Original Article

sciendo

Development and validation of a nomogram for predicting the disease progression of nonsevere coronavirus disease 2019

Xue-lian Li[1,#], Cen Wu[2,#], Jun-gang Xie[3,#], Bin Zhang[4,#], Xiao Kui[5,#], Dong Jia[6], Chao-nan Liang[7], Qiong Zhou[8], Qin Zhang[9], Yang Gao[10], Xiaoming Zhou[2,#], Gang Hou[9, 10, 11, 12, 13,14]

[1]Department of Epidemiology, School of Public Health, China Medical University, Shenyang, Liaoning Province, China;
[2]Department of Pulmonary and Critical Care Medicine, Shengjing Hospital of China Medical University, Shenyang, Liaoning Province, China;
[3]Department of Respiratory and Critical Care Medicine, National Clinical Research Center of Respiratory Disease, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei Province, China;
[4]Department of Respiratory Disease, Second Affiliated Hospital, Zhejiang University College of Medicine, Hangzhou, Zhejiang Province, China;
[5]Department of Pulmonary and Critical Care Medicine, Second Xiangya Hospital, Central South University, Changsha, Hunan Province, China;
[6]Department of Emergency Medicine, Shengjing Hospital of China Medical University, Shenyang, Liaoning Province, China;
[7]Department of Pulmonary and Critical Care Medicine, First Hospital of China Medical University, Shenyang, Liaoning Province, China;
[8]Department of Respiratory and Critical Care Medicine, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei Province, China;
[9]Department of Respiratory Disease, The Second Affiliated Hospital of Baotou Medical College, Baotou, China;
[10]Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China-Japan Friendship Hospital, Beijing, China;
[11]National Center for Respiratory Medicine, Beijing, China;
[12]Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Beijing, China;
[13]National Clinical Research Center for Respiratory Diseases, Beijing, China;
[14]Graduate School of Peking Union Medical College, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

[#]These authors contributed equally to the study.

Address for Correspondence:
Dr. Gang Hou, Department of Pulmonary and Critical Care Medicine, First Affiliated Hospital of China Medical University, Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China-Japan Friendship Hospital, Beijing, China; National Center for Respiratory Medicine, Beijing, China; Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Beijing, China; National Clinical Research Center for Respiratory Diseases, Beijing, China; Graduate School of Peking Union Medical College, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China.
E-mail: hougangcmu@163.com

Access this article online

Website: www.intern-med.com

DOI: 10.2478/jtim-2021-0030

Quick Response Code:

ABSTRACT

Background and Objectives: The majority of coronavirus disease 2019 (COVID-19) cases are nonsevere, but severe cases have high mortality and need early detection and treatment. We aimed to develop a nomogram to predict the disease progression of nonsevere COVID-19 based on simple data that can be easily obtained even in primary medical institutions. **Methods:** In this retrospective, multicenter cohort study, we extracted data from initial simple medical evaluations of 495 COVID-19 patients randomized (2:1) into a development cohort and a validation cohort. The progression of nonsevere COVID-19 was recorded as the primary outcome. We built a nomogram with the development cohort and tested its performance in the validation cohort. **Results:** The nomogram was developed with the nine factors included in the final model. The area under the curve (AUC) of the nomogram scoring system for predicting the progression of nonsevere COVID-19 into severe COVID-19 was 0.875 and 0.821 in the development cohort and validation cohort, respectively. The nomogram achieved a good concordance index for predicting the progression of nonsevere COVID-19 cases in the development and validation cohorts (concordance index of 0.875 in the development cohort and 0.821 in the validation cohort) and had well-fitted calibration curves showing good agreement between the estimates and the actual endpoint events. **Conclusions:** The proposed nomogram built with a simplified index might help to predict the progression of nonsevere COVID-19; thus, COVID-19 with a high risk of disease progression could be identified in time, allowing an appropriate therapeutic choice according to the potential disease severity.

Key words: coronavirus disease 2019, nomogram, risk factors, worsening, progression, prediction, nomogram.

INTRODUCTION

Since the global spread of severe acute respiratory syndrome coronavirus 2 disease (SARS-CoV-2, previously temporarily named 2019 novel coronavirus, or COVID-19), it has been designated the sixth public health emergency of international concern.[1, 2] To date, the severity of most cases of COVID-19 has been nonsevere,[3] and most patients with nonsevere COVID-19 have a good prognosis.[4, 5] However, due to the widespread transmission of the virus, a relatively large number of deaths have been reported due to acute respiratory distress syndrome and multiple organ failure.[4, 5] Medical resources required to treat patients with nonsevere COVID-19 are relatively limited, while those required by severe and critical patients have increased significantly.[6]

According to the China data in 2020, 81% of cases of COVID-19 were classified as nonsevere, 14% were classified as severe, and 5% were classified as critical illness.[7] Based on these characteristics and situations, Fangcang shelter hospitals were developed and used for the first time in China to tackle the coronavirus disease 2019 (COVID-19) outbreak.[8] However, inevitably, some patients with COVID-19 were initially diagnosed with nonsevere cases that eventually progressed. Thus, identifying the progression in the early stage would be significant.

By using a predictive model based on the characteristics of hospitalized patients with nonsevere COVID-19 at the time of first medical evaluation to predict disease progression of COVID-19 from a nonsevere type into a severe type, it might be feasible to identify high-risk patients at the very beginning of the disease and helpful to support the early allocation of medical resources and evidence-based decision-making. However, many prediction models or risk factors have been identified based on the characteristics and outcome of COVID-19.[4, 7, 9-15] The results varied, which might contribute to the complexity and the different variants and regional distribution of COVID-19. However, screening the possibility of the progression of nonsevere COVID-19 into severe COVID-19 based on the simplified index available in most medical institutions independent of the regions would be helpful for the early treatment and better prognosis of COVID-19. Such predictive tools would be especially useful in Fangcang shelter hospitals and general practitioners' offices. Nomograms are easy to use and can facilitate management-related decision-making.[13]

In this study, we aimed to build a nomogram with very simplified variables obtained at the time of the first evaluation of patients with COVID-19 (including symptoms, comorbidities, routine blood tests, C-reactive protein [CRP] levels, and chest radiography) that can provide an individualized, evidence-based, highly accurate risk estimation,[16] and these variables were available for most temporary medical institutions, especially in Fangcang shelter hospitals. To the best of our knowledge, we have established the first nomogram for predicting the risk of the progression of COVID-19 from a nonsevere to a severe type based on the data extracted from Fangcang shelter hospitals.

METHODS

Study design and participants

This was a multicenter retrospective cohort study. This study was performed to develop and validate a prediction tool as a nomogram for evaluating the progression of nonsevere COVID-19 cases into severe COVID-19 based on simplified data from the patients' first visit to a medical institution due to COVID-19, which was limited to history, physical examination, routine blood tests, CRP, and chest radiology. The study was approved by the research ethics committees of the First Hospital of China Medical University and followed the Declaration of Helsinki. The primary outcome of this study was defined as the progression of COVID-19 from a nonsevere type to a severe type during the follow-up.[17, 18]

For diagnosis of severe COVID-19 group, at least one of the following conditions should be met according to WHO guidelines, complemented by the COVID-19 Diagnosis and Treatment Guidance (2020) of China (version 6.0)[18]: (1) respiratory rate (RR) ≥ 30 times/min; (2) arterial oxygen saturation (resting status) ≤ 93% in the resting state; or (3) the ratio of partial pressure of oxygen to fraction of inspiration O_2 (PaO_2/FiO_2) ≤ 300 mmHg (1 mm Hg = 0.133 kPa), as the other studies did. [19]

The inclusion criteria were as follows: (1) diagnosis of COVID-19 confirmed by positive specific RT-PCR for SARS-CoV-2; (2) diagnosis of nonsevere COVID-19 according to the 7th edition of the medical guidelines from the National Health Commission of the People's Republic of China; (3) age ≥ 18 years; (4) patients with detailed clinical records at the first medical evaluation[18]; and (5) patients with at least a 28-day follow-up period.

The exclusion criteria were as follows: (1) incomplete data from the first visit to the medical institutions; (2) worsening of the clinical condition mainly due to other diseases but not COVID-19; and (3) classification of the severe or critically severe type of COVID-19 according to the data achieved at the first medical evaluation. If missing data for a certain feature or sample is more than 5%, then we leave that features or sample out.

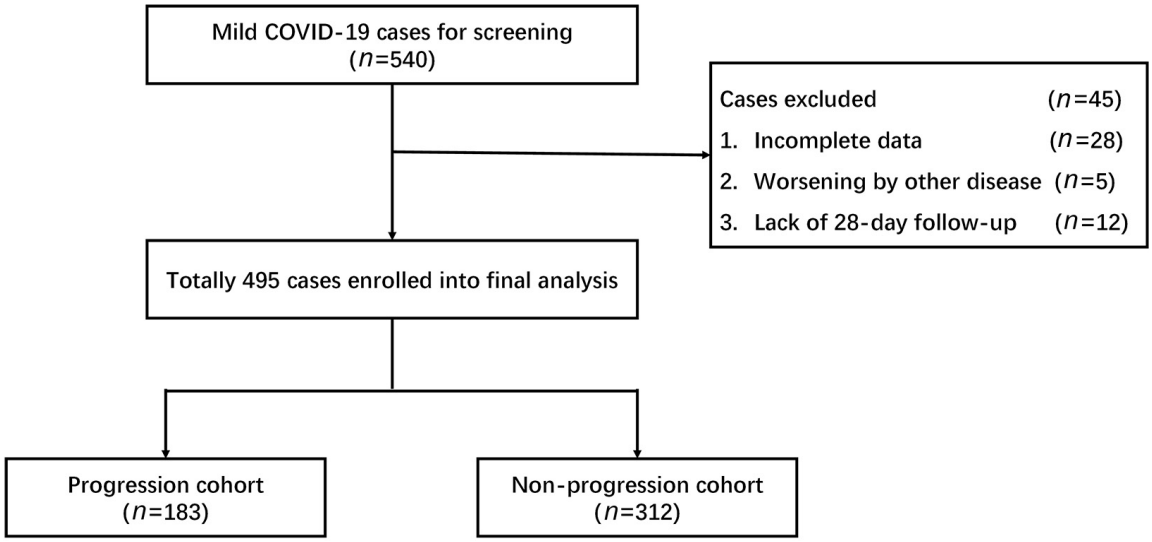

Figure 1: Study flow chart.

We collected data from a total of 495 patients with COVID-19 from different designated hospitals: Wuchang Fangcang Shelter Hospital, Tongji Hospital of Tongji Medical College and Union Hospital of Tongji Medical College in Wuhan, Hubei, China, between February 10, 2020 and March 8, 2020 (Figure 1). Data were extracted from electronic medical records.

The development and validation of the nomogram adhered to the guidelines in the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement (Supplementary Table S1).[16]

Definitions and predictors

See SECTION 1 of the online supplement.

Statistical analysis

The Shapiro–Wilk test was used to determine whether the data were normally distributed. Data are expressed as the mean ± standard deviation for continuous normally distributed data or median and interquartile range (IQR, 25–75th percentiles) for non-normal continuous variables. Categorical data are presented as frequencies (percentages). Differences between the two groups were determined by the *t*-test for parametric data and the Mann–Whitney *U* test for nonparametric data. The χ^2 test was used to analyze the differences in categorical variables. Missing data were handled by multiple imputations.[20]Imputation for missing variables was considered if missing values were less than 20%, while in our study, the missing data for each variable were < 5%.[21] Logistic regression analysis was used for univariate and multivariate analyses. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated. All cases were randomly divided into a development cohort and an internal validation cohort at a ratio of 2:1. A nomogram was formulated based on the final model selection, which was performed in the development cohort by a backward step-down selection process with the Akaike information criterion. Models were internally validated using bootstraps, and 1000 resamples were used for these activities.[22] The concordance index (C-index) was used to measure the performance of the nomogram. Calibration curves were plotted based on bootstrapping to obtain bias-corrected estimates of predicted versus observed values. Receiver operating characteristic (ROC) curves and calibration curves were drawn for each dataset to predict the disease progression of nonsevere COVID-19.[23] The Youden index was defined as an optimal cutoff based on the ROC curve.[24] The positive and negative predictive values (PPV and NPV, respectively) and the mean risk stratification (MRS) were calculated. Statistical significance was set at a two-tailed $P < 0.05$. The statistical analysis was performed with R software version 3.6.3 (http://www.R-project.org).

RESULTS

Baseline characteristics of COVID-19 patients

Among the 495 patients enrolled in our study, there were 242 males and 253 females, with a median age of 50.55 years (18–87 years). During the follow-up period, 183 (37.0%) patients experienced progression of nonsevere COVID-19 [119 (36.1%) in the development cohort and 64 (38.8%) in the validation cohort]. The median duration from the onset of the disease to the examination of routine blood tests and CRP levels was 5 days (0–38 days) and 6

Table 1: Baseline characteristics between progression cohort and nonprogression cohort

	Total (n = 495)	Progression cases (n = 183)	Nonprogression cases (n = 312)	P value
Sex (male/female)	246/262	101/82	141/171	0.040
Age (years)	51(19)	58(18.5)	48(19)	<0.001
Hypertension	87(17.6%)	42(23.0%)	45(14.4%)	0.022
Diabetes mellitus	46(9.1%)	28(15.3%)	18(5.8%)	<0.001
CVD	7(1.4%)	3(1.6%)	5(1.6%)	0.93
COPD	1(0.2%)	0(0.0%)	1(0.3%)	1.00
Tumor	1(0.4%)	0(0.0%)	1(0.3%)	0.53
Current smoking	53(10.7%)	25(13.7%)	28(9.0%)	0.14
Cough	364(73.4)	139(76.0%)	225(72.1%)	0.78
Sputum	117(23.6%)	45(24.6%)	72(23.1%)	0.70
Dyspnea	98(19.8%)	64(35.0%)	34(10.9%)	<0.001
Hemoptysis	15(3.0%)	6(3.3%)	9(2.9%)	0.90
Myalgia	64(12.9%)	27(14.8%)	37(11.8%)	0.43
Fatigue	136(27.4%)	44(24.0%)	92(29.5%)	0.29
Nausea or vomiting	32(6.5)	11(6.0%)	21(6.7%)	0.90
Fever	365(73.7%)	155(84.7%)	210(67.3%)	<0.001
Systolic blood pressure (mmHg)	127(15)	127(19)	127(14)	0.653
Heart rate (beats/min)	85(16)	85(15)	85(17)	0.908
Tachycardia (heart rate >100 beats/min)	62(12.5%)	41(22.4%)	21(6.7%)	<0.001
Respiratory rate (breaths/min)	19(2)	19(2)	20(2)	0.042
Tachypnea (respiratory rate >24 breaths/min)	11(2.2%)	8(4.4%)	3(1.0 %)	0.022
Leukocyte count ($\times 10^9$)	5.20(2.55)	5.35(2.36)	5.12(2.68)	0.134
Neutrophil count ($\times 10^9$)	3.40(2.05)	3.48(2.13)	3.34(2.03)	0.257
Lymphocyte count ($\times 10^9$)	1.19(0.68)	0.98(0.505)	1.33(0.80)	<0.001
Neutrophil-to-lymphocyte ratio	2.80(2.24)	3.52 (2.75)	2.45 (1.79)	<0.001
Hemoglobin (g/L)	136(22)	129(23)	139(21)	<0.001
Anemia (hemoglobin level <120 g/L in males and <110 g/L in females)	33(6.7%)	21(11.5%)	12(3.8%)	0.002
Platelet count ($\times 10^9$)	182(94.5)	178(98)	185(90.5)	0.936
Thrombocytopenia (platelet count <100 $\times 10^9$)	25(5.0%)	7(3.8%)	18(5.8%)	0.46
CRP level (mg/L)	22.73(80.5)	36.54(70.69)	13.36(79.48)	<0.001
Multilobar involvement	289(58.4%)	152(83.1%)	137(43.9%)	<0.001

COPD: chronic obstructive pulmonary disease; CRP: C-reactive protein; CVD: cardiovascular disease.

days (0–38 days), respectively. The median duration from the onset of disease to deterioration of disease was 11 days (1–49 days).

The characteristics in different cohorts

In total, 183 patients with disease progression and 312 patients without disease progression were enrolled in the development and validation cohorts of our study. There were significant differences in sex distribution, age, comorbidities of hypertension and diabetes mellitus, symptoms of dyspnea, heart rate, tachycardia, respiratory rate, tachypnea, lymphocyte count, hemoglobin, anemia, CRP level, and multilobar involvement (≥ 3 lobes) between the cases with disease progression and cases with no disease progression (Table 1). There were no significant differences in the predictors between the development cohort and the validation cohort (Table 2).

Development of the nomogram

All the risk factors were included in the univariate analysis to analyze their association with the progression of nonsevere COVID-19 (Table 3). The results of the univariate analysis with $P < 0.10$ were included in the multivariate analysis (Table 3). The nomogram was developed with the nine factors included in the final model (Figure 2 and Supplementary Figure S1). The formula for the total risk point was as follows:

1.12 × Age (years) - 13.33 × Sex (Male 0 or Female 1) + 21.19 × Diabetes mellitus (No 0 or Yes 1) + 32.1 × Dyspnea (No 0 or Yes 1) + 20 × Tachycardia (No 0 or Yes 1) - 39.62 × (Lymphocyte count - 3.5) + 18.33 × Anemia (No 0 or Yes 1) + 0.1 × CRP (mg/L) + 35.9 × Multilobar involvement (No 0 or Yes 1).

Table 2: Comparison of baseline characteristics between the development and validation cohorts

	Development cohort (n = 330)	Validation cohort (n = 165)	P value
Sex (male/female)	157/173	85/80	0.73
Age (years)	50(19)	54(20)	0.09
Hypertension	53(16.1%)	34(20.6%)	0.26
Diabetes mellitus	27(8.2%)	19(11.5%)	0.3
CVD	5(1.5%)	3(1.8%)	0.80
COPD	1(0.3%)	0(0.0%)	0.48
Tumor	1(0.3%)	1(0.6%)	0.61
Current smoking	36(10.9%)	17(10.3%)	0.96
Cough	243(73.6%)	121(73.3%)	1
Sputum	73(22.1%)	44(26.7%)	0.31
Dyspnea	68(20.6%)	30(18.2%)	0.6
Hemoptysis	12(3.6%)	3(1.8%)	0.4
Myalgia	47(14.2%)	17(10.3%)	0.28
Fatigue	96(29.1%)	40(24.2%)	0.3
Nausea or vomiting	20(6.1%)	12(7.3%)	0.75
Fever	244(73.9%)	121(73.3%)	0.97
Systolic blood pressure	127(15)	127(16)	0.94
Tachycardia	42(12.7%)	20(12.1%)	0.96
Tachypnea	9(2.7%)	2(1.2%)	0.45
Leukocyte count ($\times 10^9$)	5.17(2.43)	5.31(2.98)	0.29
Neutrophil count ($\times 10^9$)	3.4(1.855)	3.4(2.39)	0.78
Lymphocyte count ($\times 10^9$)	1.205(0.695)	1.15(0.63)	0.59
Neutrophil-to-lymphocyte ratio	2.892(2.136)	2.722(2.300)	0.83
Anemia	24(7.3%)	9(5.5%)	0.57
Thrombocytopenia	18(5.5%)	7(4.2%)	0.72
CRP level (mg/L)	22.815(80.4)	21.12(80.5)	0.98
Multilobar involvement	201(60.9%)	88(53.3%)	0.13
Progression case	119(36.1%)	64(38.8%)	0.62

COPD: chronic obstructive pulmonary disease; CRP: C-reactive protein; CVD: cardiovascular disease.

Table 3: Univariate and multivariate analyses for association with progression of COVID-19 from a nonsevere type to a severe type

	Univariate analysis		Multivariate analysis	
	OR (95% CI)	P value	OR (95% CI)	P value
Sex (male/female)	0.71 (0.45–1.12)	0.143	1.06 (1.03–1.08)	< 0.001
Age (years)	1.06 (1.04–1.08)	< 0.001	0.54 (0.29–1)	0.049
Hypertension	1.74 (0.96–3.15)	0.068	0.78 (0.35–1.71)	0.534
Diabetes mellitus	2.82 (1.26–6.31)	0.011	2.92 (0.96–8.85)	0.059
Cough	1.45 (0.86–2.46)	0.163	..	..
Dyspnea	4.57 (2.6–8.04)	< 0.001	4.33 (2.06–9.09)	< 0.001
Fever	2.08 (1.19–3.62)	0.01	1.38 (0.64–2.98)	0.414
Tachycardia	4.33 (2.18–8.61)	< 0.001	2.58 (1.03–6.46)	0.044
Tachypnea	6.53 (1.33–31.97)	0.021	1.28 (0.18–9)	0.803
Leukocyte count ($\times 10^9$)	0.99 (0.89–1.1)	0.823	..	..
Lymphocyte count ($\times 10^9$)	0.19 (0.11–0.33)	< 0.001	0.26 (0.13–0.52)	< 0.001
Anemia	3.24 (1.37–7.65)	0.007	2.38 (0.76–7.44)	0.134
Thrombocytopenia	0.88 (0.32–2.41)	0.804	1.01 (1–1.01)	0.211
CRP level ⩾ 20 mg/L	1.01 (1–1.02)	< 0.001	5.43 (2.72–10.83)	< 0.001
Multilobar involvement	8.15 (4.45–14.93)	< 0.001	0.26 (0.13–0.52)	< 0.001

COVID-19: Coronavirus disease 2019; CRP: C-reactive protein.

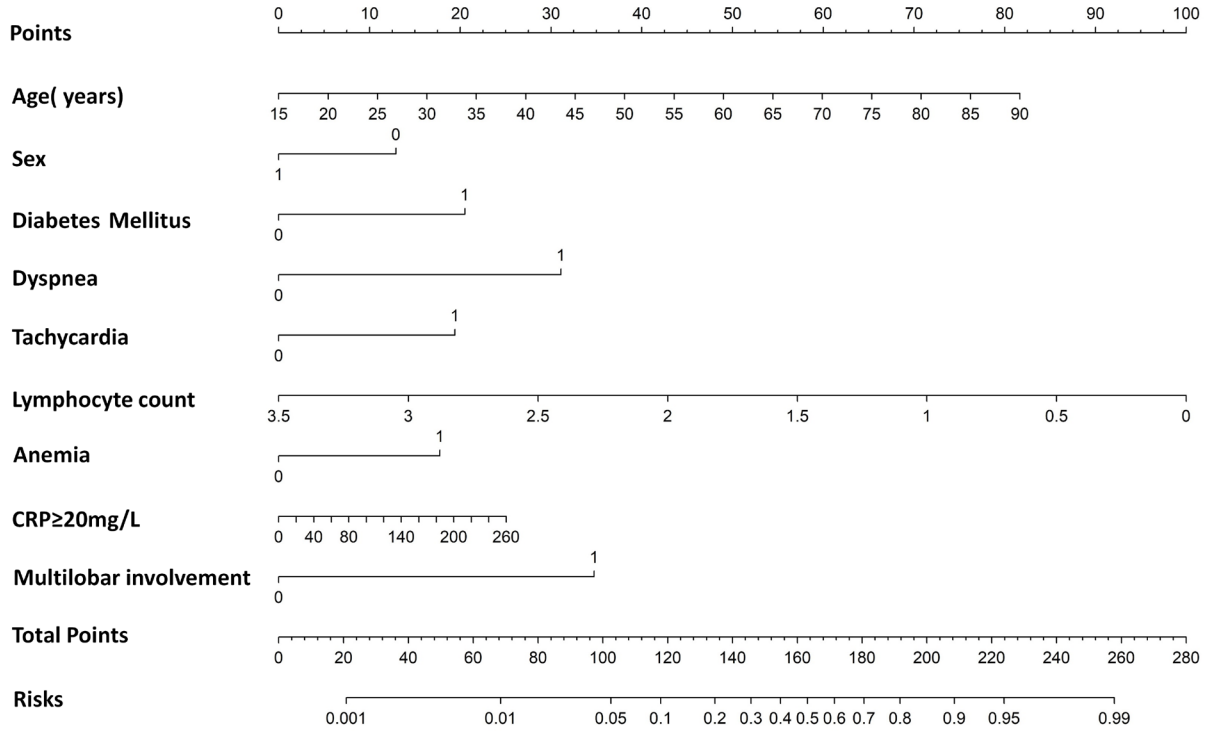

Figure 2: A nomogram to predict the risk of progression nonsevere coronavirus disease 2019. To use the nomogram, draw a vertical line to identify the corresponding points of each variable according to their actual status. Then, add the points for all variables and find the position on the total point axis. With the same line mentioned above, you can determine the risk of progression nonsevere COVID-19 with the initial medical evaluation results at the lower line of the nomogram. Tachycardia is defined as a heart rate ≥ 100 beats per minute. Tachypnea is defined as a respiration rate ≥ 24 breaths per minute. Anemia is defined as a hemoglobin level < 120 g/L for males and < 110 g/L for females. Multilobar involvement is defined as the involvement ≥ 3 lobes on a CT scan. Using the cutoff score of 129.9, the sensitivity and specificity for discriminating between those with a high and low risk of developing the progression of COVID-19 in the validation cohort were 65.6% and 85.7%, respectively.

The risk of each individual was assessed according to the above equation. The probability of a case with no disease progression transforming into a case with disease progression could be calculated as follows: $1/(1 + e^{-z})$ (e = 2.718 and z = –7.439).

Validating performance of the nomogram

Based on the nomogram built in our study, a score was calculated for each patient, and these scores were used for the following analyses. In the development cohort, the area under the curve (AUC) of the nomogram scoring system for predicting the progression of nonsevere COVID-19 was 0.875 (95% CI 0.836–0.913) (Figure 3A).

In the validation cohort, the AUC of the nomogram scoring system for predicting the progression of nonsevere COVID-19 was 0.821 (95% CI 0.754–0.888) (Figure 3B). Using the threshold score of 178, the sensitivity and specificity for discriminating between those at high and low risk of disease progression were 85.1% and 65.6%, respectively. The Youden index J (J = Sens + Spec−1) was 0.506. PPV and NPV were 62.65% and 85.37%, respectively. MRS [MRS = 2 * (Sens + Spec–1) * π * (1-π), where π = prevalence of nonsevere to severe progression, which is 37.9%] was 24.0%. The nomogram shows a high-risk stratification ability for COVID-19, which has a higher disease progression prevalence.

Incorporating these variables, the nomogram achieved a good concordance index for predicting the progression of nonsevere COVID-19 cases in the development and validation cohorts (concordance index of 0.875 in the development cohort and 0.821 in the validation cohort) and had well-fitted calibration curves showing good agreement between the estimates based on the nomogram and the actual endpoint events (Figure 4A and B). To evaluate the clinical applicability of our risk prediction nomogram, clinical impact curve analysis (CICA) and decision curve analysis (DCA) were performed. The CICA (Figure 5A) and DCA (Figure 5B) visually showed that the nomogram had a superior overall net benefit within the wide and practical ranges of threshold probabilities and impacted patient outcomes.

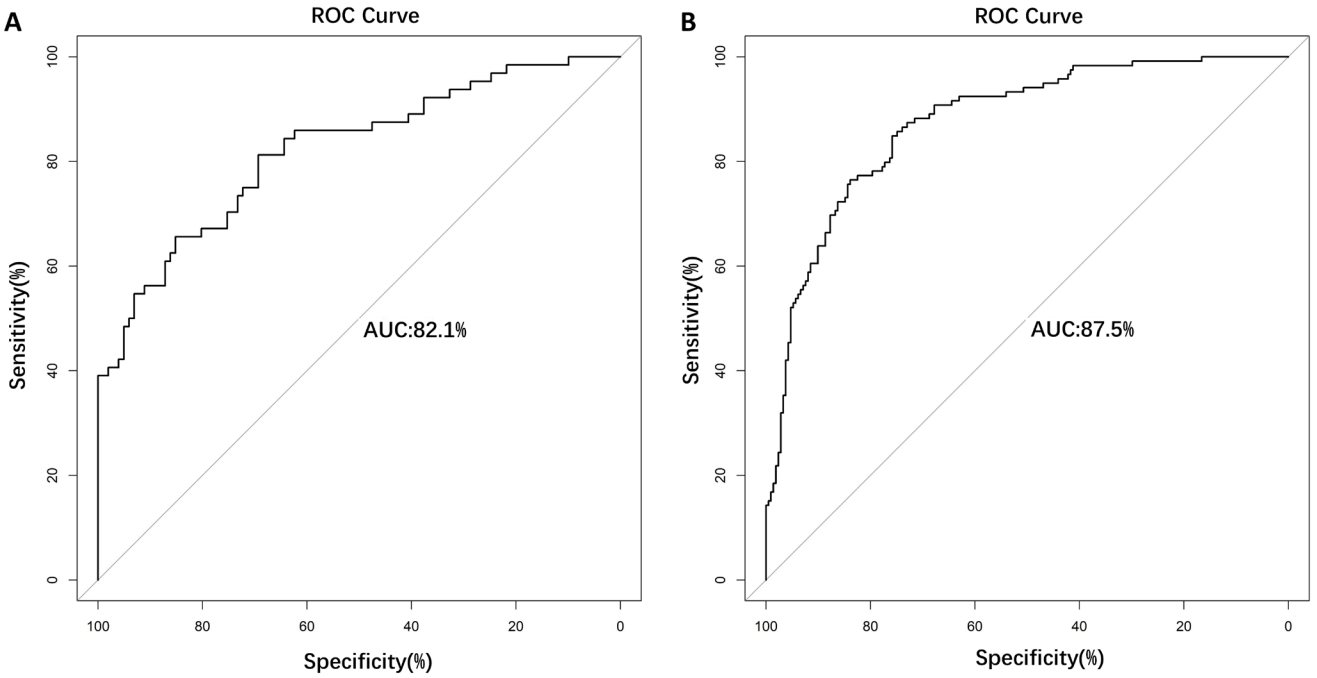

Figure 3: Receiver operating characteristic curve of the prediction nomogram. (A) In the development cohort, the AUC of the nomogram scoring system for predicting the progression of COVID-19 was 0.893 (95% CI 0.858–0.928); (B) In the internal validation cohort, the AUC of the nomogram scoring system for predicting the progression of COVID-19 was 0.847 (95% CI 0.787–0.906).

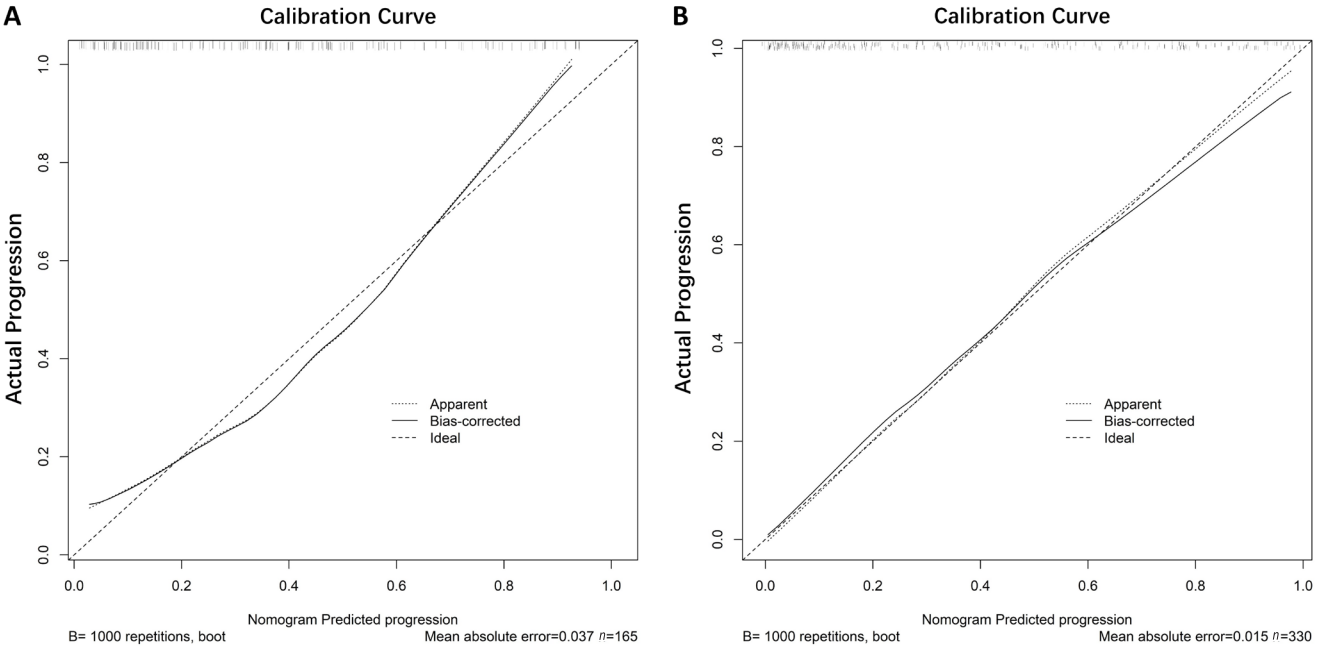

Figure 4: Calibration plot showing the predicted probability of the risk of progression nonsevere Coronavirus Disease 2019. Bootstrapping was used to obtain bias-corrected (overfitting-corrected) estimates of the predicted versus observed values based on nonparametric smoothers. The three lines represented the ideal accuracy, the apparent accuracy, and the bias-corrected estimate of predictive accuracy. The bias was estimated due to overfitting or the "optimism" in the final model fit. After the optimism was estimated, it can be subtracted from the index of accuracy derived from the original sample to obtain a bias-corrected or overfitting-corrected estimate of predictive accuracy. (A) Development cohort; (B) Validation cohort.

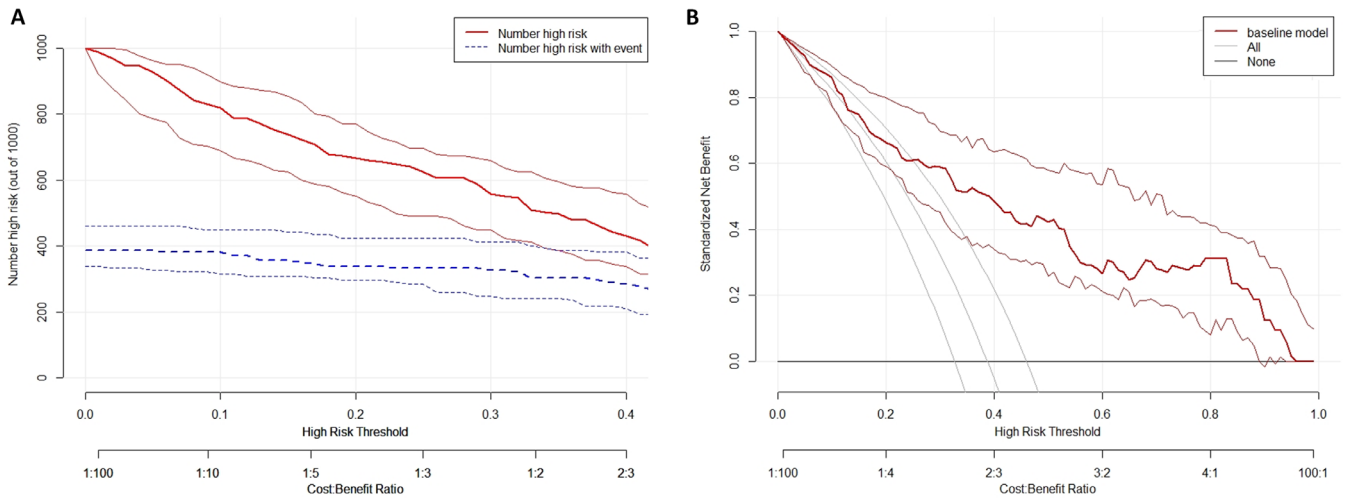

Figure 5: Clinical impact curve analysis (CICA) and decision curve analysis (DCA) of the prediction nomogram. (A) Clinical impact curve analysis (CICA). The clinical impact curve of the nomogram plots the number of COVID-19 patients classified as high risk, and the number of cases classified high risk with severe NCAP at each high-risk threshold. (B) Decision curve analysis (DCA). DCA compares the net clinical benefits of three scenarios in predicting the severe COVID-19 probability: a perfect prediction model (gray line), screen none (horizontal solid black line), and screen based on the nomogram (red line).

DISCUSSION

In this study, we built a nomogram that showed good performance in predicting the progression of nonsevere COVID-19 and good agreement between the prediction and observations in identifying the probability of progression of nonsevere COVID-19. As the manifestations and severity of COVID-19 vary from asymptomatic cases to severe cases[4, 25], and if the progression of the nonsevere type into the severe type could be predicted earlier just based on the simplified risk factors, the medical resources could be allocated more reasonably, the identified high-risk patients could be given critical care more immediately, and a better balance could be achieved by epidemic control, disease severity evaluation, and the allocation of medical resources. Because this method is simple and easily practicable based on physical evaluation and the limited associated examinations, such as routine blood tests and radiological changes, which are readily available, it is suitable and helpful for use in temporary hospitals such as Fangcang shelter hospitals and primary health-care centers during the pandemic.

The median time from illness onset of COVID-19 to the first recorded medical evaluation was 5–6 days, while the median time from illness onset to the progression of the disease was 11 days in our study. In a previous study, the median time from illness onset to acute respiratory distress syndrome (ARDS) was 12 days.[26] Thus, if the nomogram was used, there would be approximately 5 days in advance to predict progression and intervene in patients likely to experience progression of the disease.

Many prediction models have been built for the severity determination or the progression risks of COVID-19,[26, 27] but the predictors and methods differ based on the index involved in different studies based on different clinical settings with different sample sizes.[28] In addition, all the prediction models were appraised to have a high risk of bias owing to a combination of poor reporting and poor methodological conduct for participant selection, predictor description, and statistical methods used, and 16.67% of the 66 reported prediction models had external validation, and calibration was rarely assessed.[28] Thus, in our study, we enrolled COVID-19 patients in the development and external validation cohorts to develop a nomogram for the evaluation of the progression of nonsevere COVID-19 into severe COVID-19. In addition, we only employed the readily available index acquired in the Fangcang shelter hospitals based on the first simplified evaluation of the patients. Thus, only disease history, symptoms, vital signs, routine blood tests, CRP, and radiological images were involved.

In total, seven predictors in multivariate analyses were independently associated with the progression of COVID-19 from a nonsevere COVID-19 to a severe type in our study. Based on a backward step-down selection process, nine variables were included in the nomogram in our study according to the complex evaluation of the prediction risk factors. Diabetes mellitus and anemia, which were not fully significant in the initial multivariate analyses, were also included in the final model. Comorbidities are risk factors for a poor outcome.[15, 26, 29] Regarding comorbidities, hypertension (17.5%) and diabetes mellitus (9.1%) were the

most common[29] and may be risk factors for patients with severe disease compared with those with the nonsevere disease,[30] which is in accordance with the comorbidity predictors from the univariate analysis associated with the progression COVID-19. However, in our study, in the multivariate logistic regression analysis performed to develop the nomogram, hypertension was not selected, which might be because of an overlapping association of these comorbidities with elderly age. Diabetes outweighed hypertension, possibly because diabetes mellitus not only influences the immune status of patients but also overlaps with older age.

In our study, we built a nomogram with age and sex, and male sex increased the weight of the probability of disease progression. Age is reported to be one of the significant risk factors for a poor outcome,[15, 26, 29] while the sex impact varies. Based on a systematic review with a pool of 31 studies, age and sex were involved in 12 and 6 prediction models, respectively, targeting the prediction of the risk of hospital admission, diagnosis, and prognosis of COVID-19.[28] In one study, it was found that the optimal cutoff of age for predicting imaging progression on chest CT was 51 years, with a sensitivity, specificity, PPV, and NPV of 0.65, 0.58, 0.35, and 0.83, respectively, and the ROC curve revealed that the AUC of age in the prediction model was 0.6.[31] Earlier studies indicated that males might be more common in COVID-19 patients due to occupational exposure, but subsequent studies revealed sex equivalence.[21] Furthermore, androgen-regulated gene, transmembrane protease, and serine 2 (TMPRSS2) expression in lung tissue expressed mainly in the adult prostate may explain the increased susceptibility of males to severe COVID-19 complications. Moreover, angiotensin-converting enzyme 2 (ACE-2) acts as a functional receptor for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and male hormones are effective in the ACE-2 passageway and simplify SARS-CoV-2 entry into host cells.[32] Thus, we included both age and sex in the nomogram for the prediction of the progression of nonsevere COVID-19 to severe COVID-19, with older age and male sex as the weighing factors for the progression of COVID-19.

In many studies, dyspnea was not the most common symptom of COVID-19 due to the various clinical spectra of COVID-19 ranging widely from nonsevere illness to ARDS. On admission, the proportion of patients with symptoms of dyspnea (70.6% vs. 24.7% in the nonsurvivor group and recovered groups, respectively) has been reported to vary.[33] However, it was suggested that patients with dyspnea should be closely monitored, especially 1–2 weeks after symptom onset.[34] Based on the univariate and multivariate analyses in our study, dyspnea was selected into the nomogram for the evaluation of the risk for the progression of COVID-19, especially in the early stage, and it is important to evaluate the symptoms of dyspnea.

Tachycardia was identified as one of the predictors in the nomogram. Surely, tachycardia is an index of and response to stress and hypoxia. Furthermore, the cardiac injury was identified as a common condition (19.7%) in hospitalized patients with COVID-19 that was associated with worse severity, mortality, and multiple organ dysfunction, and greater proportions of patients with cardiac injury required noninvasive mechanical ventilation or invasive mechanical ventilation than those without cardiac injury. [35]Additionally, abnormally higher hypersensitive troponin I (>0.04 pg/mL; OR = 4.388) was associated with unfavorable clinical outcomes.[36] Myocarditis is depicted as another cause of morbidity among COVID-19 patients.[37] In addition, drugs currently used to treat COVID-19 are known to prolong the QT interval and can have a proarrhythmic propensity.[38] Thus, the monitoring of tachycardia might be helpful for the prediction of the severity of the disease and cardiac complications.

The weight of lymphopenia agrees with the findings of a previous study that showed that 85% of critically ill patients with COVID-19 presented with lymphopenia.[4, 26, 39] The longitudinal hematologic variation of the progression of COVID-19 showed that initial levels of absolute lymphocyte count were significantly lower in nonsurvivors regardless of the initial disease severity.[40] Endothelial dysfunction has been shown to induce disassembly of intercellular junctions, endothelial cell death, and blood–tissue barrier disruption, along with enhanced leukocyte adhesion and extravasation, which could contribute to the lymphopenia observed in patients with severe COVID-19.[41] It has been hypothesized that successful survival may be related to the adequate replenishment of lymphocytes that are killed by SARS-CoV-2.[42] In addition, anemia, which was one of the predictors in our nomogram, has been distinguished as one of the risk factors for mortality in inpatients in Wuhan, although hemoglobin has not.[26] Hemoglobin levels were lower with older age, a higher percentage of subjects with diabetes, hypertension and overall comorbidities, and admission to intensive care,[43] which is in accordance with the risk factors in our study to evaluate the progression of the disease. Compared to moderate cases, severe COVID-19 cases had lower hemoglobin and red blood cell counts and higher ferritin and red cell distribution widths.[43] Interestingly, thrombocytopenia is associated with an increased risk of severe disease and mortality in patients with COVID-19 with a greater than fivefold increased risk;[26, 44] however, only lymphopenia and anemia were included in the nomogram in this study, which outweighs the platelet count.

It is common for COVID-19 patients to have bilateral lung involvement and multilobar involvement on CT scans from the very beginning (28%), in the intermediate stage (76%), and in the late stage (88%).[45] Thirteen prediction models of the 66 prediction models reviewed were proposed to support the diagnosis of COVID-19 or COVID-19 pneumonia (and monitor progression) based on CT images.[28] The predictive performance varied widely, with estimated C-index values ranging from 0.81 to nearly 1. Importantly, the severity and area of opacification assessed on initial CT scans were significantly related to the progression of opacification on follow-up CT scans.[46] The ROC curve showed that the sensitivity and specificity of the CT score were 80.0% and 82.8%, respectively, for discriminating the nonsevere type and severe type.[47] Thus, the initial involvement of the segments on the CT scan indicated the risk of disease progression, which corresponded with our results.

A nomogram could supply a quantitative tool for risk stratification instead of categorized classification. In some studies, the progression or severity of COVID-19 patients was categorized into high risk or low risk, while some continuous variables, such as CRP, were divided by cutoff values, such as CRP >41.8 mg/L and were more likely to have severe complications.[26] We did not classify the risks or continuous variables into categories such as high risk or low risk because it was suggested that risks calculated by the accumulative points of the risk factors would be more flexible for individual decision-making in different clinical settings with various case volumes and medical resources. [14] In another nomogram building study, a nomogram was found to have good performance for the early prediction of severe COVID-19 (AUC 0.912/0.853 in the training and validation cohorts, respectively),[48] but the parameters varied from ours without the enrollment of radiography in the analysis and a higher requirement for laboratory tests such as direct bilirubin, albumin, and lactate dehydrogenase, which are not quite readily available in some primary medical settings.

Some limitations of this study should also be acknowledged. First, this was a retrospective study in which selection biases were unavoidable, although we included a large number of patients from multiple sites. The study was designed as a prospective study. However, the rapid control of COVID-19 in 2020 left the COVID-19 patients scattered in different regions and restricted the prospective external validation; thus, we reregulated the design as a retrospective validation. Nevertheless, the validation performance of the nomogram revealed a superior overall net benefit within the wide and practical ranges of threshold probabilities and impacted patient outcomes. However, if possible, a prospective study is needed to validate the reliability and generalizability of the nomogram. Second, the model was developed and validated using data from patients in Wuhan, China. There might be significant regional differences in the mortality rate and percentage of severe cases in different regions. However, based on one meta-analysis, with a wide range of severe types of COVID-19 (7.2–63.3%), with respect to the severity cohort, nonsurvivors had more significant decreases in lymphocyte count and reduced hemoglobin values than survivors, with no significant difference in mean difference between groups.[49] To ensure simplicity and practicability, we did not take all the laboratory indices into account, but instead, we employed the simplified indices available in the primary medical institution, and some might be relatively more important with regard to the severity of the disease. In addition, the symptoms and histories were mainly from patients' self-report at admission, which might lead to recall bias. Finally, the exclusion criteria might lead to bias that originated from those of even sicker patients with severe complications and was classified as progression due to other diseases but not COVID-19 and those of the nonsevere type without the minimum requirement of the follow-up period. However, our study used bootstrap resampling and an internal validation cohort and showed an ideal C-index and well-fitted calibration curves, indicating good consistency between the prediction and observation.

Herein, a nomogram was established to predict the progression of nonsevere COVID-19 built with a simplified index of age ≥ 65 years, comorbid diabetes, dyspnea, tachycardia, absolute lymphocyte count, anemia, CRP level ≥ 20 mg/L, and multilobar involvement (≥ 3 lobes) on chest CT. The proposed nomogram might aid in predicting the progression of nonsevere COVID-19; thus, COVID-19 with a high risk of disease progression could be identified in time, allowing an appropriate stratification therapeutic choice according to the potential disease severity. Future studies should be performed to confirm or explore the nomogram with a higher sample size.

Author Contributions

Conception and design: G.H. and X.L.L.; Administrative support: G.H.; Provision of study materials or patients: J.G.X., B.Z., X.K., Q. Zhou, D.J., and C.W.; Collection and assembly of data: J.G.X., B.Z., D.J., Q. Zhou, C.W., C.N.L., Q. Zhang, Y.G., and H.G.; Data analysis and interpretation: G.H., J.X., B.Z., K.X., C.W., and X.L.L.; Manuscript writing: all authors; Final approval of manuscript: all authors.

Acknowledgments

We thank Prof. Bin Cao (Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, National Clinical Research Center for Respiratory Diseases,

Beijing, China) for the review of the manuscript. We thank Prof. Hong Shang (First Hospital of China Medical University, Shenyang, China) for giving us the opportunity and support to work on the frontline of fighting against COVID-19. We thank Prof. Chen Wang (Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China; National Clinical Research Center for Respiratory Diseases, Beijing, China; Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China-Japan Friendship Hospital, Beijing, China; Chinese Academy of Engineering, Beijing, China) for the conception of Fangcang Shelter Hospitals in China. We thank Wei Chong (First Hospital of China Medical University, Shenyang, China), Xu Zhang (First Hospital of China Medical University, Shenyang, China), Quan Cai (First Hospital of China Medical University, Shenyang, China), Xin-ping Ji (Shengjing Hospital of China Medical University, Shenyang, China), Jia-hong Fu (Shengjing Hospital of China Medical University, Shenyang, China), and all the other team members working together in the Wuchang Fangcang Shelter Hospital.

Source of Funding

This work was supported by China Medical University [grant number: 02].

Conflict of Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethical Approval

The research protocol was approved by the Ethics Committees of the First Hospital of China Medical University (Ethics number 2020-13-2) and the Ethics Committees of the Tongji Hospital, Tongji Medical College (Ethics number TJ-C20200144), which waived the requirement for informed consent from the patients.

REFERENCES

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020; 395: 470-3.
2. Zhang Y, Xu J, Li H, Cao B. A Novel Coronavirus (COVID-19) Outbreak: A Call for Action. Chest 2020; 157: e99-101.
3. Xu XW, Wu XX, Jiang XG, Xu KJ, Ying LJ, Ma CL, *et al.* Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. BMJ 2020; 368: m606.
4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020; 395: 497-506.
5. Wang D, Hu B, Hu C, *et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020; 323: 1061-9.
6. Wax RS, Christian MD. Practical recommendations for critical care and anesthesiology teams caring for novel coronavirus (2019-nCoV) patients. Can J Anaesth 2020; 67: 568-76.
7. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, *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: 1081-9.
8. Chen S, Zhang Z, Yang J, Wang J, Zhai X, Bärnighausen T, *et al.* Fangcang shelter hospitals: a novel concept for responding to public health emergencies. Lancet 2020; 395: 1305-14.
9. Leung C. Risk factors for predicting mortality in elderly patients with COVID-19: A review of clinical data in China. Mech Ageing Dev 2020; 188: 111255.
10. Tan C, Huang Y, Shi F, Tan K, Ma Q, Chen Y, *et al.* C-reactive protein correlates with computed tomographic findings and predicts severe COVID-19 early. J Med Virol 2020; 92: 856-62.
11. Liu F, Li L, Xu M, Wu J, Luo D, Zhu Y, *et al.* Prognostic value of interleukin-6, C-reactive protein, and procalcitonin in patients with COVID-19. J Clin Virol 2020; 127: 104370.
12. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, *et al.* A Tool for Early Prediction of Severe Coronavirus Disease 2019 (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020; 71: 833-40.
13. Huang J, Cheng A, Lin S, Zhu Y, Chen G. Individualized prediction nomograms for disease progression in mild COVID-19. J Med Virol 2020; 92: 2074-80.
14. Shang W, Dong J, Ren Y, Tian M, Li W, Hu J, *et al.* The value of clinical parameters in predicting the severity of COVID-19. J Med Virol 2020; 92: 2188-92.
15. Wang K, Zuo P, Liu Y, Zhang M, Zhao X, Xie S, *et al.* Clinical and Laboratory Predictors of In-hospital Mortality in Patients With Coronavirus Disease-2019: A Cohort Study in Wuhan, China. Clin Infect Dis 2020; 71: 2079-88.
16. Collins GS, Reitsma JB, Altman DG, Moons KGM, members of the Tg. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. Eur Urol 2015; 67: 1142-51.
17. Cheng B, Hu J, Zuo X, Chen J, Li X, Chen Y, *et al.* Predictors of progression from moderate to severe coronavirus disease 2019: a retrospective cohort. Clin Microbiol Infect 2020; 26: 1400-5.
18. China N. China's National Heath Commission. Diagnosis and Treatment Scheme of 2019-nCoV(Version 6.0). Available from: http://www.nhc.gov.cn/jkj/s3577/202003/4856d5b0458141fa9f376853224d41d7.shtml. Accessed on Mar 7, 2020.
19. Fan G, Tu C, Zhou F, Liu Z, Wang Y, Song B, *et al.* Comparison of severity scores for COVID-19 patients with pneumonia: a retrospective study. Eur Respir J 2020; 56: 2002113
20. Park SY, Freedman ND, Haiman CA, Le Marchand L, Wilkens LR, Setiawan VW. Association of Coffee Consumption With Total and Cause-Specific Mortality Among Nonwhite Populations. Ann Intern Med 2017; 167: 228-235.
21. Wahl S, Boulesteix AL, Zierer A, Thorand B, van de Wiel MA. Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. BMC Med Res Methodol 2016; 16: 144.
22. Wang ZX, Qiu MZ, Jiang YM, Zhou ZW, Li GX, Xu RH. Comparison of prognostic nomograms based on different nodal staging systems in patients with resected gastric cancer. J Cancer 2017; 8: 950-958.
23. Yap WK, Shih MC, Kuo C, Pai PC, Chou WC, Chang KP, *et al.* Development and Validation of a Nomogram for Assessing Survival in Patients With Metastatic Lung Cancer Referred for Radiotherapy for Bone Metastases. JAMA Netw Open 2018; 1: e183242.

24. Youden WJ. Index for rating diagnostic tests. Cancer 1950; 3: 32-5.
25. Wang J, Wang L, Li L, Xu J, Xu C, Li X, *et al*. Enlightenments of Asymptomatic Cases of SARS-CoV-2 Infection. J Transl Int Med 2020; 8: 112-4.
26. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, *et al*. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020; 395: 1054-62.
27. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, *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.
28. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, *et al*. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020; 369: m1328.
29. Zhang JJ, Dong X, Cao YY, Yuan YD, Yang YB, Yan YQ, *et al*. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy 2020; 75: 1730-41.
30. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, *et al*. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis 2020; 94: 91-5.
31. Yang Z, Shi J, He Z, Lü Y, Xu Q, Ye C, *et al*. Predictors for imaging progression on chest CT from coronavirus disease 2019 (COVID-19) patients. Aging (Albany NY) 2020; 12: 6037-48.
32. Moradi F, Enjezab B, Ghadiri-Anari A. The role of androgens in COVID-19. Diabetes Metab Syndr 2020; 14: 2003-6.
33. Deng Y, Liu W, Liu K, Fang YY, Shang J, Zhou L, *et al*. Clinical characteristics of fatal and recovered cases of coronavirus disease 2019 in Wuhan, China: a retrospective study. Chin Med J (Engl) 2020; 133: 1261-7.
34. Goh KJ, Choong MC, Cheong EH, Kalimuddin S, Duu Wen S, Phua GC, *et al*. Rapid Progression to Acute Respiratory Distress Syndrome: Review of Current Understanding of Critical Illness from COVID-19 Infection. Ann Acad Med Singap 2020; 49: 10818.
35. Shi S, Qin M, Shen B, Cai Y, Liu T, Yang F, *et al*. Association of Cardiac Injury With Mortality in Hospitalized Patients With COVID-19 in Wuhan, China. JAMA Cardiol 2020; 5: 802-10.
36. Hu L, Chen S, Fu Y, Cai Y, Liu T, Yang F, *et al*. Risk Factors Associated With Clinical Outcomes in 323 Coronavirus Disease 2019 (COVID-19) Hospitalized Patients in Wuhan, China. Clin Infect Dis 2020; 71: 2089-98.
37. Babapoor-Farrokhran S, Gill D, Walker J, Rasekhi RT, Bozorgnia B, Amanullah A. Myocardial injury and COVID-19: Possible mechanisms. Life Sci 2020; 253: 117723.
38. Kochi AN, Tagliari AP, Forleo GB, Fassini GM, Tondo C. Cardiac and arrhythmic complications in patients with COVID-19. J Cardiovasc Electrophysiol 2020; 31: 1003-8.
39. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, *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: 475-81.
40. Chen R, Sang L, Jiang M, Yang Z, Jia N, Fu W, *et al*. Longitudinal hematologic and immunologic variations associated with the progression of COVID-19 patients in China. J Allergy Clin Immunol 2020; 146: 89-100.
41. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis 2020; 93: 284-6.
42. 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: 1021-8.
43. Taneri PE, Gomez-Ochoa SA, Llanaj E, *et al*. Anemia and iron metabolism in COVID-19: a systematic review and meta-analysis. Eur J Epidemiol 2020; 35: 763-73.
44. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. Clin Chim Acta 2020; 506: 145-8.
45. Bernheim A, Mei X, Huang M, *et al*. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. Radiology 2020; 295: 200463.
46. 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: 332-9.
47. 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: 327-31.
48. Chen R, Liang W, Jiang M, *et al*. Risk Factors of Fatal Outcome in Hospitalized Subjects With Coronavirus Disease 2019 From a Nationwide Analysis in China. Chest 2020; 158: 97-105.
49. Henry BM. COVID-19, ECMO, and lymphopenia: a word of caution. Lancet. Respir Med 2020; 8: e24.

How to cite this article: Li XL, Wu C, Xie JG, Zhang B, Kui X, Jia D, *et al*. Development and validation of a nomogram for predicting the disease progression of nonsevere coronavirus disease 2019. J Transl Intern Med 2021; 9: 131-42.