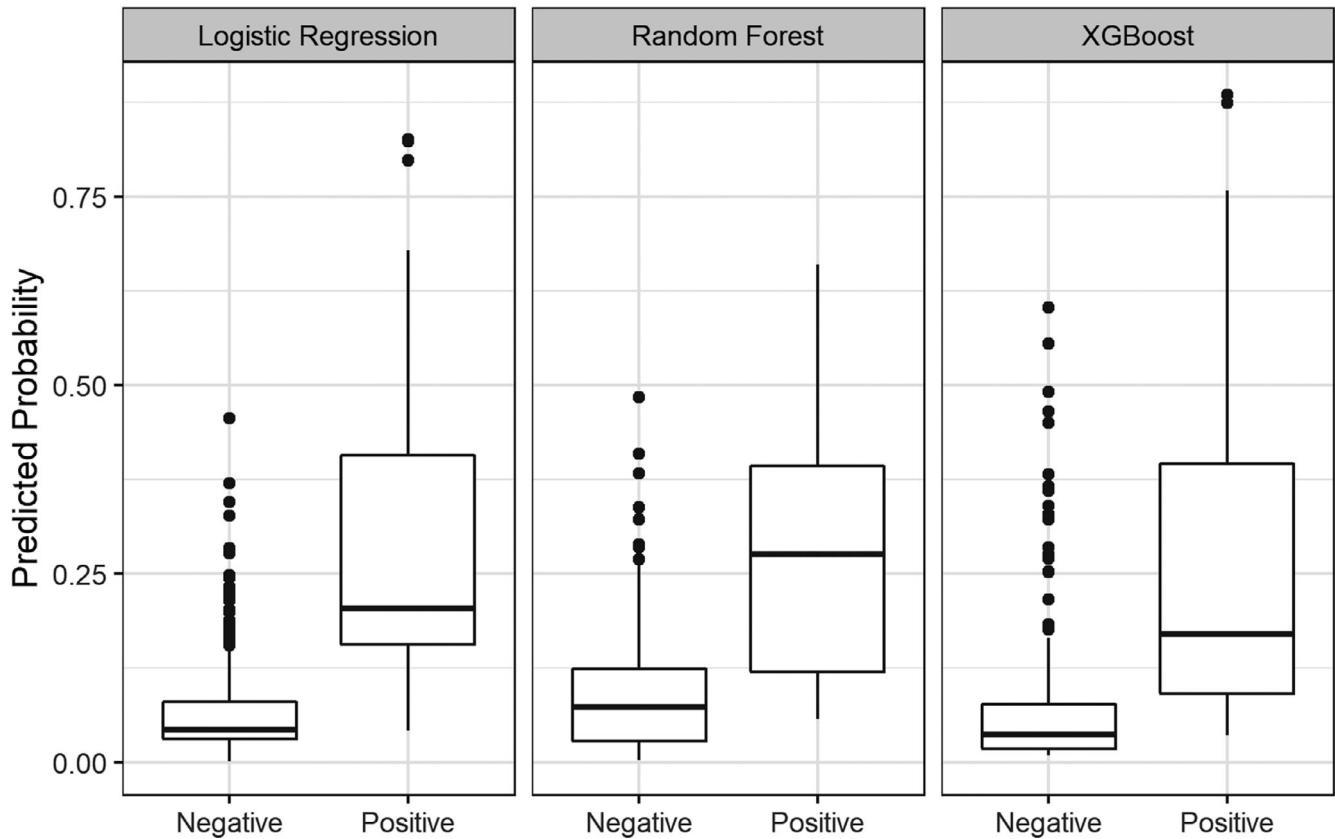
Table 4

Presenting the Full Logistic Regression Model With Intercept and Associated Odds Ratios

Intercept and Predictors	β-Coefficient	OR	95% CI
Intercept	-36		
WBC	-0.59	0.6	0.4-0.8
Temperature	0.44	1.6	1.2-2.0
Known exposure	1.51	4.5	2.4-8.7
Positive CXR	1.69	5.4	3-10

## LIMITATIONS

There are several limitations to our study. First and foremost, this was a retrospective analysis that is prone to bias. In addition, this resulted in much missing data though very few observations had a significant amount of missingness. Our institution’s testing capacity, uniquely, has been available to ED



**Figure 3.** Box plots for the three derived models with predicted probabilities stratified by the true outcome.

physicians since COVID-19 arrived in the North Texas area. Our broad inclusion criteria, institutional approach to liberal testing, and test availability likely aided to reduce spectrum bias typical in these retrospective derivation studies that overly sample the more ill patients.

Another limitation was that this was a single site-derived model in a unique patient population with a presumed inherently high-risk patient baseline. While generalization of this model would presumably show reduction in discriminatory capacity, the simplicity of the logistic regression model seems less prone to overfitting than a model with a significantly higher number of predictors. One feature that may have higher variance elsewhere would be the categorization of the radiology read because a discrete result was provided by our radiologists following a systemwide protocol. If left to emergency physicians or other radiologists without strict guidance, the predictive nature of the CXR may vary.

Additionally, with more data, a more optimally fit model could be constructed with the addition of other high value predictors. Future work intends to externally validate on data from multiple other institutions.

It is possible that the information lost due to the necessary discretization of the continuous variables within the tree-based methods affected their performance when compared to the logistic regression model. Additionally, these ensemble prediction methods are typically “data-hungry”<sup>25</sup> and could have benefitted from additional observations for training that were unavailable at a single-site study. It is possible future work with more observations would find that these methods offer better discrimination than we have found in our cohort.

Finally, defining an outcome for modeling is difficult when criterion standards are not well validated. The PCR test has been used as criterion standard for diagnosis of COVID, but some of these assays have been found to have a nontrivial false-negative rate.<sup>20,21</sup> While we could not expect all patients to get both a COVID PCR testing and CT chest in normal ED operations in the United States, we attempted to mitigate this false-negative rate by allowing for repeat testing performed within a reasonable time horizon to their initial visit. Regardless, at present and the foreseeable future, the COVID PCR test is clinically being used as the source of truth.



## CONCLUSION

Based on the analysis of a consecutive population of patients arriving to an ED for the evaluation of COVID-19, the derived and internally validated model has good discriminatory capacity for COVID-19 disease utilizing four easily obtained variables: exposure history, temperature, WBC, and chest X-ray result. This model offers an easily interpretable and immediately usable to the average emergency clinician—from a rule-of-thumb method to implementation of the model within the EHR—to help drive high-quality patient and system-level care for COVID-19 disease.

## References

- Omer SB, Malani P, Del Rio C. The COVID-19 pandemic in the US: a clinical update. *JAMA* 2020;323:1767–8.
- Tolia VM, Chan TC, Castillo EM. Preliminary results of initial testing for coronavirus (COVID-19) in the emergency department. *West J Emerg Med* 2020;21:503–6.
- Pulia MS, O'Brien TP, Hou PC, Schuman A, Sambursky R. Multi-tiered screening and diagnosis strategy for COVID-19: a model for sustainable testing capacity in response to pandemic. *Ann Med* 2020;52:207–14.
- Kobia F, Gitaka J. COVID-19: are Africa's diagnostic challenges blunting response effectiveness? *AAS Open Res* 2020;3:4.
- Chan JF, Yip CC, To KK, et al. Improved molecular diagnosis of COVID-19 by the novel, highly sensitive and specific COVID-19-RdRp/Hel real-time reverse transcription-PCR assay validated in vitro and with clinical specimens. *J Clin Microbiol* 2020;58:e00310–20.
- Phelps MA. Pretest probability estimates: a pitfall to the clinical utility of evidence-based medicine? *Acad Emerg Med* 2004;11:692–4.
- West CP, Montori VM, Sampathkumar P. COVID-19 testing: the threat of false-negative results. *Mayo Clin Proc* 2020;95:1127–9.
- Fang Y, Zhang H, Xie J, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* 2020;296:E115–7.
- Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev Mol Diagn* 2020; 20:453–4.
- Winichakoon P, Chaiwarith R, Liwsrisakun C, et al. Negative nasopharyngeal and oropharyngeal swabs do not rule out COVID-19. *J Clin Microbiol* 2020;58:e00297–320.
- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55–63.
- Willett DL, Kannan V, Chu L, et al. SNOMED CT concept hierarchies for sharing definitions of clinical conditions using electronic health record data. *Appl Clin Inform* 2018;9:667–82.
- Henry BM, de Oliveira MH, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. *Clin Chem Lab Med* 2020;58:1021–8.
- Wang Y, Wang Y, Chen Y, Qin Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *J Med Virol* 2020;92:568–76.
- Mayer M. Package 'missRanger': Fast Imputation of Missing Values. Vienna: The Comprehensive R Archive Network; 2019.
- Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2011;28:112–8.
- Harrell FE. *Regression Modeling Strategies*. Berlin: Springer, 2006.
- Wright MN, Ziegler A. ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw* 2017;77:17.
- Chen T, He T, Benesty M, et al. xgboost: Extreme Gradient Boosting. R Package Version 1.0.0.2 ed2020. Vienna: The Comprehensive R Archive Network, 2020.
- Rhoads DD, Cherian SS, Roman K, Stempak LM, Schmotzer CL, Sadri N. Comparison of Abbott ID Now, Diasorin Simplexa, and CDC FDA EUA methods for the detection of SARS-CoV-2 from nasopharyngeal and nasal swabs from individuals diagnosed with COVID-19. *J Clin Microbiol* 2020;58:e00760–820.
- Harrington A, Cox B, Snowdon J, et al. Comparison of Abbott ID Now and Abbott m2000 methods for the detection of SARS-CoV-2 from nasopharyngeal and nasal swabs from symptomatic patients. *J Clin Microbiol* 2020;58:e00798–820.
- Feng C, Huang Z, Wang L, et al. A novel triage tool of artificial intelligence assisted diagnosis aid system for suspected COVID-19 pneumonia in fever clinics. *medRxiv* 2020:2020.03.19.20039099.
- Meng Z, Wang M, Song H, et al. Development and utilization of an intelligent application for aiding COVID-19 diagnosis. *medRxiv* 2020:2020.03.18.20035816.
- Kolachalama VB, Garg PS. Machine learning and medical education. *NPJ Digit Med* 2018;1:54.
- van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014;14:137.
- Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 2018;19:270.