Received: 9 July 2021 | Accepted: 22 October 2021

DOI: 10.1002/jmv.27428

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

Development and validation of a clinical prediction model to estimate the risk of critical patients with COVID-19

Wenyu Chen[1] | Ming Yao[2] | Lin Hu[3] | Ye Zhang[4] | Qinghe Zhou[2] | Hongwei Ren[5] | Yanbao Sun[6] | Ming Zhang[1] | Yufen Xu[7]

[1]Department of Respiration, Affiliated Hospital of Jiaxing University/The First Hospital of Jiaxing, Jiaxing, China

[2]Department of Pain Medicine Center, Affiliated Hospital of Jiaxing University/The First Hospital of Jiaxing, Jiaxing, China

[3]Department of Oncology, Tianyou Hospital Affiliated to Wuhan University of Science&Technology, Wuhan, China

[4]Department of General Medicine, Affiliated Hospital of Jiaxing University/The First Hospital of Jiaxing, Jiaxing, China

[5]Department of Radiology, Tianyou Hospital Affiliated to Wuhan University of Science&Technology, Wuhan, China

[6]Department of Radiology, Affiliated Hospital of Jiaxing University/The First Hospital of Jiaxing, Jiaxing, China

[7]Department of Oncology, Affiliated Hospital of Jiaxing University/The First Hospital of Jiaxing, Jiaxing, China

Correspondence

Wenyu Chen, Department of Respiration in Affiliated Hospital of Jiaxing University/The First Hospital of Jiaxing. 1882 Zhonghuan South Rd, Jiaxing City, 314000 Zhejiang, China.
Email: 00135116@zjxu.edu.cn

Yufen Xu, Department of Oncology in Affiliated Hospital of Jiaxing University/The First Hospital of Jiaxing. 1882 Zhonghuan South Rd, Jiaxing City, 314000 Zhejiang, China.
Email: 135880607@163.com

Funding information

Jiaxing Key Laboratory of Lung Cancer Precise Treatment and A Project Supported by Scientific Research Fund of Zhejiang Provincial Education Departmen (Y202043729); Key Discipline of Jiaxing Respiratory Medicine Construction Project (No. 2019-zc-04); Jiaxing Fight Novel Coronavirus Pneumonia Emergency Technology Attack Special Project in 2020 (2020GZ30001)

Abstract

The outbreak of coronavirus disease 2019 (COVID-19) has globally strained medical resources and caused significant mortality. This study was aimed to develop and validate a prediction model based on clinical features to estimate the risk of patients with COVID-19 at admission progressing to critical patients. Patients admitted to the hospital between January 16, 2020, and March 10, 2020, were retrospectively enrolled, and they were observed for at least 14 days after admission to determine whether they developed into severe pneumonia. According to the clinical symptoms, all patients were divided into four groups: mild, normal, severe, and critical. A total of 390 patients with COVID-19 pneumonia were identified, including 212 severe patients and 178 nonsevere patients. The least absolute shrinkage and selection operator (LASSO) regression reduced the variables in the model to 6, which are age, number of comorbidities, computed tomography severity score, lymphocyte count, aspartate aminotransferase, and albumin. The area under curve of the model in the training set is 0.898, and the specificity and sensitivity were 89.7% and 75.5%. The prediction model, nomogram might be useful to access the onset of severe and critical illness among COVID-19 patients at admission, which is instructive for clinical diagnosis.

KEYWORDS

COVID-19, nomogram, prediction model

Abbreviations: ARDS, acute respiratory distress syndrome; AUC, area under curve; CICA, clinical impact curve analysis; COPD, chronic obstructive pulmonary disease; COVID-19, coronavirus disease 2019; DCA, decision curve analysis; GGO, ground glass opacity; ICU, intensive care unit; LASSO, least absolute shrinkage and selection operator; MERS, Middle East respiratory syndrome; RT-PCR, real-time reverse transcription polymerase chain reaction; SARS, severe acute respiratory syndrome; SVM, support vector machine; TSS, total severity score; WHO, World Health Organization.

1 | INTRODUCTION

A novel coronavirus was first discovered in Wuhan, China in December 2019, and has caused an unprecedented global health emergency.[1,2] As of December 25, 2020, the World Health Organization (WHO) counted more than 79.39 million confirmed cases and more than 1.74 million deaths.[3] Most patients with coronavirus disease 2019 (COVID-19) have mild clinical features, some even show no obvious symptoms at the time of infection, but then develop fatal complications, including severe pneumonia, acute respiratory distress syndrome, and multiple organ failure.[4] At present, there is no specific anti-coronavirus therapy for patients with severe symptoms, and the case fatality rate is about 20 times that of noncritically ill patients, which means the survival rate of these patients is very low. Treatment of these patients usually requires a large amount of medical resources.[5] Therefore, it has important clinical significance to identify severe patients with COVID-19 in the early.

The clinical classification of COVID-19 patients ranged from mild to severe, which is closely related to computed tomography (CT) findings.[6] A retrospective study showed that the probability of COVID-19 patients at admission developing severe disease ranged from 15.7% to 26.1%. These cases are usually associated with abnormal chest CT manifestations and clinical laboratory data.[7] Guan et al.[8] showed that patients with severe COVID-19 were more prone to ground-glass opacity (GGO), local or bilateral patchiness, and Interstitial lung abnormalities on CT, which can reflect the clinical progress of the disease, and also provide opportunities for the research of CT clinical utility, as the risk stratification of patients with a forecasting tool. In addition, CT has predictive value in the prognosis of patients with COVID-19 to ensure effective treatment and control the spread of disease. From previous experience, higher CT scores in patients with Middle East respiratory syndrome (MERS) lead to poor prognosis.[9] CT has been suggested as an auxiliary method, used for screening individuals suspected of COVID-19 pneumonia during epidemics and monitoring treatment responses based on dynamic radiological changes.[10] In addition, the guidelines of the WHO and the European Centre for Disease Control and Prevention recommend that people aged 70 and over or who have basic diseases (e.g., cardiovascular disease, hypertension, cancer, chronic obstructive pulmonary disease (COPD), asthma, and diabetes) are thought to be at higher risk of developing severe COVID-19. It has become an inevitable trend to predict the prognosis of patients with COVID-19 pneumonia by combining CT imaging results with epidemiological characteristics and laboratory examination results.

At present, some early prediction models using machine learning have been reported to predict whether COVID-19 patients may develop into severe or critical illnesses.[11–13] These models are usually evaluated by statistical methods for identification and calibration. At the same time, decision curve analysis (DCA) can evaluate the clinical utility of decision models, and the risk models it identifies can help us make better clinical decisions.[14] The purpose of this study is to develop and validate a prognostic machine learning model based on the clinical, laboratory, and imaging characteristics of COVID-19 patients at the time of admission to identify the risk of serious/critical complications in hospitalized COVID-19 patients. We also use DCA and clinical impact curve analysis (CICA) to evaluate the clinical utility and net benefit of predictive models in supporting clinical decision-making. The model can be used as a tool for the early identification of high-risk patients with a poor prognosis for COVID-19 during hospitalization.

2 | METHODS

2.1 | Study design and participants

This was a retrospective study done at several centers, including the First Affiliated Hospital of Zhejiang University School of Medicine, Jiaxing First Hospital, and Tianyou Hospital Affiliated to Wuhan University of Science and Technology. A total of 390 patients with confirmed COVID-19 pneumonia between January 16, 2020, and March 10, 2020, were retrospectively enrolled (previously exclude patients younger than 18 years old). This study was conducted in accordance with the guidelines of the Declaration of Helsinki and the principles of good clinical practice. The study was approved by the medical ethics committees of various hospitals. For emerging infectious diseases, the Ethics Committee of hospitals gave written informed consent.

2.2 | Diagnosis and clinical classification

All patients enrolled met the following diagnostic criteria: high-throughput sequencing of nasal cavity and throat swab specimens or real-time reverse transcription polymerase chain reaction test results are positive.[1] Patients are observed for at least 14 days after admission to determine whether they develop into severe pneumonia. According to the clinical symptoms, such as pneumonia, respiratory failure, shock, and other organ failures, all patients are divided into four groups: mild, normal, severe, and critical.

(1) Mild type, with mild clinical symptoms, and no pneumonia manifestations in imaging.

(2) Normal type, the patient has a fever, respiratory symptoms, and other symptoms, and pneumonia can be seen on imaging.

(3) Severe type, the patient has one of the following symptoms: (a) respiratory distress (respiratory rate ≥30 breaths/min); (b) hypoxia (oxygen saturation ≤93% in a resting state); (c) low oxyemia (arterial blood oxygen partial pressure/inspired oxygen concentration ≤300 mmHg).

(4) Critical type, have one of the following conditions: (a) respiratory failure requires mechanical ventilation; (b) shock occurs; (c) combined other organ failure requires intensive care unit.

In this study, no mild patients and no patients developed into critical illness, so all patients were divided into severe and nonsevere cases.

2.3 | Inclusion and evaluation criteria

The professional clinical team reviews, extracts, and cross-checks the data, and two clinicians check the recorded results independently.

The data extracted from the electronic medical records are as follows: demographic information (age, gender), contact history (defined as a contact in Wuhan within two weeks before the onset of the disease, or contact with local residents diagnosed with SARS-CoV-2 infection), smoking history, comorbidities (mainly including diabetes, hypertension, cardiovascular disease, COPD, malignant tumors, and chronic liver disease, scored according to the number of comorbidities, the maximum is not more than 5 points if there is no score, 0 points), clinical symptoms and signs include categorical and continuous variables: body temperature at admission, the main symptoms include fever, cough/sputum, hemoptysis, chest tightness/shortness of breath, other symptoms include fatigue, loss of appetite, dyspnea, malaise, diarrhea, headache, and so forth (score based on the number of symptoms, up to 7 points). Laboratory examination results are as follows: blood routine indicators include platelet, white blood cell, neutrophil, lymphocyte, hematocrit, monocyte, eosinophil, hemoglobin, and red blood cell. Biochemical indicators include alanine aminotransferase, aspartate aminotransferase, albumin, creatinine, serum sodium, serum potassium, and C-reactive protein. Blood gas analysis includes PaO_2, $PaCO_2$, and pH. In addition, erythrocyte sedimentation rate, fibrinogen, procalcitonin, and length of hospital stay were also recorded. The imaging findings include abnormalities in CT and its severity.

In this study, two radiologists with more than 10 years of experience in chest imaging conducted an independent review of all CT images. When the two radiologists diverged in interpreting the imaging results, a third experienced radiologist gave his opinion. The CT recorded imaging features included GGO, pulmonary consolidation, crazy paving, and air bronchogram.[15,16] GGO is defined as lung blur aggravating attenuation and preservation of bronchi and blood vessel edges, while lung consolidation is defined as blurring and opacity of blood vessel edges and airway walls. Crazy paving refers to the thickening of the lobular septum and the appearance of thickened ground glass in the interlobular septum. The distribution pattern of the lesions and the involvement of the lobes and segments were also evaluated. The CT appearance of the outer third of the lung is defined as the periphery, and the CT appearance of two-thirds of the lung is defined as the center. In addition, the presence of discrete nodules, swollen lymph nodes, and pleural effusion was also recorded.[17] Each of the five lung lobes was examined, and the degree of lesions in each lobe was evaluated semiquantitatively from 0 to 5 according to the degree of involvement: score 0, no participation; score 1, participation ≤5%; Score 2, participation rate 6%–25%; score 3, participation rate 26%–50%; score 4, participation 51%–75%; score 5, participation rate >75%.[16] The total score is obtained by summing the scores of all five leaves to provide a CT severity score between 0 and 25.

2.4 | Variable selection and model construction

Use least absolute shrinkage and selection operator (LASSO) regression analysis to filter the variables we need. Using the R "caret" package, 390 patients were randomly divided into training sets and validation sets at a ratio of 1:1. Apply LASSO regression to minimize the potential collinearity between different variables and variable overfitting. If the missing value is <20%, consider filling in the missing variable. We use predictive mean matching to estimate numerical features, logistic regression to estimate binary variables, and Bayesian polymorphic regression to estimate factor features. We use L1 minimized LASSO regression for multivariate analysis, and use 10-fold cross-validation for internal validation. This is a logistic regression model, which penalizes the absolute size of the regression model's coefficients based on the value of λ. When the penalty is greater, the estimate of the weaker factor will approach zero, so only the strongest predictor variable remains in the model. The best variable is selected by the smallest λ value. The R software package "glmnet" statistical software (R Foundation) was used to perform LASSO regression. Subsequently, the variables determined by the LASSO regression analysis were subjected to logistic regression analysis to construct our risk prediction model.

2.5 | Model validation

The calibration of the nomogram is evaluated by the calibration curve, and the Hosmer-Lemeshow test is performed to evaluate the goodness of fit. The area under the ROC curve (AUC) is used to quantify the discriminant performance of the nomogram. The external verification of the nomogram is carried out through the verification queue. DCA can estimate the net benefit of the model based on the difference between true positive and false-positive results, weighted by the probability of the selected threshold risk probability. If within a reasonable risk threshold, the predictive model's net profit curve is higher than "full treatment" or "no treatment," then the model has clinical utility. On this basis, we further drew the CICA of the model. CICA will display the estimated number of people declared high-risk for each risk threshold, and visually display the proportion of cases (true positives).

2.6 | Statistical analysis

All statistical analysis was performed with R statistical software (version 4.0.1). Continuous variables are expressed as averages and ranges, categorical variables are expressed as counts and percentages, and nonparametric rank-sum tests are used. Factors with significant differences are used in LASSO regression analysis. $p < 0.05$ is considered to have a significant difference.

3 | RESULTS

3.1 | Baseline of patients

According to the inclusion criteria, a total of 390 patients with COVID-19 pneumonia were identified, including 212 severe patients and 178 nonsevere patients. Compared with nonsevere patients, the age of severe patients is significantly older (average age, 61.4 vs. 52.1, $p = 0.003$). Significant differences were also found in gender, comorbidities, exposure history, clinical symptoms, CT score, oxygen saturation at admission, WBC count, neutrophil count, lymphocyte count, AST, ALT, ALB, Na^+, CRP, PaO_2, $PaCO_2$, pH, ESR, FIB, and PCT ($p < 0.05$) (Table 1).

3.2 | Risk factor screening and model construction

A total of 32 independent variables were collected from the original data, and 11 insignificant factors were eliminated through the nonparametric rank-sum test. We used 1:1 nonrepetitive random sampling on the original data and divided it into a training set and a validation set (Figure 1). The remaining 21 independent variables were used for LASSO regression analysis. Finally, six independent variables were selected for multivariate logistic regression analysis (Figure 2). The results showed that age, number of comorbidities, CT severity score, lymphocyte count, aspartate aminotransferase, and albumin are risk factors for the progression of COVID-19 disease (Table 2). The constructed nomogram of COVID-19 disease progression risk prediction is shown in Figure 3. Based on the above factors, the basic characteristics of severe and noncritical patients in the training set and validation set were compared (Table 3).

3.3 | Model verification

The ROC curve shows the accuracy of the nomogram for COVID-19 disease progression risk prediction. The AUC of the nomogram in the training set is as high as 0.898, which can distinguish severe COVID-19 patients from nonsevere patients, and the specificity and sensitivity were 89.7% and 75.5%, respectively. Consistent with the training set, the AUC of the nomogram in the validation set was 0.903, and the specificity and sensitivity were 84.0% and 86.0%, respectively (Figure 4). The calibration curve of the nomogram for the risk prediction of COVID-19 disease progression shows that there is good agreement between the prediction and actual observation in the training set and the validation set. The Hosmer-Lemeshow test shows that there is no statistical difference between the training set and the validation set (training set, $p = 0.995$; validation set, $p = 0.886$), which proves that the diagnostic accuracy of the nomogram is high (Figure 5).

To evaluate the clinical applicability of our risk prediction nomogram, we conducted DCA and CICA. DCA and CICA intuitively show that within a wide and practical threshold probability range (3%–89% in the training set, 7%–63% in the validation set) and within the range that affects the prognosis of patients, the nomogram has good overall net income (Figure 6).

4 | DISCUSSION

With the increasing number of COVID-19 diagnosed and critically ill patients, the management of critically ill patients has become one of the most challenging issues. Gong et al.[18] established a predictive model for the disease progression of patients with COVID-19 pneumonia by logistic regression, decision tree, random forest, and support vector machine (SVM), and verified that the prediction performance of logistic regression, random forest, and SVM had no significant difference and that the nomogram constructed based on logistic regression had high accuracy. However, in the selection of influencing factors, the clinical symptoms and CT examination results of patients during hospitalization were not taken into consideration. Xu et al.[19] confirmed that chest CT was an important diagnostic tool for COVID-19, but it was mentioned in their study that they did not find any correlation between imaging and disease progression. In this multicenter study, we retrospectively evaluated the clinical and CT characteristics of COVID-19 patients and identified clinical baseline risk factors related to disease progression. Our results show that older age at admission, a larger number of complications, a higher lymphocyte count, a higher level of aspartate aminotransferase, a lower albumin level, and a higher CT severity score promote the progression of the course of COVID-19 pneumonia patients key prognostic factors.

At present, the origin of COVID-19 pneumonia is not clear. Early scientists completed the sequencing of the genome of the virus that caused the disease and laid the foundation for the development of effective antiviral drugs and vaccines in the future.[20] On the other hand, a large number of studies currently analyze the clinical data of patients with COVID-19 pneumonia, which can predict the patient's disease progression and take targeted treatment measures to effectively prevent further aggravation of the disease. Age is the most common factor affecting disease progression. Previously, older age was reported as an important independent predictor of mortality from severe acute respiratory syndrome (SARS) and MERS, which is also applicable in COVID-19.[21–23] A recent retrospective study showed that older age, comorbidities, and severe CT scores are risk factors for COVID-19 severe/critical pneumonia.[24] Another factor that is more closely related to it is comorbidities. It is not difficult to explain that the physical fitness of the elderly is significantly reduced, and they are more likely to suffer from multiple complications. Different from the underlying disease, a large cohort study suggested that cancer patients are more likely to be infected with SARS-CoV-2 and are prone to develop severe

TABLE 1 Baseline of patients

Characteristics	Total (*n* = 390)	Sever (*n* = 212)	Nonsever (*n* = 178)	*p* value
Age (year)	57.1 (18–89)	61.4 (18–89)	52.1 (30–88)	0.003
Sex				0
Male	211 (54.1)	122 (57.5)	89 (50)	
Female	179 (45.9)	90 (42.5)	89 (50)	
Num of comorbidity				0
0	151 (38.7)	26 (12.3)	125 (70.2)	
1	107 (27.4)	91 (42.9)	16 (9)	
2	63 (16.2)	56 (26.4)	7 (3.9)	
3	34 (8.7)	18 (8.5)	16 (9)	
4	20 (5.1)	10 (4.7)	10 (5.6)	
5	15 (3.8)	11 (5.2)	4 (2.2)	
Smoking history				0.725
Yes	26 (6.7)	15 (7.1)	11 (6.2)	
No	364 (93.3)	197 (92.9)	167 (93.8)	
Wuhan contact history				0
Yes	250 (64.1)	136 (64.2)	114 (64)	
No	140 (35.9)	76 (35.8)	64 (36)	
Clinical symptom score				0.025
0–1	57 (14.6)	26 (12.3)	31 (17.4)	
2–3	167 (42.8)	81 (38.2)	86 (48.3)	
4–5	110 (28.2)	72 (34)	38 (21.3)	
6–7	56 (14.4)	33 (15.6)	23 (12.9)	
CT score				0.001
0–7	61 (15.6)	4 (1.9)	57 (32)	
8–14	215 (55.1)	110 (51.9)	105 (59)	
15–21	97 (24.9)	82 (38.7)	15 (8.4)	
22–28	17 (4.4)	16 (7.5)	1 (0.6)	
Temperature (°C)	37.8 (35.3–40.5)	37.8 (36.1–40.5)	37.7 (35.5–40.1)	0.585
Oxygen saturation on admission	0.93 (0.51–0.998)	0.92 (0.61–0.998)	0.94 (0.51–0.996)	0
Laboratory results				
WBC ($\times 10^9$/L)	6.21 (0.88–25)	6.64 (1.60–25.00)	5.7 (0.88–20.54)	0.007
NEUT ($\times 10^9$/L)	4.71 (0.34–23.8)	5.17 (1.20–23.80)	4.16 (0.34–18.86)	0.004
LYM ($\times 10^9$/L)	1.13 (0.20–3.88)	1.03 (0.20–3.08)	1.25 (0.20–3.88)	0
MONO ($\times 10^9$/L)	0.39 (0.02–1.67)	0.38 (0.04–1.67)	0.49 (0.02–1.47)	0.545
EOS ($\times 10^9$/L)	0.04 (0–0.4)	0.04 (0–0.40)	0.04 (0–0.36)	0.914
RBC ($\times 10^{12}$/L)	4.20 (2.24–5.75)	4.20 (2.66–5.59)	4.29 (2.24–5.75)	0.127
HGB (g/L)	129.8 (62–171)	128.6 (85–170)	131.2 (62–171)	0.122
PCV (%)	38.9 (20.5–52.0)	38.6 (22.8–51.7)	39.3 (20.5–52.0)	0.114

TABLE 1 (Continued)

Characteristics	Total (*n* = 390)	Sever (*n* = 212)	Nonsever (*n* = 178)	*p* value
PLT ($\times 10^9$/L)	213.1 (10.8–478)	217.1 (27.0–462)	208.3 (10.8–478)	0.277
ALT(U/L)	31.3 (5.0–144.0)	34.2 (8.0–144.0)	27.8 (5.0–125.0)	0.011
AST (U/L)	30.1 (8–80)	32.8 (13.0–80.0)	26.8 (8.0–78.0)	0
ALB (g/L)	39.0 (18.5–53.0)	37.7 (18.5–52.2)	40.5 (29.7–53.0)	0
CR (μmol/L)	70.7 (33.0–164.0)	71.4 (33.0–164.0)	69.8 (35.0–147.0)	0.478
K^+ (mEq/L)	4.0 (2.7–7.4)	4.0 (2.8–7.4)	4.0 (2.7–5.8)	0.541
Na^+ (mEq/L)	141.2 (129.0–154.2)	141.5 (129.0–154.2)	140.7 (131.6–153.5)	0.047
CRP (mg/L)	32.3 (0.1–221)	38.8 (0.3–221.0)	24.61 (0.1–180.28)	0
PaO_2 (kPa)	98.59 (2.5–276)	97.2 (34.0–276.0)	100.2 (2.5–247.0)	0.015
$PaCO_2$ (kPa)	39.2 (20.8–234.9)	39.2 (20.8–234.9)	39.3 (23.7–72.3)	0.001
pH	7.42 (7.11–7.64)	7.43 (7.12–7.64)	7.41 (7.11–7.59)	0.001
ESR (mm/h)	40.2 (2.0–128.9)	44.1 (2.0–128.9)	35.5 (2.0–119.2)	0.001
FIB (mg/dl)	4.30 (0.70–8.82)	4.48 (0.70–8.82)	4.09 (1.34–8.62)	0.004
PCT (ng/ml)	0.06 (0–1.21)	0.09 (0–1.21)	0.04 (0–1.19)	0.001
Days in hospital	18.1 (1.0–58.0)	18.5 (2.0–50.0)	17.6 (1.0–58.0)	0.519

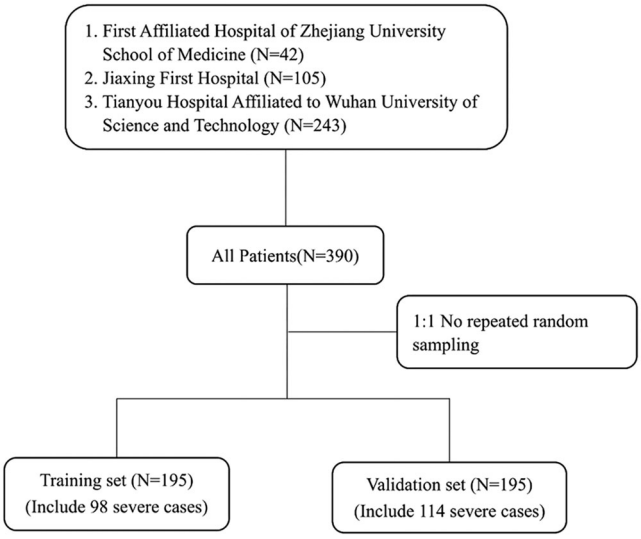

FIGURE 1 Flow chart of study participants in train and validation groups

illness.[25] The reason may be related to the type of cancer, the stage of the cancer, and the treatment of the cancer. In addition, a study developed a prognostic model to predict the severity of COVID-19 based on comorbidities. This is consistent with our results. The greater the number of comorbidities, the greater the risk of developing severe illness.[26]

Inflammation can stimulate the production of neutrophils and accelerate the apoptosis of lymphocytes. Immunity cell responses and the resulting immunological abnormalities are generally considered to play an important role in the severity of virus-induced diseases.[27] A combined study explored blood, biochemical, and immune biomarkers related to the severe disease and mortality of COVID-19, and found that WBC count, lymphocyte count, platelet count, IL-6, and serum ferritin can be used as potential Signs of progression to severe disease.[28] Our research only takes lymphocyte count as our important factor, and also includes biochemical indicators such as ALB and AST.

CT indicators show advantages in assessing the severity of COVID-19 pneumonia. For example, a cohort study determined the CT characteristics of critically ill patients with COVID-19 pneumonia, and the results showed that chest CT can quickly and accurately assess the severity of COVID-19 pneumonia, especially for critical cases.[29] A large amount of evidence shows that CT imaging provides an important reference for the early diagnosis and treatment of patients with COVID-19 pneumonia.[30] With reference to the previous experience of MERS and SARS, the imaging results will help us better judge the disease progression of patients with COVID-19 pneumonia.[31] It is worth noting that a study reported that the total severity score of CT has high diagnostic performance in evaluating severe and general patients, but it is not suitable as an independent diagnostic factor, which is different from clinical patients. The distribution ratio of

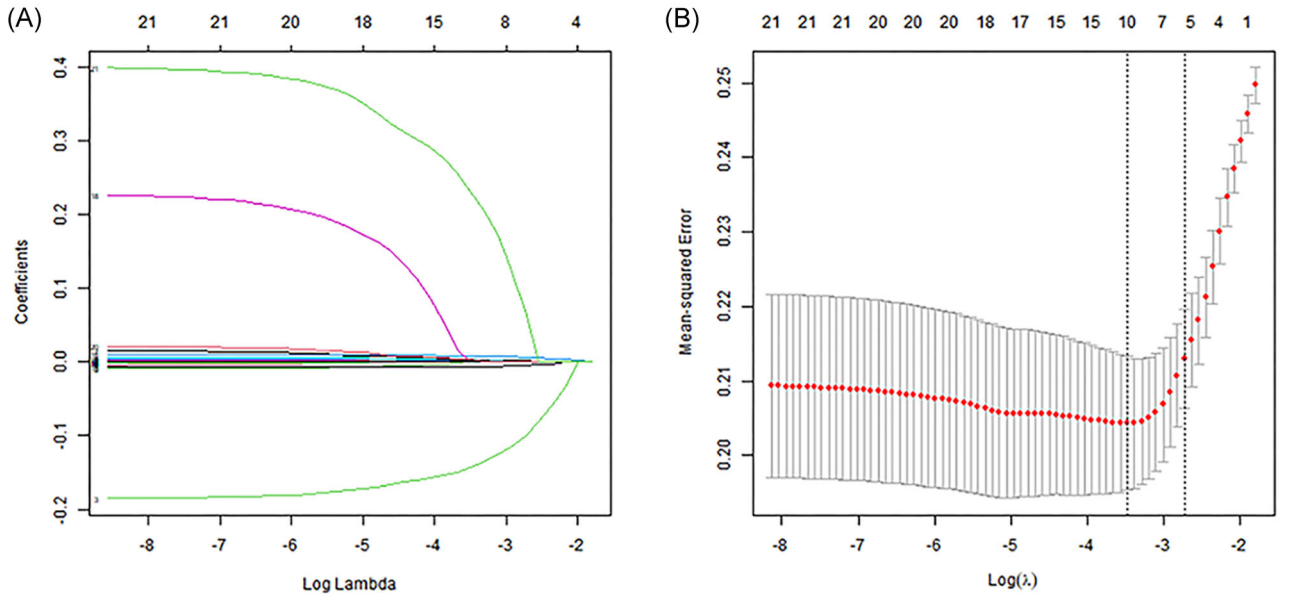

FIGURE 2 Variables selection: LASSO multiple logistic regression model; coefficient distribution map construction: the log (λ) sequence. (A) By deriving the best λ, six variables with nonzero coefficients were selected. (B) After verifying the best parameter (λ) in the LASSO model, we draw a partial likelihood deviation (binomial deviation) curve and pair number (λ), and draw a vertical dashed line based on 1 SE

TABLE 2 Logistic regression of risk factors for COVID-19 progression

Variable	β	Z	p	Odds ratio	95% CI
Intercept	−1.178	−1.050	0.292		
Age	0.028	3.260	0.001	1.860	1.281–2.700
Num of comorbidity	0.468	4.750	<0.001	2.549	1.733–3.749
CT score	0.196	5.580	<0.001	3.240	2.144–4.896
Lym	−0.680	−2.860	0.004	0.639	0.471–0.869
ALT	0.017	2.080	0.037	1.340	1.017–1.766
ALB	−0.074	−3.450	<0.001	0.535	0.374–0.763

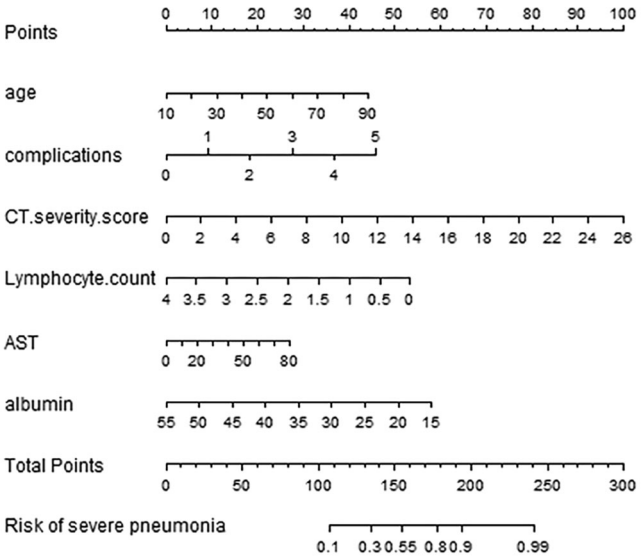

FIGURE 3 Nomogram for COVID-19 progression risk prediction

the classification is related.[32] In this study, we combined CT scores and clinical indicators to judge, which helps us to enhance the accuracy of the model. A recent meta-analysis integrated the current diagnosis and prognosis prediction models of COVID-19 infection. However, many models lack further verification, which is difficult to apply in clinical practice.[33] Our predictive model not only verifies its diagnostic efficacy but also further confirms its clinical benefit.

Inevitably, this study has some limitations. First of all, this study is a retrospective study. The distribution of patients is uneven. In the cohort of our study, the number of critically ill patients is similar to that of noncritically ill patients, which may be related to the type of patients admitted to the hospital. Second, we have not analyzed the treatment methods of the patients, and whether the patient's previous treatment has an impact on the progression of the disease is unknown. Third, this study uses CT as a risk predictor of disease progression, but this method is not widely used abroad. Fourth, the training and verification of the model are limited to a small number of domestic populations, and further verification using foreign populations will improve the generality of the model. Finally, we only use existing indicators for analysis to build a model, and the results may be biased.

TABLE 3 Basic characteristics of severe and nonsevere patients in the training and validation set

Characteristics	Training			Validation		
	Total (*n* = 195)	Severe (*n* = 98)	Nonsevere (*n* = 97)	Total (*n* = 195)	Severe (*n* = 114)	Nonsevere (*n* = 81)
Age	57.77 (24–89)	63.97 (24–89)	51.52 (31–87)	56.50 (18–88)	59.16 (18–88)	52.75 (30–88)
Num of comorbidity						
0	87 (44.6%)	16 (16.3%)	71 (73.2%)	64 (32.8%)	10 (8.8%)	54 (66.7%)
1	53 (27.2%)	44 (44.9%)	9 (9.3%)	54 (27.7%)	47 (41.2%)	7 (8.6%)
2	25 (12.8%)	22 (22.4%)	3 (3.1%)	38 (19.5%)	34 (29.8%)	4 (4.9%)
3	13 (6.7%)	8 (8.2%)	5 (5.2%)	21 (10.8%)	10 (8.8%)	11 (13.6%)
4	11 (5.6%)	5 (5.1%)	6 (6.2%)	9 (4.6%)	5 (4.4%)	4 (4.9%)
5	6 (3.1%)	3 (3.1%)	3 (3.1%)	9 (4.6%)	8 (7.0%)	1 (1.2%)
CT score						
0–7	38 (19.5%)	1 (1.0%)	37 (38.1%)	23 (11.8%)	4 (3.5%)	19 (23.5%)
8–14	68 (34.9%)	25 (25.5%)	43 (44.3%)	147 (75.4%)	85 (74.6%)	62 (76.5%)
15–21	75 (38.5%)	60 (61.2%)	15 (15.5%)	22 (11.3%)	21 (18.4%)	1 (1.2%)
22–28	14 (7.2%)	13 (13.3%)	1 (1.0%)	3 (1.5%)	3 (2.6%)	0 (0)
Lym ($\times 10^9$/L)	1.18 (0.20–3.88)	1.12 (0.20–2.34)	1.24 (0.50–3.88)	1.08 (0.20–3.08)	0.96 (0.20–3.08)	1.25 (0.20–2.59)
ALT (U/L)	28.63 (8–79)	31.68 (13–79)	25.55 (8–78)	31.52 (9–80)	33.75 (13–80)	28.38 (9–78)
ALB (g/L)	39.3 (29.4–52.2)	38.2 (29.4–52.2)	40.4 (29.7–46.3)	38.7 (18.5–53.0)	37.4 (18.5–50.1)	40.6 (29.7–53.0)

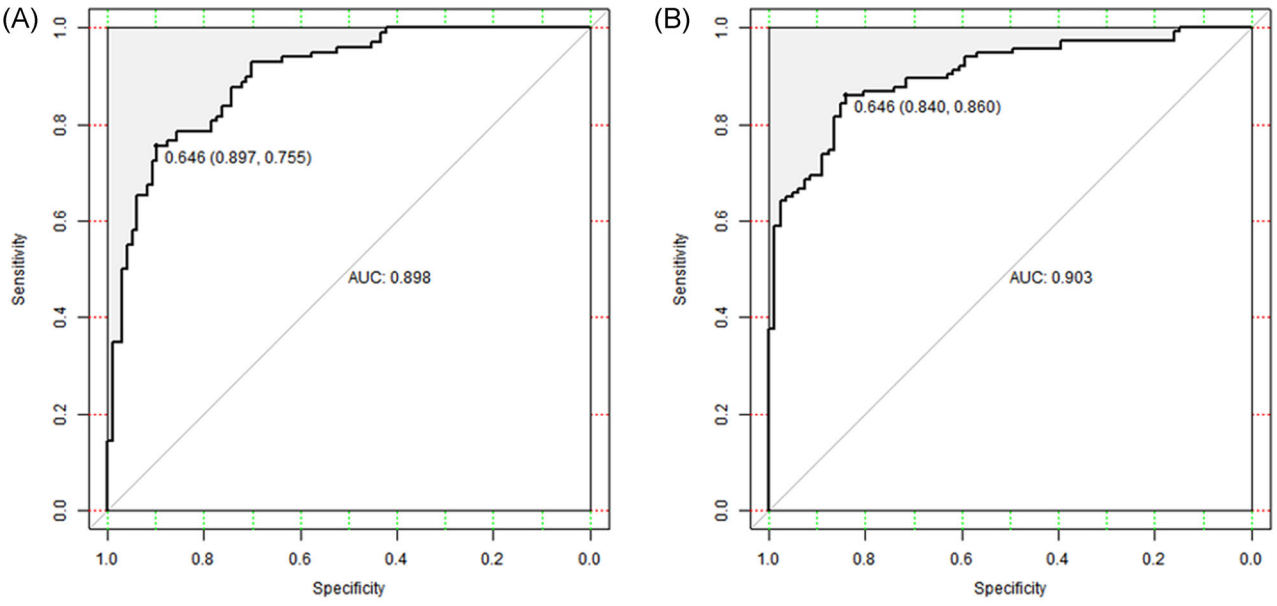

FIGURE 4 ROC of the nomogram for COVID-19 progression risk prediction. The thick black line represents the performance of the nomogram in the training set (A) and validation set (B)

5 | CONCLUSIONS

In summary, our results show that age, number of comorbidities, CT severity score, lymphocyte count, aspartate aminotransferase, and albumin are risk factors for the clinical progress of COVID-19 patients. In the training set and validation set, based on this the nomogram of the six risk factors showed good prediction accuracy. In the future, more data sets will be needed to validate our model so that it can be applied clinically.

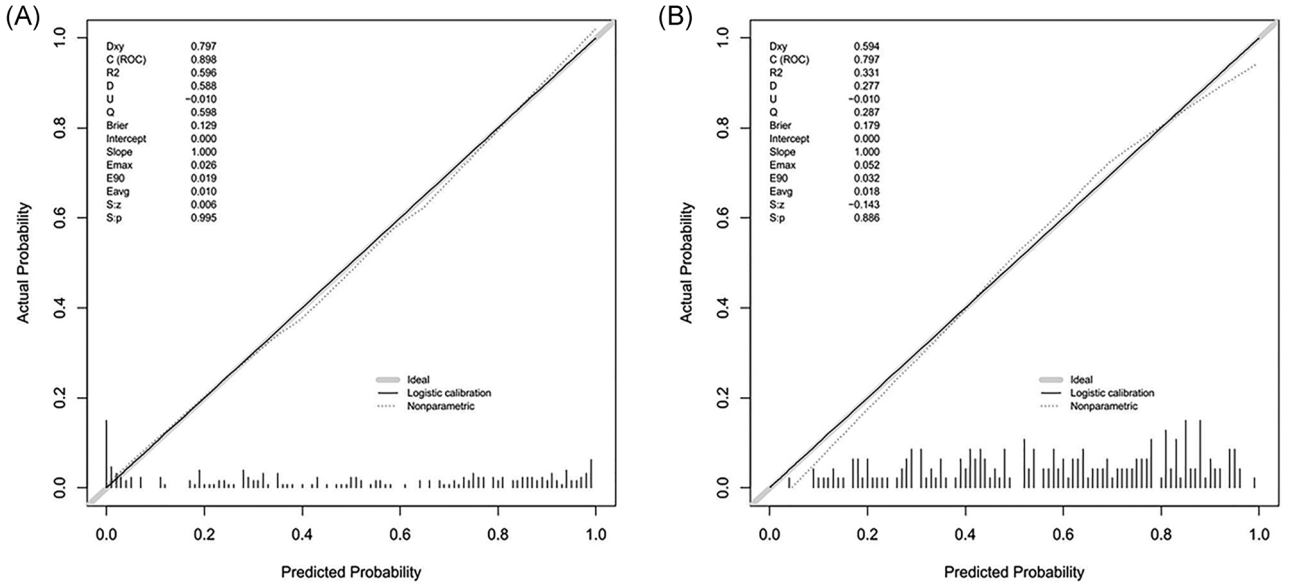

FIGURE 5 Calibration curve of the nomogram for COVID-19 progression risk prediction. The dotted line on the diagonal represents the perfect prediction of the ideal model, and the solid line represents the performance of the training set (A) and the validation set (B). The closer to the diagonal dashed line, the better the prediction effect. The *y* axis represents the actual diagnosed cases of COVID-19 disease progression, and the *x* axis represents the predicted risk of COVID-19 disease progression

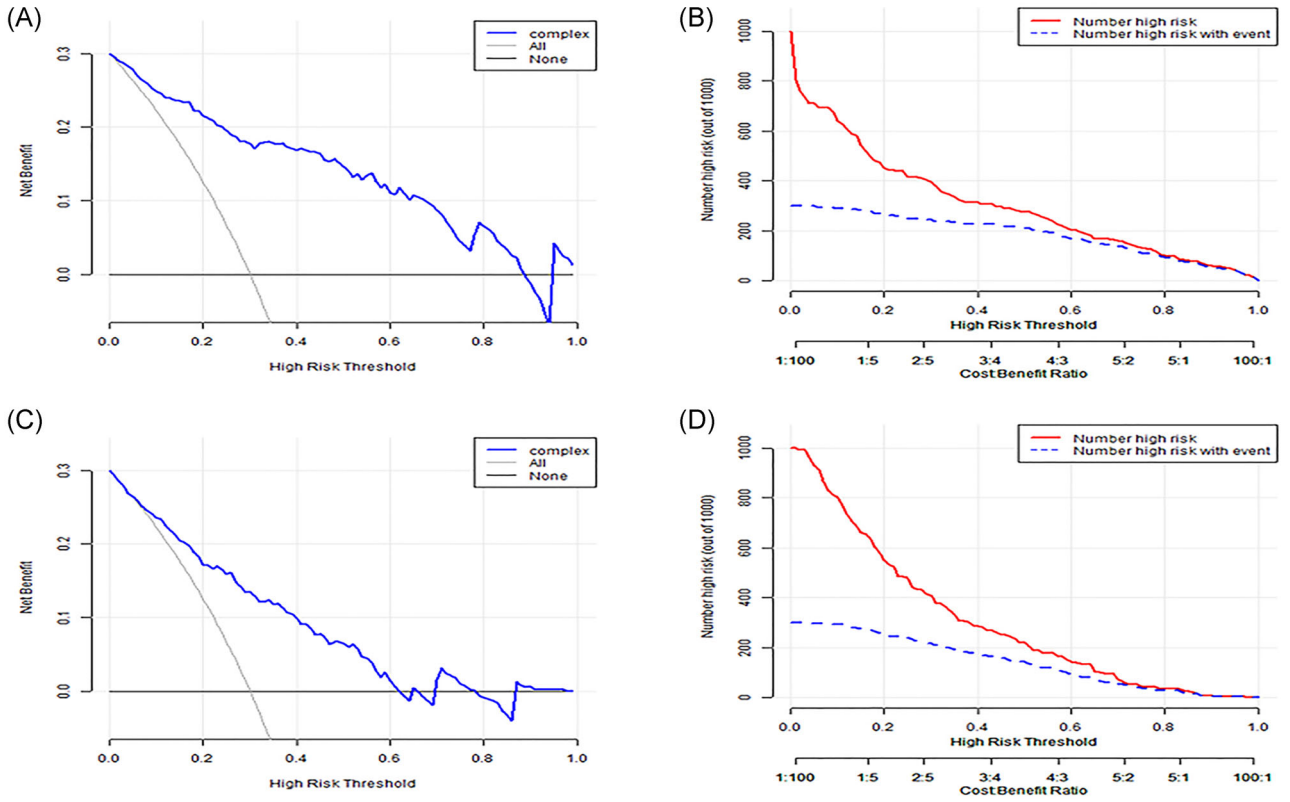

FIGURE 6 DCA and CICA. A, B: training set; C, D: validation set. The decision curve (left) compares the net clinical benefits of predicting the probability of severe COVID-19 under three scenarios: perfect prediction model (grey line), no screening (horizontal black solid line), and screening based on a nomogram (Thick blue solid line). The clinical impact curve (right) plots the number of high-risk cases of COVID-19 patients and the number of high-risk cases of severe COVID-19 under each high-risk threshold. CICA, clinical impact curve analysis; DCA, decision curve analysis

ACKNOWLEDGEMENTS

The authors gratefully acknowledge contributions from Jiaxing Fight Novel Coronavirus Pneumonia Emergency Technology Attack Special Project in 2020 (2020GZ30001), the Key Discipline of Jiaxing Respiratory Medicine Construction. Project (No.: 2019-zc-04), Jiaxing Key Laboratory of Lung Cancer Precise Treatment and General Research Project of Zhejiang Provincial Department of Education (Y202043729).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Wenyu Chen and Yufen Xu 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. Wenyu Chen, Ming Yao, Lin Hu, and Ye Zhang contributed equally. Concept and design: Wenyu Chen and Yufen Xu. Experiments and data collection: Qinghe Zhou. Data analysis and interpretation: Wenyu Chen, Ming Yao, Lin Hu, Ye Zhang, Hongwei Ren, and Yanbao Sun. Drafting of the manuscript: Wenyu Chen and Ming Zhang. All authors have read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The data analyzed during the current study are available from the corresponding author on reasonable request.

ETHICS STATEMENT

This study was approved by the Medical Ethics Committee of Jiaxing First Hospital (LS2020-009). All the patients were provided with written informed consent before enrollment.

ORCID

Wenyu Chen http://orcid.org/0000-0003-0412-2913

REFERENCES

1. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497-506.
2. Wu Y, Wang F, Shen C, et al. A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science*. 2020;368(6496):1274-1278.
3. Li AW, Sobral MC, Badrinath S, et al. A facile approach to enhance antigen response for personalized cancer vaccination. *Nat Mater*. 2018;17(6):528-534.
4. 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(20):2052-2059.
5. Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med*. 2020;8(5):475-481.
6. Xiong Y, Sun D, Liu Y, et al. Clinical and high-resolution CT features of the COVID-19 infection: comparison of the initial and follow-up changes. *Invest Radiol*. 2020;55(6):332-339.
7. 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(11):1061-1069.
8. Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382(18):1708-1720.
9. Das KM, Lee EY, Enani MA, et al. CT correlation with outcomes in 15 patients with acute Middle East respiratory syndrome coronavirus. *AJR Am J Roentgenol*. 2015;204(4):736-742.
10. Shi H, Han X, Jiang N, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis*. 2020;20(4):425-434.
11. Wu G, Yang P, Xie Y, et al. 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*. 2020;56:2.
12. Liang W, Liang H, Ou L, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med*. 2020;180(8):1081-1089.
13. Dong Y, Zhou H, Li M, et al. A novel simple scoring model for predicting severity of patients with SARS-CoV-2 infection. *Transbound Emerg Dis*. 2020;67(6):2823-2829.
14. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and interpreting decision curve analysis: a guide for investigators. *Eur Urol*. 2018;74(6):796-804.
15. Koo HJ, Lim S, Choe J, Choi SH, Sung H, Do KH. Radiographic and CT features of viral pneumonia. *Radiographics*. 2018;38(3):719-739.
16. Chung M, Bernheim A, Mei X, et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology*. 2020;295(1):202-207.
17. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology*. 2008;246(3):697-722.
18. Gong J, Ou J, Qiu X, et al. Multicenter development and validation of a novel risk nomogram for early prediction of severe 2019-novel coronavirus pneumonia. *Lancet*. 2020. https://ssrn.com/abstract=3551365
19. Xu X, Yu C, Qu J, et al. Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2. *Eur J Nucl Med Mol Imaging*. 2020;47(5):1275-1280.
20. Zhu N, Zhang D, Wang W, et al. China novel coronavirus investigating and research team. a novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727-733.
21. Choi KW, Chau TN, Tsang O, et al. Outcomes and prognostic factors in 267 patients with severe acute respiratory syndrome in Hong Kong. *Ann Intern Med*. 2003;139(9):715-723.
22. Hong KH, Choi JP, Hong SH, et al. Predictors of mortality in Middle East respiratory syndrome (MERS). *Thorax*. 2018;73(3):286-289.
23. 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(10229):1054-1062.
24. Li K, Wu J, Wu F, et al. The clinical and chest CT features associated with severe and critical COVID-19 pneumonia. *Invest Radiol*. 2020;55(6):327-331.
25. Dai M, Liu D, Liu M, et al. Patients with cancer appear more vulnerable to SARS-CoV-2: a multicenter study during the COVID-19 outbreak. *Cancer Discov*. 2020;10(6):783-791.
26. Gude-Sampedro F, Fernández-Merino C, Ferreiro L, et al. Development and validation of a prognostic model based on comorbidities to predict COVID-19 severity: a population-based study. *Int J Epidemiol*. 2021;50(1):64-74.
27. Channappanavar R, Perlman S. Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. *Semin Immunopathol*. 2017;39(5):529-539.
28. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated

with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. *Clin Chem Lab Med*. 2020;58(7):1021-1028.
29. Lyu P, Liu X, Zhang R, Shi L, Gao J. The performance of chest CT in evaluating the clinical severity of COVID-19 pneumonia: identifying critical cases based on CT characteristics. *Invest Radiol*. 2020;55(7):412-421.
30. Xu YH, Dong JH, An WM, et al. Clinical and computed tomographic imaging features of novel coronavirus pneumonia caused by SARS-CoV-2. *J Infect*. 2020;80(4):394-400.
31. Hosseiny M, Kooraki S, Gholamrezanezhad A, Reddy S, Myers L. Radiology perspective of coronavirus disease 2019 (COVID-19): lessons from severe acute respiratory syndrome and Middle East respiratory syndrome. *AJR Am J Roentgenol*. 2020;214(5):1078-1082.
32. Li K, Fang Y, Li W, et al. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). *Eur Radiol*. 2020;30(8):4407-4416.
33. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369:m1328.

How to cite this article: Chen W, Yao M, Hu L, et al. Development and validation of a clinical prediction model to estimate the risk of critical patients with COVID-19. *J Med Virol*. 2022;94:1104-1114. doi:10.1002/jmv.27428