Clinical Infectious Diseases

MAJOR ARTICLE

A Tool for Early Prediction of Severe Coronavirus Disease 2019 (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong, China

Jiao Gong,[1,a] Jingyi Ou,[2,a] Xueping Qiu,[3,a] Yusheng Jie,[4,5] Yaqiong Chen,[1] Lianxiong Yuan,[6] Jing Cao,[4] Mingkai Tan,[2] Wenxiong Xu,[4] Fang Zheng,[3,b] Yaling Shi,[2,b] and Bo Hu[1,b]

[1]Department of Laboratory Medicine, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, People's Republic of China, [2]Department of Laboratory Medicine, Guangzhou Eighth People's Hospital, Guangzhou Medical University, Guangzhou, People's Republic of China, [3]Center for Gene Diagnosis, Department of Laboratory Medicine, Zhongnan Hospital of Wuhan University, Wuhan, People's Republic of China, [4]Department of Infectious Diseases, Key Laboratory of Liver Disease of Guangdong Province, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, People's Republic of China, [5]Department of Infectious Diseases, Third Affiliated Hospital of Sun Yat-sen University Yuedong Hospital, Meizhou, People's Republic of China, and [6]Department of Science and Research, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, People's Republic of China

Background. Because there is no reliable risk stratification tool for severe coronavirus disease 2019 (COVID-19) patients at admission, we aimed to construct an effective model for early identification of cases at high risk of progression to severe COVID-19.

Methods. In this retrospective multicenter study, 372 hospitalized patients with nonsevere COVID-19 were followed for > 15 days after admission. Patients who deteriorated to severe or critical COVID-19 and those who maintained a nonsevere state were assigned to the severe and nonsevere groups, respectively. Based on baseline data of the 2 groups, we constructed a risk prediction nomogram for severe COVID-19 and evaluated its performance.

Results. The training cohort consisted of 189 patients, and the 2 independent validation cohorts consisted of 165 and 18 patients. Among all cases, 72 (19.4%) patients developed severe COVID-19. Older age; higher serum lactate dehydrogenase, C-reactive protein, coefficient of variation of red blood cell distribution width, blood urea nitrogen, and direct bilirubin; and lower albumin were associated with severe COVID-19. We generated the nomogram for early identifying severe COVID-19 in the training cohort (area under the curve [AUC], 0.912 [95% confidence interval {CI}, .846–.978]; sensitivity 85.7%, specificity 87.6%) and the validation cohort (AUC, 0.853 [95% CI, .790–.916]; sensitivity 77.5%, specificity 78.4%). The calibration curve for probability of severe COVID-19 showed optimal agreement between prediction by nomogram and actual observation. Decision curve and clinical impact curve analyses indicated that nomogram conferred high clinical net benefit.

Conclusions. Our nomogram could help clinicians with early identification of patients who will progress to severe COVID-19, which will enable better centralized management and early treatment of severe disease.

Keywords. COVID-19; nomogram; severe COVID-19 prediction; risk stratification.

Since the outbreak of novel coronavirus pneumonia (coronavirus disease 2019 [COVID-19]) in December 2019, the number of reported cases has surpassed 260 000 with > 11 180 deaths worldwide as of 22 March 2020. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a member of coronaviruses known to cause common colds and severe illnesses, is the cause of COVID-19 [1]. Compared with much higher overall case-fatality rates for the severe acute respiratory syndrome and Middle East respiratory syndrome, COVID-19 is responsible for more total deaths because of the increased transmission speed and the growing numbers of cases [2]. The World Health Organization has raised the COVID-19 outbreak risk to "very high", and SARS-CoV-2 infection has become a serious threat to public health.

According to a report recently released by the Chinese Center for Disease Control and Prevention that included approximately 44 500 confirmed cases of SARS-CoV-2 infection, up to 15.8% were severe or critical. Most COVID-19 patients have a mild disease course, but some patients experience rapid deterioration (particularly within 7–14 days) from onset of symptoms into severe COVID-19 with or without acute respiratory distress syndrome (ARDS) [3]. Current epidemiological data suggest that the mortality rate of patients with severe COVID-19 is higher than that of patients with nonsevere COVID-19 [4, 5]. This situation highlights the need to identify COVID-19 patients at risk of progressing to severe COVID-19. Patients with severe illness often require utilization of intensive medical resources. Therefore, early identification of patients at high risk for progression to severe

Received 23 March 2020; editorial decision 12 April 2020; accepted 14 April 2020; published online April 15, 2020.

[a]J. G., J. O., and X. Q. contributed equally to this work.

[b]B. H., Y. L., and F. Z. are joint corresponding authors.

Correspondence: B. Hu, Department of Laboratory Medicine, Third Affiliated Hospital of Sun Yat-sen University, Tianhe Road 600, Guangzhou 510630, People's Republic of China (hubo@mail.sysu.edu.cn).

Clinical Infectious Diseases® 2020;71(15):833–40

DOI: 10.1093/cid/ciaa443

COVID-19 will facilitate appropriate supportive care and reduce the mortality rate, as well as unnecessary or inappropriate healthcare utilization, via patient prioritization.

At present, an early-warning model for predicting COVID-19 patients at risk of developing a costly condition is lacking [3, 6]. So far, prognostic factors of COVID-19 mainly focus on the immune cells. In our study, we found that older age; higher lactate dehydrogenase (LDH), C-reactive protein (CRP), coefficient of variation of red blood cell distribution width (RDW), direct bilirubin (DBIL), and blood urea nitrogen (BUN); and lower albumin (ALB) on admission correlated with higher odds of severe COVID-19. Based on these indexes, we developed and validated an effective prognostic nomogram with high sensitivity and specificity for accurate individualized assessment of the incidence of severe COVID-19. Among these indices, the prognostic role of RDW in COVID-19 is underestimated, which is associated with the increased turnover of erythrocytes. Our results hinted that the turnover of red blood cells (RBCs) might be involved in severe illness.

MATERIALS AND METHODS

Data Collection

Data on COVID-19 inpatients between 20 January 2020 and 2 March 2020 was retrospectively collected from 3 clinical centers: Guangzhou Eighth People's Hospital, Zhongnan Hospital of Wuhan University, and Third Affiliated Hospital of Sun Yat-sen University. A total of 381 patients with COVID-19 were enrolled; 9 patients < 15 years of age were excluded from the study. Clinical laboratory test results, including SARS-CoV-2 RNA detection results, biochemical indices, and blood routine results, were collected from routine clinical practice. Written informed consent was waived by the ethics commission of each hospital for emerging infectious diseases. The study was approved by the Ethics Committee of the Eighth People's Hospital of Guangzhou, the Ethics Commission of the Third Affiliated Hospital of Sun Yat-sen University, and the Ethics Commission of Zhongnan Hospital.

The diagnosis of SARS-CoV-2 infection was based on the *Guidelines for Diagnosis and Treatment of Novel Coronavirus Pneumonia* (fifth version), released by the National Health Commission of China. Suspected cases of COVID-19 must meet any of the following epidemiological history criteria or any 2 of the following clinical manifestations:

1. Epidemiological history: A history of travel to or residence in Wuhan in the last 14 days prior to symptom onset; contact with a confirmed or suspected case of SARS-CoV-2 infection in the last 14 days prior to symptom onset; or aggressive disease onset.
2. Clinical manifestations: Fever and/or respiratory infection, or with normal/decreased white blood cell counts and normal/decreased lymphocyte counts.

In the absence of the above-mentioned criteria for epidemiological history, the suspected case should meet all of the above-mentioned criteria for clinical manifestations. A confirmed case was defined as an individual with laboratory confirmation of SARS-CoV-2, which required positive results of SARS-CoV-2 RNA, irrespective of clinical signs and symptoms. For diagnosis of severe COVID-19, at least 1 of the following conditions should be met: (1) shortness of breath (respiratory rate ≥ 30 breaths per minute); (2) arterial oxygen saturation (resting status) $\leq 93\%$; or (3) the ratio of partial pressure of oxygen to fraction of inspired oxygen (PaO_2/FiO_2) ≤ 300 mm Hg.

Laboratory Methods

Clinical laboratory test results, including SARS-CoV-2 RNA detection results, biochemical indices, and blood routine results, were collected from routine clinical practice. Clinical laboratory test results included ALB, aspartate aminotransferase, alanine aminotransferase, BUN, creatine kinase, creatine kinase-MB, creatinine, CRP, total bilirubin, DBIL, globulin, LDH, procalcitonin, total bile acid, hemoglobin, lymphocyte count, monocyte count, neutrophil count, platelet distribution width (PDW), platelet count, RBCs, and RDW. SARS-CoV-2 RNA was detected using real-time quantitative polymerase chain reaction on nucleic acid extracted from upper respiratory swab samples, which were collected on all suspected cases of SARS-CoV-2 infection on admission and immediately placed into sterile tubes with viral transport medium. All biochemical and hematologic parameters were obtained via standard automated laboratory methods and using commercially available kits following the manufacturers' protocols.

Statistical Analysis

Categorical variables were expressed as frequency and percentages, and Fisher exact test was performed to analyze the significance. Continuous variables were expressed as mean (standard deviation) or median (interquartile range), as appropriate. Parametric test (*t* test) and nonparametric test (Mann-Whitney *U* test) were used for continuous variables with or without normal distribution, respectively. A value of $P < .05$ was considered statistically significant. Except for filling missing values, all of the statistical analyses were analyzed using R (version 3.6.2) with default parameters.

Of all potential predictors in the dataset, 0.09% of the fields had missing values. Predictor exclusion was limited to those with > 7% missing rate to minimize the bias of the regression coefficient [7]. The Little missing completely at random (MCAR) test (R package, BaylorEdPsych) was used to assess the suitability of the remaining missing values for imputation. This test is used to test whether MCAR or biased. The missing values were imputed by expectation-maximization method using SPSS statistical software, version 25 (IBM SPSS, Chicago, Illinois).

To identify the relative importance of each feature, feature selection was performed using the least absolute shrinkage and selection operator (LASSO) regression method, and prediction models were built using logistic regression, decision tree, random forest, and support vector machine using R package mlr, using 3-fold cross-validation for diverse parameter conditions, respectively. As described previously, nomograms were established with the rms package and the performance of nomograms was evaluated by area under the receiver operating characteristic curve (AUC) and calibration (calibration plots and Hosmer-Lemeshow calibration test) in R. During the external validation of the nomogram, the total points for each patient in the validation cohort were calculated based on the established nomogram.

RESULTS

Clinicopathologic Characteristics of COVID-19 Patients

The selection of the study population is illustrated in Figure 1. A total of 372 COVID-19 patients were enrolled after admission from 3 centers in Guangzhou and Wuhan (Figure 1). All patients with nonsevere COVID-19 during hospitalization were followed for > 15 days after admission. Patients who deteriorated to severe or critical COVID-19 and patients who maintained a nonsevere state were assigned to the severe and nonsevere groups, respectively. There were no significant differences in age, sex, or disease type between the training and validation cohorts (Table 1). In the training cohort, the nonsevere COVID-19 group consisted of 161 (85.2%) patients, with a median age of 45 years (range, 33–62 years); 28 patients (14.8%) with a median age of 63.5 years (range, 54.5–72 years) progressed to severe COVID-19. By the end of 25 February, 1 patient with severe COVID-19 in the training group died. None of the 189 patients from the training group had a history of exposure to the Huanan seafood market in Wuhan; 58 of them (30.7%) had not left Guangzhou recently, but had a close exposure history with COVID-19 patients, and the rest (69.3%) were Wuhan citizens or visited Wuhan recently. Other baseline characteristics of the training cohort are shown in Table 2.

Prognostic Factors of Severe COVID-19

A total of 49 features were collected from each patient in the training cohort. After excluding irrelevant and redundant features, 39 features remained for LASSO regression analysis. The results of the 189 patients showed that age, DBIL, RDW, BUN, CRP, LDH, and ALB were predictive factors for severe COVID-19 with maximal AUC (Figure 2A and 2B). We then built prediction models using logistic regression, decision tree, random forest, and support vector machine, and evaluated their performance by the receiver operating characteristic curve and the precision-recall curve (Supplementary Figure 1). There were no large differences in performance of these models except for decision tree. Therefore, the logistic regression model was used for further analysis owing to its high predictive power and interpretability.

Nomogram Construction

The predictive nomogram that integrated 7 selected features for the incidence of severe COVID-19 in the training cohort is shown in Figure 2C. To evaluate clinical applicability of our risk prediction nomogram, decision curve analysis (DCA) and clinical impact curve analysis (CICA) were performed. The DCA and CICA visually showed that the nomogram had a superior overall net benefit within the wide and practical ranges of threshold probabilities and impacted patient outcomes (Figure 2D and 2E). The calibration plot for severe illness probability showed a good agreement between the prediction by nomogram and actual observation in the training cohort and validation cohort 1, respectively (Figure 3A and 3B).

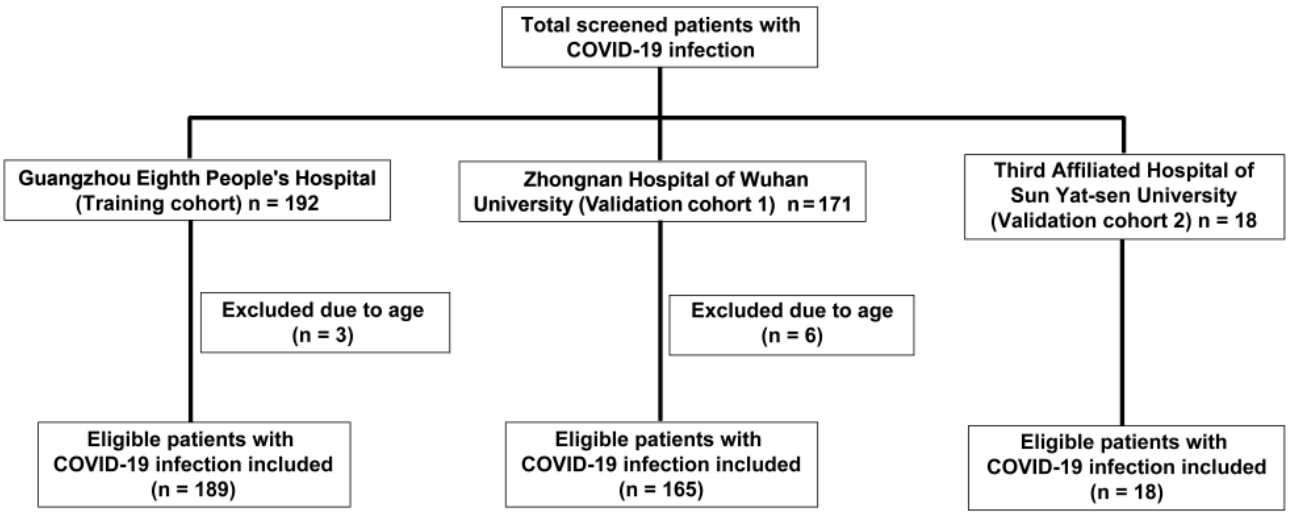

Figure 1. Flowchart of study participants in the training and validation groups. Abbreviation: COVID-19, coronavirus disease 2019.

Table 1. Baseline Characteristics of the Study Cohort

Characteristic	Training Cohort (n = 189)	Validation Cohort 1 (n = 165)	Validation Cohort 2 (n = 18)	*P* Value
Age, y, median (IQR)	49.0 (35.0–63.0)	52.0 (37.0–64.0)	41.5 (29.0–50.0)	.05
Sex				.77
Female	101 (53.4)	93 (56.4)	9 (50.0)	
Male	88 (46.6)	72 (43.6)	9 (50.0)	
Basic disease				.91
No	134 (70.9)	117 (70.9)	14 (77.8)	
Yes[a]	55 (29.1)	48 (29.1)	4 (22.2)	
Disease type				.07
Nonsevere	161 (85.2)	125 (75.8)	14 (77.8)	
Severe	28 (14.8)	40 (24.2)	4 (22.2)	

Data are presented as no. (%) unless otherwise indicated.
Abbreviation: IQR, interquartile range.
[a]Patients with 1 of the following: hypertension, diabetes, cardiovascular disease, chronic respiratory disease, or tuberculosis disease.

Validation of the Predictive Accuracy of Nomogram

In the training cohort, the nomogram had a significantly high AUC (0.912 [95% confidence interval, .846–.978]) to discriminate individuals with severe COVID-19 from those with nonsevere COVID-19, with a sensitivity of 85.7% and specificity of 87.6% (Figure 3C; Table 3). Cutpoint R package was used to calculate optimal cutpoints by bootstrapping the variability of the optimal cutpoints, which was 188.6358 for our nomogram (corresponding to a threshold probability of 0.190). Patients in the validation cohorts were then divided into the low group (score ≤ 188.6358) and the high group (score > 188.6358) for further analysis. Consistent with the training cohort, in validation cohort 1, the AUC was 0.853 for patients with severe COVID-19 vs nonsevere COVID-19 with a sensitivity of 77.5 % and specificity of 78.4% (Figure 3D; Table 3). In validation cohort 2, the sensitivity and the specificity of the nomogram were observed to be 75% and 100%, respectively.

DISCUSSION

Early identification of COVID-19 patients at risk of progression to severe disease will lead to better management and optimal use of medical resources. In this research, we identified that older age; higher LDH, CRP, DBIL, RDW, and BUN; and lower ALB on admission were correlated with higher odds of severe COVID-19. Furthermore, we developed an effective prognostic nomogram composed of 7 features, which had significantly high sensitivity and specificity to distinguish individuals with severe COVID-19 from nonsevere COVID-19. DCA and CICA further indicated that our nomogram conferred significantly high clinical net benefit, which is of great value for accurate individualized assessment of the incidence of severe COVID-19.

So far, several studies have reported some risk factors for severe COVID-19. However, a nomogram could present a quantitative and practical predictor tool for risk stratification of patients with nonsevere COVID-19 admission. Though Liu et al developed a nomogram from a single center with a small sample size and no external validation [6], our nomogram has a significantly higher AUC in the training and validation cohorts than Liu et al's nomogram (0.912/0.853 vs 0.849, respectively). Our nomogram predicted a total of 188.6358 points at a 19.0% probability threshold, which was close to the prevalence of severe COVID-19 (14.8%) in the training cohort and hence consistent with the reality. This cutoff value may lead to a slight increase of false-positive rates but, in the setting of this COVID-19 outbreak, a few high false-positive rates are acceptable to minimize risks of missed diagnosis. Meanwhile, application of the nomogram in the training cohort and validation cohort showed good differentiation with AUC values of 0.912 and 0.853, respectively, as well as high sensitivity and specificity.

Furthermore, only 7 easy-access features were included in our nomogram, including older age; higher LDH, CRP, DBIL, RDW, and BUN; and lower ALB. Age, neutrophil-to-lymphocyte ratio (NLR), and LDH have been reported to be risk factors for patients with severe COVID-19 [3, 6, 8, 9]. NLR, a widely used marker for the assessment of systemic inflammation, was not identified by LASSO as an important feature instead of LDH and CRP, which are associated with the systemic inflammatory response [10]. However, LDH could predict severity of tissue damage in the early stage of diseases as an auxiliary marker [11]. These might be reasons why the LASSO model did not identify NLR as a more important feature. Consistent with other reports, our results indicate that patients with higher levels of inflammation at admission might be at higher risk for severe COVID-19 as well.

Interestingly, we found that RDW was also an important prognostic predictor for severe COVID-19. RDW, which reflects the variation in the size of RBCs, has been reported to be significantly correlated with critical disease [12–14] but neglected in previous studies on COVID-19. It is a robust predictor of the risk of all-cause patient mortality and bloodstream infection in the critically ill [13–16], including acute exacerbation of interstitial pneumonia and ARDS [7, 12]. RDW also can

Table 2. Demographics and Characteristics of Patients With Coronavirus Disease 2019 in the Training Cohort

Characteristic	Nonsevere (n = 161)	Severe (n = 28)	*P* Value
Demographics			
Age, y	45.0 (33.0–62.0)	63.5 (54.5–72.0)	< .01
Sex, No. (%)			.3
Female	89 (55.3)	12 (42.9)	
Male	72 (44.7)	16 (57.1)	
Exposure, No. (%)			.66
Patients in Guangzhou	51 (27.0)	7 (3.7)	
Close contact with Wuhan	110 (58.2)	21 (11.1)	
Basic disease, No. (%)			.01
No	120 (74.5)	14 (50.0)	
Yes	41 (25.5)	14 (50.0)	
BMI, kg/m^2	23.4 (21.4–25.7)	23.4 (22.3–24.4)	.9
Signs and symptoms			
Fever, No. (%)			.11
No	119 (73.9)	16 (57.1)	
Yes	42 (26.1)	12 (42.9)	
Diarrhea, No. (%)			.1
No	156 (96.9)	25 (89.3)	
Yes	5 (3.1)	3 (10.7)	
Respiratory rate (breaths/min)	20.0 (20.0–20.0) (n = 160)	20.0 (20.0–22.0) (n = 28)	.04
Laboratory test			
PaO_2, kPa	12.9 (10.7–15.7) (n = 155)	10.9 (9.6–13.0) (n = 27)	.04
SaO_2, %	97.9 (96.7–98.8) (n = 157)	96.8 (95.2–97.8) (n = 28)	.02
WBC, $10^9/L$	4.6 (3.7–5.6)	5.2 (4.4–6.7)	.03
RBC, $10^{12}/L$	4.5 (0.6)	4.2 (0.6)	.02
Hemoglobin, g/L	136.8 (16.7)	128.9 (17.3)	.02
Platelets, $\times 10^9/L$	180.0 (147.0–221.0)	167.0 (139.5–200.0)	.09
Neutrophils, $\times 10^9/L$	2.8 (2.0–3.6)	3.7 (2.8–5.2)	< .01
Monocytes, $\times 10^9/L$	0.4 (0.3–0.5)	0.3 (0.3–0.4)	.51
Lymphocytes, $\times 10^9/L$	1.3 (1.0–1.8)	1.0 (0.8–1.4)	< .01
NLR	1.9 (1.4–2.9)	3.7 (2.0–6.7)	< .01
PLR	131.0 (96.6–177.4)	174.8 (117.7–210.0)	.05
SII	360.5 (229.1–562.9)	561.7 (320.1–1019.8)	< .01
RDW-SD, fL	39.9 (38.5–42.0)	42.7 (39.6–44.1)	< .01
RDW, %	12.2 (11.8–12.7)	12.8 (12.3–13.1)	< .01
PDW	11.9 (10.6–14.1)	14.9 (10.9–16.2)	.03
AST, U/L	20.8 (17.4–27.1)	33.5 (27.4–46.5)	< .01
ALT, U/L	21.0 (14.2–32.4)	23.0 (15.1–40.5)	.33
Albumin, g/L	39.7 (4.3) (n = 158)	34.2 (5.1) (n = 28)	< .01
Globulin, g/L	28.3 (26.2–30.2) (n = 157)	29.3 (27.8–32.0) (n = 26)	.07
BUN, mmol/L	3.9 (3.2–4.6)	4.7 (3.1–7.2)	.08
Creatine, μmol/L	58.8 (47.6–76.7)	57.0 (42.5–80.7)	.52
TBIL, μmol/L	9.6 (6.5–14.1) (n = 158)	12.3 (8.6–20.4) (n = 28)	.03
DBIL, μmol/L	3.9 (2.7–5.2) (n = 157)	5.2 (3.4–7.8) (n = 26)	< .01
TBA, μmol/L	2.7 (1.5–4.1) (n = 157)	3.9 (2.3–7.7) (n = 26)	.01
CK-MB, U/L	11.6 (5.0) (n = 150)	16.4 (16.8) (n = 27)	< .01
Creatine kinase, U/L	76.5 (50.0–111.0) (n = 160)	111.5 (72.5–168.5) (n = 28)	< .01
LDH, U/L	175.5 (148.5–219.5) (n = 160)	296.0 (203.0–407.0) (n = 28)	< .01
CRP, mg/L	5.0 (5.0–19.5)	35.5 (21.6–72.3)	< .01
Glucose, mmol/L	6.1 (2.4) (n = 149)	8.2 (4.4) (n = 26)	< .01
Lactate, mmol/L	1.8 (1.4–2.1) (n = 152)	1.9 (1.4–2.3) (n = 25)	.19
INR	1.0 (1.0–1.1) (n = 159)	1.1 (1.0–1.1) (n = 28)	.59
APTT, sec	39.1 (4.4) (n = 159)	40.0 (5.4) (n = 28)	.32
D-dimer, μg/L	990.0 (600.0–1380.0) (n = 158)	1225.0 (6.6–1720.0) (n = 28)	.25
SAA, mg/L	1.0 (0.0–2.0) (n = 104)	4.0 (4.0–4.0) (n = 20)	< .01
PCT, ng/mL	0.0 (0.0–0.1) (n = 32)	0.2 (0.1–0.3) (n = 12)	< .01

Data are presented as mean (SD) or median (IQR), as appropriate unless otherwise indicated. All features with missing values are labeled with a specific number of samples.

Abbreviations: ALT, alanine aminotransferase; APTT, activated partial thromboplastin time; AST, aspartate aminotransferase; BMI, body mass index; BUN, blood urea nitrogen; CK-MB, creatine kinase myocardial band; CRP, C-reactive protein; DBIL, direct bilirubin; INR, international normalized ratio; IQR, interquartile range; LDH, lactate dehydrogenase; NLR, neutrophil-to-lymphocyte ratio; PaO_2, partial pressure of oxygen; PCT, procalcitonin; PDW, platelet distribution width; PLR, platelet-to-lymphocyte ratio; RBC, red blood cell; RDW, red blood cell distribution width–coefficient variation; RDW-SD, red blood cell distribution width–standard deviation; SAA, serum amyloid A; SaO_2, oxygen saturation; SD, standard deviation; SII, systemic immune inflammation index; TBA, total bile acids; TBIL, total bilirubin; WBC, white blood cell.

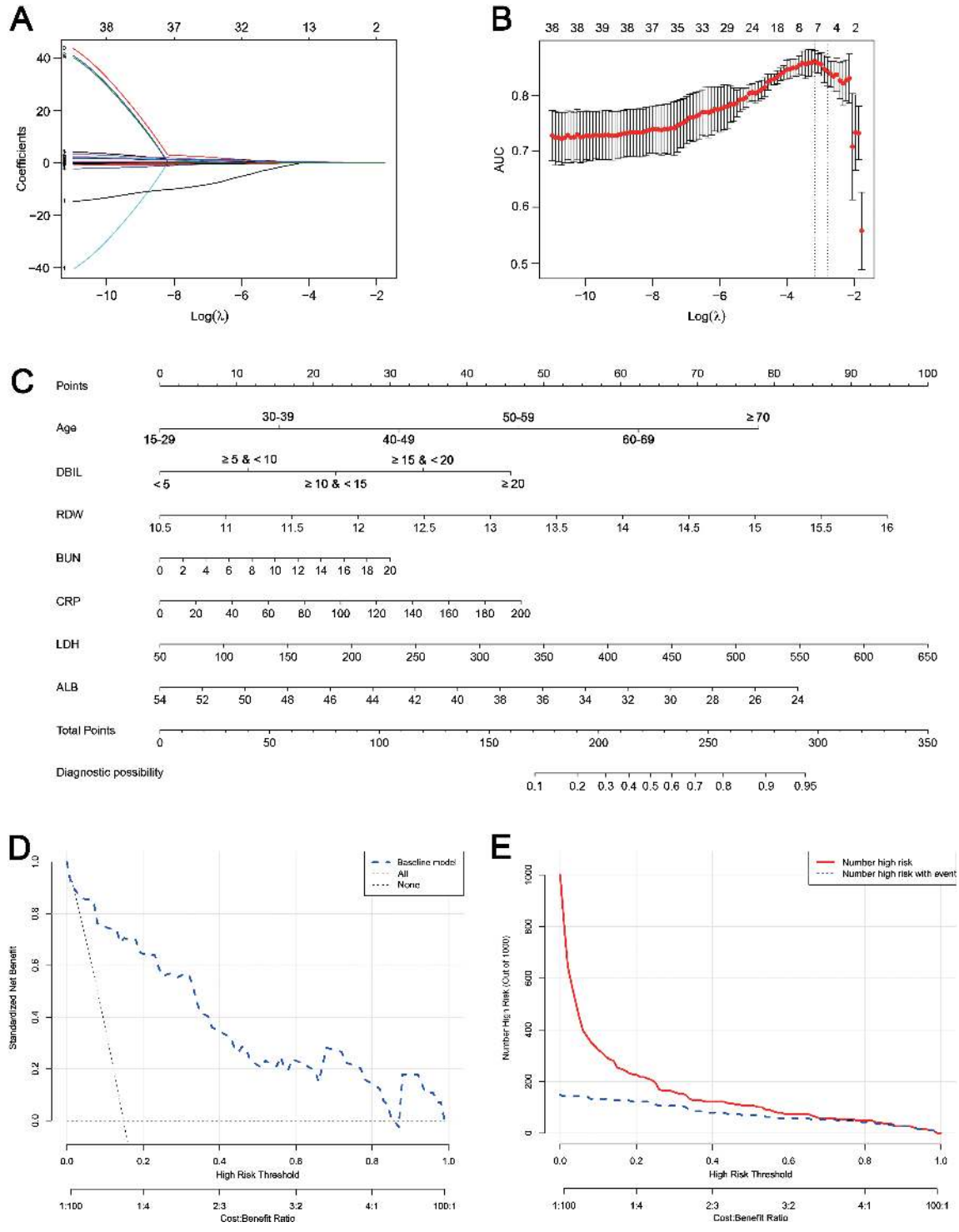

Figure 2. Construction of prediction nomogram in patients with coronavirus disease 2019 (COVID-19). The nomogram is composed of age, direct bilirubin (DBIL), red blood cell distribution width–coefficient variation (RDW), blood urea nitrogen (BUN), C-reactive protein (CRP), lactate dehydrogenase (LDH), and albumin (ALB). *A*, Least absolute shrinkage and selection operator (LASSO) coefficient profiles (y-axis) of the 39 features. The top x-axis shows the average number of predictors. *B*, Identification of the optimal penalization coefficient (λ) in the LASSO model was performed via 3-fold cross-validation based on minimum criteria. The area under the receiver operating characteristic curve (AUC) was plotted verse log(λ). Red dots represent average AUC for each model with a given λ, and vertical bars through the red dots show the upper and lower values of the AUC. The dotted vertical lines represent the optimal values of λ. When the optimal λ value of 0.042 with log(λ) = –3.17 was selected, the AUC reached the peak. The upper and lower x-axes indicate the same meaning as in (*A*). *C*, Nomogram predicting severe COVID-19 probability in patients with COVID-19 infection. To use this nomogram in clinical management, an individual patient's value is located on each variable axis, and a line is plotted upward to calculate the number of points received for each variable value. The sum of these scores is located on the total points axis and draws a line straight down to find the probability of severe COVID-19. *D*, Decision curve compares the net clinical benefits of 3 scenarios in predicting the probability of severe COVID-19: a perfect prediction model (gray line), screen none (horizontal solid black line), and screen based on the nomogram (blue thick dash line). *E*, Clinical impact curve of the nomogram plots the number of COVID-19 patients classified as high risk, and the number of cases classified as high risk with severe COVID-19 at each high risk threshold.

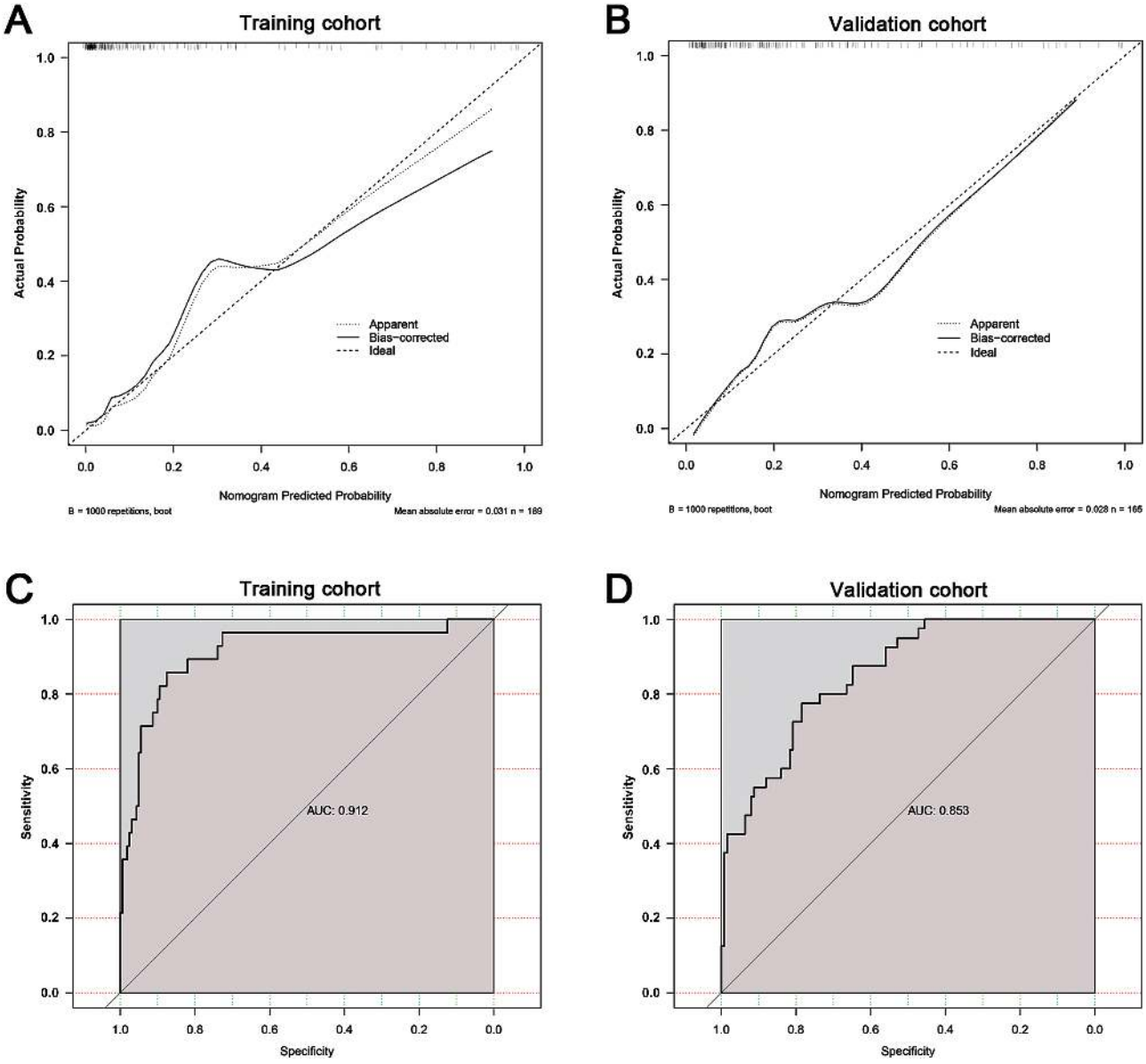

Figure 3. Calibration curve and receiver operating characteristic curve for performance to distinguish individuals with severe coronavirus disease 2019 (COVID-19) from those with nonsevere COVID-19 in the training cohort (*A* and *C*) and validation cohort 1 (*B* and *D*), respectively. Abbreviation: AUC, area under the receiver operating characteristic curve.

predict prognosis of sepsis, which was tied to poor COVID-19 outcomes and death [17].

The increased RDW in COVID-19 patients may be due to the increased turnover of erythrocytes: (1) Proinflammatory states may be responsible for insufficient erythropoiesis with structural and functional alteration of RBCs, such as decreased deformability leading to more rapid clearing of RBCs. (2) Plasma cytokines such as interleukin 1 and tumor necrosis factor–α can not only attenuate renal erythropoietin (EPO) production, but also blunt the erythroid progenitor response to EPO. In addition, interferin-γ contributes to apoptosis of the erythroid progenitors and decrease the EPO receptor expression [18]. (3) RBCs are dynamic reservoirs of cytokines [19]. Decreased deformability of RBCs in severe illness leads to RBC lysis and release of intracellular contents into the circulation [20], including some inflammatory cytokines. This positive feedback could greatly promote the apparent shortened RBC survival and ultimately more morphological variations in cell sizes (ie, elevated RDW), increase inflammatory response, and lead to severe illness. RDW can be regarded as an index of enhanced patient fragility and higher vulnerability to adverse outcomes [21]. The elevated RDW may explain fatigue experienced by patients with severe COVID-19.

Our study has several strengths. First, we provide a practical quantitative prediction tool based on only 7 features, which are relatively inexpensive and easily obtained from routine blood

Table 3. Performance of Nomogram for Early Prediction of Severe Coronavirus Disease 2019

	Severe vs Nonsevere COVID-19		
Cohort	AUC (95% CI)	Sensitivity, %	Specificity, %
Training cohort (n = 189)	0.912 (.846–.978)	85.7	87.6
Validation cohort 1 (n = 165)	0.853 (.790–.916)	77.5	78.4
Validation cohort 2[a] (n = 18)	…	75	100

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; COVID-19, coronavirus disease 2019.
[a]Owing to the limited sample size, AUC was not calculated in validation cohort 2.

tests. Second, to guarantee the robustness of the conclusion, we included data from 3 centers with a large sample size and validation in independent cohorts. The performance of our nomogram was efficient for clinical practice.

There are some limitations to this study. First, this is a retrospective study, including 372 patients with nonsevere COVID-19 on admission. ACE2, a receptor for SARS-CoV-2, has been reported to be differentially expressed in different populations [22]. The differences in patient profiles and healthcare might have an effect on the performance of the nomogram in other populations outside China. Further studies on different populations with larger patient cohorts are required to validate our findings. Second, some patients included in our study have not been discharged yet, so their condition may change with follow-up. The final survival outcome is lacking. Third, the study did not include detection of immunoglobulin M or immunoglobulin G. More comprehensive investigations need to be conducted to explain the characteristics of the 7 parameters included in the nomogram.

In summary, our data suggest that our nomogram can identify and predict COVID-19 patients at risk of severe disease, and RDW was also valuable for this prediction. Our nomogram is especially valuable for risk stratification management, which will be helpful for alleviating insufficiency of medical resources and reducing mortality.

Supplementary Data

Supplementary materials are available at *Clinical 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

Author contributions. B. H., Y. S., and F. Z. designed the study, had full access to all of the data in the study, and take responsibility for the integrity of the data and the accuracy of the data analysis. J. G., J. O., X. Q., Y. J., Y. C., and M. T. contributed to collection of data, analysis of data, and writing of the manuscript. L. Y., J. C., M. T., and W. X. contributed to the statistical analysis. All authors contributed to data interpretation and reviewed and approved the final version.

Financial support. This work is funded by the Science and Technology Program of Guangzhou, China (grant number 201903010039).

Potential conflicts of interest. The authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

References

1. Mo P, Xing Y, Xiao Y, et al. Clinical characteristics of refractory COVID-19 pneumonia in Wuhan, China. Clin Infect Dis **2020**.
2. 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 [manuscript published online ahead of print 24 February 2020]. JAMA **2020**. doi:10.1001/jama.2020.2648.
3. Feng Z, Yu Q, Yao S, et al. Early prediction of disease progression in 2019 novel coronavirus pneumonia patients outside Wuhan with CT and clinical characteristics. medRxiv. doi:10.1101/2020.02.19.20025296, posted online 23 February 2020, preprint: not peer reviewed.
4. Ji D, Zhang D, Xu J, et al. Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. Clin Infect Dis **2020**.
5. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet **2020**; 395:507–13.
6. Liu J, Liu Y, Xiang P, et al. Neutrophil-to-lymphocyte ratio predicts severe illness patients with 2019 novel coronavirus in the early stage. medRxiv. doi:10.1101/2020.02.10.20021584, posted online 12 February 2020, preprint: not peer reviewed.
7. Nei Y, Momose M, Nishiyama K, et al. RDW at hospital admission may predict prognosis of the patient with acute exacerbation of interstitial pneumonia. In: American Thoracic Society International Conference Abstracts **2019**:199.1.
8. Yan L, Zhang HT, Xiao Y, et al. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv. doi:10.1101/2020.02.27.20028027, posted online 3 March 2020, preprint: not peer reviewed.
9. Mo P, Xing Y, Xiao Y, et al. Clinical characteristics of refractory COVID-19 pneumonia in Wuhan, China [manuscript published online ahead of print 16 March 2020]. Clin Infect Dis **2020**. doi:10.1093/cid/ciaa270.
10. Manikhas GM, Beliak NP, Kutukova SI, Zhukova NV, Popova NV, Erdniev S. The prognostic value of systemic inflammatory factors in patient with metastatic gastric cancer. J Clin Oncol **2017**; 35(15 Suppl):e15505-e.
11. Nyakas M, Aamdal E, Jacobsen KD, et al. Prognostic biomarkers for immunotherapy with ipilimumab in metastatic melanoma. Clin Exp Immunol **2019**; 197:74–82.
12. Wang B, Gong Y, Ying B, Cheng B. Relation between red cell distribution width and mortality in critically ill patients with acute respiratory distress syndrome. Biomed Res Int **2019**; 2019:1942078.
13. Havens JM, Seshadri AJ, Salim A, Christopher KB. Red cell distribution width predicts out of hospital outcomes in critically ill emergency general surgery patients. Trauma Surg Acute Care Open **2018**; 3:e000147.
14. Schepens T, De Dooy JJ, Verbrugghe W, Jorens PG. Red cell distribution width (RDW) as a biomarker for respiratory failure in a pediatric ICU. J Inflamm (Lond) **2017**; 14:12.
15. Said AS, Spinella PC, Hartman ME, et al. RBC distribution width: biomarker for red cell dysfunction and critical illness outcome? Pediatr Crit Care Med **2017**; 18:134–42.
16. Bazick HS, Chang D, Mahadevappa K, Gibbons FK, Christopher KB. Red cell distribution width and all-cause mortality in critically ill patients. Crit Care Med **2011**; 39:1913–21.
17. Mahmood NA, Mathew J, Kang B, DeBari VA, Khan MA. Broadening of the red blood cell distribution width is associated with increased severity of illness in patients with sepsis. Int J Crit Illn Inj Sci **2014**; 4:278–82.
18. Sarkar M, Rajta PN, Khatana J. Anemia in chronic obstructive pulmonary disease: prevalence, pathogenesis, and potential impact. Lung India **2015**; 32:142–51.
19. Karsten E, Breen E, Herbert BR. Red blood cells are dynamic reservoirs of cytokines. Sci Rep **2018**; 8:3101.
20. Janz DR, Ware LB. The role of red blood cells and cell-free hemoglobin in the pathogenesis of ARDS. J Intensive Care **2015**; 3:20.
21. Fava C, Cattazzo F, Hu ZD, Lippi G, Montagnana M. The role of red blood cell distribution width (RDW) in cardiovascular risk assessment: useful or hype? Ann Transl Med **2019**; 7:581.
22. Cao Y, Li L, Feng Z, et al. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. Cell Discov **2020**; 6:11.