MAJOR ARTICLE

CoVA: An Acuity Score for Outpatient Screening that Predicts Coronavirus Disease 2019 Prognosis

Haoqi Sun,[1,2,3,a] Aayushee Jain,[1,3,a] Michael J. Leone,[1,3,a] Haitham S. Alabsi,[1,2] Laura N. Brenner,[2,4,5] Elissa Ye,[1,3] Wendong Ge,[1,2,3] Yu-Ping Shao,[1,3] Christine L. Boutros,[1] Ruopeng Wang,[6,7] Ryan A. Tesh,[1,3] Colin Magdamo,[1] Sarah I. Collens,[1] Wolfgang Ganglberger,[1,3] Ingrid V. Bassett,[2,8] James B. Meigs,[2,5] Jayashree Kalpathy-Cramer,[2,6,7] Matthew D. Li,[2,6,7] Jacqueline T. Chu,[2,5,8,9] Michael L. Dougan,[2,10] Lawrence W. Stratton,[1,2] Jonathan Rosand,[1,2] Bruce Fischl,[2,6,7,11] Sudeshna Das,[1,2] Shibani S. Mukerji,[1,2,b] Gregory K. Robbins,[2,8,b] and M. Brandon Westover[1,2,3,b]

[1]Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA, [2]Harvard Medical School, Boston, Massachusetts, USA, [3]Clinical Data AI Center, Massachusetts General Hospital, Boston, Massachusetts, USA, [4]Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, [5]Division of General Internal Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, [6]Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts, USA, [7]Athinoula A. Martinos Center for Biomedical Imaging, Boston, Massachusetts, USA, [8]Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts, USA, [9]MGH Chelsea HealthCare Center, Chelsea, Massachusetts, USA, [10]Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts, USA, and [11]Massachusetts Institute of Technology Health Sciences & Technology Program/Computer Science and Artificial Intelligence Laboratory, Cambridge, Massachusetts, USA

Background. We sought to develop an automatable score to predict hospitalization, critical illness, or death for patients at risk for coronavirus disease 2019 (COVID-19) presenting for urgent care.

Methods. We developed the COVID-19 Acuity Score (CoVA) based on a single-center study of adult outpatients seen in respiratory illness clinics or the emergency department. Data were extracted from the Partners Enterprise Data Warehouse, and split into development (n = 9381, 7 March–2 May) and prospective (n = 2205, 3–14 May) cohorts. Outcomes were hospitalization, critical illness (intensive care unit or ventilation), or death within 7 days. Calibration was assessed using the expected-to-observed event ratio (E/O). Discrimination was assessed by area under the receiver operating curve (AUC).

Results. In the prospective cohort, 26.1%, 6.3%, and 0.5% of patients experienced hospitalization, critical illness, or death, respectively. CoVA showed excellent performance in prospective validation for hospitalization (expected-to-observed ratio [E/O]: 1.01; AUC: 0.76), for critical illness (E/O: 1.03; AUC: 0.79), and for death (E/O: 1.63; AUC: 0.93). Among 30 predictors, the top 5 were age, diastolic blood pressure, blood oxygen saturation, COVID-19 testing status, and respiratory rate.

Conclusions. CoVA is a prospectively validated automatable score for the outpatient setting to predict adverse events related to COVID-19 infection.

Keywords. COVID-19; prognosis; risk prediction; outpatient.

The coronavirus disease 2019 (COVID-19) pandemic has presented unparalleled challenges for healthcare systems around the world [1–8]. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) first appeared in Wuhan, China, in December 2019. The first case in the United States (US) was confirmed on 20 January [9], followed by exponential spread within the US [2]. By the end of April, Massachusetts was the third hardest hit state, trailing only New York and New Jersey [10]. Within Massachusetts, Boston and Chelsea were epicenters for the spread of COVID-19.

In anticipation of a surge in COVID-19 patients and to help limit viral spread and preserve personal protective equipment, Massachusetts General Hospital (MGH) closed most outpatient and urgent care clinics and set up new respiratory illness clinics (RICs) on 7 March 2020. These clinics were staffed by clinicians and nurses reassigned from other areas, many with little urgent care experience. COVID-19 suspected cases were also screened in the emergency department (ED). Due to high volumes and precautions for staff and patients, visit duration and extent of clinical assessments for most patients were curtailed. In addition to limited clinical assessment, triage decisions were complicated by COVID-19's biphasic clinical course: Patients who initially present with mild symptoms often later return for admission, and many subsequently suffer adverse events including intensive care unit (ICU) transfer, mechanical ventilation (MV), or death. Various prediction rules have been proposed to predict these adverse events, but to our knowledge few have been prospectively validated, and most were developed for the inpatient [11–14] rather than for outpatient screening.

To help frontline clinicians appropriately triage and plan follow-up care for patients presenting for COVID-19 screening,

we developed an outpatient screening score, the COVID-19 Acuity Score (CoVA), using data from MGH's RICs and ED. CoVA is designed in a way that automated scoring could be incorporated into electronic medical record (EMR) systems, including Epic. CoVA assigns acuity levels based on demographic, clinical, radiographic, and medical history variables, and provides predicted probabilities for hospital admission, ICU admission or MV, or death within 7 days.

METHODS

Study Population

The Partners Institutional Review Board approved the study and granted a waiver of consent. We extracted EMR data from patients seen in the MGH RICs and ED between 7 March and 14 May 2020. Patients were divided into 2 mutually exclusive cohorts: a development cohort (7 March to 2 May, n = 9381), and a prospective cohort (3 May to 14 May, n = 2205). All patients were admitted either through the emergency room (ER) or through an RIC clinic, among which 484 patients were initially evaluated in RIC and subsequently sent to the ER and then admitted. We split the cohorts on 2 and 3 May as the ratio of patients in development vs prospective cohorts was the closest to 80% vs 20%, which was the same ratio as in cross-validation inside the development cohort. This was prespecified before data analysis was started.

The flowchart to generate the development and prospective cohorts is illustrated in Figure 1. Inclusion criteria were (1) an MGH RIC visit between 7 March and 14 May 2020, or an ED visit in the same timeframe where the reason for visit was considered possible COVID-19 related (cough, fever, shortness of breath) or other specific reasons listed in Supplementary Table 10; and (2) age ≥18 years. Exclusion criteria were (1) lack of any clinical assessment of vital signs (systolic/diastolic blood pressure, body temperature, heart rate, and respiratory rate), even if a COVID-19 diagnostic test was performed (often done for low-risk asymptomatic patients who needed only COVID-19 testing); and (2) patients with visits were excluded if they were

in both development and prospective cohorts, to ensure that the model was not biased to those patients. Note that 1 patient contributes only once in either cohort. In the development cohort, we randomly chose 1 visit for each patient to avoid patients with multiple visits contributes unevenly to the model coefficients. In the prospective cohort, we chose the most recent visit from each patient. This was done to focus more on prediction performance on dates further into the future relative to the development cohort.

Primary Outcome

The primary outcome was the occurrence of an adverse event within 7 days following an outpatient medical encounter, including either hospitalization at MGH, critical illness (defined as ICU care and/or MV), or death. The prediction horizon was set to 7 days because this period was considered meaningful by our frontline teams for clinical decision-making, and because empirically, within the model development cohort most adverse events occurred within 7 days of initial presentation. A secondary analysis was done using a 4-week (28-day) window.

Predictors

We selected 98 variables that were routinely available in the outpatient setting during the COVID-19 pandemic to serve as candidate predictors. These included demographic variables: age, sex, tobacco use history, most recent body mass index (BMI) (represented as binary variables designated as high BMI [>35 kg/m^2], and low BMI [<18.5 kg/m^2]), the most recent vital signs (blood pressure, respiratory rate, heart rate, temperature, blood oxygen saturation level [SpO$_2$]) within the preceding 3 days; COVID-19 testing status (based on a nasopharyngeal swab testing for SARS-CoV-2 polymerase chain reaction at the time of sampling; see Supplementary Data, "Predictor encoding," for encoding details); specific symptoms associated with COVID-19 infection (anosmia, dysgeusia based on *International Statistical Classification of Diseases, Tenth Revision* [*ICD-10*] codes), and preexisting medical diagnoses, coded as present or absent based on groups of billing codes (*ICD-10* codes;
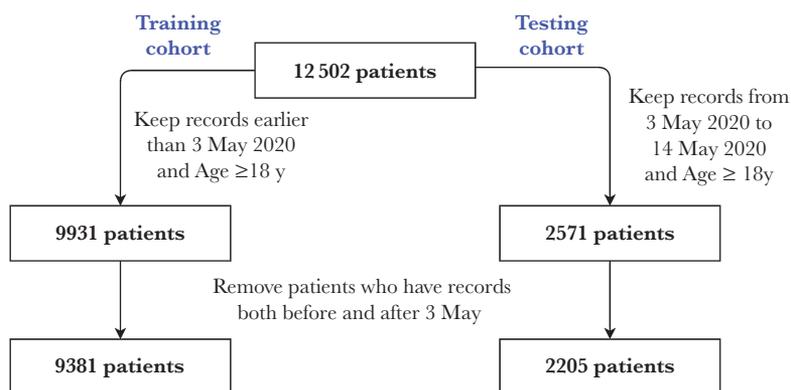

Figure 1. Data flowchart for the development and prospective cohorts.

Supplementary Table 7) in the EMR. To be coded as present, the diagnostic code had to be recorded on or before the day of presentation. The weighted Charlson Comorbidity Index (CCI) was computed based on groups of *ICD* codes [11]. We omitted race and ethnicity as predictors because (1) the composition of race/ethnicity our dataset does not necessarily generalize to populations outside MGH. Therefore, the coefficients for race/ethnicity based on our dataset may not reflect the actual (or causal) value, instead they are biased by the particular composition of race/ethnicity in our dataset; and (2) we have found that these variables were often unavailable or inaccurately recorded, especially ethnicity.

For patients who underwent chest radiograph (CXR) imaging during these encounters, we identified groups of common findings based on radiology reports that correlate with COVID-19 infection. These groups were identified in 2 steps. First, we manually reviewed 50 CXR reports and liberally extracted key words, phrases, and word patterns describing abnormal findings. Next, a pulmonary and critical care medicine physician (L. B.) categorized these phrases into groups. Five groups were identified: multifocal, patterns typical for COVID-19 (pneumonia, bronchopneumonia, acute respiratory distress syndrome), patchy consolidation, peripheral or interstitial opacity, or hazy or airspace opacities. The phrases and groupings are shown in Supplementary Table 8. CXR findings suggestive of COVID-19 were coded as present, not present, or unavailable.

Data Preprocessing and Selection of Predictor Variables

We treated predictors outside of physiologically plausible ranges as unavailable (Supplementary Table 9). Unavailable values were imputed using K-nearest neighbors [15], where the value of K was determined by minimizing the imputation error on a subset with available data by randomly masking variables according to the pattern of unavailability in the overall data. Importantly, we did not impute CXR predictors or COVID-19 status for patients who did not have either of them available. Instead, for these we coded unavailable (see details in the Supplementary Data, "Predictor encoding conventions"). We have provided the performance of CoVA depending on the availability of CXR and COVID-19 status in Supplementary Table 4.

Predictors were standardized to zero mean and unit standard deviation using *z* score transformation (Supplementary Table 1). Predictors were selected for inclusion in the CoVA model in 2 stages. First, we used analysis of variance (ANOVA) to identify predictors associated with hospitalization, ICU care or MV, or death. At an α = .05 significance level, we identified 65 predictors to carry forward into the model fitting procedure (*P* values are shown in Supplementary Table 11). A 1-way ANOVA test was used to test the null hypothesis, if the average feature value was the same across the 4 groups: hospitalization, critical illness, death, and none of the above. Here we treated ordinal outcome as categorical outcome). Second, we used least absolute shrinkage and selection regression during the model fitting procedure to select a reduced subset of highly predictive and relatively uncorrelated variables.

Model Development

We assigned an ordinal scale to adverse events, including no event, hospitalized, ICU care and/or MV, and death. We performed pairwise learning to rank an implementation of ordinal regression [16, 17]. Training involves learning to predict which of a pair of patients will have a worse outcome. On biological grounds, and to address co-linearity among predictors, we constrained the model optimization to allow only nonnegative coefficients for CXR predictors, medical comorbidities, CCI, and history of current or past tobacco use; and unconstrained for other predictors. Model training and preliminary evaluation of model performance was performed using the development cohort, using nested 5-fold cross-validation (Supplementary Figure 2). The final model provides acuity scores between 0 and 100 and predicted probabilities for hospitalization and for critical illness or death within 7 days (in any situation including patients more generally in other populations). The final model was tested on the prospective validation cohort. Performance is reported both for the development cohort (cross-validation) and for the prospective cohort.

Statistical Analysis of CoVA Predictive Performance

We summarized the distribution of cohort characteristics and adverse events using event counts and proportions for categorical predictors, and the mean and standard deviation for continuous variables. For model calibration, an expected-to-observed ratio (E/O) of 1.0 indicates that the number of expected events equals the number of observed events; and the calibration slope (CS) is defined as the linear correlation between the observed O and expected probabilities E, where the expected (predicted) values (E) are binned into quintiles [18].

We calculated the area under the receiver operating characteristic curve (AUC) to quantify how well CoVA scores discriminated between individuals who were hospitalized within 7 days vs those who were not hospitalized, and between those who had critical illness (ICU care, MV, or death) within 7 days vs those who did not. We considered an AUC between 0.50 and 0.55 to be poor; between 0.55 and 0.65, moderate; between 0.65 and 0.75, acceptable; and >0.75, excellent. We also calculated specificity, positive predictive value, and negative predictive value at the 90% sensitivity level. Finally, we looked at the ability of the CoVA score to predict adverse events over a longer 4-week (28-day) window.

RESULTS

Cohort Characteristics

From 7 March to 2 May 2020, 9381 patients met inclusion criteria and were included in the development cohort. The

average age was 51 years old, with 49% being female. Among these, 3344 (35.6%) had adverse events within 7 days of presentation: 2562 (27.3%) were hospitalized, 679 (7.2%) received ICU care and/or were mechanically ventilated, and 103 (1.1%) died.

From 3 to 14 May 2020, 2205 additional patients met inclusion criteria and were included in the prospective cohort. The average age was 53 years old, with 49% being female. Among these, 726 (32.9%) had adverse events: 575 (26.1%) were hospitalized, 139 (6.3%) received ICU care and/or were mechanically ventilated, and 12 (0.5%) died.

Cohort characteristics are summarized in Table 1. Compared to the model development period, during the prospective validation period there were modest increases in the proportion of patients with CXRs (prospective 52.3% vs development 41.0%), the proportion of outpatient evaluations performed in RIC clinics (prospective 26.5% vs development 21.4%), and in testing rates for COVID-19 (prospective 78.8% vs development 59.5%). Several other small but likely clinically insignificant differences between cohorts also reached statistical significance, due to the large cohort sizes. COVID-19 infections and clinical adverse events by age and decade of life are shown in Supplementary Figure 1.

Table 1. Characteristics of Study Participants

Characteristic	Development Cohort	Prospective Cohort	P Value
No. of patients	9381	2205	
Respiratory illness clinic	2013 (21.5%)	584 (26.5%)	<.01**
Emergency department	7368 (78.5%)	1621 (73.5%)	<.01**
COVID-19 positive			<.01**
Yes	1404 (15.0%)	243 (11.0%)	
No	4178 (44.5%)	1494 (67.8%)	
Untested or unknown	3799 (40.5%)	468 (21.2%)	
Outcome in 7 days			NS
Hospitalization	2562 (27.3%)	575 (26.1%)	
ICU or MV	679 (7.2%)	139 (6.3%)	
Death	103 (1.1%)	12 (0.5%)	
Age, y, mean (SD)	51.1 (19.2)	52.6 (18.6)	<.01**
Female sex	4587 (48.9%)	1087 (49.3%)	.735
BMI, kg/m^2, mean (SD)	28.4 (7.0)	28.4 (6.8)	.792
Race			<.01**
Asian	373 (4.0%)	79 (3.6%)	
Black	1023 (10.9%)	218 (9.9%)	
Pacific Islander	15 (0.2%)	4 (0.2%)	
Native American	6 (0.1%)	1 (0.0%)	
White	5673 (60.5%)	1434 (65.0%)	
Other or unknown	2291 (24.4%)	469 (21.3%)	
Ethnicity			<.01**
Hispanic	1879 (20.0%)	377 (17.1%)	
Non-Hispanic	6698 (71.4%)	1635 (74.1%)	
Unavailable	804 (8.6%)	193 (8.8%)	
Smoking			<.01**
Yes or quit	3190 (34.0%)	834 (37.8%)	
Never or passive	5245 (55.9%)	1165 (52.8%)	
Not asked/unknown	946 (10.1%)	206 (9.3%)	
Weighted Charlson score, mean (SD)	1.5 (2.4)	1.6 (2.6)	.031*
SpO$_2$, %, mean (SD)	97.3 (2.3)	97.3 (2.2)	.152
CXR (percentage is among available patients except the first one)			<.01**
CXR available	3851 (41.1%)	1154 (52.3%)	
Multifocal	1214 (31.5%)	242 (21.0%)	
Typical pattern for COVID-19	548 (14.2%)	119 (10.3%)	
Patchy consolidation	739 (19.2%)	245 (21.2%)	
Peripheral/interstitial opacity	72 (1.9%)	8 (0.7%)	
Hazy or airspace opacity	480 (12.5%)	101 (8.8%)	

Data are presented as No. (%) unless otherwise indicated.

Abbreviations: BMI, body mass index; COVID-19, coronavirus disease 2019; CXR, chest radiograph; ICU, intensive care unit; MV, mechanical ventilation; NS, not significant; SD, standard deviation; SpO$_2$, oxygen saturation.

*P < .05.

**P < .01.

Predictive Performance of the CoVA Score

CoVA showed excellent calibration and discrimination in the development cohort for hospitalization (E/O: 1.00 [.98–1.02], CS: 0.99 [.98–.99], AUC: 0.80 [.79–.81]); for critical illness (E/O: 1.00 [.93–1.06], CS: 0.98 [.96–.99], AUC: 0.82 [.80–.83]); and for death (E/O: 1.00 [.84–1.21], CS: 1.32 [1.01–1.60], AUC: 0.87 [.83–.91]) (all values in brackets indicate the 95% confidence interval). Performance generalized to the prospective validation cohort, with similar results for hospitalization (E/O: 1.01 [.96–1.07], CS: 0.99 [.98–1.00], AUC 0.76 [.73–.78]); for critical illness (E/O: 1.03 [.89–1.20], CS: 0.98 [.94–1.00], AUC: 0.79 [.75–.82]); and for death (E/O: 1.63 [1.03–3.25], CS: 0.91 [.35–1.47], AUC: 0.93 [.86–.98]) (all values in brackets indicate the 95% confidence interval). Additional performance metrics are reported in Table 2. We also trained a model for 28-day horizon and provided the performances in Supplementary Table 2. The model achieved numerically similar results compared to that for the 7-day horizon. To explore whether including patients with visits in both cohorts would make a difference, we also randomly assigned half of them to the development cohort and half to the prospective cohort. The results are in Supplementary Table 3, which do not differ significantly with that from excluding these patients.

We also investigated performance over different time periods during the pandemic under investigation using 3-day nonoverlapping windows (Figure 5). The linear fit shows that there is no trend in AUC for hospitalization ($P = .6$) or for critical illness ($P = .1$); no trend in E/O for hospitalization ($P = .9$) or for critical illness ($P = .3$).

Properties of the CoVA Score

For the prospective cohort, the proportion of patients with adverse events at 7 days increases with higher CoVA scores, rising from 18% with CoVA scores in the 0–20 range, to 88% for those with scores in the 80–100 range. The proportion of patients with critical illness or death also rises, from 2% for scores between 0 and 20, to 32% with scores between 80 and 100 (Figure 2A).

We examined the ability of the CoVA score to predict events over a 4-week (28 days) window following initial presentation to the RIC or ED in the prospective cohort. Curves for the cumulative incidence of adverse events over time for different levels of CoVA scores are shown in Figure 2B and additional performance data are provided in Supplementary Table 2. By 28 days, 751 (34%) of patients experienced hospitalization, critical illness, or death. Of these, 699 (32%) occurred within 1 day, and 726 (33%) had occurred by 7 days. Critical illness or death occurred in 175 (8%) of patients within 28 days of presentation. Of these, 111 (5%) occurred within 1 day, and 151 (7%) occurred within 7 days. Therefore, a high percentage of patients experiencing adverse events did so within both 1 day and 7 days. These numbers support our choice of 7 days as a clinically meaningful event prediction horizon, and reflect a high percentage of patients who were admitted directly from the RIC or ED.

Predictors of Adverse Events

Thirty predictors were selected by the data-driven model training procedure (Table 3). All but 4 predictors increase the predicted probability of adverse events when present. SpO_2, diastolic and systolic blood pressure, and low BMI inversely correlated with the probability of adverse events. Predictors from CXR reports included in the model were multifocal patterns (diffuse opacities, ground glass) and patterns typical for COVID-19.

Table 2. Prediction Performance for 7-Day Horizon

Metric	Hospitalization, ICU, MV, or Death	ICU, MV, or Death	Death
Concurrent validation (based on the development cohort but cross-validated)			
Patients, No. (%)	3344 (35.6%)	782 (8.3%)	103 (1.1%)
Calibration (E/O)	1.00 (.98–1.02)	1.00 (.93–1.06)	1.00 (.84–1.21)
Calibration slope	1.02 (.99–1.06)	0.86 (.79–.93)	1.32 (1.05–1.58)
AUC	0.80 (.80–.81)	0.82 (.80–.83)	0.87 (.83–.91)
Specificity at 90% sensitivity	0.56 (.53–.59)	0.44 (.41–.48)	0.35 (.24–.45)
PPV at 90% sensitivity	0.47 (.45–.49)	0.16 (.14–.17)	0.03 (.02–.04)
NPV at 90% sensitivity	0.89 (.88–.90)	0.98 (.98–.99)	1.00 (1.00–1.00)
Prospective validation			
Patients, No. (%)	726 (32.9%)	151 (6.8%)	12 (0.54%)
Calibration (E/O)	1.01 (.96–1.07)	1.03 (.89–1.20)	1.63 (1.03–3.25)
Calibration slope	0.92 (.83–.99)	0.76 (.60–.95)	0.84 (.37–1.40)
AUC	0.76 (.73–.78)	0.79 (.75–.82)	0.93 (.86–.98)
Specificity at 90% sensitivity	0.66 (.62–.70)	0.53 (.43–.63)	0.29 (.01–.31)
PPV at 90% sensitivity	0.40 (.38–.43)	0.11 (.09–.14)	0.017 (.01–.22)
NPV at 90% sensitivity	0.87 (.86–.89)	0.98 (.98–.99)	1.00 (1.00–1.00)

Data in parentheses indicate 95% confidence intervals.

Abbreviations: AUC, area under the receiver operating characteristic curve; E/O, ratio of expected to number of observed adverse events; ICU, intensive care unit; MV, mechanical ventilation; NPV, negative predictive value; PPV, positive predictive value.
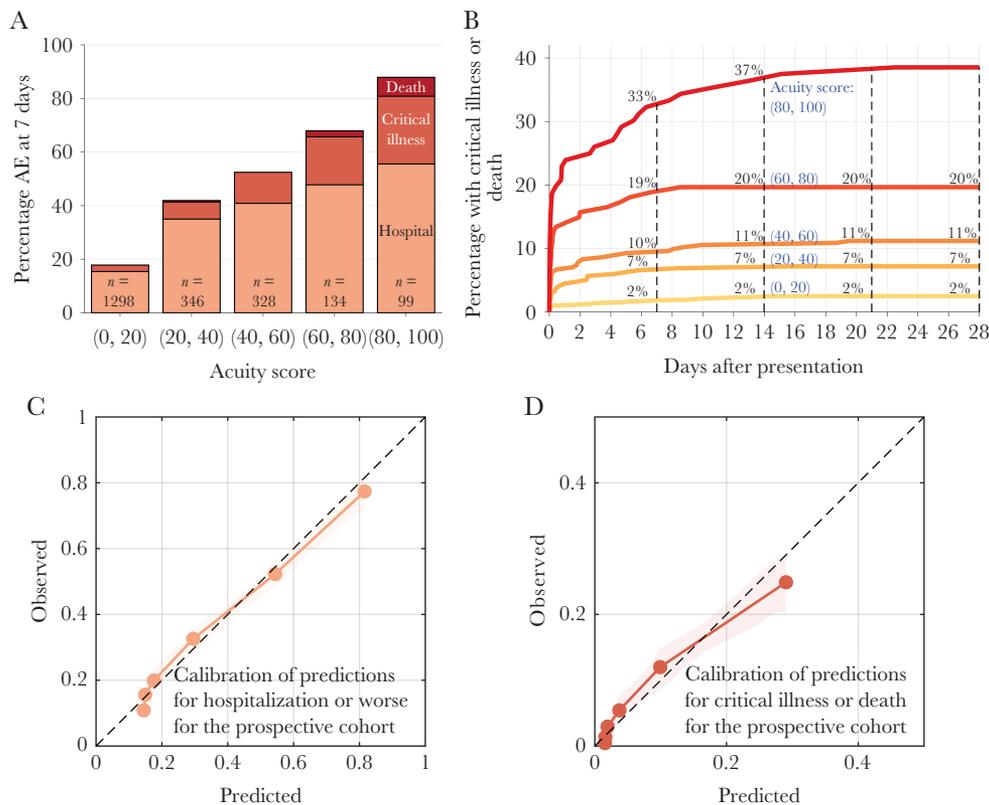
Figure 2. *A*, Distributions of adverse events (AEs) within 7 days after initial outpatient evaluation in the respiratory illness clinical/emergency department, binned by acuity score. Colors from light to dark represent distinct AEs: hospitalization, intensive care unit/mechanical ventilation, or death. *B*, Cumulative incidence of critical illness or death up to 17 days following initial evaluation, based on initial acuity score. Curves are computed based on cross-validation in the development cohort. *C* and *D*, Calibration curves: predicted probability of AEs vs observed rate of AEs. *C*, Calibration for predicting hospitalization (dashed line for the development cohort; and solid line for the prospective cohort). *D*, Calibration for predicting critical illness or death (dashed line for the development cohort; and solid line for the prospective cohort). The overall calibration (ratio of expected to number of observed AEs) and calibration slopes are reported in Table 2.

Impact of Chest Radiographs and COVID-19 Testing on Predicting Critical Illness

CXRs and testing for COVID-19 were not universal, and testing rates evolved over time. One potential use of CoVA is to help determine whether to perform these tests. We therefore investigated the impact of CXR findings on the predicted probability of critical illness or death. As shown in Supplementary Figure 3, positive CXR findings are most informative when the pre-CXR probability is 30% (critical illness or death), in which case they increase the predicted probability of an adverse event by 4%.

We also examined the effect of COVID-19 testing results on the predicted probability of severe adverse events (critical illness or death). The largest effect of a positive test result occurs when pretest probability of a severe adverse event is 28%, in which case the posttest probability increases by 8%.

DISCUSSION

Risk stratification of outpatients presenting for COVID-19 is important for medical decisions regarding testing, hospitalization, and follow-up. We developed and prospectively validated the

COVID-19 Acuity Score (CoVA) as an automatable outpatient screening score that can be implemented in EMRs including EPIC. The model exhibits excellent calibration, discrimination, and negative predictive value both in concurrent validation (n = 9381, 7 March–2 May) and in large-scale prospective validation (n = 2205, 3–14 May). While several COVID-19 risk prediction models have been proposed for the inpatient setting [12–14], CoVA fills an unmet need for a prospectively validated risk score designed for outpatient screening and could easily be deployed before future COVID-19 surges.

A few critical thresholds are needed to practically guide the decision for admission and level of follow-up with CoVA. The outpatient clinic plans to provide follow-up calls to reassess remotely when the CoVA score is from 10 to 20 (CoVA score ranges from 0 to 100) (probability of hospitalization or worse in 7 days ranges between 20% to 30%) and to bring patients back to reassess in the outpatient RIC or ED when the CoVA score is >20 (probability of hospitalization or worse in 7 days >30%).

Several predictors selected by the model have been identified in prior studies, including advanced age [4, 5, 9, 19]; preexisting pulmonary [4, 5], kidney [4], and cardiovascular disease

Table 3. Coefficients of the Coronavirus Disease 2019 Acuity Score (CoVA) Model

ID	Predictor	Coefficient	ID	Predictor	Coefficient
1	Age	0.7353	16	Renal cancer[a]	0.0681
2	Diastolic blood pressure	−0.4724	17	Pancreatitis[a]	0.0608
3	SpO$_2$	−0.3776	18	Cystic fibrosis[a]	0.0492
4	Ever COVID-19 positive up to event	0.2750	19	Cardiac arrest[a]	0.0491
5	Respiratory rate	0.2746	20	Seizure disorder	0.0437
6	Acute ischemic stroke[a]	0.1746	21	Amyolateral sclerosis[a]	0.0405
7	CXR: Multifocal	0.1293	22	Metabolic acidosis[a]	0.0385
8	Heart rate	0.1215	23	Myasthenia gravis[a]	0.0374
9	Body temperature	0.1206	24	Pneumothorax[a]	0.0300
10	Systolic blood pressure	−0.1151	25	Spinal muscular atrophy[a]	0.0241
11	Weighted Charlson score	0.1142	26	Pericarditis[a]	0.0144
12	Intracranial hemorrhage[a]	0.1087	27	High BMI (>35 kg/m^2)	0.0028
13	Subarachnoid hemorrhage[a]	0.0919	28	CXR: Typical for COVID-19	0.0001
14	Male sex	0.0808	29	Low BMI (<18.5 kg/m^2)	−0.0001
15	Hematologic malignancy[a]	0.0765	30	ARDS[a]	0.0001

Diagnoses were based on past medical history and were coded as present (eg, pneumothorax = 1) if recorded in the electronic medical record at any time before the date of presentation for COVID-19 screening. All coefficients were applied to rescaled predictors, where the rescaling was done by subtracting the mean and then dividing the standard deviation in Supplementary Table 1.

Abbreviations: ARDS, acute respiratory distress syndrome; BMI, body mass index; COVID-19, coronavirus disease 2019; CXR, chest radiograph; SpO$_2$, oxygen saturation.

[a]Preexisting conditions documented in the electronic medical record.

[4, 9, 20]; obesity [21]; and increased respiratory or heart rate or hypoxia [4]. We found that other preexisting medical conditions also increased risk for adverse outcomes, including cancer and pancreatitis. Hypertension and diabetes mellitus did not emerge as predictors in the CoVA model, despite being identified in prior studies [4, 5, 9]. These comorbidities are correlated with outcomes in univariate analysis (Supplementary Table 11) and correlation analysis shows that both are strongly associated with older age, higher CCI, high BMI, and other comorbidities already selected as predictors, helping to explain their nonselection by the model (Supplementary Figure 4). Human immunodeficiency virus/AIDS was not significantly associated with the outcome, but this may have been due to low numbers (85/9381, <1% of developmental cohort).

Several studies have documented neurological manifestations of COVID-19 [22–24], but the association of past neurological history with COVID-19 outcomes remains poorly understood. We were surprised to find that a variety of neurological diseases surfaced as robust predictors of adverse outcomes in COVID-19 infection, including ischemic stroke, intracranial hemorrhage, subarachnoid hemorrhage, epilepsy, amyotrophic lateral sclerosis, myasthenia gravis, and spinal muscular atrophy. It is unclear if these neurological diseases are merely markers of health or if the worsened outcomes are due to the interaction of COVID-19 with neurological disorders that amplifies the pathology.

Prior work on predicting outcomes in COVID-19 patients is summarized in Supplementary Table 6. Like our study, most have attempted to predict critical illness or death. However, most are based on small cohorts (median, n = 189 [range, n = 26–577]); focus on inpatients; and utilize laboratory values,

which were rarely available for our outpatient cohort. Only 3 included prospective or external validation. By contrast, CoVA is designed for the outpatient setting. In this setting, the availability of COVID-19 test results were variable (results were available for 60% in the development and 80% in the prospective cohort) and other laboratory results were rarely available. To ensure generalizability, we used a large development cohort of 9381 patients, and trained the model using a rigorous approach. We ensured clinical interpretability by utilizing a linear model with positivity constraints on predictors expected to increase risk. Finally, we validated our model on a large (n = 2205), prospectively collected patient cohort, providing an assessment of model generalizability with minimal (or reduced) bias.

Strengths of this study include its large sample size, careful EMR phenotyping, and rigorous statistical approach, and all variables required by the model are available and automatically extractable within most electronic health record systems. The study also has limitations. First, our study is from a single center, with demographics specific to MGH patients, who mostly come from greater Boston and the surrounding regions. Second, we did not have complete follow-up on all patients, especially those who were not admitted. As a result, we may not have captured hospitalizations and death that occurred outside MGH. However, we believe most patients would have been readmitted to MGH and the EMR would have captured most deaths. Third, the model used data from both the RICs and patients seen in the ED for a variety of reasons. Nevertheless COVID-19 was, and for now, remains a significant concern for patients seen in the ED. We also excluded patients prior to the onset of COVID-19 in Boston (7 March 2020); therefore, our model is relevant for screening

during times of high alert for COVID-19. We also did not include laboratory test values, since they were typically not available during initial outpatient assessments. Finally, the study was conducted in the first few months of the epidemic in Boston; further research is needed to see if these results hold more broadly and as the epidemic matures.

In conclusion, the COVID-19 Acuity Score (CoVA) is a well-calibrated, discriminative, prospectively validated, and interpretable score that estimates the risk for adverse events among outpatients presenting with possible COVID-19 infection.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. JAMA **2020**; 323:2052–9.
2. Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the United States. N Engl J Med **2020**; 382:929–36.
3. Fauci AS, Lane HC, Redfield RR. Covid-19—navigating the uncharted. N Engl J Med **2020**; 382:1268–9.
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**; 395:1054–62.
5. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA **2020**; 323:1239–42.
6. Centers for Disease Control and Prevention COVID-19 Response Team. Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:343–6.
7. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med **2020**; 26:672–5.
8. Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med **2020**; 172:577–82.
9. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA **2020**; 323:1061–9.

10. The Covid-19 tracker. STAT **2020**. https://www.statnews.com/feature/coronavirus/covid-19-tracker/. Accessed 5 June 2020.

11. Gasparini A. comorbidity: an R package for computing comorbidity scores. J Open Source Soft **2018**; 3:648.

12. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit Care **2020**; 24:1–4.

13. Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. PLoS One **2020**; 15:e0230548.

14. Wang S, Zha Y, Li W, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. Eur Respir J **2020**; 56:2000775.

15. Batista GE, Monard MC. A study of K-nearest neighbour as an imputation method. In: HIS, Santiago, Chile, 1–4 December **2002**.

16. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res **2011**; 12:2825–30.

17. Liu T-Y. Learning to rank for information retrieval. In: Foundations and Trends in Information Retrieval **2009**; 3:225–331.

18. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Validity of prognostic models: when is a model clinically useful? Semin Urol Oncol **2002**; 20:96–107.

19. Guan W-J, Ni Z-Y, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med **2020**; 382:1708–20.

20. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med **2020**; 46:846–8.

21. Cai Q, Chen F, Wang T, et al. Obesity and COVID-19 severity in a designated hospital in Shenzhen, China. Diabetes Care **2020**; 43:1392–8.

22. Filatov A, Sharma P, Hindi F, Espinosa PS. Neurological complications of coronavirus disease (COVID-19): encephalopathy. Cureus **2020**; 12:e7352.

23. Mao L, Jin H, Wang M, et al. Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan, China. JAMA Neurol **2020**; 77:683–90.

24. Wu Y, Xu X, Chen Z, et al. Nervous system involvement after infection with COVID-19 and other coronaviruses. Brain Behav Immun **2020**; 87:18–22.