

Development of Severe COVID-19 Adaptive Risk Predictor (SCARP), a Calculator to Predict Severe Disease or Death in Hospitalized Patients With COVID-19

Shannon Wongvibulsin, PhD; Brian T. Garibaldi, MD, MEHP; Annukka A.R. Antar, MD, PhD; Jiyang Wen, BS; Mei-Cheng Wang, PhD; Amita Gupta, MD, MHS; Robert Bollinger, MD, MPH; Yanxun Xu, PhD; Kunbo Wang, MS; Joshua F. Betz, MS; John Muschelli, PhD; Karen Bandeen-Roche, PhD; Scott L. Zeger, PhD; and Matthew L. Robinson, MD

Background: Predicting the clinical trajectory of individual patients hospitalized with coronavirus disease 2019 (COVID-19) is challenging but necessary to inform clinical care. The majority of COVID-19 prognostic tools use only data present upon admission and do not incorporate changes occurring after admission.

Objective: To develop the Severe COVID-19 Adaptive Risk Predictor (SCARP) (https://rsconnect.biostat.jhsph.edu/covid_trajectory/), a novel tool that can provide dynamic risk predictions for progression from moderate disease to severe illness or death in patients with COVID-19 at any time within the first 14 days of their hospitalization.

Design: Retrospective observational cohort study.

Settings: Five hospitals in Maryland and Washington, D.C.

Patients: Patients who were hospitalized between 5 March and 4 December 2020 with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) confirmed by nucleic acid test and symptomatic disease.

Measurements: A clinical registry for patients hospitalized with COVID-19 was the primary data source; data included demographic characteristics, admission source, comorbid conditions, time-varying vital signs, laboratory measurements, and clinical severity. Random forest for survival, longitudinal, and multivariate (RF-SLAM) data analysis was applied to predict the 1-day and 7-day risks for progression to severe disease or death for any given day during the first 14 days of hospitalization.

Results: Among 3163 patients admitted with moderate COVID-19, 228 (7%) became severely ill or died in the next 24 hours; an additional 355 (11%) became severely ill or died in the next 7 days. The area under the receiver-operating characteristic curve (AUC) for 1-day risk predictions for progression to severe disease or death was 0.89 (95% CI, 0.88 to 0.90) and 0.89 (CI, 0.87 to 0.91) during the first and second weeks of hospitalization, respectively. The AUC for 7-day risk predictions for progression to severe disease or death was 0.83 (CI, 0.83 to 0.84) and 0.87 (CI, 0.86 to 0.89) during the first and second weeks of hospitalization, respectively.

Limitation: The SCARP tool was developed by using data from a single health system.

Conclusion: Using the predictive power of RF-SLAM and longitudinal data from more than 3000 patients hospitalized with COVID-19, an interactive tool was developed that rapidly and accurately provides the probability of an individual patient's progression to severe illness or death on the basis of readily available clinical information.

Primary Funding Source: Hopkins inHealth and COVID-19 Administrative Supplement for the HHS Region 3 Treatment Center from the Office of the Assistant Secretary for Preparedness and Response.

Ann Intern Med. doi:10.7326/M20-6754

Annals.org

For author, article, and disclosure information, see end of text.

This article was published at Annals.org on 2 March 2021.

Worldwide cases of coronavirus disease 2019 (COVID-19) continue to increase and are now responsible for over 100 million infections and 2 million deaths (1). COVID-19 stresses the limits of health care systems, even in high-income countries such as the United States, where 2021 began with more than 125 000 patients hospitalized with COVID-19 (2). Clinicians and health care systems urgently need reliable tools to determine which patients hospitalized with COVID-19 are at highest risk for severe disease or death, to ensure that patients are treated in a setting with the appropriate level of care. Such tools can also assist with communication to patients and their families about their prognosis and inform clinical decision making. Large observational cohorts of patients with COVID-19 have reported clinical characteristics that are associated with severe illness and death, but these findings are difficult to translate to the quantification of absolute risk in the context of informing care at the level of individual patients (3-5). Although clinical risk

calculators have been created (6-9), there are currently no patient-level prediction tools in widespread use for COVID-19 because of the limitations of existing prediction tools. These limitations include disregarding clinical information beyond baseline data present upon hospital admission, not accounting for censoring, requiring values for all variables in the risk calculator to be entered to obtain a result, and using limited classification approaches rather than more robust survival analysis methodology.

We sought to develop a novel tool that can provide dynamic risk predictions for progression to severe illness

See also:

Web-Only
Supplement

or death in patients hospitalized with COVID-19. To overcome the shortcomings of current approaches and respond to the clinical demand for robust and individualized prediction of COVID-19 severity, we used a well-curated registry of patients admitted with COVID-19 to a 5-hospital health care system and used a machine learning approach, called random forest for survival, longitudinal, and multivariate (RF-SLAM) data analysis (10), to develop the Severe COVID-19 Adaptive Risk Predictor (SCARP).

METHODS

The data for this study were collected at the 5 hospitals (Johns Hopkins Hospital and Johns Hopkins Bayview Medical Center, Baltimore, Maryland; Howard County General Hospital, Columbia, Maryland; Suburban Hospital, Bethesda, Maryland; and Sibley Memorial Hospital, Washington D.C.) that comprise Johns Hopkins Medicine, a system licensed to operate 2513 beds and 354 intensive care unit beds serving a population of approximately 7 million people. The institutional review boards of the participating hospitals approved this study as minimal risk and waived requirement for informed consent. The methods and reporting of results adhere to TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guidelines (11).

Study Design, Participants, and Data Collection

Johns Hopkins created the JH-CROWN: COVID-19 Precision Medicine Analytic Platform Registry to serve as a comprehensive projection of structured clinical data for patients in the Johns Hopkins Medicine system who test positive for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The JH-CROWN database is constructed directly from the clinical electronic health record (as described in the **Supplement**, available at [Annals.org](#)) and serves as the data source for this study (12). Hospitalized patients diagnosed with COVID-19 who were admitted between 5 March and 4 December 2020 were considered for cohort inclusion. Diagnosis of COVID-19 was defined as the detection of SARS-CoV-2 from any nucleic acid test of any specimen type with an Emergency Use Authorization from the U.S. Food and Drug Administration and an International Classification of Diseases, 10th revision (ICD-10), code indicating the presence of symptomatic disease (**Supplement Table 1**, available at [Annals.org](#)). Given our goal of predicting future risk for severe illness or death, patients who developed severe illness or died within the first 6 hours of hospitalization were excluded. The remaining study cohort included adult patients (aged ≥ 18 years) at risk for severe disease or death after the first 6 hours of hospitalization. Some of the patients in this cohort have been included in other studies (13, 14).

Outcome and Variable Definitions

The JH-CROWN registry includes every vital sign, supplemental oxygen recording, and laboratory measurement that compose clinical flowsheets. In addition, structured demographic information and comorbid

conditions derived from ICD-10 codes were included (**Supplement Table 2**, available at [Annals.org](#)) (15). Disease severity was defined with the World Health Organization (WHO) COVID-19 severity scale, a 10-point ordinal scale ranging from uninfected (0 = no viral RNA detected) and ambulatory mild disease (1 = asymptomatic; 2 = symptomatic, independent; 3 = symptomatic requiring assistance) to hospitalized with moderate disease (4 = room air, 5 = nasal cannula or facemask oxygen), hospitalized with severe disease (6 = high-flow nasal cannula or noninvasive ventilation, 7 = intubation and mechanical ventilation [$\text{PaO}_2/\text{FiO}_2$ ratio >150 or $\text{SpO}_2/\text{FiO}_2$ ratio >200], 8 = intubation and mechanical ventilation [$\text{PaO}_2/\text{FiO}_2$ ratio <150 or $\text{SpO}_2/\text{FiO}_2$ ratio <200] or vasopressor use, 9 = intubation and mechanical ventilation [$\text{PaO}_2/\text{FiO}_2$ ratio <150] and vasopressors, dialysis or extracorporeal membrane oxygenation), and 10 = death. The primary outcome was progression to severe illness or death, according to the WHO COVID-19 score of 6 or greater, which includes patients who have died or are hospitalized with respiratory support by noninvasive ventilation, high-flow nasal cannula, or mechanical ventilation (16). The health care workers recording the data used as variables in this clinical registry were blinded to subsequent patient outcomes.

We considered both fixed baseline clinical information and time-varying covariates as potential predictors of COVID-19 severity. Baseline clinical characteristics included demographic characteristics, admission source (for example, from a nursing home), behavioral risk factors (for example, tobacco use, alcohol use), comorbid conditions, and body mass index. Do-not-resuscitate and do-not-intubate statuses were not included as predictor variables, given their relationship with the outcome of interest in our analysis. Time-varying covariates included vital signs, components of a complete blood count with differential, basic metabolic panel, liver function tests, and markers of inflammation and coagulation (ferritin, lactate dehydrogenase, C-reactive protein, interleukin-6, D-dimer, fibrinogen, and international normalized ratio). All covariates and outcomes were binned into 6-hour intervals to account for the time-varying nature of the patient's health state (10). When a single interval contained multiple observations of the same time-varying covariate, the value that represented the most extreme deviation from the normal range was taken to represent the time-varying covariate for that interval. For example, the highest temperature in a 6-hour interval was recorded, whereas for oxygen saturation, the lowest measurement was recorded. Extreme values corresponding to the past 24 hours were also considered for vital signs. Laboratory measurements corresponding to the most abnormal value in the preceding 24 hours were used for the analysis, with the exception of laboratory measurements that are not typically performed daily; for these laboratory variables, the most abnormal values recorded in the past 72 hours were used. Ratios of common laboratory values were calculated and were considered as potential predictor variables. The 72-hour trajectory of vital signs and common laboratory values were reported as the slope calculated by fitting a linear regression model to the data

available for the preceding 72 hours. For the analysis, missing data were handled in an adaptive manner during the tree construction process where imputed values for missing data are randomly drawn from the nonmissing data as the trees are constructed in the random forest algorithm (17). Further details regarding variable selection and methods for addressing missing data are found in the **Supplement**.

Statistical Analysis

We used the machine learning approach RF-SLAM to develop SCARP. A total of 105 variables were considered as inputs into the RF-SLAM predictive algorithm as potential predictors. By learning from data, RF-SLAM allows for the creation of a risk calculator to predict an individual's risk for developing severe illness or death in patients hospitalized with COVID-19. More specifically, RF-SLAM is a method that builds on the concept of decision trees for risk stratification and extends the ensemble learning methodology of random forest (a collection of decision trees) to the analysis of right-censored survival data with time-varying covariates (10). Risk predictions are determined from the ensemble hazard rates from the Bayes estimate of the event rate. The analyses were conducted by using a modification of the `randomForestSRC` package in the R statistical software (R Foundation) (17-19).

Model Performance

Model performance for the predicted risk for progression to severe illness or death in the next 1 day or 7 days was assessed in 3 ways: 1) estimates of cross-validated performance for predictions developed by using data across all hospital sites between 5 March and 4 December 2020; 2) hospital-specific, cross-validation of performance; and 3) prospective validation, where data from 5 March to 4 July 2020 were used for training and data from 5 July to 4 December 2020 were used to evaluate performance. Model discrimination was evaluated with the cross-validated time-varying area under the receiver-operating characteristic curve (AUC) (20). Model calibration was evaluated with the decile method, in which predicted risks are grouped into 10 deciles and plotted along with loess-based calibration curves (21). Additional details are found in the **Supplement**.

Interpretability: Variable Importance and Summary Trees

Although random forests are notable for their impressive predictive ability, they have minimal interpretability. This often limits their adoption in clinical practice because of the lack of ability to communicate how the predictions were generated (22). To provide an interpretable visual display as a summary of the RF-SLAM predictions, we created summary regression trees for the 1-day and 7-day predictions as simplified visualizations of the algorithm. In addition, although random forests can handle a large number of predictor variables, inherent variable selection is performed during the tree building process. Each of the many trees that compose the ensemble model (forest) uses only a small subset of

available variables. We report the importance of each individual variable to the model by quantifying the percentage of trees that use the variable. To visualize the relationship between each top predictor variable and the predicted risk, we used a dependence plot. This is a scatter plot of the predictions against the predictor variable along with a smooth curve to show the dependence.

Predictor Variables and SCARP

The predictor variables identified in the 1-day and 7-day summary regression trees were included as part of the dynamic web-based risk calculator. The calculator was designed to dynamically provide progressively accurate estimates of the RF-SLAM predictions as users input additional variables. The order of variable input was determined by following the path an individual patient traverses in a combined summary regression tree of 1-day and 7-day predictions to maximize the improvement in prediction accuracy for each variable input. A collection of regression trees was created to best summarize the 1-day and 7-day RF-SLAM predictions for each possible sequence of variable inputs. The interactive, online SCARP tool was developed by using the `shiny`, `shinyjs`, and `shinydashboard` R packages. All analyses were conducted by using the R statistical software, version 3.6.2.

Role of the Funding Source

The data utilized were part of the JH-CROWN: The COVID PMAP Registry, which is based on the contribution of many patients and clinicians and is funded by Hopkins inHealth, the Johns Hopkins Precision Medicine Program. Drs. Garibaldi, Muschelli, Robinson, and Gupta received funding from the COVID-19 Administrative **Supplement** for the HHS Region 3 Treatment Center from the Office of the Assistant Secretary for Preparedness and Response. The funders had no role in the design, analysis, or conduct of the study or in the decision to submit the manuscript for publication.

RESULTS

There were 3494 adults admitted between 5 March and 4 December 2020 who tested positive for SARS-CoV-2 by a nucleic acid test and had an ICD-10 code indicative of symptomatic disease. In the first 6 hours of hospitalization, 331 patients developed severe COVID-19; of these, 13 died. The median age of the cohort was 61 years (interquartile range, 46 to 74 years), and 13% were admitted from nursing homes (**Table 1**). Patients who developed severe illness in the first 6 to 24 hours had an elevated respiratory rate, low SpO₂:FiO₂ ratio, and elevated C-reactive protein level. There remained 3163 patients at risk for severe illness or death after the first 6 hours of hospitalization (**Appendix Figure 1** and **Supplement Figure 1**, available at [Annals.org](https://annals.org)).

Completeness of laboratory data varied among individual laboratory tests (**Table 1**). Components of the complete blood count and basic metabolic panel were the most frequently reported laboratory results (**Supplement Figure 2**, available at [Annals.org](https://annals.org)). At least 1 measurement of inflammatory markers in the first 7 days of hospitalization

Table 1. Baseline Clinical Characteristics of Patients, by Disease Trajectory*

Clinical Characteristic	All Patients (n = 3494)	Patients With Missing Data, n (%)	Progression to Severe Illness or Death			No Progression to Severe Illness or Death in > 14 d (n = 2524)
			Within First 6 h (n = 331)	Between 6 h and 1 d (n = 228)	During Days 1 to 7 (n = 354)	
Age, y	61 (46-74)	0 (0)	65 (53-75)	63 (52-73)	67 (56-81)	58 (43-72)
BMI, kg/m ²	29 (24.7-34.4)	391 (11)	29.6 (25.1-34.9)	28.6 (25.7-34.5)	29.8 (25.5-35.8)	28.9 (24.6-34.3)
Male, n (%)	1778 (51)	0 (0)	200/1778	128/1778	195/1778	1223/1778
Admission from skilled nursing facility†	450 (13)	5 (0)	72/450	23/450	62/450	260/450
Former smoker†	724 (21)	24 (1)	76/724	56/724	97/724	478/724
Current smoker†	360 (10)	24 (1)	33/360	15/360	30/360	278/360
Race/ethnicity†						
Non-Hispanic White	1082 (31)	0 (0)	96/1082	65/1082	125/1082	764/1082
Black	1250 (36)	0 (0)	112/1250	75/1250	127/1082	917/1082
Hispanic	809 (23)	0 (0)	84/809	59/809	62/809	600/809
Asian	181 (5)	0 (0)	23/181	14/181	18/181	125 (5)
Other/unknown	172 (5)	0 (0)	16/172	15/172	22/172	118/172
Mean arterial pressure, mm Hg	76.3 (68.7-84.7)	24 (1)	66 (57.3-74.3)	69.7 (62.3-78.1)	74.7 (67.3-83)	77.7 (71-85.3)
Pulse, beats/min	96 (85-108)	23 (1)	109 (94.5-126)	106 (94-120.2)	99 (89-110)	94 (83-104)
Respiratory rate, breaths/min	22 (20-30)	32 (1)	36 (30-42)	37 (31-43.2)	26 (20-32)	20 (20-25)
Temperature, °C	37.5 (37-38.2)	46 (1)	37.7 (37.2-38.6)	37.7 (37.1-38.6)	37.9 (37.2-38.6)	37.4 (37-38.1)
SpO ₂ :FiO ₂ ratio	339 (252-470)	34 (1)	100 (90-151.5)	103 (91-155)	272 (211-334.2)	422 (324-475)
Absolute lymphocyte count, × 10 ⁹ cells/L	0.9 (0.6-1.4)	125 (4)	0.8 (0.5-1.1)	0.7 (0.5-1.1)	0.8 (0.5-1.1)	1 (0.7-1.4)
Hemoglobin level, g/L	123 (108-136)	35 (1)	117 (102-132)	122 (108-138)	123 (107-135)	123 (109-135)
Red cell distribution width, %	13.6 (12.9-14.9)	36 (1)	14.3 (13.2-15.7)	13.9 (13-15.5)	13.9 (13-15.2)	13.5 (12.8-14.7)
Leukocyte count, × 10 ⁹ cells/L	7 (5.1-9.7)	35 (1)	10.5 (7.5-14.6)	9.1 (6.7-12.2)	6.9 (5.1-9.3)	6.6 (4.9-8.9)
BUN level						
mg/dL	16 (11-27)	41 (1)	25 (15-41)	21 (14.8-35)	20 (13-34)	15 (10-24)
mmol/L	5.7 (3.9-9.6)	41 (1)	8.9 (5.4-14.6)	7.5 (5.3-12.5)	7.1 (4.6-12.1)	5.4 (3.6-8.6)
Bicarbonate level, mmol/L	25 (23-27)	41 (1)	24 (21-27)	24 (21-27)	25 (22-27)	25 (23-27)
GFR, mL/min per 1.73 m ²	77 (47-99)	46 (1)	59 (32-87.8)	64 (36.8-87.2)	65 (39-91)	80 (51-102)
Albumin level, g/L	0.4 (0.3-0.4)	109 (3)	0.3 (0.3-0.3)	0.3 (0.3-0.4)	0.3 (0.3-0.4)	0.4 (0.3-0.4)
ALT level, U/L	29 (18-49)	133 (4)	39 (22.8-64)	37 (22-64)	30 (20-45)	28 (18-46)
Total bilirubin level						
mg/dL	0.5 (0.3-0.7)	111 (3)	0.6 (0.4-0.9)	0.6 (0.4-0.8)	0.5 (0.4-0.7)	0.5 (0.3-0.7)
μmol/L	8.6 (5.1-12)	111 (3)	10.3 (6.8-15.4)	10.3 (6.8-13.7)	8.6 (6.8-12)	8.6 (5.1-12)
CRP level, mg/L	66 (26.1-127)	913 (26)	142.5 (87.2-223)	124.5 (74.4-215.3)	94.2 (62-154)	49 (19.1-98)
IL-6 level, pg/mL	37.6 (15.4-83.4)	2409 (69)	96.4 (44-275.8)	96.1 (44.6-189.8)	55.2 (27.3-114)	26.8 (11.7-53.5)
D-dimer level, nmol/L	5 (2.8-10.4)	853 (24)	9.4 (5.1-21.9)	6.6 (4.1-12.9)	5.3 (3.3-9.4)	4.4 (2.6-8.8)
Procalcitonin level, ng/mL	0.2 (0-0.4)	1917 (55)	0.5 (0.2-1.8)	0.3 (0.1-0.9)	0.2 (0.1-0.5)	0.1 (0-0.2)
Fibrinogen level, g/L	5 (4-6.2)	2596 (74)	6.1 (4.7-7.5)	6 (5-7)	5.5 (4.4-6.4)	4.7 (3.9-5.9)
Ferritin level, μg/L	550.5 (253-1059)	1098 (31)	882 (392-1607)	895 (553.5-1464.5)	718 (419-1337)	470 (203-885.8)
ESR, mm/h	48 (28-72.5)	2939 (84)	66 (39.5-95)	62 (44-89)	50 (40-85)	43 (25-65.5)
Lactate level, mmol/L	0.2 (0.1-0.3)	1962 (56)	0.3 (0.2-0.4)	0.2 (0.2-0.3)	0.2 (0.1-0.2)	0.2 (0.1-0.2)
Pro-BNP level, ng/L	182.5 (45.2-1033.2)	1662 (48)	601 (172-2176)	342.5 (126.5-1917.5)	251 (92.5-1576.5)	127.5 (33.2-644.2)
INR	1.1 (1-1.1)	1399 (40)	1.1 (1-1.2)	1.1 (1-1.2)	1.1 (1-1.2)	1 (1-1.1)

ALT = alanine aminotransferase; BMI = body mass index; BUN = blood urea nitrogen; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate; IL-6 = interleukin-6; INR = international normalized ratio; GFR = glomerular filtration rate; pro-BNP = pro-brain natriuretic peptide.

* Unless otherwise indicated, data are presented as the median (interquartile range).

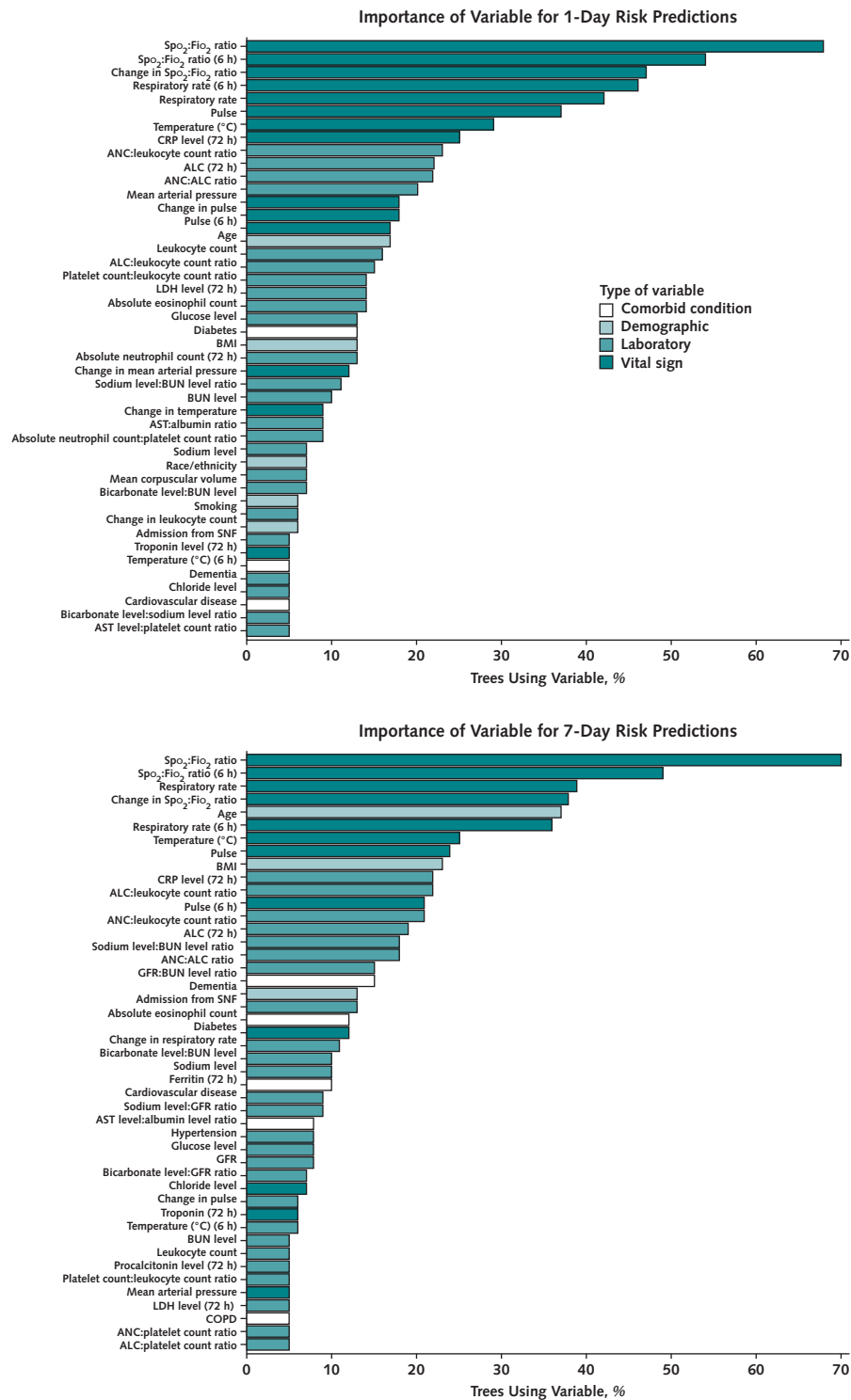
† Data are presented as the number (percentage) of patients or number/number of patients.

before onset of severe disease or death was available for C-reactive protein in 2629 patients (83%), interleukin-6 in 1237 patients (39%), D-dimer in 2678 patients (85%), ferritin in 2492 patients (79%), and lactate dehydrogenase in 1996 patients (63%). Among vital signs, data were missing for 7275 (3%) of 2549206-hour patient intervals. Baseline, time-varying, and derived variables were inputs into the RF-SLAM predictive algorithm. The variable importance, determined by the percentage of trees using the variable, was highest for SpO₂:FiO₂ ratio, respiratory rate, age, and pulse (Figure 1). We then visualized the relationship of the individual variables with the risk predictions (Figure 2).

Performance

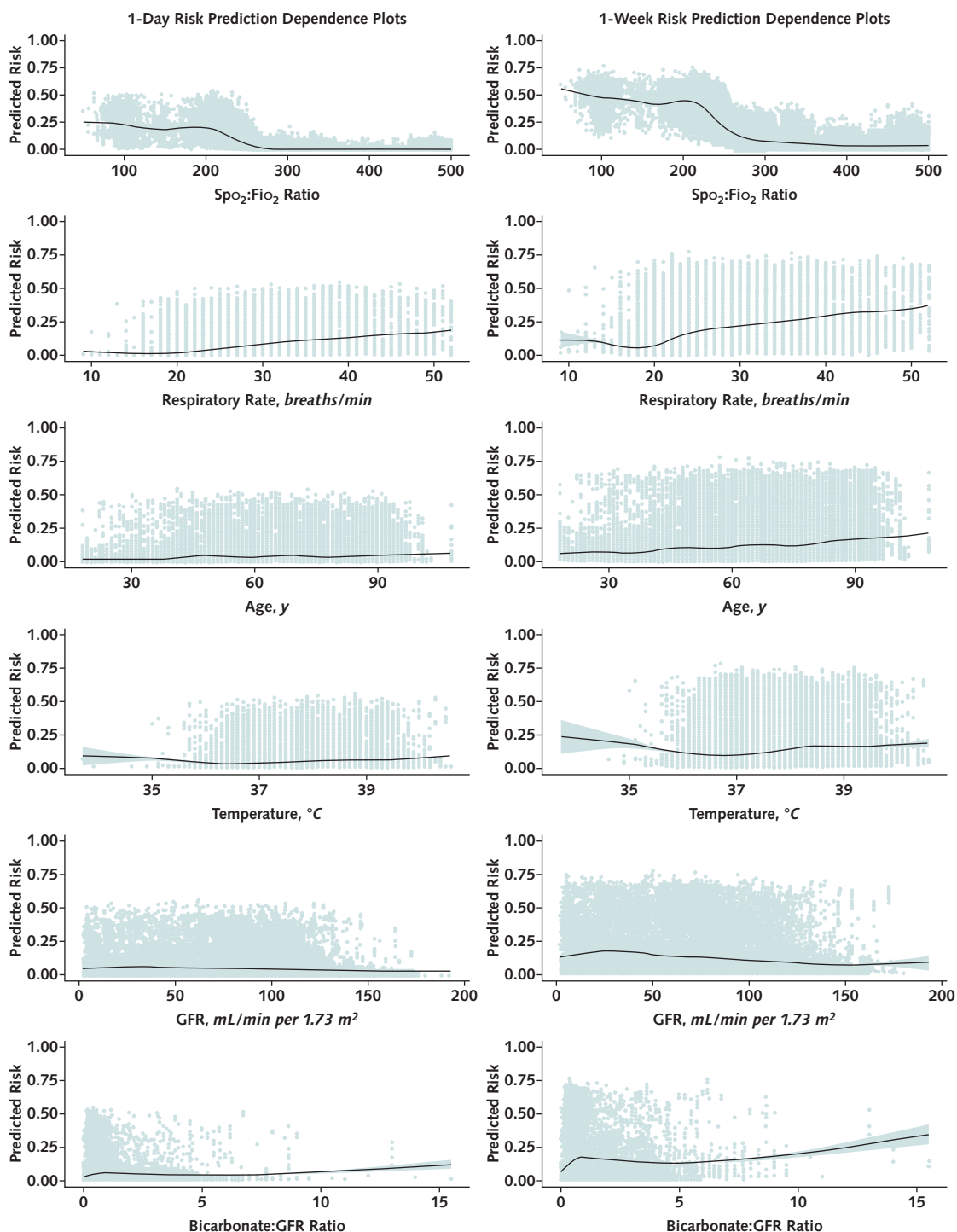
We reported the AUC of the RF-SLAM predictive algorithm separately for the first week and the second week to communicate model performance to clinicians and how performance varies during hospitalization (Table 2). The AUC for 1-day predictions was 0.89 (95% CI, 0.88 to 0.90) during the first week of hospitalization and 0.89 (CI, 0.87 to 0.91) during the second week of hospitalization. The AUC for 7-day predictions was 0.83 (CI, 0.83 to 0.84) during the first week of hospitalization and 0.87 (CI, 0.86 to 0.89) during the second week of hospitalization. Model calibration assessed by risk decile demonstrated that the 1-day and 7-day risk predictions

Figure 1. Variable importance.



The percentage of trees incorporating each of the variables is used as a simple and interpretable measure of variable importance. The variables used by 5% or more of the trees are shown in the plots. Laboratory and vital sign values correspond to values obtained in the past 24 hours unless otherwise specified (for example, "6 h" indicates that the value corresponds to one obtained in the past 6 hours). ALC = absolute lymphocyte count; ANC = absolute neutrophil count; AST = aspartate aminotransferase; BMI = body mass index; BUN = blood urea nitrogen; COPD = chronic obstructive pulmonary disease; CRP = C-reactive protein; GFR: glomerular filtration rate, LDH = lactate dehydrogenase; SNF = skilled nursing facility.

Figure 2. Dependence plots for top predictors of progression to severe illness or death.



Individual data points versus the predicted risk are shown, and the line shows the relationship between the variable and predicted risk.

were well calibrated (Appendix Figure 2, available at Annals.org).

Interpretability

To provide an interpretable overview to clinicians of the logic underlying RF-SLAM predictions, summary

decision trees were created for 1-day and 7-day predictions of progression to severe illness or death. The 1-day risk prediction summary tree captured 89% of the variance of the 1-day RF-SLAM risk predictions and the 7-day risk prediction summary tree captured 90% of the variance of the 7-day RF-SLAM risk predictions of severe

illness or death (Appendix Figure 3, available at [Annals.org](https://annals.org)).

Dynamic Risk Score and Web Application Risk Predictor

For the web-based tool, summary regression trees fitted to the RF-SLAM predictions included at least 15 variables. However, risk prediction for an individual patient requires input of only the variables in the direct path from the start of the summary tree to its terminal leaf. Therefore, for any patient, only a maximum of 8 variables would be required as inputs. The web-based risk prediction calculator SCARP (https://rsconnect.biostat.jhsph.edu/covid_trajectory/) uses adaptive data entry, whereby upon entry of a variable, the algorithm prompts the user for the next variable that best increases the performance of the prediction for that particular patient. This allows users to input a limited number of variables but still benefit from the predictive power of the random forest, which incorporated 105 variables. Detailed instructions for using SCARP are provided in Supplement Figure 3 to 5 (available at [Annals.org](https://annals.org)).

DISCUSSION

We developed SCARP, a novel risk calculator that provides clinically meaningful predictions of whether patients hospitalized with COVID-19 will progress to severe illness or death on the basis of variables that are readily available to treating clinicians. The purpose of this tool is to assist frontline clinicians with real-time prognostication of an individual patient's 1-day and 7-day risk for developing severe illness or death. This tool can inform decisions regarding appropriate level of care and use of scarce hospital resources and can assist clinicians with conversations with patients and family members about a patient's prognosis. Our novel risk calculator was developed with rigorous methodology, including the use of highly granular clinical observations, time-dependent covariates, and discrete time survival analysis in a registry from a diverse 5-hospital health care system. We found that readily available clinical information, such as pulse oximetry, oxygen supplementation, respiratory rate, and pulse, and their trends are highly predictive of progression to severe illness or death. The resulting risk calculator allows clinicians to input a sparse set of

adaptively chosen variables for an individual patient into a web-based tool to obtain an accurate probability of progression to severe disease or death in time frames that are highly relevant to patient care.

Knowledge of the immediate risk for progression to severe COVID-19 within the next 24 hours is important to ensure that patients have access to life-saving care and therapeutics. Some patients with moderate COVID-19 progress rapidly from mild hypoxia to respiratory failure requiring intubation (23). As the COVID-19 pandemic has spread from large cities in high-income countries with advanced tertiary care centers to every region of the United States as well as to low- and middle-income countries, patients with COVID-19 are more likely to receive care in settings without access to advanced supportive care. Mortality from COVID-19 is higher in hospitals with fewer intensive care unit beds and in poorer neighborhoods. Accurate prediction of the risk for clinical deterioration in the next 24 hours may guide transfer of patients to the most appropriate setting within an individual hospital or between hospitals in larger health care systems.

Most patients with COVID-19, even those who are hospitalized, recover with only supportive care. For patients with low risk for progression to severe illness or death, the benefit of unproven therapeutics may be limited. Some emerging therapeutics for COVID-19 are most efficacious when administered early in the disease course (24). Clinicians therefore face the challenge of striving to prescribe therapeutics as early as possible, but given resource limitations, may only do so for a subset of patients with COVID-19 (23, 25). The ability to identify the minority of patients with the greatest risk for progression to severe illness or death will allow clinicians to make informed decisions regarding the use of therapeutics and provide clinical trialists with tools to target patients who may benefit most from novel interventions.

Numerous clinical prediction tools for COVID-19 have already been described, but these have several limitations (6, 26–28). Liang and colleagues (6, 26) created calculators to predict severe illness or death that achieved AUCs of 0.88 by using the least absolute shrinkage and selection operator (LASSO) and logistic regression to develop a predictive risk score (called COVID-GRAM) and 0.91 by using a deep learning survival Cox model. However, these methods depend on baseline variables and have limited tolerance to missing data.

Table 2. AUCs for 1-Day and 1-Week Risk Predictions*

Validation approach	1-Day Risk Predictions		1-Week Risk Predictions	
	AUC for First Week	AUC for Second Week	AUC for First Week	AUC for Second Week
All 5 hospital sites and time	0.89 (0.88–0.90)	0.89 (0.87–0.91)	0.83 (0.83–0.84)	0.87 (0.86–0.89)
Leave out hospital 2	0.90 (0.88–0.92)	0.96 (0.95–0.98)	0.82 (0.80–0.83)	0.93 (0.91–0.95)
Leave out hospital 3	0.92 (0.90–0.93)	0.83 (0.75–0.91)	0.84 (0.83–0.86)	0.79 (0.75–0.83)
Leave out hospital 4	0.86 (0.84–0.87)	0.79 (0.76–0.83)	0.83 (0.82–0.84)	0.87 (0.85–0.89)
Leave out hospital 5	0.91 (0.89–0.93)	0.88 (0.80–0.95)	0.83 (0.81–0.85)	0.89 (0.83–0.94)
Prospective validation	0.89 (0.87–0.90)	0.90 (0.85–0.95)	0.83 (0.83–0.84)	0.78 (0.73–0.82)

AUC = area under the receiver-operating characteristic curve.

* The first row corresponds to the estimate of the cross-validated AUC including data from all hospital sites. The following rows correspond to hospital-specific performance, from leave-one-hospital-out cross validation. For the hospital-site cross validation, the main hospital (site 1) was always included in the training data. The last row corresponds to prospective validation in which data from 5 March to 4 July 2020 were used for training and data from 5 July to 4 December 2020 were used for validation. Further details are provided in the Supplement (available at [Annals.org](https://annals.org)).

Furthermore, when applied to a large United Kingdom cohort, COVID-GRAM did not outperform CURB-65, a simple pneumonia severity score (29). The Coronavirus Clinical Characterisation Consortium (4C) mortality score used observations from more than 35 000 patients in the United Kingdom to derive a score that predicts inpatient mortality with an AUC of 0.77. However, the 4C mortality score only provides predictions at hospital admission, and values for all variables must be available for the calculator to function, including C-reactive protein, a laboratory value that is not always available (29). The Northwell COVID-19 Survival Calculator derived from patients in New York has a stable AUC of approximately 0.86 over the first 10 days of hospitalization to predict 7-day mortality, but it also requires input of all variables, including components of a complete blood cell count with differential, which are not always readily available (27). Our prior tool provides accurate prediction of progression to severe disease or death at the time of admission, but it requires input of 23 variables including symptoms, Charlson Comorbidity Index score, and a full suite of laboratory values with limited tolerance for missingness and is applicable only on the day of admission (8). These existing tools do not take advantage of the predictive potential of time-dependent covariates, including updated measurements of vital signs and the changes in these values, which are among the most important predictors we found for progression to severe illness or death.

The SCARP tool substantially advances the performance and reliability of clinical prediction of COVID-19 severity by using time-varying covariates, highly granular clinical information, and robust survival analysis methods. It is designed to provide interpretable and personalized risk prediction for severe disease or death in patients hospitalized with COVID-19 at any time in the first 2 weeks of their hospitalization. The tool is adaptive and interactive in that it allows users to sequentially input variables to increase the usability and interpretability of clinical prediction. The prediction windows for this tool were based on the clinical demand for determining the trajectory of hospitalized patients at an individual level to inform key decisions regarding clinical care in real time. Clinical features measured after or concomitant with the onset of severe illness do not contribute to prediction; for that reason, we chose to exclude patients who became severely ill or died within the first 6 hours of presentation because their presence would probably unfairly favor the model performance. To further aid in the transparency and interpretability of the predictions from SCARP, we use summary trees to provide visual representations underlying the predictions and further understanding of how the estimation of risk is influenced by each additional variable.

Our study has limitations. Although the JH-CROWN registry includes 5 hospitals serving a mix of urban and suburban populations, the hospitals are all members of a single health system, which may limit the generalizability of SCARP to patients in other health systems. However, as in our previous analyses (8), we observed limited heterogeneity in performance across the hospital sites in this study, suggesting that our tool has the potential to be more broadly applicable to other hospitals. By design, the focus of this work is on the prediction of

severe illness or death and is therefore not intended to predict the trajectory of patients who are already severely ill and only predicts outcomes up to a 1-week time frame in the future. Because almost all patients who develop severe illness or die do so in the first 2 weeks of their hospital admission (8), we believe that a 1-week forward-looking predictor will adequately predict progression of disease within a clinically useful timeframe. The cohort is observational and encompasses a period of more than 9 months, during which supportive care standards evolved, treatment options expanded, and clinical trials enrolled participants. As treatment options for COVID-19 expand and outcomes improve, the risk probabilities reported here may overstate the risk for progression to severe disease or death in future patients. We are committed to preserving the long-term utility of this risk calculator by continuously updating the tool and reporting its prospective performance as treatment options for COVID-19 expand and clinical outcomes improve.

In conclusion, we developed SCARP, a novel, easy-to-use clinical prediction tool that adaptively and dynamically reports the probability that an individual patient with COVID-19 will progress to severe illness or death. The tool uses readily available clinical information and has undergone internal and temporal validation. Further studies with national-level, external data sets have been planned for larger-scale validation and generalizability. In addition, work is under way to integrate a simplified version of SCARP into the electronic medical record and assess its utility in clinical practice. SCARP has the potential to serve as a quantitative tool to help guide clinicians managing patients hospitalized with COVID-19, whose clinical courses are complex and seemingly unpredictable, and inform hospital operations to best use resources in meeting the ever-changing demand for intensive care.

From Johns Hopkins University School of Medicine, Baltimore, Maryland (S.W., B.T.G., A.A.A., A.G., R.B., M.L.R.); Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (J.W., M.W., J.F.B., J.M., K.B., S.L.Z.); and Johns Hopkins University, Baltimore, Maryland (Y.X., K.W.).

Disclaimer: The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Sherrilyn and Ken Fisher Center for Environmental Infectious Diseases, or Johns Hopkins University School of Medicine.

Financial Support: JH-CROWN received funding from Hopkins inHealth, the Johns Hopkins Precision Medicine Program, the Sherrilyn and Ken Fisher Center for Environmental Infectious Diseases, Division of Infectious Diseases of the Johns Hopkins University School of Medicine. Dr. Wongvibulsin is funded by National Institutes of Health grants F30HL142131 and 5T32GM007309. Drs. Garibaldi, Gupta, and Robinson received funding from the COVID-19 Administrative Supplement for the HHS Region 3 Treatment Center from the Office of the Assistant Secretary for Preparedness and Response.

Disclosures: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M20-6754.

Reproducible Research Statement: *Study protocol:* Available in the Supplement (available at [Annals.org](https://www.annals.org)). *Statistical code:* Available at <https://github.com/JH-SCARP/SCARP> and from Dr. Robinson (e-mail, mrobin85@jhmi.edu). *Data set:* Individual patient-level data are not available.

Corresponding Author: Matthew L. Robinson, MD, Center for Clinical Global Health Education, Johns Hopkins School of Medicine, 600 North Wolfe Street, Phipps 521, Baltimore, MD 21287; e-mail, mrobin85@jhmi.edu.

Current author addresses and author contributions are available at [Annals.org](https://www.annals.org).

References

- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time [Letter]. *Lancet Infect Dis.* 2020;20:533-534. [PMID: 32087114] doi:10.1016/S1473-3099(20)30120-1
- The COVID Tracking Project. The Atlantic. Accessed at <https://covidtracking.com/data/national/hospitalization> on 10 January 2021.
- Richardson S, Hirsch JS, Narasimhan M, et al; the Northwell COVID-19 Research Consortium. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA.* 2020;323:2052-2059. [PMID: 32320003] doi:10.1001/jama.2020.6775
- Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature.* 2020; 584:430-436. [PMID: 32640463] doi:10.1038/s41586-020-2521-4
- Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* 2020;395:1054-1062. [PMID: 32171076] doi:10.1016/S0140-6736(20)30566-3
- Liang W, Liang H, Ou L, et al; China Medical Treatment Expert Group for COVID-19. 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;180:1081-1089. [PMID: 32396163] doi:10.1001/jamainternmed.2020.2033
- Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ.* 2020;369:m1966. [PMID: 32444366] doi:10.1136/bmj.m1966
- Garibaldi BT, Fiksel J, Muschelli J, et al. Patient trajectories among persons hospitalized for COVID-19: a cohort study. *Ann Intern Med.* 2021;174:33-41. doi:10.7326/M20-3905
- Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of Covid-19 infection: systematic review and critical appraisal. *BMJ.* 2020;369:m1328. [PMID: 32265220] doi:10.1136/bmj.m1328
- Wongvibulsin S, Wu KC, Zeger SL. Clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis. *BMC Med Res Methodol.* 2019;20:1. [PMID: 31888507] doi:10.1186/s12874-019-0863-0
- Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162:55-63. doi:10.7326/M14-0697
- Johns Hopkins Institute for Clinical and Translational Research. COVID-19 Precision Medicine Analytics Platform Registry (JH-CROWN). Accessed at <https://ictr.johnshopkins.edu/coronavirus/jh-crown/> on 10 January 2021.
- Ignatius EH, Wang K, Karaba A, et al. Tocilizumab for the treatment of COVID-19 among hospitalized patients: a matched retrospective cohort analysis. *Open Forum Infect Dis.* 2021;8:ofaa598. [PMID: 33537364] doi:10.1093/ofid/ofaa598
- Andersen KM, Mehta HB, Palamuttam N, et al. Association between chronic use of immunosuppressive drugs and clinical outcomes from coronavirus disease 2019 (COVID-19) hospitalization: a retrospective cohort study in a large US health system. *Clin Infect Dis.* 2021. [PMID: 33410884] doi:10.1093/cid/ciaa1488
- Gasparini A. comorbidity: an R package for computing comorbidity scores. *J Open Source Softw.* 2018;3:648.
- WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection. A minimal common outcome measure set for COVID-19 clinical research. *Lancet Infect Dis.* 2020;20:e192-e197. [PMID: 32539990] doi:10.1016/S1473-3099(20)30483-7
- Ishwaran H, Kogalur UB, Blackstone EH, et al. Random survival forests. *Ann Appl Stat.* 2008;2:841-60.
- Ishwaran H, Kogalur UB. Random survival forests for R. *R News.* 2007;7:25-31.
- Ishwaran H, Kogalur UB. Fast unified random forests for survival, regression, and classification (RF-SRC). R package, version 2.9.3. 2020. Accessed at <https://cran.r-project.org/package=randomForestSRC> on 19 December 2020.
- Bansal A, Heagerty PJ. A comparison of landmark methods and time-dependent ROC methods to evaluate the time-varying performance of prognostic markers for survival outcomes. *Diagn Progn Res.* 2019;3:14. [PMID: 31367681] doi:10.1186/s41512-019-0057-6
- Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med.* 2014;33:517-35. [PMID: 24002997] doi:10.1002/sim.5941
- Wongvibulsin S, Wu KC, Zeger SL. Improving clinical translation of machine learning approaches through clinician-tailored visual displays of black box algorithms: development and validation. *JMIR Med Inform.* 2020;8:e15791. [PMID: 32515746] doi:10.2196/15791
- Gupta S, Hayek SS, Wang W, et al; STOP-COVID Investigators. Factors associated with death in critically ill patients with coronavirus disease 2019 in the US. *JAMA Intern Med.* 2020. [PMID: 32667668] doi:10.1001/jamainternmed.2020.3596
- Joyner MJ, Senefeld JW, Klassen SA, et al. Effect of convalescent plasma on mortality among hospitalized patients with COVID-19: initial three-month experience. medRxiv. Preprint posted online 12 August 2020. doi:10.1101/2020.08.12.20169359
- Adhikari S, Pantaleo NP, Feldman JM, et al. Assessment of community-level disparities in coronavirus disease 2019 (COVID-19) infections and deaths in large US metropolitan areas. *JAMA Netw Open.* 2020;3:e2016938. [PMID: 32721027] doi:10.1001/jamanetworkopen.2020.16938
- Liang W, Yao J, Chen A, et al. Early triage of critically ill COVID-19 patients using deep learning. *Nat Commun.* 2020;11:3543. [PMID: 32669540] doi:10.1038/s41467-020-17280-8
- Levy TJ, Richardson S, Coppa K, et al. Development and validation of a survival calculator for hospitalized patients with COVID-19. medRxiv. Preprint posted online 2 June 2020. doi:10.1101/2020.04.22.20075416
- Xiao LS, Zhang WF, Gong MC, et al. Development and validation of the HNC-LL score for predicting the severity of coronavirus disease 2019. *EBioMedicine.* 2020;57:102880. [PMID: 32645614] doi:10.1016/j.ebiom.2020.102880
- Knight SR, Ho A, Pius R, et al; ISARIC4C investigators. Risk stratification of patients admitted to hospital with Covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. *BMJ.* 2020;370:m3339. [PMID: 32907855] doi:10.1136/bmj.m3339

Current Author Addresses: Dr. Wongvibulsin: Medical Scientist Training Program, Johns Hopkins University School of Medicine, 1830 East Monument Street, Suite 2-300, Baltimore, MD 21205.

Dr. Garibaldi: Division of Pulmonary and Critical Care, Johns Hopkins University School of Medicine, 1830 East Monument Street, Fifth Floor, Baltimore, MD 21205.

Dr. Antar: Division of Infectious Diseases, Miller Research Building, Room 869, 733 North Broadway, Baltimore, MD 21205.

Drs. Wang, Muschelli, Bandeen-Roche, and Zeger; Mr. Wen; and Mr. Betz: Division of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205-2179.

Drs. Bollinger, Gupta, and Robinson: Center for Clinical Global Health Education, Johns Hopkins School of Medicine, 600 North Wolfe Street, Phipps 521, Baltimore, MD 21287.

Dr. Xu and Mr. Wang: Department of Applied Mathematics and Statistics, Johns Hopkins University, 3300 North Charles Street, Baltimore, MD 21218.

Author Contributions: Conception and design: S. Wongvibulsin, R. Bollinger, S.L. Zeger, M.L. Robinson.

Analysis and interpretation of the data: S. Wongvibulsin, B.T. Garibaldi, A.A.R. Antar, A. Gupta, R. Bollinger, Y. Xu, J. Muschelli, S.L. Zeger, M.L. Robinson.

Drafting of the article: S. Wongvibulsin, B.T. Garibaldi, A.A.R. Antar, A. Gupta, R. Bollinger, J.F. Betz, S.L. Zeger, M.L. Robinson.

Critical revision for important intellectual content: S. Wongvibulsin, B.T. Garibaldi, A.A.R. Antar, J. Wen, A. Gupta, R. Bollinger, J.F. Betz, K. Bandeen-Roche, S.L. Zeger, M.L. Robinson.

Final approval of the article: S. Wongvibulsin, B.T. Garibaldi, A.A.R. Antar, J. Wen, M.C. Wang, A. Gupta, R. Bollinger, Y. Xu, K. Wang, J.F. Betz, J. Muschelli, K. Bandeen-Roche, S.L. Zeger, M.L. Robinson.

Provision of study materials or patients: B.T. Garibaldi.

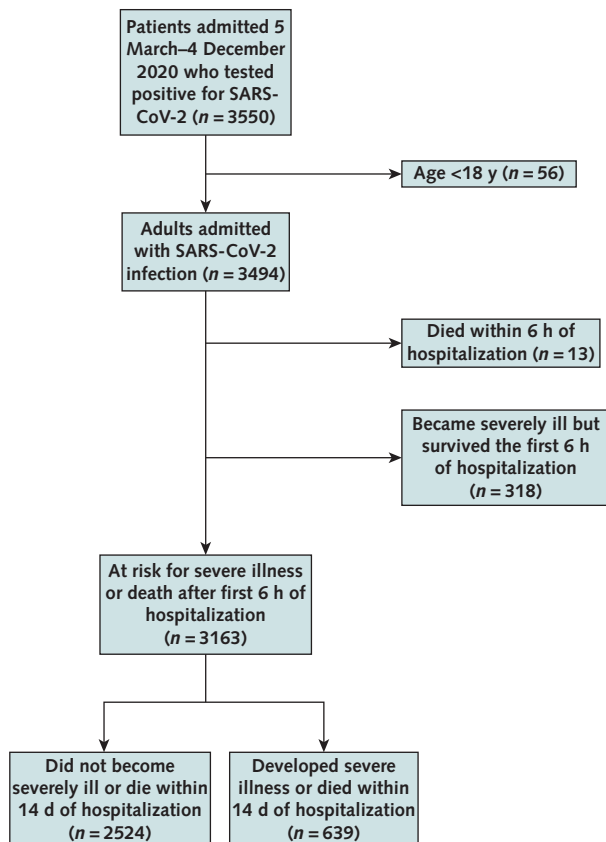
Statistical expertise: S. Wongvibulsin, J. Wen, M.C. Wang, Y. Xu, K. Wang, J.F. Betz, J. Muschelli, S.L. Zeger, M.L. Robinson.

Obtaining of funding: B.T. Garibaldi.

Administrative, technical, or logistic support: S. Wongvibulsin, B.T. Garibaldi, A. Gupta, S.L. Zeger, M.L. Robinson.

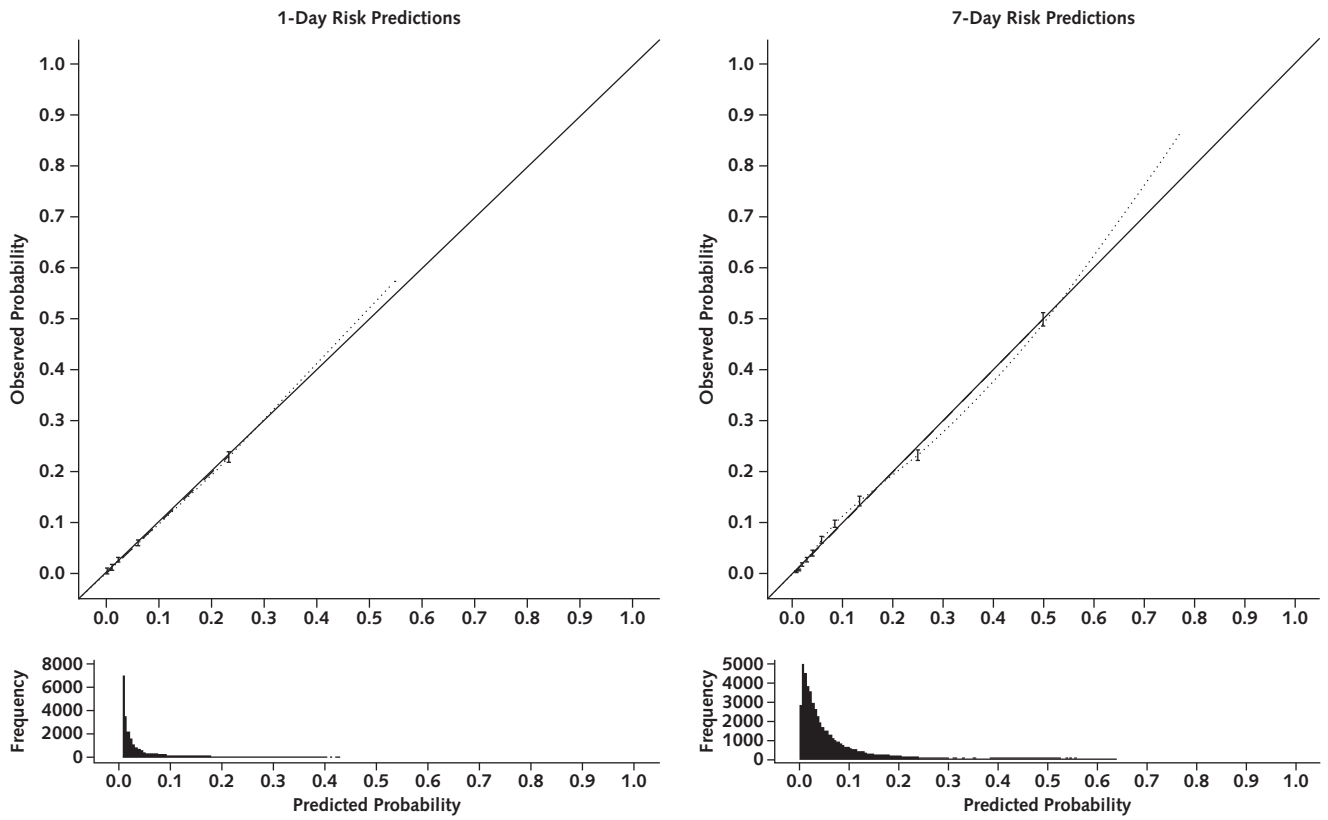
Collection and assembly of data: B.T. Garibaldi, J. Muschelli, M.L. Robinson.

Appendix Figure 1. Study flow diagram.



Among patients who were at risk for severe illness or death after the first 6 hours of hospitalization, 639 developed severe illness or death within 14 days of hospitalization. SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2.

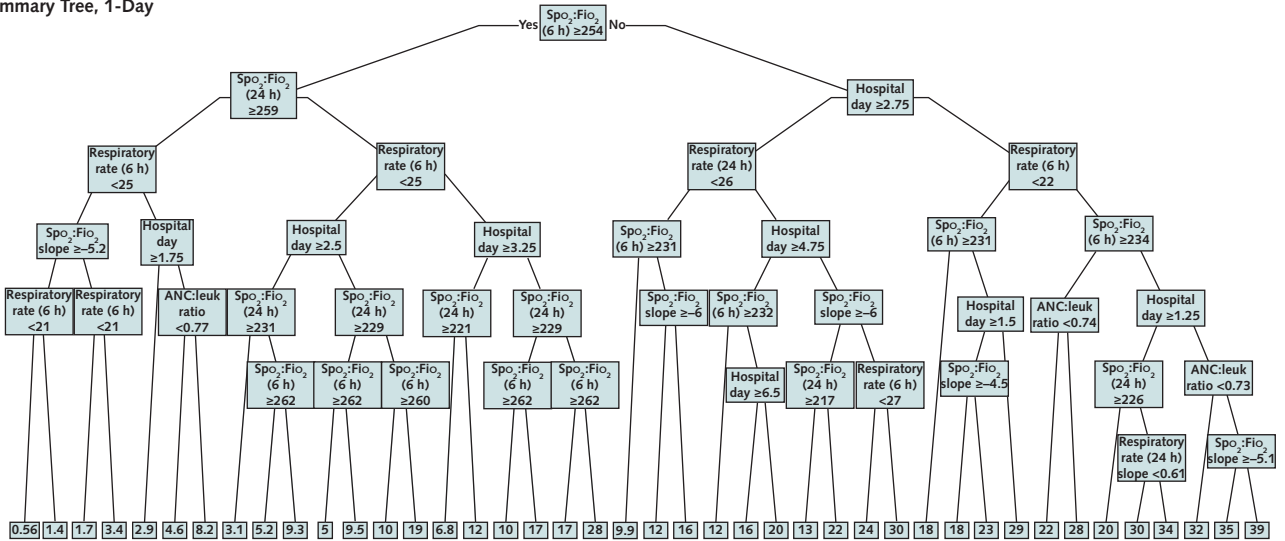
Appendix Figure 2. Calibration curves for predictions of severe illness or death in the next 1 day and 7 days.



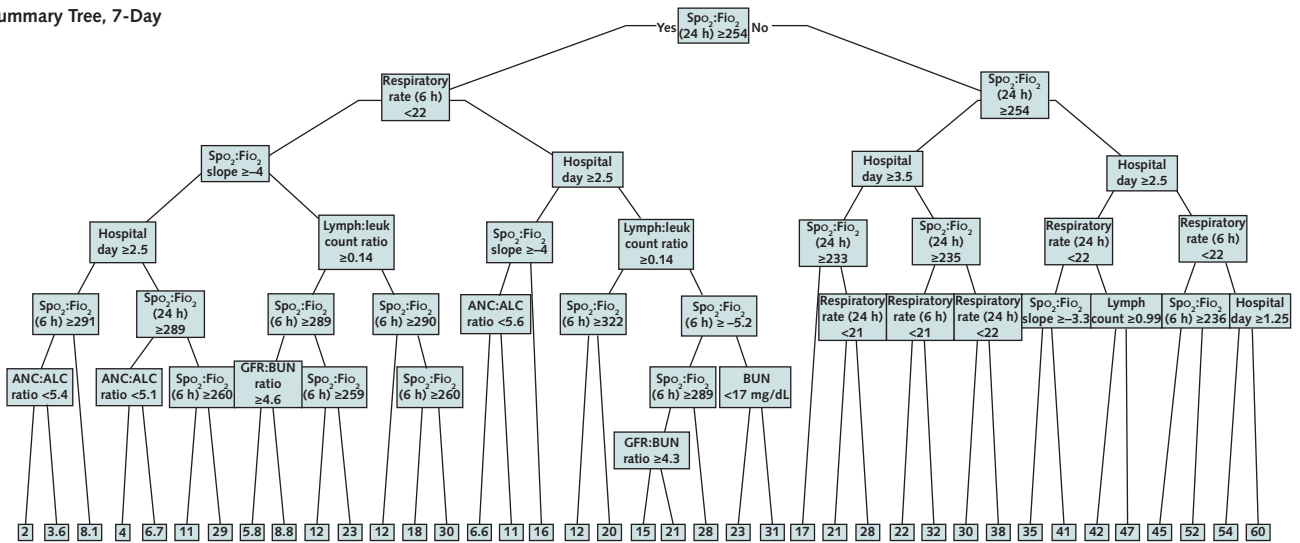
The points and 95% CIs show calibration by decile. Dashed lines indicate linear fit through the decile points. The dotted line shows the locally weighted smoothing (loess) curve through the predicted probabilities versus observed outcomes. Histograms of the distribution of predicted values are shown below the calibration plots.

Appendix Figure 3. Summary trees of random forest for survival, longitudinal, and multivariate predictions of 1-day and 1-week risk for severe disease or death.

Summary Tree, 1-Day



Summary Tree, 7-Day



The predicted probabilities are expressed in the terminal nodes and shaded according to predicted probability from lowest risk (0%) to highest risk (100%). ALC = absolute lymphocyte count; ANC = absolute neutrophil count; BUN = blood urea nitrogen; GFR = glomerular filtration rate; leuk = leukocyte; lymph = lymphocyte.