

# Development and Validation of a Clinical Risk Score to Predict Hospitalization Within 30 Days of Coronavirus Disease 2019 Diagnosis

Maya Aboumrada, MPH<sup>b\*</sup>; Gabrielle Zwain, BA<sup>\*</sup>; Jeremy Smith, MPH<sup>\*</sup>; Nabin Neupane, BS<sup>\*</sup>; Ethan Powell, BA<sup>\*</sup>; Brendan Dempsey, MD<sup>\*</sup>; Carolina Reyes, PhD<sup>†</sup>; Sacha Satram, PhD<sup>†</sup>; Yinong Young-Xu, ScD, MA, MS<sup>\*,‡</sup>

---

## ABSTRACT

### Introduction:

Early identification of patients with coronavirus disease 2019 (COVID-19) who are at risk for hospitalization may help to mitigate disease burden by allowing healthcare systems to conduct sufficient resource and logistical planning in the event of case surges. We sought to develop and validate a clinical risk score that uses readily accessible information at testing to predict individualized 30-day hospitalization risk following COVID-19 diagnosis.

### Methods:

We assembled a retrospective cohort of U.S. Veterans Health Administration patients (age  $\geq 18$  years) diagnosed with COVID-19 between March 1, 2020, and December 31, 2020. We screened patient characteristics using Least Absolute Shrinkage and Selection Operator logistic regression and constructed the risk score using characteristics identified as most predictive for hospitalization. Patients diagnosed before November 1, 2020, comprised the development cohort, while those diagnosed on or after November 1, 2020, comprised the validation cohort. We assessed risk score discrimination by calculating the area under the receiver operating characteristic (AUROC) curve and calibration using the Hosmer–Lemeshow (HL) goodness-of-fit test. This study was approved by the Veteran’s Institutional Review Board of Northern New England at the White River Junction Veterans Affairs Medical Center (Reference no.:1473972-1).

### Results:

The development and validation cohorts comprised 11,473 and 12,970 patients, of whom 4,465 (38.9%) and 3,669 (28.3%) were hospitalized, respectively. The independent predictors for hospitalization included in the risk score were increasing age, male sex, non-white race, Hispanic ethnicity, homelessness, nursing home/long-term care residence, unemployed or retired status, fever, fatigue, diarrhea, nausea, cough, diabetes, chronic kidney disease, hypertension, and chronic obstructive pulmonary disease. Model discrimination and calibration was good for the development (AUROC = 0.80; HL *P*-value = .05) and validation (AUROC = 0.80; HL *P*-value = .31) cohorts.

### Conclusions:

The prediction tool developed in this study demonstrated that it could identify patients with COVID-19 who are at risk for hospitalization. This could potentially inform clinicians and policymakers of patients who may benefit most from early treatment interventions and help healthcare systems anticipate capacity surges.

---

## INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious respiratory illness caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).<sup>1,2</sup> As of August 15, 2021, the USA continues to report the highest number of COVID-19 cases (36.6 million) and deaths (618,591) globally.<sup>2</sup> COVID-19 has also contributed to significant health-care capacity challenges, with a cumulative hospitalization

rate of 597 per 100,000 population.<sup>3</sup> Strategies to mitigate this disease and healthcare system burden include policies centered on reducing viral transmission (e.g., social distancing, wearing masks, and testing and tracing), deployment of SARS-CoV-2 vaccines, and increasing both hospital beds and staff.<sup>4,5</sup> Although these efforts have helped to flatten the curve and reduce the acute need for hospital resources, the long-term healthcare capacity requirements remain unclear as restrictions on public gatherings begin to relax and 49% of the U.S. population remains unvaccinated to date.<sup>6–8</sup> Moreover, the increase in severe cases due to new highly infectious SARS-CoV-2 variants (e.g., Delta) could further strain healthcare resources and lead to more hospitalizations and deaths.<sup>8</sup> Therefore, development of predictive tools that can assist with early identification of patients at risk for severe illness that leads to hospitalization is a high priority.<sup>9–12</sup> Such tools may help to guide clinical decision-making and allocation of limited resources, as well as help to identify patients that

---

\*Clinical Epidemiology Program, White River Junction Veterans Affairs Medical Center, White River Junction, VT 05009, USA

<sup>†</sup>Division of Health Economics and Outcomes Research, VIR Biotechnology Inc., San Francisco, CA 94158, USA

<sup>‡</sup>Department of Psychiatry, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA  
doi:<https://doi.org/10.1093/milmed/usab415>

Published by Oxford University Press on behalf of the Association of Military Surgeons of the United States 2021. This work is written by (a) US Government employee(s) and is in the public domain in the US.

may benefit most from early treatment interventions (e.g., monoclonal antibodies).<sup>13–16</sup>

Prediction models of varying quality have been developed for a wide range of COVID-19 outcomes, including hospitalization.<sup>9–12,17</sup> However, a recent systematic review reported that they all have a high or uncertain risk of bias.<sup>12</sup> This may suggest that the performance of these models will be optimistic at best in new samples and could potentially increase the risk of poor patient outcomes owing to insufficient resource and logistical planning at the hospital level.<sup>12</sup> These prediction models were created during the early months of the pandemic and have not been externally validated to date. Moreover, the limited sample sizes and number of outcome events derived from local and/or regional healthcare systems may increase the risk of biasing the model. This may be partially due to differential COVID-19 burden by geography, as well as potential systemic differences in the characteristics of patients requiring hospitalization.<sup>5</sup> Development and validation of a prediction model at the national level is warranted to help improve its generalizability and predictive performance.

Our primary objective was to develop and validate a clinical risk score for predicting 30-day hospitalization among a national cohort of Veterans Health Administration (VHA) patients with COVID-19. As the largest integrated healthcare system in the USA, the VHA was required to respond to COVID-19 in all geographic regions. This offers a unique opportunity to improve our understanding of risk factors for hospitalization given the VHA's diverse operating environments and patient population.

## METHODS

This study was approved by the Veteran's Institutional Review Board of Northern New England at the White River Junction Veterans Affairs Medical Center (Reference no.: 1473972-1). All study procedures were carried out in compliance with federal and institutional ethical guidelines. The requirement to obtain informed consent from study participants was waived as the Institutional Review Board deemed this study to involve no more than a minimal risk to the privacy of individuals. The research data used in this study are not shared publicly to protect the confidentiality of personal health information.

### Data Source and Study Population

The VHA is composed of over 170 medical centers and 1,250 community-based outpatient clinics.<sup>18</sup> It provides comprehensive medical care to more than 9 million veterans, including primary and specialty care.<sup>18</sup> The VHA has an electronic medical record system with a centralized Corporate Data Warehouse, which contains longitudinal information on receipt of all services provided by VHA facilities including outpatient and inpatient visits, pharmacological and non-pharmacological treatments, and laboratory results, as well as patients' sociodemographic and clinical characteristics. We also used publicly available data from The New York Times COVID-19 tracker to evaluate county-level case burden based

on reports from state and local health agencies across the 50 states and District of Columbia.<sup>19</sup> These reports contain standard geographic identifiers (i.e., Federal Information Processing Standards codes), which allowed us to match this information to our dataset.

Our cohort consisted of all VHA-enrolled patients (age  $\geq 18$  years) diagnosed with COVID-19 at a VHA facility between March 1, 2020, and December 31, 2020. We focused on patients diagnosed within the VHA due to lack of testing and diagnosis data from non-VHA facilities. Laboratory-confirmed infections were identified by positive polymerase chain reaction assays for nasal and nasopharyngeal swab specimens. Patients entered the cohort at the time of specimen collection (index date) to account for any lags in testing results. We excluded patients that were hospitalized for more than 1 day at the time of specimen collection to ensure our cohort comprised community-acquired rather than hospital-acquired infections. We excluded any patients residing outside of the 50 states and District of Columbia due to county-level case burden data availability. We also excluded patients that did not have at least one healthcare encounter at a VHA facility within 2 years before COVID-19 diagnosis. We implemented this exclusion criterion to ensure patients were active users of VHA medical services in order to more accurately capture both hospitalization rates and clinical profiles. Lastly, we excluded any patients that received a SARS-CoV-2 vaccine in the pre-testing and post-testing period due to its potential mitigating effect on disease severity.<sup>20</sup>

Our primary outcome of interest was 30-day hospitalization, defined as hospital admissions occurring within 30 days of COVID-19 diagnosis. We followed patients through January 31, 2021, to allow for sufficient assessment time of our outcome. Patients remained in the cohort until the date of hospitalization or end of 30-day follow-up, whichever occurred first.

### Study Variables

We selected our study variables based on current literature describing risk factors for COVID-19 hospitalization and severity.<sup>10,21–23</sup> These include patients' sociodemographic characteristics, COVID-19 symptoms, clinical history, and comorbidities.

We examined county-level case burden and patients' age, sex, race, ethnicity, urbanicity of residence, and VHA priority rating (1-8) at the index date. County-level case burden represents the average, daily case rate per 10,000 persons in each county across the USA for which COVID-19 testing occurred. The VHA priority group served as an indirect proxy for socioeconomic status as it takes into consideration patients' income, financial security, Medicaid eligibility, receipt of VHA assistance benefits (e.g., pension, assisted-living, and Adult Day Care), and capacity for gainful employment.<sup>24–26</sup> It also accounts for health-related factors including severity of service-connected disabilities, and environmental and/or

other exposures.<sup>24-26</sup> Priority group ratings range from 1 to 8, with lower ratings assigned higher priority by the VHA. Priority groups 1-4 include patients with service-connected disabilities rated between 10% and 100%, those who are unemployable as a result of service-related injury, those with a recognized status (e.g., Purple Heart recipient), and/or those who receive pension and/or other assistance benefits from the VHA.<sup>24-26</sup> Priority groups 5-6 include patients with service-connected disabilities rated as 0%, those eligible for Medicaid, and/or those with an annual income below the VHA national income threshold.<sup>24-26</sup> Priority groups 7-8 include patients with an annual income above the VHA national income threshold who agree to pay copayments.<sup>24-26</sup> We examined patients' smoking status and housing status (i.e., homeless, nursing home/long-term care residence, or other housing type) within 2 years before and including the index date. The "other housing type" group included patients that did not have any record of homelessness or long-term care residence during that time. We defined the group this way because we were unable to confidently distinguish between residence types (e.g., apartment, single-family houses, or other) from the address of primary residence recorded in patients' electronic medical records. Lastly, we examined patients' most recent record of employment status. We did not restrict the lookback period for this variable as it often remains static until changes are reported.

We examined the presence of commonly reported mild-to-moderate COVID-19 symptoms within 14 days before and including the index date through a combination of International Classification of Diseases, Tenth Revision (ICD-10) codes and natural language processing of clinician notes in patients' electronic medical records. These included fever, cough, fatigue, nausea, diarrhea, headache, loss of taste or smell, myalgia, and/or sore throat.<sup>27</sup> We focused on mild-to-moderate symptoms due to the biphasic clinical course of COVID-19. Patients may present with mild, moderate, or no symptoms at testing and often return at a later time with severe illness that requires hospitalization.<sup>10,11</sup>

We examined the presence of clinical comorbidities within 2 years before and including the index date, as indicated by the presence of  $\geq 2$  outpatient or  $\geq 1$  inpatient visit(s) with an ICD-10 code for the condition. These included asthma, diabetes, cancer, cerebrovascular disease, chronic obstructive pulmonary disease, congestive heart failure, coronary artery disease, dementia, hypertension, liver disease, myocardial infarction, peripheral vascular disease, and chronic kidney disease. Additionally, we used patients' body mass index (BMI,  $\text{kg}/\text{m}^2$ ) to generate obesity categories based on the Centers for Disease Control and Prevention definition for adult obesity (normal:  $\text{BMI} < 25$ ; overweight:  $\text{BMI} 25-29$ ; obese:  $\text{BMI} \geq 30$ ).<sup>28</sup> We generated a composite immunocompromised category through the presence of immunocompromising illnesses and/or immunosuppressive therapies.<sup>29</sup> Moreover, we calculated patients' illness burden according to the Charlson Comorbidity Index. The Charlson Comorbidity

Index score is a validated, weighted measure that predicts 1-year mortality, with higher scores indicating greater illness burden.<sup>30-32</sup> Lastly, we examined receipt of any influenza vaccination before the index date using Current Procedural Terminology codes given that a few recent studies demonstrated prior vaccination was protective against COVID-related hospitalization.<sup>10,33</sup> The timeframe of assessment for vaccination was between July 1st and December 31st for the 2019-2020 and 2020-2021 influenza seasons.

### Statistical Analysis

We used chi-squared test to compare the proportion of categorical variables and Student's *t*-test to compare the mean with SD of continuous variables between patients that were hospitalized compared to those who were not. We reported the proportion of missing data where applicable.

We used least absolute shrinkage and selection operator (LASSO) logistic regression to screen all study variables and identify important predictors for 30-day hospitalization. This method penalizes the absolute size of regression coefficients by using a tuning parameter ( $\lambda$ ) to shrink the estimates of weaker predictors toward zero so that only the strongest predictors are retained. It also reduces the potential for collinearity of variables and overfitting the model. We used 10-fold cross-validation to select the  $\lambda$  that resulted in the most regularized model while controlling the overall model mean square error within one standard error of the minimum. We subsequently entered the identified set of predictors from this model into a multivariable logistic regression model and constructed the risk score using predictors that remained statistically significant. Data were missing for three categorical variables (i.e., smoking status, employment status, and BMI). We included these "unknown" classifications as indicator variables in the model to preserve our sample size and make use of all available data, as well as assess whether LASSO selection would identify these variables as measures of significant importance.

We identified a total of 24,443 VHA patients with COVID-19 during the study period. We built our model using a development cohort, which comprised 11,473 patients tested for COVID-19 before November 1, 2020. To assess our model's validity and functionality over time, we "prospectively" tested it in a validation cohort composed of 12,970 patients tested for COVID-19 on or after November 1, 2020. We determined the division of patients into development and validation cohorts based on the timeframe that provided us with approximately equal sample sizes, allowing us to more confidently develop and test the model. We used this approach to examine whether the relationship between various risk factors and COVID-19 hospitalization could have changed over time, and to align with methods used in other studies that developed and validated risk scores for COVID-19 outcomes (e.g., hospitalization and/or mortality).<sup>9,10,17,34</sup> We assessed model discrimination in both the development and validation

cohorts by calculating the area under the receiver operating characteristic (AUROC) curve. We also assessed model calibration in both cohorts using the Hosmer–Lemeshow (HL) goodness-of-fit test, which rejects the null hypothesis that the fit is poor if  $P \geq .05$ . Lastly, we calculated sensitivity, specificity, positive predictive value, and negative predictive value at varying cutoffs of predicted risk. All analyses were performed using Stata/MP version 15.1 software (StataCorp, 2015).

### Sensitivity Analysis

In order to assess the potential for bias introduced by the use of temporal training in the primary analysis, we tested our development model's performance in a second validation cohort that represented a random sample comprising 50% ( $N = 12,222/24,443$ ) of all patients with no timeframe restriction.

## RESULTS

### Patient Characteristics

Among the development and primary validation cohorts, a total of 4,465 (38.9%) and 3,669 (28.3%) were hospitalized, respectively. Patient characteristics for both cohorts are presented in Table S1. Among the secondary validation cohort in our sensitivity analysis, a total of 4,125 (33.8%) patients were hospitalized. Patient characteristics for the secondary validation cohort may be found in Table S2.

Hospitalized compared to nonhospitalized patients in the development cohort were older (mean = 68.5, SD = 13.2 vs. mean = 57.3, SD = 15.5) and more likely to be male (94.6% vs. 87.7%). Although both groups were largely white, the hospitalized group had a higher frequency of black patients (36.0% vs. 30.6%). Approximately five times as many hospitalized patients were nursing home/long-term care residents (12.1% vs. 2.4%), and they were more likely to be homeless (8.1% vs. 4.5%), unemployed (45.6% vs. 35.8%), or retired (29.1% vs. 16.4%). Hospitalized patients had a higher frequency of COVID-19 symptoms including fever (54.4% vs. 39.6%) and fatigue (17.1% vs. 6.6%). They also had more comorbidities including hypertension (72.7% vs. 47.0%), diabetes (47.0% vs. 26.4%), chronic obstructive pulmonary disease (24.4% vs. 12.2%), immunocompromised status (27.5% vs. 14.3%), coronary artery disease (25.1% vs. 11.0%), and chronic kidney disease (23.2% vs. 7.2%). Patient characteristics in the validation cohort were largely similar to the development cohort (Table S1). Of note, we observed differences for race between the development (black 36.0%, white 56.7%) and validation (black 25.0%, white 67.9%) cohorts, as well as county-level case burden.

### Predictor Selection and Risk Score Development

The variables identified as predictors for hospitalization by LASSO selection and included in the adjusted logistic model were age, male sex, non-white race, Hispanic ethnicity,

homelessness, nursing home/long-term care residence, unemployed or retired status, fever, fatigue, diarrhea, nausea, cough, diabetes, all stages of chronic kidney disease (1-3 and 4-6), hypertension, and chronic obstructive pulmonary disease (Table I). The strongest predictors for 30-day hospitalization risk were age  $\geq 65$  years, nursing home/long-term care residence, and chronic kidney disease stages 4-6. All variables remained statistically significant and were included in the risk score. The total sum of risk score points corresponds to an individualized, predicted risk of 30-day hospitalization. For example, a patient that is male (2-points), 57 years old (6-points), black (3-points), non-Hispanic (0-points), febrile (5-points), experiencing diarrhea (2-points), and has diabetes (3-points) will have a total risk score point value of 21 and predicted 30-day hospitalization risk between 31% and 40%. The predicted 30-day hospitalization risk associated with varying ranges of total risk score points may be found in Table II. The sensitivity, specificity, positive predictive value, and negative predictive value of the risk score at varying cutoffs of predicted risk may be found in Table III.

### Risk Score Performance

Risk score discrimination in the development cohort was good, with an AUROC of 0.80 (95% CI: 0.78-0.81). The predicted and observed 30-day hospitalization risks were well-aligned, indicating successful calibration (HL  $P$ -value = .05). Risk score performance was identical in the primary validation cohort, with an AUROC of 0.80 (95% CI: 0.78-0.81) and HL  $P$ -value of .31.

From our sensitivity analysis, the development model performed identically in the secondary validation cohort (AUROC = 0.80, 95% CI: 0.78-0.81; HL  $P$ -value = .41).

## DISCUSSION

To the best of our knowledge, this is the first study using a national cohort of U.S. patients to develop and validate a predictive risk score that may provide an individualized, estimated risk of hospitalization within 30 days of COVID-19 diagnosis. Our results demonstrated an elevated hospitalization risk among patients that were 65 years and older, male, non-white, Hispanic, homeless, nursing home/long-term care residents, unemployed, retired, and/or experiencing symptoms of fever, fatigue, diarrhea, nausea, and/or cough. Additionally, our results confirmed that the presence of preexisting medical conditions is predictive of hospitalization risk including diabetes, chronic kidney disease, hypertension, and chronic obstructive pulmonary disease. The predictive performance of our risk score was good, as evidenced by successful discrimination and calibration in both the development and validation cohorts.

There are a few prediction tools available for hospitalization risk among patients with COVID-19.<sup>9-12</sup> However, none of these tools are currently recommended for widespread use in clinical practice due to high risk of bias.<sup>10,12</sup> Jehi et al. developed a risk score calculator for 30-day hospitalization

**TABLE I.** Adjusted Odds Ratios with 95% CIs and Number of Risk Score Points for 30-day Hospitalization, U.S. Veterans Health Administration Development Cohort (March 1, 2020-October 31, 2020)

Characteristic	Odds ratio (95% CI)	Risk score points
<b>Sociodemographic</b>		
Age in years		
≤44	1.00 (reference)	0
45-54	1.56 (1.27-1.93)	4
55-64	2.20 (1.81-2.68)	6
65-74	3.86 (3.18-4.70)	10
≥75	7.50 (6.00-9.37)	16
Male (vs. female)	1.25 (1.03-1.52)	2
Race		
White	1.00 (reference)	0
Other	1.20 (1.02-1.49)	1
Black	1.44 (1.28-1.61)	3
Hispanic (vs. non-Hispanic)	1.54 (1.27-1.87)	3
Housing status		
Nursing home/long-term care	3.93 (3.08-5.03)	11
Homeless	2.16 (1.72-2.70)	6
Other housing type	1.00 (reference)	0
Employment Status		
Employed	1.00 (reference)	0
Retired	1.29 (1.11-1.50)	2
Unemployed	1.58 (1.41-1.78)	4
<b>COVID-19 Symptoms</b>		
Fever	2.00 (1.79-2.22)	5
Cough	1.14 (1.03-1.28)	1
Fatigue	2.07 (1.75-2.44)	6
Diarrhea	1.36 (1.19-1.55)	2
Nausea	1.45 (1.24-1.69)	3
<b>Clinical Comorbidities</b>		
Diabetes	1.48 (1.32-1.65)	3
Hypertension	1.22 (1.08-1.37)	2
Chronic obstructive pulmonary disease	1.46 (1.27-1.67)	3
Chronic kidney disease		
Stages 1-3	1.48 (1.23-1.79)	3
Stages 4-6	3.68 (2.63-5.16)	10

using a cohort of 4,536 patients diagnosed with COVID-19 between March 8 and June 5, 2020, from one healthcare system in Ohio and Florida.<sup>10</sup> Similar to our findings, the authors reported that increasing age, male sex, non-white race, hypertension, diabetes, fever fatigue, and diarrhea were significant predictors of hospitalization.<sup>10</sup> They also identified other predictors of hospitalization including certain pre-testing laboratory values (e.g., alanine aminotransferase, blood urea nitrogen, chloride, and potassium) and additional COVID-related symptoms (e.g., shortness of breath and vomiting).<sup>10</sup> Our goal was to develop a predictive risk score that can identify at-risk patients early in the disease course to help prevent progression to severe illness. Therefore, we

**TABLE II.** Range of Total Risk Score Points and Associated 30-day Hospitalization Risk following Coronavirus Disease 2019 Diagnosis

Total risk score points	Predicted 30-day hospitalization risk (%)
0-8	≤10
9-14	11-20
15-18	21-30
19-22	31-40
23-25	41-50
26-28	51-60
29-32	61-70
33-37	71-80
38-44	81-90
≥45	91-100

focused on mild rather than severe (e.g., shortness of breath) symptoms given that severe symptoms are an indication for hospitalization in the previously described biphasic clinical course of COVID-19.<sup>11,35</sup> We believe this could potentially improve the utility of our risk score as a meaningful prediction tool and increase opportunities for early intervention. Furthermore, we did not include pre-testing laboratory results as they may not be available at the time of risk assessment, thereby limiting feasibility of use. The discriminatory ability of the risk score developed by Jehi et al. is slightly better than ours, with an AUROC of 0.90 and 0.81 in the development and validation cohorts, respectively.<sup>10</sup> This difference may be expected due to some variation in included predictors, populations, timeframes of assessment, and geographic scope (national vs. regional).

We validated our development model using two different cohorts (i.e., “prospective” and 50% random sample), which allowed us to assess the model’s performance in settings with varying levels of absolute COVID-19 risk. We found that our development model was able to identify patients at risk for hospitalization at the same level of accuracy in each validation cohort. This could suggest that the risk factors for hospitalization identified in this study will likely remain consistent despite possible temporal and/or regional differences. Although this may provide additional support for the validity and clinical utility of our risk score, future work should assess its applicability to other populations including those who are vaccinated for SARS-CoV-2 and/or infected with a SARS-CoV-2 variant.

There are several advantages of the risk score for 30-day COVID-19 hospitalization developed in this study. It may be readily administered by both clinical and nonclinical staff in an outpatient setting, and the information needed to calculate an individual’s risk of severe illness and/or hospitalization is easily attainable and timely (i.e., does not require laboratory testing). Patients may also be able to assess their risk at home and discuss the results with their clinician via telemedicine to avoid further community exposure, as well as unnecessary

**TABLE III.** Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value of the 30-day Hospitalization Risk Score at Varying Cutoffs of Predicted Risk, U.S. Veterans Health Administration Validation Cohort (November 1, 2020, to December 31, 2020)

Predicted risk	Sensitivity	Specificity	Positive predictive value	Negative predictive value
10%	94.5%	34.9%	33.7%	94.8%
30%	66.2%	77.1%	50.3%	86.7%
50%	34.8%	93.1%	63.7%	80.3%
70%	13.7%	98.8%	78.7%	76.6%
90%	2.1%	99.9%	83.3%	74.5%

hospitalization. Additionally, clinical studies have shown that outpatient use of monoclonal antibodies are efficacious for early treatment of patients with mild-to-moderate illness who are at risk for developing severe illness and/or hospitalization, reducing the risk of hospitalization or death by 70%-87%.<sup>13-16</sup> The risk score developed in this study may help to identify and inform policymakers of patients who could benefit from early COVID-19 treatment interventions as more information regarding safety and efficacy becomes available. Lastly, this predictive risk score could potentially be used to help healthcare systems anticipate capacity surges and allocate resources accordingly. Future research could build on this work through development of models that may predict necessary staffing levels and medical supplies, as well as investigate the impact of adequate staffing and treatment capacity on COVID-19 outcomes.

### Limitations

Our study included patients that were tested for SARS-CoV-2 within the VHA healthcare system. It is important to acknowledge that the VHA typically treats a population that is older, predominantly male, clinically complex, and have greater risk behaviors compared to the non-VHA population.<sup>36</sup> Approximately one-third of the total VHA patient population resides in rural locations where telemedicine may play an integral role in care delivery due to travel distance from VHA medical centers and other access barriers.<sup>36,37</sup> Additionally, close to 150,000 veterans are homeless, one-half of which are black or Hispanic.<sup>37</sup> More than 40,000 veterans reside in VHA-funded nursing home/long-term care facilities.<sup>38</sup> During the COVID-19 pandemic, many VHA facilities were closed to routine care, and it is possible that our cohort may be biased toward patients that underwent SARS-CoV-2 testing during a healthcare encounter that could not be postponed. Therefore, patients in this study could potentially be less healthy or reflective of the overall VHA population. Reassuringly, we found that the frequency and type(s) of comorbidities in this study were similar to those reported among hospitalized and non-hospitalized patients with COVID-19 in published literature from the non-VHA population.<sup>10,12</sup> Further research is warranted to determine the generalizability of our findings to the overall VHA and non-VHA population.

Similarly, we limited our analysis to patients that were hospitalized for COVID-19 at a VHA facility. Although we

attempted to more accurately capture hospitalization rates by excluding patients that were not active users of VHA medical services, we cannot rule out the possibility of VHA-tested patients being hospitalized at non-VHA facilities given geographic differences in hospital bed capacity and other resource limitations over time. This could lead to under- or over-reporting of rates for both hospitalization and risk factors examined in this study, potentially biasing the magnitude of hospitalization risk associated with each characteristic of interest.

We attempted to exclude patients that were vaccinated for SARS-CoV-2 due to its potential mitigating effect on disease severity. However, it is possible that our data may not comprehensively capture vaccinations rendered at non-VHA facilities, including vaccines for both SARS-CoV-2 and seasonal influenza. Many VHA patients utilize non-VHA facilities for annual flu shots, which are not routinely recorded in patients' electronic medical record. This often requires supplementation of VHA data with outside sources (e.g., Centers for Medicaid and Medicare Services) that were not available to us at the time of this study. Therefore, it is likely that the rate of influenza vaccination reported in this study is underestimated. On the other hand, we believe our SARS-CoV-2 vaccination data may be less susceptible to similar issues given that vaccines were much more readily available in VHA compared to non-VHA facilities during the initial months of vaccine rollout (beginning December 14, 2020).<sup>39</sup> Nevertheless, we are unable to rule out the possibility of patients with non-VHA vaccinations for SARS-CoV-2 being included in this study. Given the timeframe under investigation (March 2020-December 2020), it is unlikely that SARS-CoV-2 vaccinations would meaningfully affect our results.

Although we selected our study variables to align with current literature on risk factors for COVID-19 hospitalization and severity, it is possible we missed clinical characteristics and lifestyle factors that are uniquely relevant to the veteran population and/or affect patients' overall health, including substance abuse disorders.<sup>40</sup> Future work is needed to externally validate our findings on risk factors for hospitalization in the general population, as well as among particular sub-populations of interest (e.g., immunocompromised).

### CONCLUSION

Early identification of patients with COVID-19 who are at high risk for hospitalization may help prevent progression to

severe illness, assist with clinical decision-making, and alleviate disease burden on a healthcare system with finite resources and capacity. We developed a well-performing prediction risk tool that can be easily implemented in the outpatient setting to assist with this process.

### ACKNOWLEDGMENT

None declared.

### SUPPLEMENTARY MATERIAL

Supplementary material is available at *Military Medicine* online.

### FUNDING

This project was funded by VIR Biotechnology Inc. The study sponsor had no role in study design or conduct; data collection, management, analysis, or interpretation; manuscript preparation, review, or approval; or decision to submit the manuscript for publication.

### CONFLICT OF INTEREST STATEMENT

C.R. and S.S. are employees and shareholders of VIR Biotechnology Inc. Y.Y.X. received funding from VIR Biotechnology Inc. to conduct this study. M.A., G.Z., J.S., N.N., E.P., and B.D. do not have any conflicts of interest to disclose.

### REFERENCES

- Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. *Int J Antimicrob Agents* 2020; 55(3): 105924.
- Centers for Disease Control and Prevention: Coronavirus disease 2019 (COVID-19): cases, data, & surveillance. Available at <https://covid.cdc.gov/covid-data-tracker>, published 2021; accessed August 15, 2021.
- Centers for Disease Control and Prevention: COVIDView: a weekly surveillance summary of US COVID-19 activity. Available at <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covid-view/past-reports/01292021.html>, published 2021; accessed August 15, 2021.
- Veterans Health Administration Office of Emergency Management: COVID-19 response plan. Available at [https://www.va.gov/opa/docs/VHA\\_COVID\\_19\\_03232020\\_vF\\_1.pdf](https://www.va.gov/opa/docs/VHA_COVID_19_03232020_vF_1.pdf), published 2020; accessed June 1, 2020.
- Miller IF, Becker AD, Grenfell BT, Metcalf CJE: Disease and health-care burden of COVID-19 in the United States. *Nat Med* 2020; 26(8): 1212–7.
- Moghadas SM, Shoukat A, Fitzpatrick MC, et al: Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proc Natl Acad Sci USA* 2020; 117(16): 9122–6.
- Lewnard JA, Lo NC: Scientific and ethical basis for social-distancing interventions against COVID-19. *Lancet Infect Dis* 2020; 20(6): 631–3.
- Del Rio C, Malani PN, Omer SB: Confronting the delta variant of SARS-CoV-2, Summer 2021. *JAMA* 2021; 326(11): 1001–2.
- Clift AK, Coupland CA, Keogh RH, et al: Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. *BMJ* 2020; 371: m3731.
- Jehi L, Ji X, Milinovich A, et al: Development and validation of a model for individualized prediction of hospitalization risk in 4,536 patients with COVID-19. *PLoS One* 2020; 15(8): e0237419.
- Sun H, Jain A, Leone MJ, et al: CoVA: an acuity score for outpatient screening that predicts coronavirus disease 2019 prognosis. *J Infect Dis* 2021; 223(1): 38–46.
- Wynants L, Van Calster B, Collins GS, et al: Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020; 369: m1328.
- Chen P, Nirula A, Heller B, et al: SARS-CoV-2 neutralizing antibody LY-CoV555 in outpatients with Covid-19. *N Engl J Med* 2021; 384(3): 229–37.
- Gottlieb RL, Nirula A, Chen P, et al: Effect of bamlanivimab as monotherapy or in combination with etesevimab on viral load in patients with mild to moderate COVID-19: a randomized clinical trial. *JAMA* 2021; 325(7): 632–44.
- Vir Biotechnology Inc.: Vir Biotechnology and GSK announce VIR-7831 reduces hospitalization and risk of death in early treatment of adults with COVID-19. Available at <https://investors.vir.bio/node/7906/pdf>, published 2021; accessed March 20, 2021.
- Weinreich DM, Sivapalasingam S, Norton T, et al: REGN-COV2, a neutralizing antibody cocktail, in outpatients with Covid-19. *N Engl J Med* 2021; 384(3): 238–51.
- King JT Jr, Yoon JS, Rentsch CT, et al: Development and validation of a 30-day mortality index based on pre-existing medical administrative data from 13,323 COVID-19 patients: the Veterans Health Administration COVID-19 (VACO) Index. *PLoS One* 2020; 15(11): e0241825.
- Eibner C, Krull H, Brown KM, et al: Current and projected characteristics and unique health care needs of the patient population served by the Department of Veterans Affairs. *Rand Health Q* 2016; 5(4): 13.
- The New York Times: Coronavirus (Covid-19) data in the United States. Available at <https://github.com/nytimes/covid-19-data>, published 2021; accessed January 1, 2021.
- Centers for Disease Control and Prevention: Benefits of getting a COVID-19 vaccine. Available at <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/vaccine-benefits.html>, published 2021; accessed January 31, 2021.
- Reilev M, Kristensen KB, Pottegård A, et al: Characteristics and predictors of hospitalization and death in the first 11 122 cases with a positive RT-PCR test for SARS-CoV-2 in Denmark: a nationwide cohort. *Int J Epidemiol* 2020; 49(5): 1468–81.
- Liang W, Liang H, Ou L, et al: Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med* 2020; 180(8): 1081–9.
- Izurieta HS, Graham DJ, Jiao Y, et al: Natural history of COVID-19: risk factors for hospitalizations and deaths among > 26 million US Medicare beneficiaries. *J Infect Dis* 2021; 223(6): 945–56.
- Trivedi AN, Jiang L, Silva G, et al: Evaluation of changes in Veterans Affairs Medical Centers' mortality rates after risk adjustment for socioeconomic status. *JAMA Netw Open* 2020; 3(12): e2024345.
- Silva A, Tarlov E, French DD, Huo Z, Martinez RN, Stroupe KT: Veterans Affairs health system enrollment and health care utilization after the Affordable Care Act: initial insights. *Mil Med* 2016; 181(5): 469–75.
- Caroff SN, Leong SH, Ng-Mak D, et al: Socioeconomic disparities and metabolic risk in veterans with serious mental illness. *Community Ment Health J* 2018; 54(6): 725–34.
- Centers for Disease Control and Prevention: ICD-10-CM official coding and reporting guidelines April 1, 2020 through September 30, 2020. Available at <https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf>, published 2020; accessed January 1, 2021.
- Centers for Disease Control and Prevention: Defining adult overweight and obesity. Available at <https://www.cdc.gov/obesity/adult/defining.html>, published 2021; accessed March 10, 2021.
- Centers for Disease Control and Prevention: US influenza vaccine effectiveness network protocol for influenza and other respiratory viruses (including COVID-19). Available at <https://www.cdc.gov/vaccines/covid-19/downloads/us-flu-vaccine-effectiveness-network-protocol-508.pdf>, published 2021; accessed March 1, 2021.

30. Charlson ME, Pompei P, Ales KL, MacKenzie CR: A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; 40(5): 373–83.
31. De Groot V, Beckerman H, Lankhorst GJ, Bouter LM: How to measure comorbidity: a critical review of available methods. *J Clin Epidemiol* 2003; 56(3): 221–9.
32. Quan H, Sundararajan V, Halfon P, et al: Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005; 43(11): 1130–39.
33. Conlon A, Ashur C, Washer L, Eagle KA, Bowman MAH: Impact of the influenza vaccine on COVID-19 infection rates and severity. *Am J Infect Control* 2021; 49(6): 694–700.
34. Aboumradi M, Shiner B, Riblet N, Huizenga H, Neupane N, Young-Xu Y: Trends in COVID-19 cases and clinical management in Veterans Health Administration medical facilities: a national cohort study. *PLoS One* 2021; 16(7): e0246217.
35. Jain V, Yuan J-M: Predictive symptoms and comorbidities for severe COVID-19 and intensive care unit admission: a systematic review and meta-analysis. *Int J Public Health* 2020; 65(5): 533–46.
36. Hoerster KD, Lehavot K, Simpson T, McFall M, Reiber G, Nelson KM: Health and health behavior differences: US Military, veteran, and civilian men. *Am J Prev Med* 2012; 43(5): 483–9.
37. Khadduri J, Culhane D, Cortes A: Veteran homelessness: a supplemental report to the 2009 Annual Homeless Assessment Report to Congress. Available at <https://www.huduser.gov/portal/sites/default/files/pdf/2009AHARVeteransReport.pdf>, published 2020; accessed March 15, 2021.
38. Clowers AN: VA Health Care: VA faces challenges in meeting demand for long-term care. Available at <https://www.gao.gov/assets/gao-20-284.pdf>, published 2020; accessed March 15, 2021.
39. Tarzian AJ, Geppert CM: The Veterans Health Administration approach to COVID-19 vaccine allocation—balancing utility and equity. *Fed Pract* 2021; 38(2): 52–4.
40. Volkow ND: Collision of the COVID-19 and addiction epidemics. *Ann Intern Med* 2020; 173(1): 61–2.