



Development and external validation of a prognostic tool for nonsevere COVID-19 inpatients

Original Paper

Cite this article: Luo E, Zhong Q, Wen Y, Cai J, Xie X, Zhou L (2023). Development and external validation of a prognostic tool for nonsevere COVID-19 inpatients. *Epidemiology and Infection*, **151**, e128, 1–10 <https://doi.org/10.1017/S0950268823000717>

Received: 15 September 2022

Revised: 26 April 2023

Accepted: 03 May 2023

Keywords:

COVID-19; development; duration of viral shedding; external validation; prediction model

Corresponding author:

Ensi Luo;

Email: luo_ensi@yeah.net

Ensi Luo¹ , Qingyang Zhong², Yongtao Wen³, Jie Cai³, Xia Xie⁴ and Lingjuan Zhou⁵

¹Department of Endocrinology, Binhaiwan Central Hospital of Dongguan, Dongguan, China; ²Medical Department, The Ninth People's Hospital of Dongguan, Dongguan, China; ³Department of Respiratory Medicine, Binhaiwan Central Hospital of Dongguan, Dongguan, China; ⁴Pain Department, Binhaiwan Central Hospital of Dongguan, Dongguan, China and ⁵Nursing Department, Binhaiwan Central Hospital of Dongguan, Dongguan, China

Abstract

To develop a machine learning model and nomogram to predict the probability of persistent virus shedding (PVS) in hospitalized patients with coronavirus disease 2019 (COVID-19), the clinical symptoms and signs, laboratory parameters, cytokines, and immune cell data of 429 patients with nonsevere COVID-19 were retrospectively reviewed. Two models were developed using the Akaike information criterion (AIC). The performance of these two models was analyzed and compared by the receiver operating characteristic (ROC) curve, calibration curve, net reclassification index (NRI), and integrated discrimination improvement (IDI). The final model included the following independent predictors of PVS: sex, C-reactive protein (CRP) level, interleukin-6 (IL-6) level, the neutrophil-lymphocyte ratio (NLR), monocyte count (MC), albumin (ALB) level, and serum potassium level. The model performed well in both the internal validation (corrected C-statistic = 0.748, corrected Brier score = 0.201) and external validation datasets (corrected C-statistic = 0.793, corrected Brier score = 0.190). The internal calibration was very good (corrected slope = 0.910). The model developed in this study showed high discriminant performance in predicting PVS in nonsevere COVID-19 patients. Because of the availability and accessibility of the model, the nomogram designed in this study could provide a useful prognostic tool for clinicians and medical decision-makers.

Introduction

Over the past 2 years, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been rapidly spreading worldwide. It has led to more than 587 million confirmed cases of coronavirus disease 2019 (COVID-19) and 6.42 million deaths worldwide [1]. Currently, COVID-19 is considered a global pandemic. With the variant iteration of SARS-CoV-2 and the implementation of active epidemic prevention measures and vaccination in China, the incidence and mortality of severe and critical COVID-19 are at low levels [2].

According to China's current epidemic prevention policy, COVID-19 can spread rapidly throughout a region, not only having a substantial impact on public health and causing a shortage of medical resources but also affecting the efficiency of the government and utilizing people's financial resources. Due to the differences in the data among regions, policies on the duration of inpatient and outpatient isolation for people with COVID-19 have been controversial. Knowledge of the duration of viral shedding (DVS) has significant implications for hospital infection prevention and control interventions, timely discharge management, and public health policies. Effective prediction of the PVS risk in patients with COVID-19 is critical for informing guidance around transmission-based isolation precautions. Recent studies indicated that PVS was associated with delirium incidence and six-month mortality in hospitalized COVID-19 patients [3]. Therefore, the continuous detection of the dynamic manifestations of viral shedding in COVID-19 patients provides appropriate clinical data for predicting the clinical course or prognosis of hospitalized patients.

Studies have shown that male, elderly, overweight COVID-19 patients have prolonged DVS [4, 5]. Abnormal inflammatory indicators, such as interleukin-6 (IL-6) levels, C-reactive protein (CRP) levels, lactate dehydrogenase (LDH) levels, neutrophil count (NC), lymphocyte count (LC), and monocyte count (MC), are important factors for an extended DVS in COVID-19 patients [6–8]. The neutrophil-lymphocyte ratio (NLR) is the ideal inflammatory marker, reflecting the systemic inflammatory response to disease [9]. Mahat et al. [10] performed a meta-analysis and showed that the NLR was significantly associated with the severity, length of stay, and prognosis of COVID-19 patients.

The Omicron variant has mutations in multiple sites, such as the spike protein, rendering a variety of neutralizing antibodies against SARS-CoV-2 ineffective or significantly less potent

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.



against this strain. The human debridement efficiency of the virus decreases, leading to a long-term infection state [2]. Studies have shown that the Omicron variant has a 13-fold increase in viral infectivity and is 2.8-fold more infectious than the Delta variant [11]. The Diagnosis and Treatment Protocol for COVID-19 (trial version 9) has corrected the boundary value of lifting isolation management and discharge standard cycle threshold (Ct) [12]. Bullard *et al.* [13] assessed the correlation of RNA Ct values with SARS-CoV-2 growth in cell culture. The data suggested that SARS-CoV-2 decreases infectivity when the Ct values are >24 and that each additional cycle with these Ct values reduces the odds ratio of restoring the virus by 32% in cell culture. La Scola *et al.* [14], assessing the correlation between SARS-CoV-2 isolation and real-time RNA Ct values in cell culture, found a significant relationship between Ct values and culture positivity. SARS-CoV-2 could not be isolated from any sample at a Ct value >34. Since February 2022, the COVID-19 epidemic has developed rapidly in Dongguan City, Guangdong Province, and the medical resources in the city have been significantly diminished. We realize that if medical personnel can effectively predict the PVS of COVID-19 patients, it could greatly improve the efficiency of medical resource use in medical institutions. Therefore, prediction models should be developed to predict PVS to help inform practical decisions related to infection control and public health policies.

Methods

Design and patients

In this study, 511 nonsevere COVID-19 patients hospitalized at the Ninth People's Hospital of Dongguan City, Guangdong Province, China, from February 2022 to June 2022 were enrolled. The COVID-19 diagnostic criteria and clinical classification criteria of the Diagnosis and Treatment Protocol for COVID-19 (trial version 9) were used [12]. All subjects had a complete medical history and information about disease evolution. The following patients were excluded according to the exclusion criteria: severe or critical COVID-19 patients; COVID-19 patients receiving Paxlovid, remdesivir, or molnupiravir; COVID-19 patients receiving convalescent plasma and high-dose glucocorticoids; COVID-19 patients with a history of glucocorticoid use; COVID-19 patients who died during treatment; COVID-19 patients with malignancy, human immunodeficiency virus (HIV) infection, and severe renal and liver disease; COVID-19 patients who were pregnant or lactating; COVID-19 patients aged below 18 years; and COVID-19 patients who were not subjected to examination and treatment. Ultimately, 429 patients, including 337 patients with local onset, were included for model development and internal validation, and 92 patients of overseas origin were used for external validation of the model (Figure 1).

Data collection

Clinical data were collected from the medical system of the Ninth People's Hospital, Dongguan City, Guangdong Province, including demographic information, COVID-19 vaccination history, typical COVID-19 symptoms (fever, dry cough, fatigue, decreased olfactory taste, nasal congestion, runny nose, sore throat, conjunctivitis, myalgia and diarrhea), hypertension, diabetes, Ct value of viral RNA, DVS, biochemical indicators, and other laboratory indicators. According to the Diagnosis and Treatment Protocol for COVID-19 (Trial Version 9), we developed a relevant

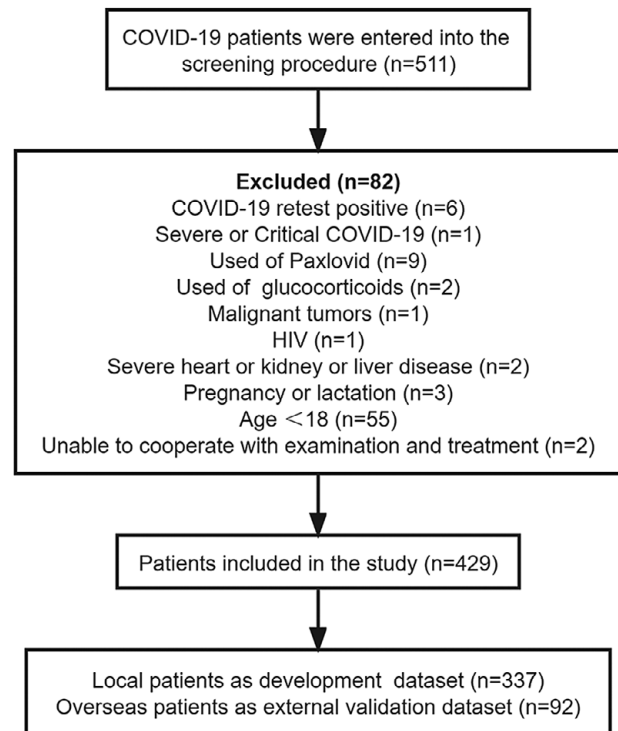


Figure 1. Patient screening and enrollment process.

definition of virus clearance to measure the outcome and time to virus clearance in patients: (i) DVS—the time from the first day of a positive RT–PCR result to the second consecutive negative viral RNA test was the duration cutoff point; (ii) Early virus clearance (EVC)—a DVS ≤ 14 days; (iii) persistent virus shedding (PVS)—a DVS >14 days; (iv) initial time of virus shedding—the first day a positive RT–PCR result was received; (v) endpoint of viral shedding—the second of two consecutive RT–PCR tests was negative, with tests performed at least 24 hours apart; (vi) negative viral RNA results—viral RNA test of the N gene and ORF gene Ct values greater than or equal to 35; and (vii) positive viral RNA result—a positive viral RNA test indicating that the ORF gene and/or N gene Ct values were lower than 35. Viral RNA testing was performed at the Dongguan CDC using commercial kits (supplementary Material). The commercial kit supplier was Shanghai BioGerm Medical Technology Co., Ltd.

Statistical analysis

Data were fully anonymized before data cleaning and analyses. In this study, the transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines were followed to develop, validate, and report multivariable prediction models. All data analyses were performed using IBM SPSS Statistics statistical software, version 24.0, and R language, version 4.0.1 (<https://www.r-project.org>). $P < 0.05$ was considered statistically significant. The sample size required for model development correlated with the number of candidate variables, the size of the total sample, and the proportion of outcome events [15]. In the early stage, we collected data from a small number of patients for statistical analysis. The incidence of an effective outcome was 40.0%, the mean absolute percentage error (MAPE) between the predicted outcome and the actual

outcome was 0.07, the number of candidate predictor variables was 11, and the required sample size was at least 312 patients. The dataset developed in the present study included 337 patients, with an effective outcome incidence of 42.1%, and the corresponding number of effective outcomes per candidate predictor variable was 12.9. First, the data were cleaned, and Multiple Imputation by Chained Equations (MICE) was used to transform the dataset into complete data [16]. The baseline data of the overall study patients, local patients, and overseas patients were statistically described. Local patient data were divided into the viral early clearance group (EVC Group) and PVS Group by a DVS of 14 days and a DVS longer than 14 days. Descriptive data were calculated and are provided as frequencies, percentages, or medians with interquartile ranges (IQRs). Patient characteristics in the model development and validation cohorts were compared through nonparametric tests. Multicollinearity was possible by correlation coefficients between independent variables. If the variance inflation factor (VIF) exceeded 10, the multicollinearity was considered significant, and the independent variable with high multicollinearity was excluded. We used the call in the R language to the restricted cubic spline (RCS) in the “rms” package. A P nonlinear >0.05 suggested a linear relationship between the outcome and independent variables. The “MASS” and “glmnet” packages in the R language were used to screen the following 11 candidate variables using the LASSO algorithm: sex, age, body mass index (BMI), ORF gene Ct value, NLR, MC, CRP levels, IL-6 levels, ALB levels, LDH levels, and K^+ levels. The model was constructed using multiple logistic regression by the “glm” package. The “PredictABEL” package was used to compare the advantages and disadvantages of the two model groups by the NRI and IDI. The calibration of the model was assessed by the C-statistic and the calibration curve. Decision curve analysis (DCA) was used to evaluate the net clinical benefit of the predictive model and was completed by the “ggDCA” package. The local patient data were used as the development dataset for the development model, and the prediction model was then internally verified by the self-service sampling method (boot = 1000). To assess the external validity of the model, overseas patient data were used as the validation set. The external validity of the model was assessed using the C-statistic, the calibration slope, and the calibration plot.

Results

Clinical characteristics of the patients

The development dataset included 337 patients of a native origin, and the validation set included 92 patients with an overseas origin. In the development dataset, the number of patients with PVS was 142 (42.1%), the median age was 37.0 years old (IQR: 29.0 to 50.0), and 143 patients were women (42.4%). The validation dataset included 92 patients, the number of patients with PVS was 36 (39.1%), the median age was 49.5 years old (IQR: 36.0 to 60.0), and the proportion of men (79.3%) was higher than that in the developed dataset ($P < 0.001$). Compared with the validation dataset, in the development dataset, fewer patients with hypertension (7.1% vs. 22.8%) and diabetes (4.5% vs. 16.3%) were present ($P < 0.001$). There were no differences in the time to admission, proportion of ACEI/ARB antihypertensive drugs used, number of vaccinations, BMI, NLR, and blood potassium levels between the development and validation datasets. From the baseline feature table, it can be seen that the mean age of the patients in the

development dataset was younger, with a large female proportion and smaller proportions of combined hypertension and diabetes compared with the validation set. The patients included in the development dataset had lower Ct values and Cr and alanine aminotransferase (ALT) levels for the primary detection of SARS-CoV-2 RNA, higher ALB levels, and higher levels of the inflammatory markers CRP and IL-6 than those included in the validation set. Full data are shown in Tables 1 and 2.

Model development

In the development dataset, 11 candidate variables were screened using the AIC and LASSO algorithms. The AIC screen identified seven candidate variables: sex, CRP levels, IL-6 levels, NLR, MC, ALB levels, and K^+ levels. The LASSO algorithm ($\lambda = \text{“lambda.1se”}$) screen identified 9 candidate variables: age, sex, Ct value of the ORF gene, CRP levels, IL-6 levels, NLR, MC, ALB levels, and K^+ levels. The above two models, Model 1 and Model 2, were fitted. We compared the performance of the two models using the NRI, IDI, C-statistic, Brier score, ROC curve, and AIC (Table 3). The NRI was 0.19% (95% CI: $-0.061, 0.065$), indicating that the accuracy of Model 2 fitted with 9 variables in distinguishing PVS was 0.19% higher than that of Model 1. The IDI was 0.61% (95% CI: $-0.002, 0.014$), indicating that the predictive power of Model 2 was 0.61% higher than that of Model 1. ROC (0.766 vs. 0.771, $P = 0.346$) and AIC (301.94 vs. 398.24) analyses were performed by the roc.test function and AIC function (Table 3). The above results showed that the number of predictor variables of Model 1 was less than that of Model 2 while maintaining good discrimination and accuracy. Therefore, sex, CRP levels, IL-6 levels, NLR, MC, ALB levels, and K^+ levels were ultimately determined to be the predictive model variables in Model 1 (Table 4). The ROC curve of the model showed excellent discrimination (C-statistic = 0.766, Brier score = 0.192) and good calibration (slope = 1) (Figure 2). The DCA plot indicated a good clinical net benefit in patients with PVS as predicted using Model 1 (Figure 3). The probability of predicting PVS occurrence in COVID-19 patients is presented as a nomogram (Figure 4a). A nomogram is a two-dimensional graphical tool that can be used to predict the probability of results, consisting of several lines arranged in scale. It shows a great advantage in simply and intuitively quantifying the risk of clinical events. The predictive model developed by our study renders specific risk prediction variables visualized through nomograms. For the final model, the probability of a DVS of >14 days in COVID-19 patients is presented as a nomogram. For example, a 46-year-old female patient with an NLR of 0.86, a CRP level of 4.4 mg/L, an IL-6 level of 5.53 $\mu\text{g/mL}$, an ALB level of 40.4 g/L, a K^+ level of 3.5 mmol/L, and an MC of $0.48 \times 10^9/\text{L}$ had a total score of 319. The probability of a viral shedding duration >14 days in this patient was 69.3% (Figure 5).

Model validation

Internal validation was performed by the bootstrap method, showing that the model had a corrected C-statistic = 0.748, a corrected Brier score = 0.201, a corrected slope = 0.910, and a Hosmer–Lemeshow test $P > 0.05$ (Table 5 and Figure 6). Therefore, Model 1 was considered not poorly fit.

The external validation dataset included 92 overseas COVID-19 patients, of whom 36 (39.1%) had PVS. Model 1 performed well in the external validation dataset, with a C-statistic = 0.845, a Brier score = 0.155, and a slope = 1 (Table 5 and Figure 7). The model

Table 1. Clinical characteristics of the patients

	Developmental dataset local patients (N = 337)	Validation dataset overseas patients (N = 92)	Total patients (N = 429)	P
Age, years	37.0 (29.0, 50.0)	49.5 (36.0, 60.0)	39.0 (30.0, 52.0)	<0.001
Sex				<0.001
Male, <i>n</i>	194 (57.6%)	73 (79.3%)	267 (62.2%)	
Female, <i>n</i>	143 (42.4%)	19 (20.7%)	162 (37.8%)	
Days since onset, days	0.0 (0.0, 1.0)	1.0 (0.0, 1.0)	0.0 (0.0, 1.0)	0.199
Comorbidities				
Hypertension, <i>n</i>	24 (7.1%)	21 (22.8%)	45 (10.5%)	<0.001
Diabetes, <i>n</i>	15 (4.5%)	15 (16.3%)	30 (7.0%)	<0.001
Use of ACEIs/ARBs	11 (3.3%)	6 (6.5%)	17 (4.0%)	0.156
COVID-19 vaccine doses				0.261
0 doses, <i>n</i>	18 (5.3%)	5 (5.4%)	23 (5.4%)	
1 dose, <i>n</i>	7 (2.1%)	2 (2.2%)	9 (2.1%)	
2 doses, <i>n</i>	129 (38.3%)	25 (27.2%)	154 (35.9%)	
3 doses, <i>n</i>	183 (54.3%)	60 (65.2%)	243 (56.6%)	
Symptomatic, <i>n</i>	198 (46.2%)	30 (32.6%)	228 (53.1%)	<0.001
Weight, kg	61.0 (55.0, 72.0)	65.5 (60.0, 72.0)	63.0 (55.0, 72.0)	0.020
Height, m	1.65 (1.58, 1.70)	1.68 (1.63, 1.71)	1.65 (1.58, 1.70)	0.001
BMI, kg/m ²	23.34 (21.11, 26.04)	23.60 (21.00, 25.87)	23.38 (21.07, 25.99)	0.902
ORF gene Ct value	23.0 (19.0, 30.0)	27.5 (21.5, 31.9)	24.0 (20.0, 30.0)	<0.001
N gene Ct value	24.0 (18.0, 30.0)	27.0 (22.0, 31.0)	24.0 (19.0, 30.0)	<0.001
ORF ≥25, <i>n</i>	155 (46.0%)	57 (62.0%)	212 (49.4%)	0.007
N ≥ 25, <i>n</i>	317 (94.1%)	85 (92.4%)	402 (93.7%)	0.558
WBC count, 10 ⁹ /L	6.01 (4.75, 7.39)	6.69 (5.39, 8.49)	6.09 (4.80, 7.73)	0.007
Neutrophil count, 10 ⁹ /L	3.96 (2.87, 5.36)	4.49 (3.23, 5.82)	4.09 (2.92, 5.47)	0.058
Lymphocyte count, 10 ⁹ /L	1.16 (0.79, 1.65)	1.47 (1.01, 1.97)	1.22 (0.84, 1.72)	<0.001
Monocyte count, 10 ⁹ /L	0.49 (0.36, 0.63)	0.49 (0.39, 0.60)	0.49 (0.38, 0.62)	0.809
NLR	3.20 (2.01, 6.03)	2.93 (1.97, 4.48)	3.19 (2.01, 5.72)	0.176
CRP, mg/L	3.89 (1.00, 6.92)	1.50 (0.45, 6.35)	3.10 (0.90, 6.92)	0.002
K ⁺ , mmol/L	3.89 (3.65, 4.09)	3.93 (3.69, 4.12)	3.90 (3.66, 4.09)	0.420
Cl ⁻ , mmol/L	104.00 (102.00, 106.00)	103.35 (101.85, 105.55)	103.80 (102.00, 106.00)	0.313
ALT, U/L	15.90 (11.50, 25.20)	18.90 (14.05, 28.65)	16.30 (11.80, 25.80)	0.024
AST, U/L	18.90 (16.00, 23.50)	19.30 (15.60, 23.00)	19.00 (15.90, 23.45)	0.976
ALB, g/L	46.10 (44.30, 48.00)	44.70 (43.05, 47.55)	45.90 (43.90, 47.90)	0.008
Cr, μmol/L	70.00 (59.00, 83.00)	79.00 (68.50, 90.00)	73.00 (59.00, 85.00)	<0.001
LDH, U/L	171.00 (152.00, 189.00)	170.00 (150.00, 187.00)	171.00 (151.50, 188.50)	0.458
CK, U/L	86.00 (64.00, 112.00)	100.50 (77.50, 138.50)	87.00 (65.00, 116.50)	0.007
CK-MB, U/L	11.70 (9.70, 14.20)	11.70 (10.10, 13.90)	11.70 (9.90, 14.00)	0.700
D-dimer, μg/mL	0.22 (0.22, 0.35)	0.22 (0.22, 0.29)	0.22 (0.22, 0.34)	0.226
IL-6, μg/mL	4.12 (1.97, 8.48)	3.26 (1.01, 5.11)	3.87 (1.68, 7.93)	0.008

ACEI, angiotensin converting enzyme inhibitor; ALB, albumin; ALT, alanine aminotransferase; ARB, angiotensin receptor blocker; AST, aspartate aminotransferase; BMI, body mass index; CK, creatine kinase; CK-MB, creatine kinase-MB; Cr, serum creatinine; CRP, C-reactive protein; IL-6, interleukin-6; LDH, lactate dehydrogenase; NLR, neutrophil-lymphocyte ratio; WBC, white blood cell.

Data are presented as the median (IQR) or *n* (%). *P* values were calculated using Wilcoxon's rank-sum test, the χ^2 test, or Fisher's exact test.

Table 2. Differences in demographics between COVID-19 patients with a duration of viral shedding ≤ 14 days and > 14 days in the developmental and validation datasets

DVS	Developmental dataset (local patients)			Validation dataset (overseas patients)		
	≤ 14 days	> 14 days	<i>P</i> value	≤ 14 days	> 14 days	<i>P</i> value
Age, years	36.0 (29.0, 45.0)	40.0 (30.0, 53.0)	0.017	42.5 (34.5, 55.5)	57.0 (46.0, 62.0)	0.003
Sex			0.955			0.033
Male, <i>n</i>	112 (57.4%)	82 (57.7%)		40 (71.4%)	33 (91.7%)	
Female, <i>n</i>	83 (42.6%)	60 (42.3%)		16 (28.6%)	3 (8.3%)	
Days since onset, days	0.0 (0.0, 1.0)	0.5 (0.0, 1.0)	0.353	1.0 (0.0, 1.0)	1.0 (0.0, 1.0)	0.581
Comorbidities						
Hypertension, <i>n</i>	12 (6.2%)	12 (8.5%)	0.418	10 (17.9%)	11 (30.6%)	0.157
Diabetes, <i>n</i>	7 (3.6%)	8 (5.6%)	0.369	7 (12.5%)	8 (22.2%)	0.218
Use of ACEIs/ARBs	7 (3.6%)	4 (2.8%)	0.766	5 (8.9%)	1 (2.8%)	0.398
COVID-19 vaccine doses			0.315			0.109
0 doses, <i>n</i>	9 (4.6%)	9 (6.3%)		5 (8.9%)	0 (0.0%)	
1 dose, <i>n</i>	2 (1.0%)	5 (3.5%)		1 (1.8%)	1 (2.8%)	
2 doses, <i>n</i>	79 (40.5%)	50 (35.2%)		18 (32.1%)	7 (28.0%)	
3 doses, <i>n</i>	105 (53.8%)	78 (54.9%)		32 (57.1%)	28 (77.8%)	
Symptomatic, <i>n</i>	111 (56.9%)	87 (61.3%)	0.424	14 (25.0%)	16 (44.4%)	0.052
Weight, kg	61.0 (55.0, 70.5)	61.0 (55.0, 72.5)	0.710	64.5 (58.0, 72.0)	69.9 (61.8, 72.5)	0.133
Height, m	1.65 (1.58, 1.70)	1.65 (1.57, 1.70)	0.425	1.67 (1.61, 1.73)	1.68 (1.65, 1.70)	0.580
BMI, kg/m ²	23.31 (20.76, 25.97)	23.44 (21.51, 26.04)	0.411	22.96 (20.81, 25.30)	24.39 (21.70, 26.47)	0.148
ORF gene Ct value	25.0 (19.0, 31.0)	23.0 (18.0, 28.0)	0.018	29.0 (24.0, 33.5)	24.0 (21.0, 28.5)	0.003
N gene Ct value	25.0 (19.0, 30.0)	23.0 (18.0, 29.0)	0.014	28.0 (24.0, 33.0)	24.0 (21.0, 28.5)	0.001
WBC count, 10 ⁹ /L	6.01 (4.58, 7.43)	5.96 (4.87, 7.43)	0.556	6.46 (4.93, 8.60)	6.98 (5.54, 8.23)	0.707
Neutrophil count, 10 ⁹ /L	3.77 (2.79, 5.31)	4.25 (3.04, 5.37)	0.150	4.29 (3.01, 5.84)	4.56 (3.30, 5.80)	0.611
Lymphocyte count, 10 ⁹ /L	1.25 (0.84, 1.74)	1.09 (0.72, 1.52)	0.004	1.54 (1.10, 2.02)	1.41 (0.92, 1.84)	0.229
Monocyte count, 10 ⁹ /L	0.49 (0.38, 0.63)	0.49 (0.35, 0.62)	0.823	0.48 (0.37, 0.62)	0.51 (0.43, 0.59)	0.218
NLR	3.02 (1.86, 5.10)	3.54 (2.30, 7.32)	0.006	2.76 (1.81, 4.37)	3.27 (2.12, 4.87)	0.183
CRP, mg/L	2.80 (0.90, 5.77)	4.25 (1.40, 9.10)	<0.001	1.00 (0.10, 3.65)	3.05 (1.03, 8.85)	0.002
K ⁺ , mmol/L	3.97 (3.76, 4.13)	3.82 (3.60, 3.98)	<0.001	3.96 (3.76, 4.18)	3.70 (3.57, 4.03)	0.010
Cl ⁻ , mmol/L	104.00 (101.85, 106.20)	103.65 (102.00, 106.00)	0.952	103.35 (101.40, 105.05)	103.55 (102.25, 106.00)	0.185
ALT, U/L	16.10 (12.05, 25.55)	15.35 (11.00, 24.30)	0.281	21.35 (14.45, 32.00)	17.00 (13.40, 22.15)	0.091
AST, U/L	19.00 (15.90, 23.70)	18.85 (16.20, 22.70)	0.756	20.60 (16.00, 23.30)	17.90 (14.55, 21.80)	0.145
ALB, g/L	46.40 (44.90, 48.40)	45.70 (43.60, 47.01)	<0.001	45.90 (43.80, 48.45)	43.95 (41.50, 46.40)	0.002
Cr, μ mol/L	70.00 (59.50, 81.50)	72.00 (56.00, 85.00)	0.626	77.50 (62.50, 89.00)	81.00 (74.00, 92.50)	0.098
LDH, U/L	169.00 (151.00, 189.00)	171.00 (153.00, 188.00)	0.779	169.00 (150.00, 182.00)	171.50 (150.00, 189.50)	0.908
CK, U/L	86.00 (65.00, 111.00)	88.00 (64.00, 113.00)	0.986	89.50 (75.00, 127.50)	106.50 (79.50, 155.50)	0.261
CK-MB, U/L	11.70 (9.90, 13.95)	11.75 (9.60, 14.20)	0.728	11.50 (10.10, 13.40)	12.35 (10.10, 14.35)	0.331
D-dimer, μ g/mL	0.22 (0.22, 0.34)	0.22 (0.22, 0.39)	0.538	0.22 (0.22, 0.28)	0.23 (0.22, 0.31)	0.616
IL-6, μ g/mL	3.61 (1.56, 7.43)	5.50 (2.47, 11.18)	0.001	2.57 (1.00, 4.62)	3.87 (1.62, 7.78)	0.066

ACEI, angiotensin converting enzyme inhibitor; ALB, albumin; ALT, alanine aminotransferase; ARB, angiotensin receptor blocker; AST, aspartate aminotransferase; BMI, body mass index; CK, creatine kinase; CK-MB, creatine kinase-MB; Cr, serum creatinine; CRP, C-reactive protein; DVS, duration of viral shedding; LDH, lactate dehydrogenase; IL-6, interleukin-6; NLR, neutrophil-lymphocyte ratio; WBC, white blood cell.

Data are presented as the median (IQR) or *n* (%). *P* values were calculated using Wilcoxon's rank-sum test, the χ^2 test, or Fisher's exact test.

Table 3. Model 1 for the odds ratios and beta coefficients in the development dataset

	Model 1	Model 2
C-statistic	0.766	0.771
Brier score	0.192	0.191
ROC curve	0.766	0.771
AIC	301.24	398.96

AIC, Akaike information criterion; ROC curve, receiver operating characteristic curve.

Table 4. Model 1 for the odds ratios and beta coefficients in the development dataset

Variable	Beta coefficients (95% CI)	Odds ratio (95% CI)
Sex	0.593 (1.063–3.123)	1.810 (1.057–3.098)
CRP	0.084 (1.040–1.142)	1.642 (1.250–2.159)
IL-6	0.059 (1.014–1.113)	1.466 (1.084–1.983)
NLR	0.096 (1.034–1.176)	1.467 (1.136–1.895)
MC	−0.332 (0.070–0.935)	0.698 (0.493–0.989)
ALB	−0.232 (0.709–0.882)	0.424 (0.284–0.635)
K ⁺	−1.643 (0.080–0.450)	0.485 (0.332–0.709)

ALB, albumin; CRP, C-reactive protein; IL-6, interleukin-6; MC, monocyte count; NLR, neutrophil-lymphocyte ratio.

was verified by the bootstrap method to have a corrected C-statistic = 0.793, a corrected Brier score = 0.190, and a corrected slope = 0.672. Model 1 overestimated individual PVS at a probability of less than 30% and underestimated individual PVS at a probability greater than 30%.

Discussion

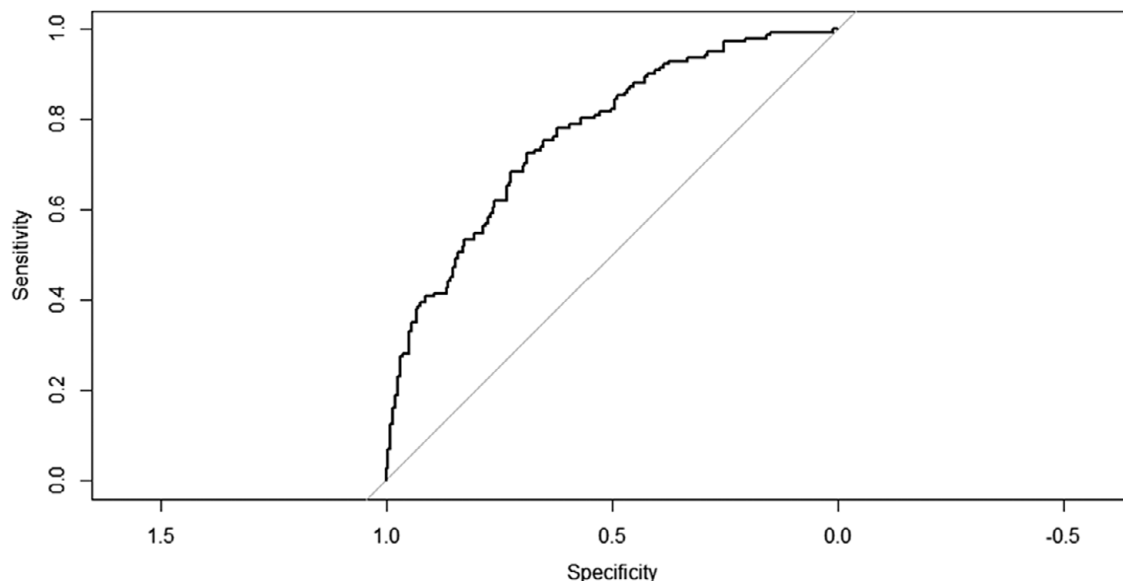
There have been very few clinical prediction model studies of clinical characteristics and DVS. In our study, the median DVS

for the development dataset was 13 days (IQR: 11 ~ 16). According to the study of the Guangdong Provincial Center for Disease Control and Prevention, the total median DVS time of COVID-19 patients in Guangdong Province is 17 days (IQR: 11–25) [17].

Many recent reports of COVID-19 have revealed that male patients have more severe symptoms and a longer DVS than female patients. Many epidemiological data and observations on COVID-19 at home and abroad share a common feature: men are more likely to show worse disease severity, mortality, and DVS than women [18]. In SARS-CoV-2 infection, the innate immune response is a major driver of viral clearance and pathogenesis. Women are considered to have a stronger immune system than men because they exhibit lower rates of infectious diseases and mortality and show a higher response to various types of vaccination than men [19]. Specific mechanisms might be related to the sex hormones that regulate immune function [20].

ACE2 is considered the counterregulator of ACE. It mainly degrades angiotensin II (Ang II), which has vasoconstriction, cell proliferation, and proinflammatory effects, into angiotensin 1–7 (Ang 17), and it has vascular dilation, cell apoptosis, and anti-inflammatory effects. ACE2 is generally expressed in the cardiovascular, respiratory, urinary, and digestive systems [21]. SARS-CoV-2 enters host cells through ACE2 and causes host ACE2 expression downregulation and membrane ACE2 shedding. This change results in elevated RAAS activity. The host develops vasoconstriction, thrombosis, inflammation, and increased angiogenesis, triggering poor prognosis [22].

Electrolyte disorder is a relatively common clinical manifestation of hospitalized COVID-19 patients. ACE2 is highly expressed in kidney tissues, and SARS-CoV-2 infection causes ACE2 shedding or ACE2 internalization, altering the ratio between angiotensin II and angiotensin and enhancing RAAS activity, leading to increased water-sodium reabsorption, blood volume, and urinary potassium excretion, thereby increasing cardiac load and kidney injury [23]. In the case of SARS-CoV-2 infection, the presenting gastrointestinal symptoms cause potassium transgastrointestinal loss secondary to anorexia with concurrent disease [24]. Hypokalemia was found in 41% of COVID-19 patients admitted to the Tertiary Teaching Hospital of Modena (Italy) [25].

**Figure 2.** Receiver operating characteristic (ROC) curve of Model 1 (C-statistic = 0.766, Brier score = 0.192, slope = 1).

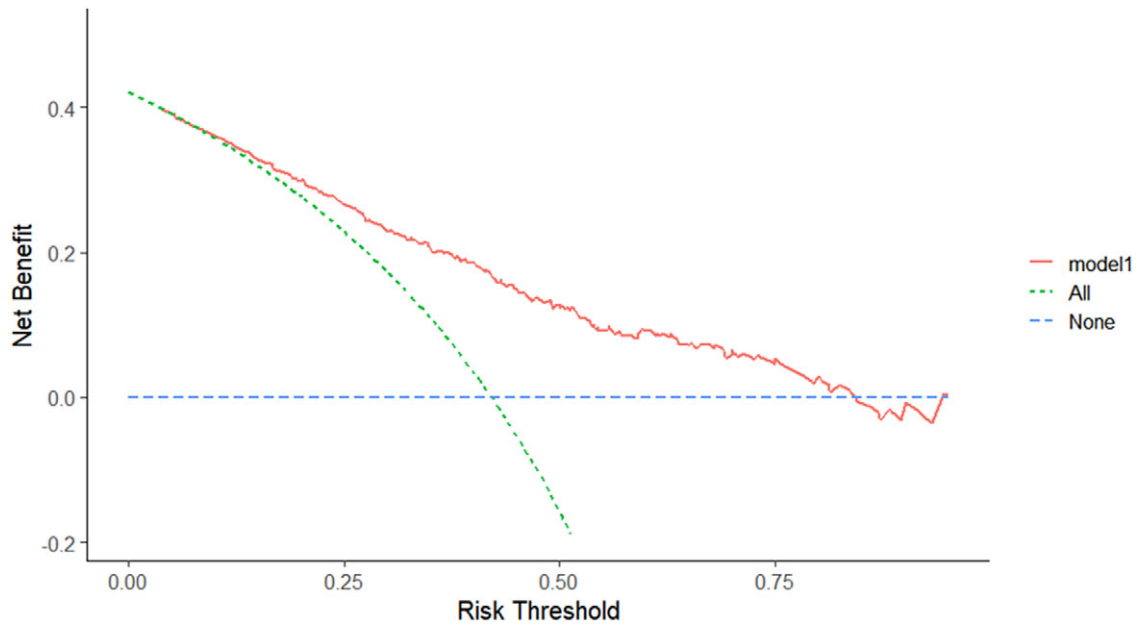


Figure 3. Decision curve analysis of Model 1. The DCA plot indicates a better clinical net benefit probability in patients with PVS as predicted using Model 1.

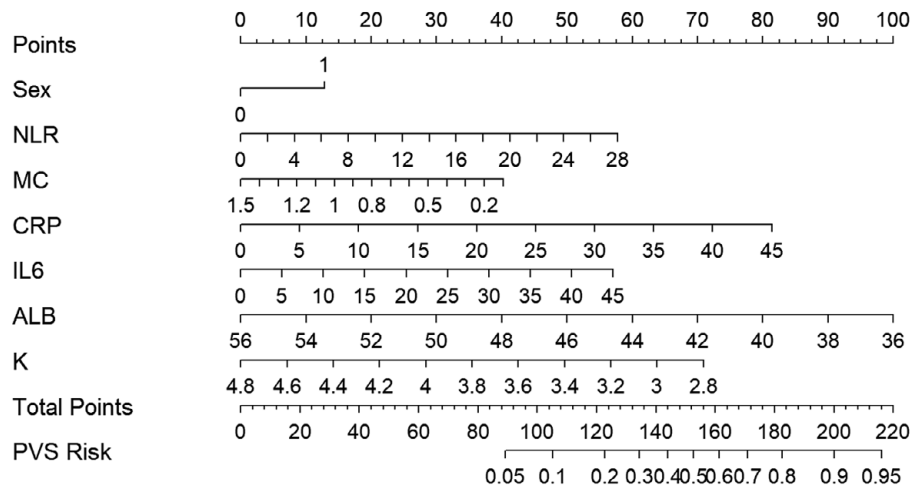


Figure 4. Nomogram of Model 1. Predicting the probability of the development of PVS in patients with COVID-19.

SARS-CoV-2 induces a hyperactivated inflammatory response that mediates multiorgan damage. IL-6 is one of the most prominent proinflammatory cytokines identified during SARS-CoV-2 infection [26]. Activation of the IL-6 pathway induces hepatocyte synthesis and the release of various acute-phase proteins, including CRP, and reduces synthesis from albumin [27]. This result indicates that the proinflammatory effect of IL-6 has a substantial effect on DVS. CRP is an acute-phase inflammatory protein produced by the liver that is regulated at the transcriptional level by the cytokines IL-6 and IL-1. It is an important indicator for the diagnosis and assessment of severe pulmonary infectious diseases [28]. In a meta-analysis, elevated IL-6 levels, CRP levels, and NLR values were associated with severity and mortality in COVID-19 patients [10].

ALB levels are associated with prognosis and DVS in COVID-19 patients. ALB levels were an independent risk factor in the study [29, 30]. Currently, it is speculated that the negative correlation between ALB levels and DVS might be related to the acute

inflammatory response, acute kidney injury, thrombosis, and the downregulation of ACE2.

Neutrophils (NEUs) are a major component of leukocyte cell populations that can be activated and migrate from the venous system to the immune organs or systems. NEUs release large amounts of reactive oxygen species that induce cellular DNA damage and release the virus from the cells. Furthermore, NEUs can be triggered by the virus-associated inflammatory factors IL-6 and TNF- α and granulocyte colony-stimulating factors produced by lymphocytes, endothelial cells, and interferons [31]. Conversely, the human immune response triggered by viral infection mainly depends on lymphocytes, and systemic inflammation significantly suppresses cellular immunity, thus significantly reducing CD4⁺ T lymphocytes and increasing CD8⁺ inhibitory T lymphocytes [32]. Lymphocytopenia in COVID-19 patients is mainly caused by CD4⁺ T lymphocyte depletion and is believed to be associated with a dysregulated immune response with more inflammatory

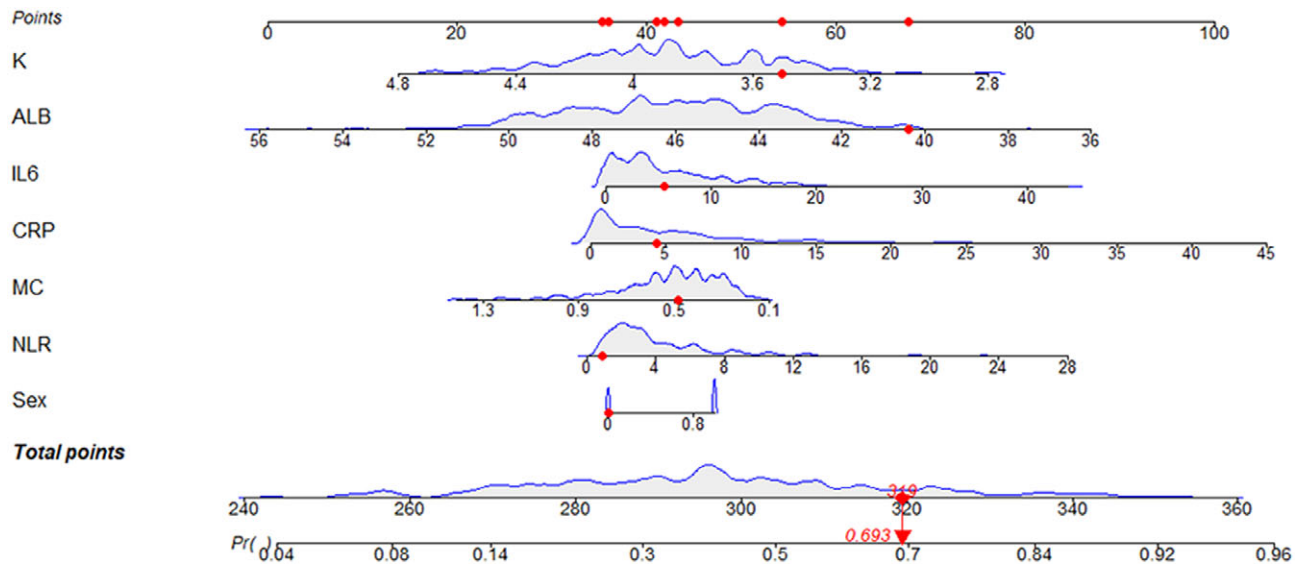


Figure 5. Risk prediction for individual patients. For example, a 46-year-old female patient with an NLR value of 0.86, a CRP level of 4.4 mg/L, an IL-6 level of 5.53 $\mu\text{g}/\text{mL}$, an ALB level of 40.4 g/L, a K^+ level of 3.5 mmol/L, and an MC of $0.48 \times 10^9/\text{L}$ had a total score of 319. The probability of a DVS > 14 days in this patient was 69.3%.

Table 5. Internal and external validation of Model 1

Measure	Internal validation: model performance			External validation: model performance		
	Original value	Optimism	Corrected	Original value	Optimism	Corrected
C-statistic	0.766	0.018	0.748	0.845	0.052	0.793
R^2	0.277	0.037	0.240	0.458	0.147	0.311
Brier score	0.192	-0.009	0.201	0.155	-0.035	0.190
Slope	1.000	0.090	0.910	1.000	0.328	0.672

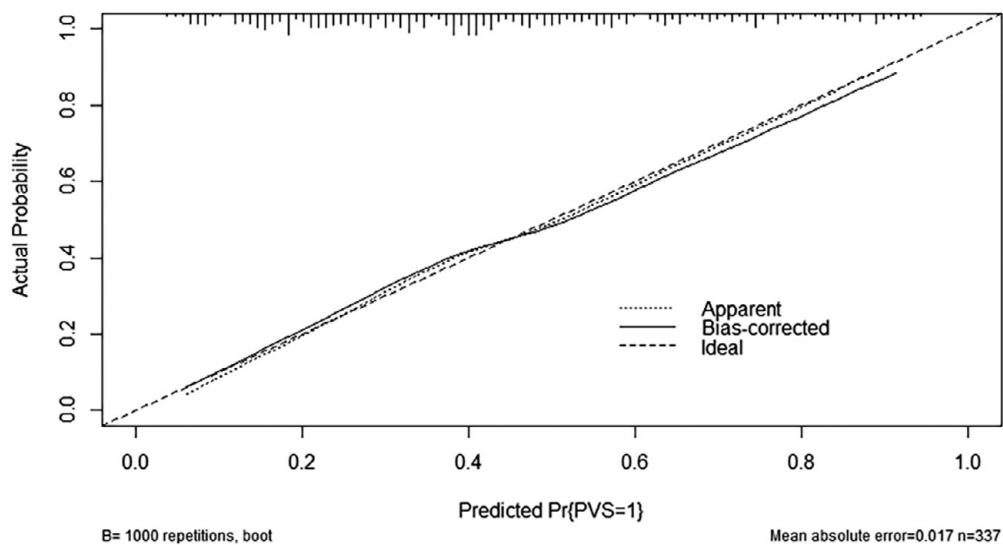


Figure 6. Calibration curve of Model 1 (corrected C-statistic = 0.748, corrected Brier score = 0.201, corrected slope = 0.910, Hosmer–Lemeshow test $P > 0.05$).

responses [33]. Ding *et al.* [34], in a retrospective study of 72 hospitalized COVID-19 patients, found significantly elevated leukocyte and NEU counts and a positive NLR value with an increased length of stay.

A nomogram is a two-dimensional graphical tool that can be used to predict the probability of an outcome, consisting of several lines arranged in scale. It demonstrates great superiority in simply and intuitively quantifying the risk of clinical events. A large

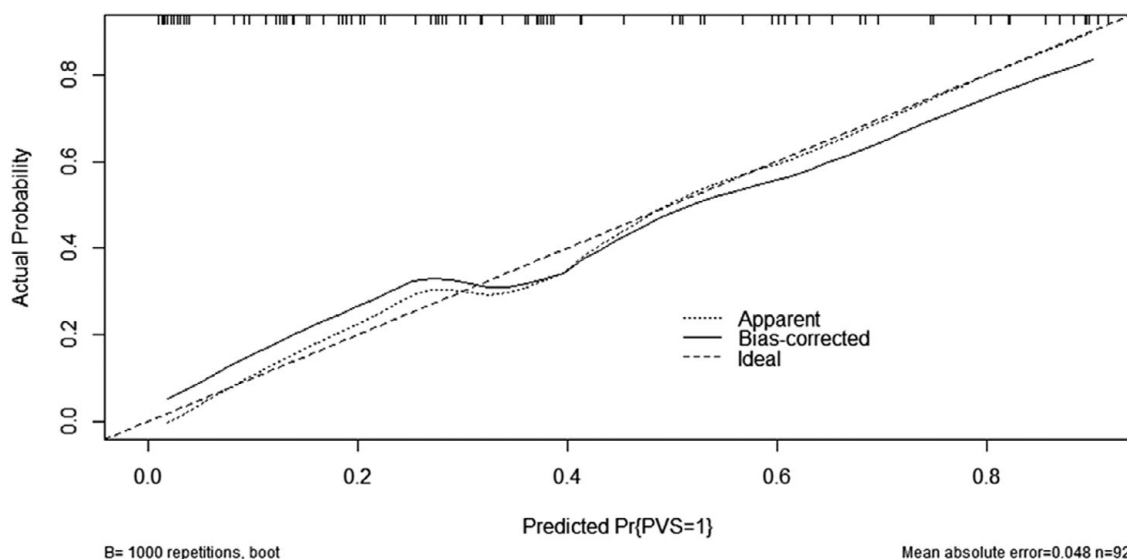


Figure 7. Calibration curve for the external validation of Model 1. Model 1 performed well in the external validation dataset, with a C-statistic = 0.845, Brier score = 0.155, and slope = 1. It was verified by the bootstrap method with a corrected C-statistic = 0.793, corrected Brier score = 0.190, and corrected slope = 0.672.

amount of clinical data is derived for COVID-19 patients during hospitalization, but it is difficult for clinicians to pay full attention to these data. Therefore, we have provided a visual description of the prediction model to help clinicians easily understand the model's prediction results and factors. The present study has several limitations. This was a retrospective study from a single center. Although we assigned local and overseas patients to different groups for comparison, the treatment scheme and isolation measures may limit the extrapolation of the research results. Although we used the bootstrap method and simplified model predictor variables to avoid overfitting, the sample size was relatively small, and there was still an inherent risk of overfitting. Our study only included patients with nonsevere COVID-19, so our results might not be generalizable to patients with severe COVID-19. Variables were removed due to more patients with missing data on the urinary leukocyte count, urinary red blood cell count, and urinary protein levels and the high collinearity of the N gene Ct values, NC, LC, and blood creatinine with other variables within the model. Therefore, the predictive model developed in this study could not be absolutely applicable for prognostic evaluation using the excluded variables. Despite these limitations, the model developed in this study showed excellent discrimination and calibration. A simple and easy-to-use nomogram could enable most clinicians to easily predict the prognosis of patients with COVID-19.

Conclusion

The predictive model developed in this study could efficiently and accurately calculate the probability of a DVS > 14 days by easily accessible laboratory examination results. It could contribute to saving limited medical resources and optimizing quarantine policies and treatment options in the future.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0950268823000717>.

Data availability statement. The data supporting the findings of this study are available from the Ninth People's Hospital of Dongguan. Restrictions apply

to the availability of these data, which were used under license for this study. Data are available from the corresponding author with the permission of the Ninth People's Hospital of Dongguan.

Acknowledgements. We thank the Ninth People's Hospital of Dongguan and Binhaiwan Central Hospital of Dongguan for their support of this study.

Author contribution. All authors have reviewed and approved the final version of the manuscript and contributed significantly to the work. Conceptualization: ESL, QYZ, YTW, and JC. Data collection: XX and LZZ. Data analysis: ESL, QYZ, and JC. Project administration: YTW and JC. Supervision: QYZ and LZZ. Writing: ESL and LZZ.

Financial support. This research received no specific grants from any funding agency in the commercial or not-for-profit sector.

Competing interest. The authors declare none.

Ethical standard. The Ninth People's Hospital, Dongguan City and Binhaiwan Central Hospital, Dongguan City approved the recollection of information.

References

- [1] **World Health Organization Coronavirus Disease Situation Dashboard.** Available at <https://covid19.who.int/> (accessed 16 July 2022).
- [2] **Guo Y, Han J, Zhang Y, He J, Yu W, Zhang X, Wu J, Zhang S, Kong Y, Guo Y, Lin Y and Zhang J (2022)** SARS-CoV-2 omicron variant: Epidemiological features, biological characteristics, and clinical significance. *Frontiers in Immunology* **13**, 877101.
- [3] **Batra A, Clark JR, Kang AK, Ali S, Patel TR, Shlobin NA, Hoffman SC, Lim PH, Orban ZS, Visvabharathy L, Graham EL, Sullivan DP, Muller WA, Chou SHY, Ungvári Z, Korálnik IJ and Liotta EM (2022)** Persistent viral RNA shedding of SARS-CoV-2 is associated with delirium incidence and six-month mortality in hospitalized COVID-19 patients. *Geroscience* **44**, 1241–1254.
- [4] **Zhang X, Lin B, Yang G, Liu L, Lu J, Lu Z and Xue Y (2021)** Delayed SARS-CoV-2 clearance in patients with obesity. *Infection and Drug Resistance* **14**, 2823–2827.
- [5] **Xu K, Chen Y, Yuan J, Yi P, Ding C, Wu W, Li Y, Ni Q, Zou R, Li X, Xu M, Zhang Y, Zhao H, Zhang X, Yu L, Su J, Lang G, Liu J, Wu X, Guo Y, Tao J, Shi D, Yu L, Cao Q, Ruan B, Liu L, Wang Z, Xu Y, Liu Y, Sheng J and Li L (2020)** Factors associated with prolonged viral RNA shedding in

- patients with coronavirus disease 2019 (COVID-19). *Clinical Infectious Diseases* **71**, 799–806.
- [6] **Chen X, Zhu B, Hong W, Zeng J, He X, Chen J, Zheng H, Qiu S, Deng Y, Chan JCN, Wang J and Zhang Y** (2020) Associations of clinical characteristics and treatment regimens with the duration of viral RNA shedding in patients with COVID-19. *International Journal of Infectious Diseases* **98**, 252–260.
- [7] **Çölkesen F, Poyraz N, Arslan Ş, Yılmaz PD, Kepenek Kurt E, Korkmaz C, Yıldız E, Çölkesen F, Aytekin G, Keleş S and Reisli İ** (2021) An association between immune status and chest CT scores in COVID-19 patients. *International Journal of Clinical Practice* **75**, e14767.
- [8] **Coomes EA and Haghbayan H** (2020) Interleukin-6 in Covid-19: A systematic review and meta-analysis. *Reviews in Medical Virology* **30**, 1–9.
- [9] **Lin SF, Lin HA, Chuang HC, Tsai HW, Kuo N, Chen SC and Hou SK** (2022) Fever, tachypnea, and monocyte distribution width predicts length of stay for patients with COVID-19: A pioneer study. *Journal of Personalized Medicine* **12**, 449.
- [10] **Mahat RK, Panda S, Rathore V, Swain S, Yadav L and Sah SP** (2021) The dynamics of inflammatory markers in coronavirus disease-2019 (COVID-19) patients: a systematic review and meta-analysis. *Clinical Epidemiology and Global Health* **11**, 100727.
- [11] **Dong M, Zhang J, Ma X, Tan J, Chen L, Liu S, Xin Y and Zhuang L** (2020) ACE2, TMPRSS2 distribution and extrapulmonary organ injury in patients with COVID-19. *Biomedicine & Pharmacotherapy* **131**, 110678.
- [12] **National Health Commission of the People's Republic of China** (2022) Diagnosis and treatment plan for COVID-19 (trial version 9). *International Journal of Epidemiology and Infectious Diseases* **49**, 73–80.
- [13] **Bullard J, Dust K, Funk D, Strong JE, Alexander D, Garnett L, Boodman C, Bello A, Hedley A, Schiffman Z, Doan K, Bastien N, Li Y, van Caesele PG and Poliquin G** (2020) Predicting infectious severe acute respiratory syndrome coronavirus 2 from diagnostic samples. *Clinical Infectious Diseases* **71**, 2663–2666.
- [14] **la Scola B, le Bideau M, Andreani J, Hoang VT, Grimaldier C, Colson P, Gautret P and Raoult D** (2020) Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards. *European Journal of Clinical Microbiology & Infectious Diseases* **39**, 1059–1061.
- [15] **van Smeden M, Moons KGM, de Groot JAH, Collins GS, Altman DG, Eijkemans MJC and Reitsma JB** (2019) Sample size for binary logistic prediction models: beyond events per variable criteria. *Statistical Methods in Medical Research* **28**, 2455–2474.
- [16] **Royston P** (2005) Multiple imputation of missing values: Update of ice. *The Stata Journal* **5**, 527–536.
- [17] **Xu H, Xie CY, Li PH, Ji ZL, Sun JF, Hu B, Li X and Fang M** (2022) Demographic, virological characteristics and prognosis of asymptomatic COVID-19 patients in South China. *Frontiers in Medicine* **9**, 830942.
- [18] **Fontana LM, Villamagna AH, Sikka MK and McGregor JC** (2021) Understanding viral shedding of severe acute respiratory coronavirus virus 2 (SARS-CoV-2): Review of current literature. *Infection Control and Hospital Epidemiology* **42**, 659–668.
- [19] **Márquez EJ, Chung CH, Marches R, Rossi RJ, Nehar-Belaid D, Eroglu A, Mellert DJ, Kuchel GA, Banchereau J and Ucar D** (2020) Sexual-dimorphism in human immune system aging. *Nature communications* **11**, 751.
- [20] **Giefing-Kröll C, Berger P, Lepperdinger G and Grubeck-Loebenstern B** (2015) How sex and age affect immune responses, susceptibility to infections, and response to vaccination. *Aging Cell* **14**, 309–321.
- [21] **Harmer D, Gilbert M, Borman R and Clark KL** (2002) Quantitative mRNA expression profiling of ACE 2, a novel homologue of angiotensin converting enzyme. *FEBS Letters* **532**, 107–110.
- [22] **Jackson CB, Farzan M, Chen B and Choe H** (2022) Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology* **23**, 3–20.
- [23] **Angeli F, Zappa M, Reboldi G, Trapasso M, Cavallini C, Spanevello A and Verdecchia P** (2021) The pivotal link between ACE2 deficiency and SARS-CoV-2 infection: one year later. *European Journal of Internal Medicine*; **93**: 28–34.
- [24] **Cheung KS, Hung IFN, Chan PPY, Lung KC, Tso E, Liu R, Ng YY, Chu MY, Chung TWH, Tam AR, Yip CCY, Leung KH, Fung AY, Zhang RR, Lin Y, Cheng HM, Zhang AJX, To KKW, Chan KH, Yuen KY and Leung WK** (2020) Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from a Hong Kong cohort: Systematic review and meta-analysis. *Gastroenterology* **159**, 81–95.
- [25] **The Modena Covid-19 Working Group (MoCo19), Alfano G, Ferrari A, Fontana F, Perrone R, Mori G, Ascione E, Magistroni R, Venturi G, Pederzoli S, Margiotta G, Romeo M, Piccinini F, Franceschi G, Volpi S, Faltoni M, Ciusa G, Bacca E, Tutone M, Raimondi A, Menozzi M, Franceschini E, Cuomo G, Orlando G, Santoro A, di Gaetano M, Puziolante C, Carli F, Bedini A, Milic J, Meschiari M, Mussini C, Cappelli G and Guaraldi G** (2021) Hypokalemia in patients with COVID-19. *Clinical and Experimental Nephrology* **25**, 401–409.
- [26] **Potere N, Batticciotto A, Vecchié A, Porreca E, Cappelli A, Abbate A, Dentali F and Bonaventura A** (2021) The role of IL-6 and IL-6 blockade in COVID-19. *Expert Review of Clinical Immunology* **17**, 601–618.
- [27] **Unver N and McAllister F** (2018) IL-6 family cytokines: Key inflammatory mediators as biomarkers and potential therapeutic targets. *Cytokine & Growth Factor Reviews* **41**, 10–17.
- [28] **Black S, Kushner I and Samols D** (2004) C-reactive protein. *Journal of Biological Chemistry* **279**, 48487–48490.
- [29] **Cogliati Dezza F, Oliva A, Cancelli F, Savelloni G, Valeri S, Mauro V, Calabretto M, Russo G, Venditti M, Turriziani O and Mastroianni CM** (2021) Determinants of prolonged viral RNA shedding in hospitalized patients with SARS-CoV-2 infection. *Diagnostic Microbiology and Infectious Disease* **100**, 115347.
- [30] **Violi F, Ceccarelli G, Cangemi R, Alessandri F, D'Etterre G, Oliva A, Pastori D, Loffredo L, Pignatelli P, Ruberto F, Venditti M, Pugliese F and Mastroianni CM** (2020) Hypoalbuminemia, coagulopathy, and vascular disease in COVID-19. *Circulation Research* **127**, 400–401.
- [31] **Yang AP, Liu JP, Tao WQ and Li HM** (2020) The diagnostic and predictive role of NLR, d-NLR and PLR in COVID-19 patients. *International Immunopharmacology*; **84**, 106504.
- [32] **Wang M, Zhu Q, Fu J, Liu L, Xiao M and du Y** (2020) Differences of inflammatory and non-inflammatory indicators in Coronavirus disease-19 (COVID-19) with different severity. *Infection, Genetics and Evolution* **85**, 104511.
- [33] **Phares TW, Stohlman SA, Hwang M, Min B, Hinton DR and Bergmann CC** (2012) CD4 T cells promote CD8 T cell immunity at the priming and effector site during viral encephalitis. *Journal of Virology* **86**, 2416–2427.
- [34] **Ding X, Yu Y, Lu B, Huo J, Chen M, Kang Y, Lou J and Liu Z** (2020) Dynamic profile and clinical implications of hematological parameters in hospitalized patients with coronavirus disease 2019. *Clinical Chemistry and Laboratory Medicine* **58**, 1365–1371.