Internal and Emergency Medicine (2022) 17:1879–1889
https://doi.org/10.1007/s11739-022-03014-9

IM - ORIGINAL

A novel evidence-based predictor tool for hospitalization and length of stay: insights from COVID-19 patients in New York city

Maan El Halabi[1] · James Feghali[2] · Jeeyune Bahk[1] · Paulino Tallón de Lara[1] · Bharat Narasimhan[1] · Kam Ho[1] · Mantej Sehmbhi[1] · Joseph Saabiye[3] · Judy Huang[2] · Georgina Osorio[3] · Joseph Mathew[4] · Juan Wisnivesky[5] · David Steiger[4]

Received: 15 February 2022 / Accepted: 20 May 2022 / Published online: 30 June 2022

Abstract

Predictive models for key outcomes of coronavirus disease 2019 (COVID-19) can optimize resource utilization and patient outcome. We aimed to design and internally validate a web-based calculator predictive of hospitalization and length of stay (LOS) in a large cohort of COVID-19-positive patients presenting to the Emergency Department (ED) in a New York City health system. The study cohort consisted of consecutive adult (> 18 years) patients presenting to the ED of Mount Sinai Health System hospitals between March 2020 and April 2020, diagnosed with COVID-19. Logistic regression was utilized to construct predictive models for hospitalization and prolonged (> 3 days) LOS. Discrimination was evaluated using area under the receiver operating curve (AUC). Internal validation with bootstrapping was performed, and a web-based calculator was implemented. From 5859 patients, 65% were hospitalized. Independent predictors of hospitalization and extended LOS included older age, chronic kidney disease, elevated maximum temperature, and low minimum oxygen saturation ($p < 0.001$). Additional predictors of hospitalization included male sex, chronic obstructive pulmonary disease, hypertension, and diabetes. AUCs of 0.881 and 0.770 were achieved for hospitalization and LOS, respectively. Elevated levels of CRP, creatinine, and ferritin were key determinants of hospitalization and LOS ($p < 0.05$). A calculator was made available under the following URL: https://covid19-outcome-prediction.shinyapps.io/COVID19_Hospitalization_Calculator/. This study yielded internally validated models that predict hospitalization risk in COVID-19-positive patients, which can be used to optimize resource allocation. Predictors of hospitalization and extended LOS included older age, CKD, fever, oxygen desaturation, elevated C-reactive protein, creatinine, and ferritin.

Keywords Coronavirus disease 2019 (COVID-19) · Hospitalization · Length of stay · Predictor model

Introduction

Ten months after the initial outbreak of the 2019 novel coronavirus (SARS-Cov-2) in Wuhan, China, the disease evolved into a global pandemic. New York City (NYC) became the

Maan El Halabi and James Feghali contributed equally to this manuscript.

✉ David Steiger
david.steiger@mountsinai.org

[1] Department of Medicine, Icahn School of Medicine at Mount Sinai, Mount Sinai Morningside and Mount Sinai West Hospitals, New York, NY, USA

[2] Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[3] Division of Infectious Disease, Department of Medicine, Mount Sinai Morningside and Mount Sinai West Medical Center, New York, NY, USA

[4] Division of Pulmonary and Critical Care Medicine, Department of Medicine, Mount Sinai Morningside and Mount Sinai West Medical Center, New York, NY 10019, USA

[5] Division of Pulmonary and Critical Care Medicine, Department of Medicine, Mount Sinai Hospital, New York, NY, USA

epicenter of the pandemic, accounting for a large proportion of infections in the United States [1].

The scale of infection imposed a major strain on medical infrastructure and resources leading to substantial shortages. Advanced age, hypoxia upon presentation, abnormal chest imaging, and elevated inflammatory markers appear to be strong predictors of worse outcomes [2–6]. Identifying predictors of hospitalization and length of stay (LOS), which are both highly relevant outcomes in patients with COVID-19, have been described. Increasing age and multimorbidity were associated with hospitalization in two studies of patients who tested positive for SARS-CoV-2 [4–7]. Understanding the determinants of the need for hospitalization and the projected LOS can optimize the utilization of hospital resources, aid in triaging, provide opportunities for timely health care assessment for patients, and improve shared decision-making [5]. This study aims to describe the demographics, clinical characteristics, and outcomes in a large COVID-19 cohort and to derive predictive models for hospitalization and prolonged LOS.

Methods

Patients

Data for the study were obtained from the Mount Sinai Data Warehouse, a registry of de-identified patient data extracted from the electronic medical record system (EPIC) across the Mount Sinai network. The database consisted of consecutive SARS-CoV-2-positive adult (> 18 years) patients presenting to an emergency department (ED) of one of the Mount Sinai Health System hospitals: Mount Sinai Hospital, Morningside, West, Brooklyn, Queens, all located in NYC between March 20 and April 23 2020. Diagnosis was confirmed by reverse transcriptase polymerase chain reaction (RT-PCR) of nasopharyngeal or oropharyngeal specimens. These data were collected starting from the initial phases of the pandemic, and all patients presented with symptoms consistent with COVID-19 including fever, cough, diarrhea, or shortness of breath. This study was approved by the Mount Sinai Institutional Review Board (IRB). Since no direct patient contact or intervention from the study group was needed, no patient consent was required.

Variables

Patient demographics, diagnosis codes (International Classification of Diseases-9/10-Clinical Modification (ICD-9/10-CM) code), and clinical data including symptoms, vital signs as well as laboratory data were collected on presentation. We defined a pre-existing condition as the presence of diagnosis codes (ICD 9/10) associated with specific diseases. The earliest available laboratory results during the first 24 h were used in the prediction analysis.

The primary outcome was the need for hospital admission after presentation. A secondary outcome included an extended LOS defined as hospitalization lasting more than 3 days among patients that completed follow-up (i.e., we excluded patients that were still hospitalized at the time of data analysis). As recommended by Hintz et al. [8], we also excluded patients who died within 3 days to avoid a potentially misleading LOS value denoting good outcome. The LOS cut-off of 3 days was selected as Public health authorities including the Centers for Disease Control and Prevention, European Centre for Disease Prevention and Control, and the National Centre for Infectious Disease states that three days of symptom resolution, namely fever and respiratory symptoms, is the cut-off for safe discharge [1, 9]. Moreover, median time to readmission [10], median time to radiographic progression [11], as well as median time to clinical deterioration following admission [12] were all 3 days. Patients were tracked for mortality, need for intensive care, and intubation.

Statistical analysis

Analyses were performed using R (R Foundation for Statistical Computing, Austria) with statistical significance set at $p \leq 0.05$. Descriptive statistics (n, % for categorical variables and mean ± SD for continuous variables) were used to summarize the baseline demographic, clinical, and laboratory variables of the study population. Continuous variables were subsequently categorized using clinically relevant cut-offs. For the first outcome of interest, a univariable analysis of factors associated with the need for hospitalization consisting of a Chi-square or Fischer exact test for categorical variables and the Student's t test for continuous variables was conducted. Characteristics with a p value < 0.1 were subsequently entered a stepwise logistic regression model. Given the amount of missing data for laboratory values, these variables were left out of this step-wise process. Collinearity between variable pairs was evaluated using a strict variance inflation factor (VIF) cut-off of 3. Model discrimination was assessed using a receiver operating characteristic (ROC) curve analysis to obtain an area under the curve (AUC). Bootstrapping with 1000 samples with replacement was utilized to calculate an optimism-corrected AUC to check for possible overfitting [13, 14]. Calibration was evaluated using the Hosmer–Lemeshow test with an adjusted number of subgroups to account for a large sample size in our study [15].

In hospitalized patients that completed follow-up and survived, a similar process consisting of univariable analysis, stepwise multivariable logistic regression for factors independently associated with an extended LOS.

Finally, we developed a web-based calculator that uses the models to predict the probability of a patient requiring hospitalization and extended LOS using readily available components of the history and vital signs on first patient encounter using the Shiny package from R.

Results

Patients

The cohort consisted of 5859 patients with a mean age of 60.5 years (SD = 17.5 years) with 3253 males (56%). Racial, ethnic groups included 1373 (24%) white, 1576 (28%) black, 229 (4%) Asian, 140 (3%) Native Hawaiian/Pacific Islander, and 2351 (41%) with other ethnicities. The baseline demographic and clinical characteristics are summarized in Table 1.

Univariable analysis, hospitalization

Out of 5859 patients, 3794 (65%) were hospitalized. Demographic factors and comorbidities that were significantly associated ($p < 0.001$) with hospitalization on univariable analysis included older age, race, male sex, ever smoking, history of chronic obstructive pulmonary disease (COPD), hypertension, obesity, diabetes, chronic kidney disease (CKD), and cancer. With regard to vital signs, maximum temperature of 38 degrees Celsius or more, systolic blood pressure < 90 mmHg, minimum percent oxygen saturation < 90%, elevated C-reactive protein and ferritin, were significantly associated with hospitalization ($p < 0.001$). Table 2 summarizes univariable analysis.

Adjusted analysis, hospitalization

Independent predictors of hospitalization included older age, male sex, COPD, hypertension, diabetes, CKD, elevated maximum temperature, and low minimum percent oxygen saturation. The optimal multivariable model resulting from stepwise logistic regression is summarized in Table 3, and consisted of age (OR = 6.29; 95% CI [1.83–2.63] for older adults (> 65 years) compared to younger adults (18–44 years), male sex (OR = 1.35 [1.17–1.55]), COPD (OR = 1.74 [1.00–3.03]), hypertension (OR = 1.39 [1.13–1.70]), diabetes (OR = 1.45 [1.16–1.81]), CKD (OR = 1.69 [1.23–2.32]), elevated maximum temperature (OR = 4.98 [4.28–5.79]), and low minimum oxygen saturation (OR = 13.40 [10.59–16.96]). The AUC for the model was 0.881 (95% CI 0.872–0.890). Bootstrap validation yielded negligible optimism of 0.0013 which translates into an optimism-corrected AUC of 0.880, indicating absence of significant overfitting. The Hosmer–Lemeshow test yielded a non-significant p value, indicating appropriate calibration.

Table 1 Patient demographics, comorbidities, and vital signs on presentation

Characteristic*	Value
Demographics	
Age, years, mean (SD)	60.5 ± 17.5
Race (*n* = 5669)	
White	1373 (24)
African American	1576 (28)
Asian	229 (4)
Native Hawaiian or Other Pacific Islander	140 (2)
Other	2351 (41)
Sex, female	2606 (44)
Body mass index, kg/m^2 (n = 4201)	29.19 ± 8.39
Ever smoked (n = 4424)	1282 (29)
Comorbidities	
Asthma	257 (4)
COPD	176 (3)
Hypertension	1627 (28)
Diabetes	1157 (20)
CKD	524 (9)
HIV	90 (2)
Cancer	322 (5)
Vital signs	
Temperature max, degrees Celsius (*n* = 5848)	38 ± 1.2
Heart rate, beats per minute (*n* = 5850)	95.7 ± 19
Respiratory rate, breaths per minute (*n* = 5844)	20.5 ± 5.3
Systolic blood pressure, mmHg (*n* = 5841)	131.6 ± 23.1
Diastolic blood pressure, mmHg (*n* = 5841)	75.5 ± 13.8
%O_2 saturation minimum (*n* = 5848)	87.5 ± 15.6

HIV human immunodeficiency virus; *COPD* chronic obstructive pulmonary disorder

*Numbers in parentheses represent number of patients without missing data

Length of stay

The course of the study population is described in Fig. 1. Out of 3794 patients requiring hospitalization, 631 (17%) were excluded as were still hospitalized at the time of analysis, leaving 3163 patients. There was a mortality rate of 28% (897/3163) among hospitalized patients and 17% (897/5228) among all patients who completed follow-up (i.e., discharged from ED or inpatient care). In addition, 16% (492/3163) of hospitalized patients required admission into an intensive care unit, and 13% (401/3163) required intubation. Among the 897 patients who died, 386 (43%) required admission into an ICU, and 315 (35%) were intubated.

After excluding patients who died within 3 days, the mean LOS was 7.3 days (SD = 5.3 days; median = 6 days;

Table 2 Univariable Predictors of need for hospitalization

Characteristic	No admission ($n=2065$)	Admission ($n=3794$)	p value
Demographics			
Age			< 0.001†
19–44 years	808 (39)	397 (10)	
45–65 years	854 (41)	1350 (36)	
> 65 years	403 (20)	2047 (54)	
Race			0.001†
White	471 (24)	902 (25)	
Black	608 (30)	968 (26)	
Asian	87 (4)	142 (4)	
Native Hawaiian or Other Pacific Islander	62 (3)	78 (2)	
Other	775 (39)	1576 (43)	
Sex, female	1038 (50)	1568 (41)	< 0.001†
Comorbidities			
Ever smoked	331 (22)	951 (33)	< 0.001†
Asthma	77 (4)	180 (5)	0.070
COPD	19 (1)	157 (4)	< 0.001†
Hypertension	303 (15)	1324 (35)	< 0.001†
Obesity (BMI ≥ 30 kg/m^2)	286 (35)	1233 (37)	0.306
Diabetes	210 (10)	947 (25)	< 0.001†
CKD	72 (3)	452 (12)	< 0.001†
HIV	25 (1)	65 (2)	0.135
Cancer	71 (3)	251 (7)	< 0.001†
Vital Signs			
Temperature maximum			< 0.001†
< 38 degrees Celsius	1638 (80)	1447 (38)	
≥ 38 degrees Celsius	416 (20)	2347 (62)	
Systolic blood pressure			< 0.001†
< 90 mmHg	23 (1)	99 (3)	
≥ 90 mmHg	2030 (99)	3687 (97)	
%O_2 saturation minimum			< 0.001†
< 90	85 (4)	1980 (53)	
≥ 90	1956 (96)	1768 (47)	
Lab values			
ALT (U/L)			0.291
1–45*	406 (73)	2493 (70)	
46–135	128 (23)	886 (25)	
> 135	23 (4)	186 (5)	
AST (U/L)			< 0.001†
1–35*	232 (45)	1248 (36)	
36–105	247 (48)	1808 (52)	
> 105	40 (8)	426 (12)	
Creatinine (mg/dL)			< 0.001†
0–1*	457 (57)	1750 (46)	
1–2	273 (34)	1258 (33)	
> 2	67 (8)	770 (20)	
CRP (mg/L)			< 0.001†
0.0–5.0*	6 (5)	23 (2)	
5.1–100	64 (52)	500 (38)	
> 100	52 (43)	784 (60)	

Table 2 (continued)

Characteristic	No admission ($n=2065$)	Admission ($n=3794$)	p value
D-dimer (mcg/ml)			0.382
0.0–0.5*	94 (9)	190 (10)	
0.6–1.5	447 (42)	772 (40)	
> 1.5	511 (49)	966 (50)	
Ferritin (ng/mL)			0.005†
30–400*	370 (33)	566 (27)	
401–500	81 (7)	145 (7)	
> 500	678 (60)	1356 (66)	
Anemic hemoglobin‡	268 (37)	537 (39)	0.378
LDH (U/L)			0.243
100–220*	65 (6)	96 (5)	
221–250	50 (5)	79 (4)	
> 250	940 (89)	1764 (91)	
Procalcitonin > 0.5 ng/mL	328 (29)	595 (29)	0.951
Troponin (ng/mL)			0.747
0.0–0.03*	531 (52)	952 (51)	
0.04–0.1	250 (24)	448 (24)	
> 0.1	250 (24)	480 (25)	
WBC count (cells/mm^3)			0.904
< 4000	156 (9)	220 (9)	
4000–10,999	1158 (70)	2534 (70)	
≥ 11,000	340 (21)	758 (21)	

ALT alanine aminotransferase; *AST* aspartate aminotransferase; *COPD* chronic obstructive pulmonary disorder; *CRP* C-reactive protein; *FEU* fibrinogen equivalent units; *HIV* human immunodeficiency virus; *LDH* lactate dehydrogenase; *WBC* white blood cell

*Numbers in parentheses represent number of patients without missing data

†Statistically significant ($p \leq 0.05$)

‡Defined as below 13 g/dL in males and below 12 g/dL in females

*Reference value

Table 3 Optimal multivariable logistic regression model predictive of need for hospitalization ($n=5787$)

Variable	Odds ratio [95% CI]
Age, 18–44 years as reference	
45–65	2.19 [1.83–2.63]
> 65	6.29 [5.15–7.69]
Sex, male vs. female	1.35 [1.17–1.55]
COPD	1.74 [1.00–3.03]
Hypertension	1.39 [1.13–1.70]
Diabetes	1.45 [1.16–1.81]
CKD	1.69 [1.23–2.32]
Temperature max ≥ 38 degrees Celsius	4.98 [4.28–5.79]
$\%O_2$ saturation minimum < 90	13.40 [10.59–16.96]

1.2% of the data were missing at least one of the above variables

COPD chronic obstructive pulmonary disorder

*Statistically significant ($p \leq 0.05$)

IQR = 6 days), and 2177 of the remaining 2892 patients (75%) required an LOS exceeding 3 days. The characteristics of hospitalized patients with comparisons between early- and late-discharge patients are summarized in Fig. 2. Factors that were independently associated with an extended LOS included older age (OR = 1.03 [1.02–1.04]), CKD (OR = 1.91 [1.35–2.71]), elevated maximum temperature (OR = 2.91 [2.40–3.53]), and low minimum oxygen saturation (OR = 3.89 [3.16–4.79]). The univariable analysis and optimal adjusted model are summarized in Table 4. Age provided better discrimination when employed as a continuous variable. The stepwise model provided an AUC of 0.770 (95% CI 0.752–0.789). Bootstrap validation yielded a negligible optimism of 0.0029 which translates into a bias-corrected AUC of 0.768. The Hosmer–Lemeshow test yielded a non-significant p value indicating appropriate calibration.

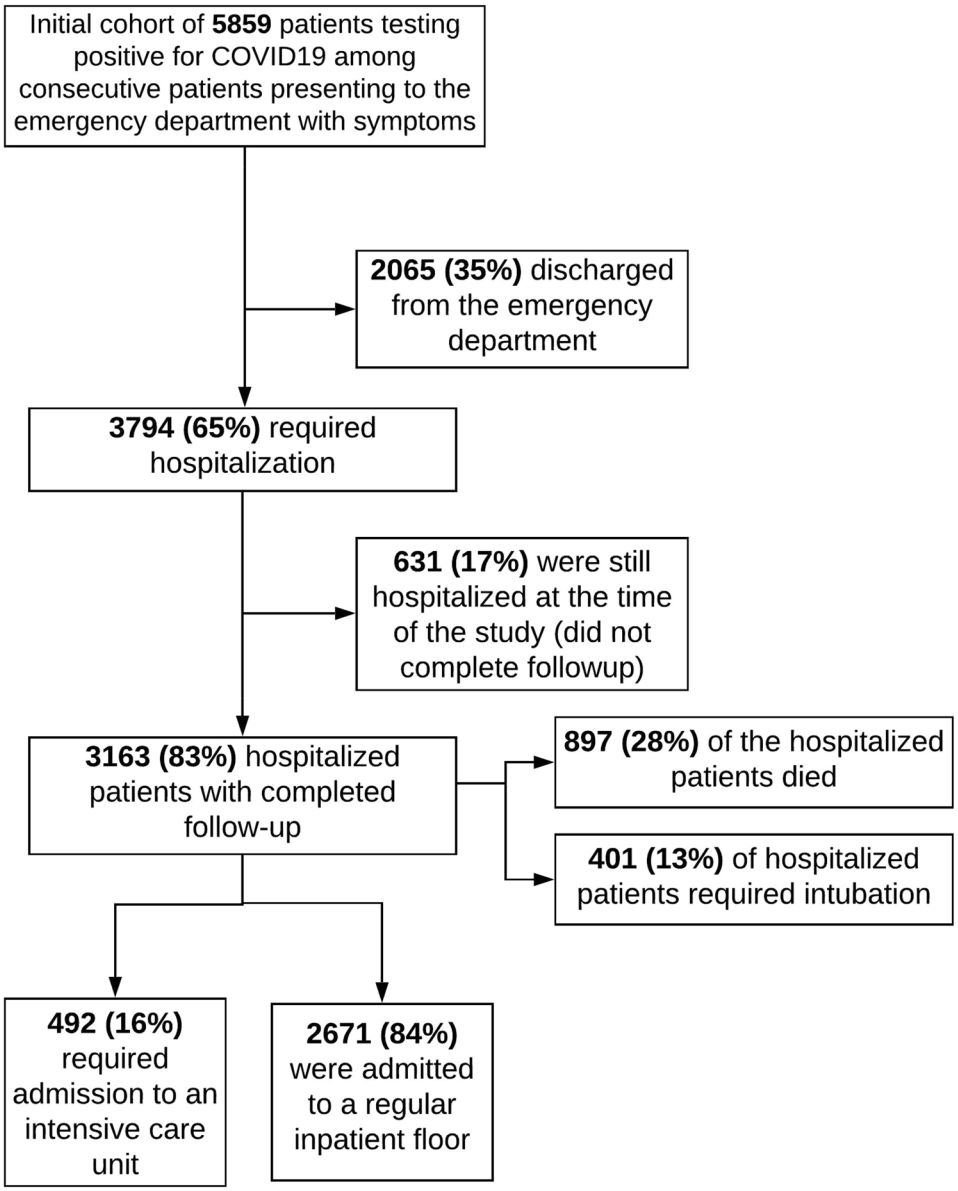

Fig. 1 Flow diagram summarizing the course of COVID-19-positive patients

A combined calculator of likelihood of hospitalization and extended LOS can be accessed with the following URL:**https://covid19-outcome-prediction.shinyapps.io/COVID19_Hospitalization_Calculator/**

Discussion

We analyzed a large cohort of 5859 COVID-19 patients in the United States. The results yielded internally validated models with good discrimination that were able to predict both the need for hospitalization, as well as the risk of prolonged hospitalization among admitted COVID-19 patients.

Hospitalization

There remains a relative paucity of US studies analyzing hospitalization rates and their determinants. Petrilli et al. analyzed factors associated with hospitalization in a cohort of 4103 COVID-19 patients presenting to New York University (NYU) Langone Health system [16]. The inclusion criteria however included both inpatient and outpatient visits rather than ED visits; hence, a lower hospitalization rate of 48.7% was reported, compared to 65% in our cohort, which

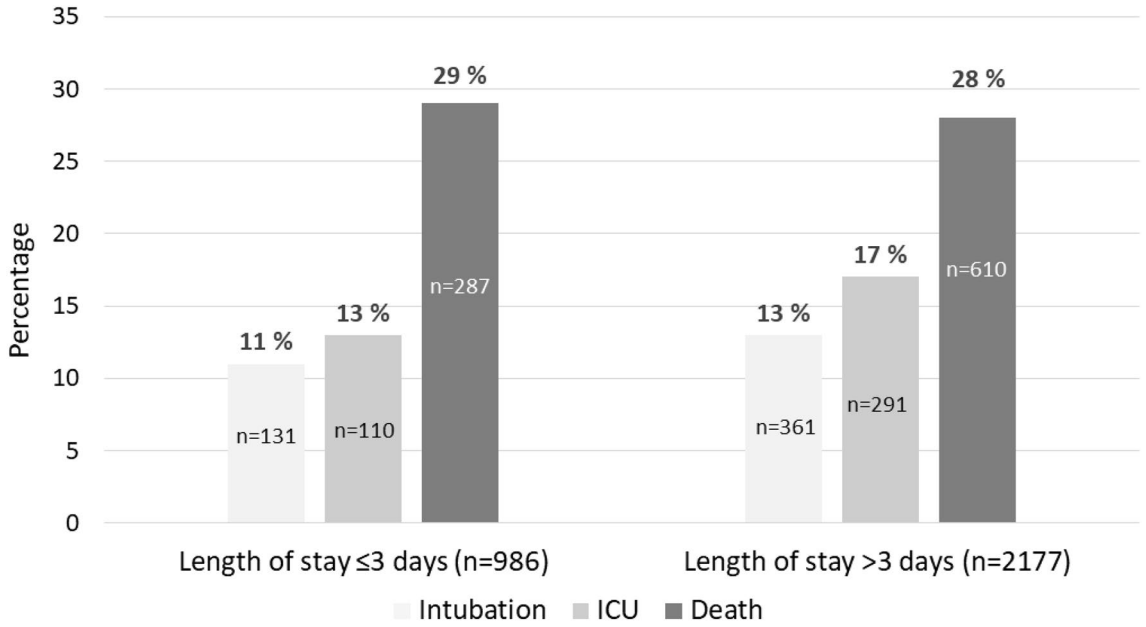

Fig. 2 Description of hospitalized patients ($n = 3163$)

only included patients evaluated in the ED. In concordance with our results, identified risk factors for hospitalization included increased age, CKD, diabetes, and male gender. Other important factors included obesity, race, heart failure, hyperlipidemia, and tobacco use. Similarly, Richardson et al. identified the characteristics of 5700 hospitalized COVID-19 patients at the Northwell Health System, but also excluded ED visits [17]. Although the authors did not construct predictive risk models in these studies, we expect our calculator to be generalizable to these cohorts and other urban settings given the significant overlap in patient demographics and risk factors, but external validation remains warranted. While admission risk calculators have been developed to estimate hospital utilization for COVID-19, they have limited applications to the US adult population since they included pediatric [18] and/or non-US [19] populations. Moreover, the NYU study and the Cleveland Clinic health system study by Jehi et al. lacked important objective predictors (e.g., vital signs, laboratory values) [20]. Our data identified key vital signs, such as fever and oxygen desaturation, and important laboratory values, such as AST, ferritin, CRP, and creatinine, but not procalcitonin and white blood cell counts, predictive of hospitalization. The advantages of our study lie in the utilization of an exceptionally large multi-ethnic NYC population at the height of the pandemic to develop a concise and practical tool with objective measures of hospitalization risk.

Length of stay

One of the largest published studies on hospitalized COVID-19 patients consists of a descriptive analysis of 5700 patients presenting to the Northwell hospital system in NYC [10]. The reported overall median LOS was 4.1 days, including patients who died during hospitalization. 3066 out of the 5700 (54%) patients were still hospitalized at the time of analysis and therefore did not complete follow-up. Among hospitalized patients discharged alive, the median LOS increased from 2.5 days in the 20–29 years age group to 4.8 day in 90 years and above group. Although no statistical significance was investigated, this trend closely matches our identified association of extended LOS with increasing age. Rees et al. reported a pooled median LOS of 14 days in China and of 5 days outside of China and attributed this difference to variation in criteria for admission and discharge as well as heterogeneity in timing in relation to the pandemic [21]. Our median LOS of 6 days is in close agreement. Worldwide variation also underscores the importance of verifying applicability of prediction models before implementation across geographical regions. To our knowledge, no internally validated predictive models for LOS derived from readily available demographic, comorbidity, and clinical data in COVID-19 patients exists in the literature. Our data highlight the key predictive value of age, CKD,

Table 4 Univariable and multivariable analysis of factors associated with prolonged length of stay (> 3 days) in surviving hospitalized patients ($n = 2892$)

	Univariable		Stepwise Multivariable
Characteristic	Odds ratio [95% CI]	*p* value	Odds ratio [95% CI]
Demographics			
Age, per year	1.03 [1.02–1.04]	< 0.001*	1.03 [1.02–1.04]
Race, White as reference		0.012*	–
Black	0.81 [0.63–1.03]	0.089	–
Asian	0.89 [0.54–1.45]	0.627	–
Native Hawaiian/Pacific Islander	0.80 [0.44–1.48]	0.479	–
Other	0.67 [0.53–0.84]	0.001*	–
Sex, male vs. female	0.98 [0.82–1.16]	0.791	–
Comorbidities			–
Ever smoked	1.18 [0.96–1.46]	0.113	–
Asthma	1.30 [0.85–1.97]	0.224	–
COPD	1.93 [1.13–3.30]	0.017*	–
Hypertension	1.42 [1.18–1.70]	< 0.001*	–
Obese (BMI ≥ 30 kg/m^2)	0.91 [0.67–1.23]	0.534	–
Diabetes	1.41 [1.14–1.73]	0.001*	–
CKD	2.21 [1.59–3.06]	< 0.001*	1.91 [1.35–2.71]
HIV	0.96 [0.52–1.78]	0.899	–
Cancer	2.17 [1.42–3.32]	< 0.001*	–
Vital Signs			
Temperature max ≥ 38 degrees Celsius	2.83 [2.38–3.36]	< 0.001*	2.91 [2.40–3.53]
%O_2 saturation minimum < 90	4.93 [4.04–6.02]	< 0.001*	3.89 [3.16–4.79]
Systolic blood pressure < 90 mmHg	1.22 [0.70–2.13]	0.478	
Lab values			
ALT, 1–45 U/L as reference		0.017*	–
46–135	0.83 [0.68–1.02]	0.076	–
> 135	0.62 [0.42–0.91]	0.014*	–
AST, 1–35 U/L as reference		0.769	–
36–105	1.03 [0.85–1.25]	0.763	–
> 105	1.12 [0.82–1.54]	0.470	–
Creatinine, 0–1 mg/dL as reference		< 0.001*	–
1–2	1.66 [1.37–2.02]	< 0.001*	–
> 2	2.43 [1.86–3.19]	< 0.001*	–
CRP, 0.0–5.0 mg/L as reference		< 0.001*	–
5.1–100	1.37 [0.53–3.52]	0515	–
> 100	2.83 [1.10–7.29]	0.031*	–
D-dimer, 0.0–0.5 mcg/ml FEU as reference		0.183	–
0.6–1.5	1.37 [0.92–2.04]	0.121	–
> 1.5	1.13 [0.77–1.67]	0.528	–
Ferritin, 30–400 ng/mL as reference		0.013*	–
401–500	2.37 [1.30–4.33]	0.005*	–
> 500	1.01 [0.78–1.31]	0.934	–
Anemic hemoglobin‡	0.99 [0.74–1.32]	0.922	–
LDH, 100–220 U/L as reference		0.059	–
221–250	2.52 [1.12–5.565]	0.026*	–
> 250	1.64 [1.01–2.69]	0.050*	–
Procalcitonin > 0.5 ng/mL	0.92 [0.72–1.18]	0.505	–
Troponin, 0.0–0.03 ng/ml as reference		0.248	–
0.04–0.1	0.92 [0.68–1.24]	0.577	–
> 0.1	0.78 [0.59–1.04]	0.095	–

Table 4 (continued)

Characteristic	Univariable		Stepwise Multivariable
	Odds ratio [95% CI]	*p* value	Odds ratio [95% CI]
WBC count, normal range as reference		0.549	–
< 4000	0.85 [0.64–1.14]	0.280	–
≥ 11,000	1.00 [0.80–1.25]	0.442	–
Medications, none as reference		< 0.001*	–
Azithromycin only	0.88 [0.63–1.23]	0.448	–
Hydroxychloroquine only	2.49 [1.84–3.37]	< 0.001*	–
Azithromycin + Hydroxychloroquine	3.67 [2.87–4.70]	< 0.001*	–

ALT alanine aminotransferase; *AST* aspartate aminotransferase; *COPD* chronic obstructive pulmonary disorder; *CRP* C-reactive protein; *FEU* fibrinogen equivalent units; *HIV* human immunodeficiency virus; *LDH* lactate dehydrogenase; *WBC* white blood cell

*Statistically significant ($P \leq 0.05$)

temperature, oxygen saturation, elevated levels of CRP, creatinine, and ferritin, which can identify patients requiring more hospital resources.

Mortality

Among 3163 hospitalized patients who completed follow-up, the mortality rate was 28% which is comparable to the rate of 21% identified in the corresponding NYC cohort of 2634 patients [10]. A study by Zhou et al. of 191 hospitalized patients in Wuhan, China similarly reported a mortality rate of 28% (54/191). Our data confirm previously reported mortality rates among hospitalized COVID-19 patients and enhances precision using the largest hospitalized cohort with completed follow-up to date. Grasselli et al. reported a mortality rate of 26% among 1581 ICU COVID-19 patients in Lombardy region of Italy [6]. In general, case fatality rates reported in the literature include 2.3% (1023 deaths of 44,672 confirmed cases) in China [22] and 7.2% (100 out of 1,625 patients) in Italy [23]. Our ICU admission rate of 16% and intubation rate of 13% mirrored those of Richardson et al. reporting rates of 14% and 12%, respectively in the Northwell cohort [10].

Previous models

A systematic review by Wynants et al. which discussed 10 previously published prognostic models for COVID-19 patients, identified 9 studies: 6 focused on mortality, 2 focused on the development of critical illness, and 1 focused on LOS [3]. All studies were based in China, and the largest sample size for model derivation was 577 patients. The LOS model was based solely on CT imaging findings and predicted LOS greater than 10 days using a limited sample of 26 COVID-19 patients [24]. The findings described in our study underscore the importance of incorporating comorbidity and clinical data, which even alone, can explain a large proportion of variance in outcomes. Otherwise, age, sex, hypertension, LDH, and CRP constituted some of the main predictors of mortality and critical illness models [25–28]. Main criticisms of previous models included the lack of calibration assessment and the absence of a readily available format for use in clinical practice [3]. Our analysis ensured adequate calibration and rendered the models easily accessible through a user-friendly web-based calculator.

Utility

The main strength of this calculator lies in its ability to provide accurate discrimination of illness requiring hospitalization and prolonged LOS based on simple clinical and laboratory variables that are readily available at the earliest point of contact. Radiological data were not included as this would limit the broad application of the model, including in resource-constrained environments. Not all patients receive imaging studies on initial ED presentation, and a significant proportion of initial chest radiography may be normal in mild or early COVID-19 [29, 30]. While the quantitative burden of consolidation or ground glass opacities at Chest CT can be used to predict clinical deterioration and death, this use of CT is not a standard of care [31]. Our model may be used for risk stratification for patients with COVID-19, particularly in outpatients setting as a decision-support tool for referring patients for emergency care. The application of the model in an outpatient setting would have to be

validated. The tool may also help to supply prognostic information that would aid in providing expectations on possible LOS for patients and families.

Limitations

Our results were extrapolated from a restricted geographic area and the lack of standardized guidelines warrant an external validation. Lower performance of our model on LOS, which is largely influenced by local standards of care, supports this. As the data were based on a particular cohort of COVID-19 (alpha variant) prior to the development of vaccines, our model's predictive performance on new variants and a vaccinated population may be limited. However, it should be noted that six hospitals of MSHS serve a racially diverse population with a wide range in incomes and chronic disease burden. Although our model can serve as an adjunct during an initial evaluation, analyses of harder outcomes (mortality, ICU transfer, need for ventilation) will be helpful in guiding further care during hospitalizations.

Conclusion

Age, comorbidities and vital signs on admission were predictors of hospitalization and length of stay in COVID-19 patients presenting to the ED. The prediction tool derived from this study can help design resource allocation during a surge of COVID-19 patients presenting to hospital EDs and outpatient settings, help guide quality of care, and assist in designing future studies on the triage and management of patients with COVID-19.

Author contributions All authors contributed to the study conception, design, material preparation, data collection, and analysis. The first draft of the manuscript was written by MEH and JF. The final draft was created by JB. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This study did not require ethics approval.

Human and animal rights statement and Informed consent This study did not require ethics approval as it did not include humans and/or animals nor informed consent.

References

1. Centers for Disease Control and Prevention (2020) United States COVID-19 Cases. Published 2020. Accessed April 30. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/summary.html. Accessed 30 Apr 2020
2. Yan L, Zhang H-T, Xiao Y et al (2020) Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic m. medRxiv. https://doi.org/10.1101/2020.02.27.20028027
3. Wynants L, Van Calster B, Bonten MMJ et al (2020) Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 369:m1328. https://doi.org/10.1136/bmj.m1328
4. Petrilli CM, Jones SA, Yang J et al (2020) Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. BMJ 369:m1966. https://doi.org/10.1136/bmj.m1966
5. Lipsitch M, Swerdlow DL, Finelli L (2020) Defining the epidemiology of Covid-19—studies needed. N Engl J Med 382(13):1194–1196. https://doi.org/10.1056/NEJMp2002125
6. Grasselli G, Zangrillo A, Zanella A et al (2020) Baseline Characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region. Italy JAMA 323(16):1574–1581. https://doi.org/10.1001/jama.2020.5394
7. Reilev M, Kristensen KB, Pottegård A et al (2020) 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. https://doi.org/10.1093/ije/dyaa140
8. Hintz SR, Bann CM, Ambalavanan N, Cotten CM, Das A, Higgins RD (2010) Predicting time to hospital discharge for extremely preterm infants. Pediatrics 125(1):e146–e154. https://doi.org/10.1542/peds.2009-0810
9. European Centre for Disease Control and Prevention. Discharge Criteria for Confirmed COVID-19 Cases. https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-Discharge-criteria.pdf. Accessed 30 Apr 2020
10. Richardson S, Hirsch JS, Narasimhan M et al (2020) Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York city area. JAMA. https://doi.org/10.1001/jama.2020.6775
11. Chen J, Qi T, Liu L et al (2020) Clinical progression of patients with COVID-19 in Shanghai. China J Infect 80(5):e1–e6. https://doi.org/10.1016/j.jinf.2020.03.004
12. Cummings MJ, Baldwin MR, Abrams D et al (2020) Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York city: a prospective cohort study. medRxiv. https://doi.org/10.1101/2020.04.15.20067157
13. Moons KGM, Altman DG, Reitsma JB et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 162(1):W1-73. https://doi.org/10.7326/M14-0698
14. von Dadelszen P, Payne B, Li J et al (2011) Prediction of adverse maternal outcomes in pre-eclampsia: development and validation of the fullPIERS model. Lancet (London, England) 377(9761):219–227. https://doi.org/10.1016/S0140-6736(10)61351-7
15. Paul P, Pennell ML, Lemeshow S (2013) Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. Stat Med 32(1):67–80. https://doi.org/10.1002/sim.5525
16. Petrilli CM, Jones SA, Yang J et al (2020) Factors associated with hospitalization and critical illness among 4103 patients with

COVID-19 disease in New York city. medRxiv. https://doi.org/10.1101/2020.04.08.20057794
17. Richardson S, Hirsch JS, Narasimhan M et al (2020) Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized With COVID-19 in the New York city area. JAMA 323(20):2052–2059. https://doi.org/10.1001/jama.2020.6775
18. Steinberg E, Balakrishna A, Habboushe J, Shawl A, Lee J (2020) Calculated decisions: COVID-19 calculators during extreme resource-limited situations. Emerg Med Pract. 22(4S):CD1–CD5
19. Wu G, Yang P, Xie Y et al (2020) Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. Eur Respir J. https://doi.org/10.1183/13993003.01104-2020
20. Jehi L, Ji X, Milinovich A et al (2020) Development and validation of a model for individualized prediction of hospitalization risk in 4536 patients with COVID-19. PLoS ONE 15(8):e0237419. https://doi.org/10.1371/journal.pone.0237419
21. Rees EM, Nightingale ES, Jafari Y et al (2020) COVID-19 length of hospital stay: a systematic review and data synthesis. BMC Med 18(1):270. https://doi.org/10.1186/s12916-020-01726-3
22. Wu Z, Mc Googan JM (2020) 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. https://doi.org/10.1001/jama.2020.2648
23. Onder G, Rezza G, Brusaferro S (2020) Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. JAMA. https://doi.org/10.1001/jama.2020.4683
24. Qi X, Jiang Z, Yu Q et al (2020) Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. medRxiv. https://doi.org/10.1101/2020.02.29.20029603
25. Caramelo F, Ferreira N, Oliveiros B (2020) Estimation of risk factors for COVID-19 mortality—preliminary results. medRxiv. https://doi.org/10.1101/2020.02.24.20027268
26. Gong J, Ou J, Qiu X et al (2020) A tool to early predict severe 2019-novel coronavirus pneumonia (COVID-19): a multicenter study using the risk Nomogram in Wuhan and Guangdong, China. medRxiv. https://doi.org/10.1101/2020.03.17.20037515
27. Lu J, Hu S, Fan R et al (2020) ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China. medRxiv. https://doi.org/10.1101/2020.02.20.20025510
28. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J (2020) Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit Care 24(1):108. https://doi.org/10.1186/s13054-020-2833-7
29. Rubin GD, Ryerson CJ, Haramati LB et al (2020) The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society. Radiology 296(1):172–180. https://doi.org/10.1148/radiol.2020201365
30. Wong HY, Lam HY, Fong AH et al (2020) Frequency and distribution of chest radiographic findings in patients positive for COVID-19. Radiology 296(2):E72–E78. https://doi.org/10.1148/radiol.2020201160
31. Grodecki K, Lin A, Cadet S et al (2020) Quantitative burden of COVID-19 pneumonia at chest CT predicts adverse outcomes: a post hoc analysis of a prospective international registry. Radiology 2(5):e200389. https://doi.org/10.1148/ryct.2020200389