Development of a risk prediction model for severe patients with COVID-19

Pengqiang Du[1], Su Shen[2], Rufei Ma[3], Quncheng Zhang[4], Hongwei Zhao[1], Mingyue Chen[1], Ming Ni[1], Qingli Shen[5], and Xingang Li[2]

[1]Henan Provincial People's Hospital, Fuwai Central China Cardiovascular Hospital, Central China Fuwai Hospital of Zhengzhou University
[2]Beijing Friendship Hospital
[3]Henan Provincial People's Hospital, People's Hospital of Zhengzhou University
[4]Henan Provincial People's Hospital, People's Hospital of Zhengzhou University
[5]Henan Provincial People's Hospital

June 5, 2020

Abstract

An outbreak of COVID-19 has spread worldwide, and it is urgent to evaluate risk factors associated with severe cases. We aimed to identify risk factors for severe patients with COVID-19. A total of 52 patients with COVID-19 admitted to Henan Provincial People's Hospital were enrolled in this study, and patients' information was extracted from electronic medical records. The severity of symptoms of COVID-19 was divided into "mild" and "severe." Univariate analysis was used to select potential risk factors. A risk model was constructed using multivariable logistic regression analysis. Advanced age (OR = 1.098, 95% CI = 1.020-1.183) and number of comorbidities (OR = 6.067, 95% CI = 1.078-34.143) were significant risk factors for severe patients with COVID-19. The comorbidities included hypertension, diabetes, cardiovascular disease, respiratory disease, and past surgical history. A risk score was developed based on this model, and the developed risk prediction model had good discriminative power and was well-calibrated. This study indicates that advanced age and comorbidities are risk factors for severe patients with COVID-19. More attention should be paid to high-risk patients during hospitalization.

Introduction

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus, the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) (Lai et al., 2020). An outbreak of COVID-19 has spread worldwide. As of June 3, 2020, SARS-CoV-2 has caused 6.35 million infections and more than 380 thousand deaths in the world, according to the World Health Organization.

Mild symptoms associated with COVID-19 are mostly fever, tiredness, and dry cough. Symptoms remain mild throughout infection in the majority of patients. However, mild cases can develop into severe cases of viral infections in the patients' respiratory tracts become worse. Severe cases will have most likely developed pneumonia, acute respiratory distress syndrome, respiratory failure, shock, and other organ failures or risk of death (Wang et al., 2020; Wang et al., 2020; Zhou et al., 2020). The case fatality rate (CFR) is persistently increasing, and it is urgent to evaluate risk factors associated with severe cases as the pandemic evolves. Earlier intensive surveillance and treatment can then be performed to decrease the CFR.

Methods

Patient selection

1

This study was approved by the institutional ethics board of the Henan Provincial People's Hospital (ethics approval No: 2020[71]). The Henan Provincial People's Hospital was assigned responsibility for the treatments of patients with severe COVID-19. We conducted an exploratory study of all the 52 patients with COVID-19 admitted to the hospital from January 10 to March 30, 2020. The diagnosis of COVID-19 was based on the China National Health Commission (NHC) Guidelines for the Diagnosis and Treatment of COVID-19. Confirmed cases had radiologic evidence of pneumonia or infiltrate on chest CT. The methodology of RT-PCR used has been previously reported (Wang et al., 2020).

Data extraction

All data including demographics (sex, age, weight), diagnosis (mild or severe cases), comorbidities (hypertension, diabetes, cardiovascular disease, respiratory system disease, past surgical history), laboratory assessments (lymphocyte, white blood cell, procalcitonin, C-reactive protein, erythrocyte sedimentation rate, D-Dimer), liver function (alanine aminotransferase ALT, aspartate aminotransferase AST), renal function (blood urea nitrogen BUN, creatinine), and length of hospital stay were extracted from electronic medical records (blood tests drawn on admission were obtained).

Definition

According to the NHC Guidelines, the severity of symptoms of COVID-19 was divided into "mild" and "severe." "Severe" was defined as dyspnea, respiratory rate [?] 30/min, blood oxygen saturation [?] 93%, PaO2 / FiO2 ratio [?] 300 mmHg, respiratory failure, septic shock, multiple organ dysfunction/failure, or admission to the Intensive Care Unit (Yu et al., 2020).

Potential risk factors

Based on previous reports (Boettler et al., 2020; Chen et al., 2020; Jordan et al., 2020; Klonoff et al., 2020; Leung, 2020; Li et al., 2020; Liu et al., 2020; Long et al., 2020; Mehra et al., 2020; Wang et al., 2020; Zhang et al., 2020; Zheng et al., 2020; Zheng et al., 2020; Zhou et al., 2020), several risk factors were selected for analysis as possible confounding variables. These risk factors included age, sex, body weight, comorbidities (hypertension, diabetes, cardiovascular disease, respiratory system disease, past surgical history), liver function (ALT, AST), and renal function (BUN and creatinine).

Data analysis

According to the NHC definition, the severity of symptoms of COVID-19 was divided into mild-group and severe-group. Univariate analysis was used to select potential risk factors, and variables were significant at $P < 0.1$ in the univariate analyses. Continuous variables were presented as mean $\pm$ standard deviation (SD) (normal distribution) or median and interquartile range (IQR) (non-normal distribution), respectively. The independent t-test (normal distribution) or Mann-Whitney U test (non-normal distribution) were employed to compare the mild-group and the severe-group. Categorical variables were described as number (percentage) and analyzed using the χ^2 test or Fisher's exact test.

The significant risk factors ($P < 0.1$) were used to construct the risk model using multivariable logistic regression analysis. The area under the receiver operating characteristic (ROC) curve was used to measure the discriminatory power of the multivariable logistic regression model. The Hosmer-Lemeshow goodness-of-fit test was used to evaluate whether the number of expected events from the logistic regression model reflects the number of observed events in the data. To facilitate clinical usage, a risk score was developed according to the coefficients of the multivariable logistic regression model (Pan et al., 2020). All data analyses were performed using SPSS version 17.0 (IBM, NY, USA). A two-sided $P < 0.05$ was considered significant.

Results

Baseline information

A total of 52 patients with COVID-19 were enrolled in this study, with 31 severe cases and 21 mild cases. All patients were eventually cured and discharge. The detailed information is listed in **Table 1.** Blood

tests drawn on admission were extracted from the electronic medical record. Although some studies have reported that certain laboratory test indicators are significantly related to the patient's disease progression and outcome (Ji et al., 2020; Lin et al., 2020; Lippi et al., 2020; Liu et al., 2020; Long et al., 2020; Zhu et al., 2020), this study failed to extract the complete changes of test indicators during hospitalization. Some laboratory indicators may change at different times, and the values are not constant.

Univariate analysis

A comparison of patient characteristics between severe and mild patients is shown in **Table 1** . A total of 12 variables were significant in the univariate analyses ($P < 0.1$). Considering that laboratory variables may change during patient's hospitalization, and length of hospital stay is the result of the degree of disease. We did not include the laboratory variables and length of hospital stay in the multivariate logistic regression analysis. Finally, a total of seven variables were entered into multivariable logistic regression analysis, including age, number of comorbidities, hypertension, diabetes, cardiovascular disease, respiratory system disease, and past surgical history.

Multivariable logistic regression

Multiple logistic regression analysis found that advanced age (OR = 1.098, 95% CI = 1.020-1.183) and number of comorbidities (OR = 6.067, 95% CI = 1.078-34.143) were significant risk factors for severe patients with COVID-19 (**Table 2**). The multiple logistic regression model is expressed as follow:

Logit $(P) = -5.185 + 0.093 \times age + 1.803 \times$ comorbidity Equation 1

where age means years of age, and comorbidity represents the number of comorbidities, including hypertension, diabetes, cardiovascular disease, respiratory system disease, and previous surgical history. This risk model had acceptable discriminative power, with the area under curve (AUC) of 0.930 (95% CI = 0.862-0.999) (**Figure 1**), and was well-calibrated with the Hosmer-Lemeshow χ^2 statistic of 7.565 ($P = 0.477$).

Risk score development

To facilitate the clinical usage, each risk factor was assigned a risk score according to the multivariable logistic regression results (**Table 2**). The smallest OR value, age variable, was selected as a reference value. Considering that the regression coefficient of age is very small, 0.093, we divided age into several age groups based on the unit of 10 years. The youngest patient in this study was 20 years old, and 2 points were assigned to the 20-29 years old patients. The score increases by 1 point for every 10 years of age, and so on. The regression system of age was used to calculate the relative multiple of the comorbidity regression coefficient. After rounding, the scores for comorbidities were clearly defined. Each additional comorbidity scores 2 points with a maximum of 6 points under comorbidity. In this study, the maximum number of comorbidities was 3, and the model's prediction range should not exceed the data range used to develop the model.**Table 3** shows the detailed score assignment rules.

The total risk score ranged from 2 to 14, with corresponding predicted probabilities of severe patients ranging from 2.11%-99.98% (**Figure 2**). The risk scores were divided into three levels (low-risk, moderate-risk, high-risk) to facilitate the clinical use of this risk model: low-risk (score 2-4, severe patients incidence 3.47%-17.92%), moderate-risk (score 5-6, severe patients incidence 35.62%-56.98%), and high-risk (score 7-14, severe patients incidence 77.05%-99.95%).

Discussion

The main results of this study indicate that advanced age and multiple comorbidities (hypertension, diabetes, cardiovascular disease, respiratory system disease, and past surgical history) are risk factors for severe patients with COVID-19. To enhance the clinical use of this risk model, we developed a risk prediction model for COVID-19 patients. To the best of our knowledge, this is the first study to attempt to establish a risk prediction model of severe patients with COVID-19.

3

Age, as a risk factor, for critically ill patients, has been reported in several studies (Jordan et al., 2020; Leung, 2020; Li et al., 2020; Liu et al., 2020; Wang et al., 2020). Older patients have reduced body immunity, reduced multi-organ function, and the risk of developing severe illness increases with age. This study also performed a ROC analysis on the risk factor of age (Figure S1), and its AUC reached 0.922 (95% CI = 0.852-0.991). When age is 55 years old, the Youden index reaches a maximum value of 0.711. When the patient is older than 55 years old, the risk of converting to severe illness increases significantly. In the process of modeling, we also tried to use age as a categorical variable and divide it into two groups, older age, and younger age. Although it is more convenient for clinical application when converting a continuous variable into a categorical variable, it will lose the prediction accuracy of the model. Therefore, continuous variables were still used for logistic regression analysis.

When performing the single factor analysis, there are significant differences between severe and mild patients with regards to hypertension, diabetes, cardiovascular disease, respiratory system disease, and past surgical history ($P < 0.1$). The above comorbidities have been reported in the literature as risk factors (Aghagoli et al., 2020; Cook, 2020; Jordan et al., 2020; Klonoff et al., 2020; Leung, 2020; Li et al., 2020; Li et al., 2020; Li et al., 2020; Liu et al., 2020; Liu et al., 2020; Long et al., 2020; Wei et al., 2020; Yang et al., 2020; Zhang et al., 2020; Zheng et al., 2020; Zhou et al., 2020; Zhou et al., 2020). However, when performing multivariate logistic regression, these factors were not included in the final model, which may have a great relationship with the small sample size of this study. Therefore, we combined these comorbidities to reduce the dimensions and analyzed the number of comorbidities as risk factors. The results showed that the number of comorbidities had a significant impact on the risk of serious illness (OR = 6.067,95%CI = 1.078-34.143). The AUC obtained by the ROC analysis reached 0.871 (95% CI = 0.772-0.970) (Figure S2). Because our patients had a maximum of three comorbidities, the risk prediction model was given a maximum of 6 points (2 points for each comorbidity). It should be noted that two of the included patients had a stroke, two had chronic gastritis, one had hepatitis B, and one had gout. Because the sample size of these complications was too small, they were not considered when performing data analysis due to concern about introducing these into the model would lead to model deviation.

Several studies have reported that biochemical indicators are risk factors for severe patients (Ji et al., 2020; Lin et al., 2020; Lippi et al., 2020; Liu et al., 2020; Long et al., 2020; Mehra et al., 2020; Yao et al., 2020; Zhu et al., 2020). When extracting the data, we also obtained some laboratory indicators of the patient at the time of admission, but we did not obtain the full data of these indicators. Because some indicators might change during hospitalization, these factors were not included in the model when the model was built. From the data of mild and severe patients, leukocytes, calcitonin, and C-reactive protein all have significant differences (P < 0.05), suggesting that clinical attention should be paid to these values. When the above indicators increase, it may indicate that the patient's risk of becoming severe is increased or it has developed into a severe case. In addition, we believe that the changes in laboratory indicators may be due to changes in the patient's condition; they may not be the cause of leading to severe coronavirus disease. That is, the changes in the index of the laboratory examination are the result of the difference in the disease condition. Due to our limited data, our view needs to be supported by more data.

It has been reported in the literature that damage to liver function (Boettler et al., 2020; Lei et al., 2020) and kidney function (Cheng et al., 2020) is a serious risk factor. However, in our study, there is only one patient with impaired renal function and three patients with mildly impaired liver function in the severe-group. The sample size is too small to obtain positive results. Univariate analysis found that the patients' creatinine levels were significantly different between mild and severe patients, and the median levels were within the normal range in these two groups.

This study has the following limitations 1) our hospital might have given priority to treat more severe patients, which might lead to a lower proportion of mild cases in our study, 2) the sample size is limited, and our results need clinical validation, and 3) some continuous variables do not conform to the normal distribution, which will affect the power of statistical results.

In summary, advanced age and multiple comorbidities are risk factors for severe patients with COVID-19.

4

This study provides a relatively practical tool for clinicians to identify severe high-risk cases. More attention should be paid to high-risk patients before they develop into severe cases.

Competing interests

We declare no competing interests.

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Abbreviations:

COVID-19, coronavirus disease; SARS-CoV-2, severe acute respiratory syndrome coronavirus-2; CFR, case fatality rate; NHC, National Health Commission; ALT, alanine aminotransferase; AST, aspartate amino-transferase; BUN, blood urea nitrogen; SD, standard deviation; IQR, interquartile range; ROC, receiver operating characteristic; AUC, area under curve

Authors' contributions

X.L. and P.D. conceived the idea, designed and supervised the study, had full access to all data and took responsibility for the integrity of the data. R.M., Q.Z., H.Z., M.C., M.N., Q.S. were responsible for the collection the clinical and laboratory data. S.S. and X.L. analysed data and performed statistical analysis. All authors reviewed and approved the final version.

References

Aghagoli, G, Gallo Marin, B, Soliman, LB, Sellke, FW. (2020). Cardiac involvement in COVID-19 patients: Risk factors, predictors, and complications: A review. *J Card Surg* . doi:10.1111/jocs.14538

Boettler, T, Newsome, PN, Mondelli, MU, Maticic, M, Cordero, E, Cornberg, M, Berg, T. (2020). Care of patients with liver disease during the COVID-19 pandemic: EASL-ESCMID position paper. *JHEP Rep* . 2(3), 100113. doi:10.1016/j.jhepr.2020.100113

Chen, R, Liang, W, Jiang, M, Guan, W, Zhan, C, Wang, T, . . . Medical Treatment Expert Group for, C. (2020). Risk Factors of Fatal Outcome in Hospitalized Subjects With Coronavirus Disease 2019 From a Nationwide Analysis in China. *Chest* . doi:10.1016/j.chest.2020.04.010

Cheng, Y, Luo, R, Wang, K, Zhang, M, Wang, Z, Dong, L, . . . Xu, G. (2020). Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int* . 97(5), 829-838. doi:10.1016/j.kint.2020.03.005

Cook, TM. (2020). The importance of hypertension as a risk factor for severe illness and mortality in COVID-19. *Anaesthesia* . 75(7), 976-977. doi:10.1111/anae.15103

Ji, HL, Zhao, R, Matalon, S, Matthay, MA. (2020). Elevated Plasmin(ogen) as a Common Risk Factor for COVID-19 Susceptibility. *Physiol Rev* . 100(3), 1065-1075. doi:10.1152/physrev.00013.2020

Jordan, RE, Adab, P, Cheng, KK. (2020). Covid-19: risk factors for severe disease and death. *BMJ* . 368, m1198. doi:10.1136/bmj.m1198

Klonoff, DC, Umpierrez, GE. (2020). Letter to the Editor: COVID-19 in patients with diabetes: Risk factors that increase morbidity.*Metabolism* . 108, 154224. doi:10.1016/j.metabol.2020.154224

Lai, CC, Shih, TP, Ko, WC, Tang, HJ, Hsueh, PR. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents* . 55(3), 105924. doi:10.1016/j.ijantimicag.2020.105924

Lei, F, Liu, YM, Zhou, F, Qin, JJ, Zhang, P, Zhu, L, . . . Yuan, Y. (2020). Longitudinal association between markers of liver injury and mortality in COVID-19 in China. *Hepatology* . doi:10.1002/hep.31301

Leung, C. (2020). Risk factors for predicting mortality in elderly patients with COVID-19: A review of clinical data in China. *Mech Ageing Dev* . 188, 111255. doi:10.1016/j.mad.2020.111255

Li, L, Zhang, B, He, B, Gong, Z, Chen, X. (2020). Critical patients with coronavirus disease 2019: Risk factors and outcome nomogram. *J Infect* . 80(6), e37-e38. doi:10.1016/j.jinf.2020.03.025

Li, T, Zhang, Y, Gong, C, Wang, J, Liu, B, Shi, L, Duan, J. (2020). Prevalence of malnutrition and analysis of related factors in elderly patients with COVID-19 in Wuhan, China. *Eur J Clin Nutr* . doi:10.1038/s41430-020-0642-3

Li, X, Xu, S, Yu, M, Wang, K, Tao, Y, Zhou, Y, . . . Zhao, J. (2020). Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. *J Allergy Clin Immunol* . doi:10.1016/j.jaci.2020.04.006

Lin, A, He, ZB, Zhang, S, Zhang, JG, Zhang, X, Yan, WH. (2020). Early risk factors for the duration of SARS-CoV-2 viral positivity in COVID-19 patients. *Clin Infect Dis* . doi:10.1093/cid/ciaa490

Lippi, G, Plebani, M. (2020). Procalcitonin in patients with severe coronavirus disease 2019 (COVID-19): A meta-analysis. *Clin Chim Acta* . 505, 190-191. doi:10.1016/j.cca.2020.03.004

Liu, W, Tao, ZW, Wang, L, Yuan, ML, Liu, K, Zhou, L, . . . Hu, Y. (2020). Analysis of factors associated with disease outcomes in hospitalized patients with 2019 novel coronavirus disease. *Chin Med J (Engl)* . 133(9), 1032-1038. doi:10.1097/CM9.0000000000000775

Liu, X, Zhou, H, Zhou, Y, Wu, X, Zhao, Y, Lu, Y, . . . Wang, Y. (2020). Risk factors associated with disease severity and length of hospital stay in COVID-19 patients. *J Infect* . doi:10.1016/j.jinf.2020.04.008

Long, L, Zeng, X, Zhang, X, Xiao, W, Guo, E, Zhan, W, . . . Lai, K. (2020). Short-term outcomes of COVID-19 and risk factors for progression. *Eur Respir J* . 55(5). doi:10.1183/13993003.00990-2020

Mehra, MR, Desai, SS, Kuy, S, Henry, TD, Patel, AN. (2020). Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. *N Engl J Med* . doi:10.1056/NEJMoa2007621

Pan, C, Wen, A, Li, X, Li, D, Zhang, Y, Liao, Y, . . . Shen, S. (2020). Development and Validation of a Risk Prediction Model of Vancomycin-Associated Nephrotoxicity in Elderly Patients: A Pilot Study.*Clin Transl Sci* . 13(3), 491-497. doi:10.1111/cts.12731

Wang, D, Hu, B, Hu, C, Zhu, F, Liu, X, Zhang, J, . . . Peng, Z. (2020). Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* . doi:10.1001/jama.2020.1585

Wang, D, Yin, Y, Hu, C, Liu, X, Zhang, X, Zhou, S, . . . Peng, Z. (2020). Clinical course and outcome of 107 patients infected with the novel coronavirus, SARS-CoV-2, discharged from two hospitals in Wuhan, China. *Crit Care* . 24(1), 188. doi:10.1186/s13054-020-02895-6

Wei, YY, Wang, RR, Zhang, DW, Tu, YH, Chen, CS, Ji, S, . . . Fei, GH. (2020). Risk factors for severe COVID-19: Evidence from 167 hospitalized patients in Anhui, China. *J Infect* . doi:10.1016/j.jinf.2020.04.010

Yang, J, Zheng, Y, Gou, X, Pu, K, Chen, Z, Guo, Q, . . . Zhou, Y. (2020). Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int J Infect Dis* . 94, 91-95. doi:10.1016/j.ijid.2020.03.017

Yao, Q, Wang, P, Wang, X, Qie, G, Meng, M, Tong, X, . . . Chu, Y. (2020). A retrospective study of risk factors for severe acute respiratory syndrome coronavirus 2 infections in hospitalized adult patients. *Pol Arch Intern Med* . 130(5), 390-399. doi:10.20452/pamw.15312

Yu, X, Sun, X, Cui, P, Pan, H, Lin, S, Han, R, . . . Fu, C. (2020). Epidemiological and clinical characteristics of 333 confirmed cases with coronavirus disease 2019 in Shanghai, China. *Transbound Emerg Dis* . doi:10.1111/tbed.13604

Zhang, J, Wang, X, Jia, X, Li, J, Hu, K, Chen, G, . . . Dong, W. (2020). Risk factors for disease severity, unimprovement, and mortality in COVID-19 patients in Wuhan, China. *Clin Microbiol Infect* . 26(6), 767-772. doi:10.1016/j.cmi.2020.04.012

Zheng, KI, Gao, F, Wang, XB, Sun, QF, Pan, KH, Wang, TY, . . . Zheng, MH. (2020). Letter to the Editor: Obesity as a risk factor for greater severity of COVID-19 in patients with metabolic associated fatty liver disease. *Metabolism* . 108, 154244. doi:10.1016/j.metabol.2020.154244

Zheng, Z, Peng, F, Xu, B, Zhao, J, Liu, H, Peng, J, . . . Tang, W. (2020). Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *J Infect* . doi:10.1016/j.jinf.2020.04.021

Zhou, F, Yu, T, Du, R, Fan, G, Liu, Y, Liu, Z, . . . Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* . 395(10229), 1054-1062. doi:10.1016/S0140-6736(20)30566-3

Zhou, Y, Zhang, Z, Tian, J, Xiong, S. (2020). Risk factors associated with disease progression in a cohort of patients infected with the 2019 novel coronavirus. *Ann Palliat Med* . 9(2), 428-436. doi:10.21037/apm.2020.03.26

Zhu, Z, Cai, T, Fan, L, Lou, K, Hua, X, Huang, Z, Gao, G. (2020). Clinical value of immune-inflammatory parameters to assess the severity of coronavirus disease 2019. *Int J Infect Dis* . 95, 332-339. doi:10.1016/j.ijid.2020.04.041

Figure legends

Figure 1

Receiver operating characteristic curves for the logistic regression model of severe patients with COVID-19. AUC = 0.930 (95% CI = 0.862-0.999).
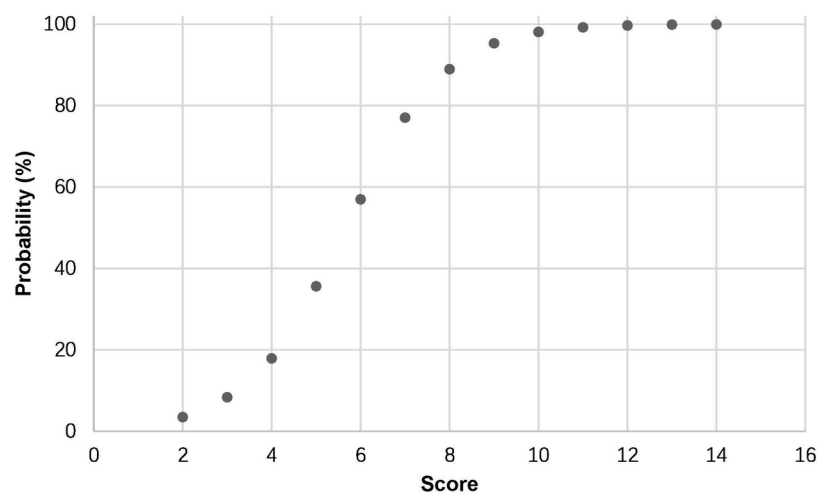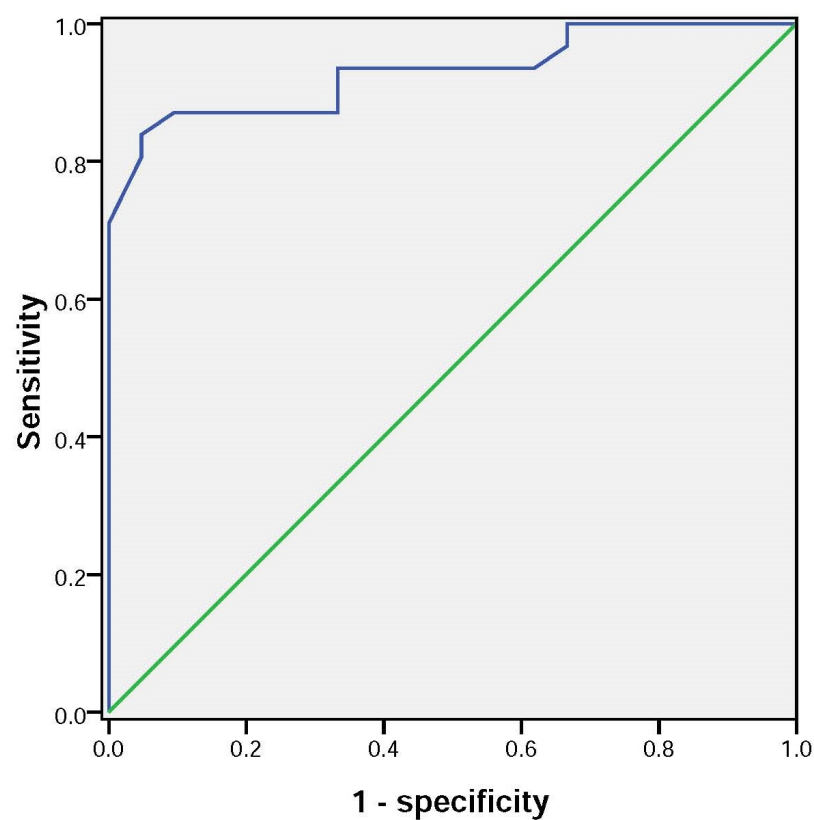
Figure 2

Scatter plot of predicted probabilities of severe patients versus risk score.

Figure S1

Receiver operating characteristic curves for age and severe patients with COVID-19. AUC = 0.922 (95%CI = 0.852-0.991)

Figure S2

Receiver operating characteristic curves for the number of comorbidities and severe patients with COVID-19. AUC = 0.871 (95%CI = 0.772-0.970).

Hosted file

Tables.docx available at https://authorea.com/users/330388/articles/457170-development-of-a-risk-prediction-model-for-severe-patients-with-covid-19