ORIGINAL RESEARCH

A Model for Predicting the Duration of Viral Shedding in Patients Who Had Been Hospitalized with Mild COVID-19: A Single-Center Retrospective Study

Ya-Da Zhang[1],*, Tai-Wen He[2],*, Yi-Ren Chen[1], Bi-Dan Xiong[1], Zhe Zhe[1], Ping Liu[3], Bin-Qing Tang[1]

[1]Department of Pneumology, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, 200071, People's Republic of China; [2]Department of Ophthalmology, Shanghai Public Health Clinical Center, Shanghai, 201500, People's Republic of China; [3]Department of Tuberculosis, Shanghai Public Health Clinical Center, Shanghai, 201500, People's Republic of China

*These authors contributed equally to this work

Correspondence: Bin-Qing Tang, Department of Pneumology, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, No. 274 of ZhiJiang Middle Road, Jing'an District, Shanghai, 200071, People's Republic of China, Tel +86 13636657240, Email tangbinqi25ng@outlook.com; Ping Liu, Department of Tuberculosis, Shanghai Public Health Clinical Center, No. 2901 of Caolong Road, Jin'shan District, Shanghai, 201500, People's Republic of China, Tel +86 18939895095, Email iupingping0517@126.com

Background: Clinical decision-making is enhanced by the development of a mathematical model for prognosis prediction. Screening criteria associated with viral shedding time and developing a prediction model facilitate clinical decision-making and are, thus, of great medical value.

Methods: This study comprised 631 patients who were hospitalized with mild COVID-19 from a single center and 30 independent variables included. The data set was randomly divided into the training set (80%) and the validation set (20%). The outcome variable included viral shedding time and whether the viral shedding time >14 days, LASSO was used to screen the influencing factors.

Results: There were 321 males and 310 females among the 631 cases, with an average age of 62.1 years; the median viral shedding time was 12 days, and 68.8% of patients experienced viral shedding within 14 days, with fever (50.9%) and cough (44.2%) being the most common clinical manifestations. Using LASSO with viral shedding time as the outcome variable, the model with lambda as 0.1592 ($\lambda = 0.1592$) and 13 variables (eg the time from diagnosis to admission, constipation, cough, hs-CRP, IL-8, IL-1β, etc.) was more accurate. Factors were screened by LASSO and multivariable logistic regression with whether the viral shedding time >14 days as the outcome variable, five variables, including the time from diagnosis to admission, CD4 cell count, Ct value of ORF1ab, constipation, and IL-8, were included, and a nomogram was drawn; after model validation, the consistency index was 0.888, the AUC was 0.847, the sensitivity was 0.744, and the specificity was 0.830.

Conclusion: A clinical model developed after LASSO regression was used to identify the factors that influence the viral shedding time. The predicted performance of the model was good, and it was useful for the allocation of medical resources.

Keywords: COVID-19, LASSO, mild, prediction model viral shedding

Introduction

COVID-19 poses a global threat to public health since it is spread through the respiratory tract and is highly contagious. It is characterized by fever, cough, runny nose, headache, hyposmia, hypogeusia, and other symptoms,[1] followed by pneumonia and involvement with the digestive, neurological, and cardiovascular systems.[2] Omicron was a novel variant identified in South Africa and Botswana in November 2021 and had replaced delta as the main circulating variant, showing high transmissibility, immune evasion, and rapid mutation ability, but significantly reduced pathogenicity, and significant resistance to vaccines and antibodies.[3]

Mathematical models are essential for predicting COVID-19. Most of the current COVID-19-related prediction models focus mostly on the risk of severe cases and death,[4,5] and a small number of articles including hospitalization are still primarily focused on severe cases. Although mild cases account for the highest proportion in COVID-19, there

are few prediction models related to them. The reason may be that clinical decision-makers are more concerned with the risk of mild cases becoming severe or fatal, while ignoring the fact that the majority of mild cases are discharged from the hospital following improvement.[6] Although some prediction models focus on risk factor analysis with hospital stay as the end variable,[7] viral shedding is the primary criterion for discharge according to the Diagnosis and Treatment Scheme for COVID-19 in China (Ninth Edition Trial). Therefore, it is essential to develop a mathematical model to predict the duration of viral shedding in hospitalized COVID-19 cases that are not severe. Fourteen days is an important time period, which has been mentioned in many studies of COVID-19. The longest incubation period from infection to symptoms of SARS-COV2 virus is 14 days, which is often used as an isolation period.[8] A meta-analysis also found that the median length of hospital stay after COVID-19 infection was 14 days.[9] Taking 14 days as a node of shedding time is helpful to determine the influencing factors of nucleic acid changes.

According to studies, the median duration of viral shedding in mild cases is approximately 5 days.[10] The duration of viral shedding in COVID-19 carriers is closely associated with age, body mass index (BMI), vaccination, timely medical visit, and underlying diseases.[11–13] In addition to these variables, indexes like glutamic-pyruvic transaminase, blood platelet count, oxygenation index, D-dimer, and hemoglobin may also affect viral shedding.[14]

Numerous factors may be associated with viral shedding time, including binary variables, continuous variables, ordered categorical variables and unordered categorical variables that are clinically collected. Using linear regression may result in overfitting due to multicollinearity. A more concise model constructed by LASSO regression eliminates the problem of linear regression overfitting, and multivariate logistic regression modeling is undertaken using the elements screened by LASSO regression to generate a nomogram predicting the relationship between variables.

Methods
Case Source
This retrospective study was approved by the Shanghai Public Health Clinical Center and comprised of 631 patients with mild COVID-19 who were hospitalized at the Shanghai Public Health Clinical Center between 1 April 2022 and 31 May 2022. This study was conducted with approval from the Ethics Committee of Shanghai Public Health Clinical Center (2023-S061-01). This study was conducted in accordance with the declaration of Helsinki. Due to the retrospective nature of the study, the requirement of patient consent for inclusion was waived. Patient personal privacy and data confidentiality have been upheld.

Inclusion and Exclusion Criteria
The inclusion and exclusion criteria were made by reference to the Diagnosis and Treatment Scheme of COVID-19 in China (Ninth Edition Trial) prepared by the National Health Commission of the People's Republic of China (http://www.nhc.gov.cn/). Inclusion criteria: 1) patients aged >18 years; 2) patients with positive nucleic acid test for COVID-19; 3) patients with symptoms such as fever, weakness, shortness of breath, dry cough, nasal congestion, runny nose, hyposmia, and hypogeusia; 4) patients with normal lung imaging or interstitial changes in extrapulmonary zone or lung consolidation.

Exclusion criteria: 1) patients with the respiratory rate at the rest state ≥30 bpm; 2) patients with the fingertip oxygen saturation (air breathing) at the rest state ≤93%; 3) patients with oxygenation index ≤300 mmHg; 4) patients with obvious lesion progression >50% within 24~48 h shown in lung imaging; 5) pregnant women and children; 6) patients who had severe cardiovascular and cerebrovascular diseases, severe liver and kidney diseases, malignancy, immunodeficiency, and other symptoms, and who were admitted to the ICU due to unstable vital signs; 7) patients with incomplete clinical data.

Data Collection
Clinical manifestations, including (the presence or absence of) insomnia, fever, cough, pharyngalgia, dyspnea, constipation, and other symptoms, were extracted and recorded; the population characteristics of the included cases, such as

gender, age, time from diagnosis to admission, frequency of vaccination, smoking history, and history of underlying diseases, were extracted. The first nasal swab nucleic acid test after admission (Ct values of N and ORF1ab genes), lymphocyte percentage, hemoglobin, hs-CRP, D-dimer, COVID-19 antibody IgG, IgM antibody (negative, weakly positive, positive), CD4+T cell count, CD4+T cell count percentage, TNF-α, IL-1β, IL-2, IL-6, IL-8, IL-17, IL-12p70, and γ interferon were recorded. Positive findings in the first nucleic acid test defined day 0. Nasal swabs were collected once daily until virus shedding. Ct values of N and ORF1ab genes ≥35 in two consecutive tests, or negative results in two consecutive nucleic acid tests (at an interval of more than 24h), were recorded as viral shedding. All data were recorded and examined by two physicians, while two physicians randomly examined the data while confirming their accuracy.

Statistical Analysis

The statistical analysis was conducted using SPSS software 21.0 and R software 4.0.1. Continuous variables were expressed by mean ± standard deviation (x±s) or median (interquartile range) and were compared by the independent-sample t-test or M-W test; categorical variables were expressed by the absolute or percentage (%) and were compared by the chi-squared test or Fisher's exact test. The outcome variable of our data was a continuous variable (time of viral shedding), but we further categorize the time of viral shedding into a binary variable (whether the time of viral shedding was >14 days) in our data analysis. The time of viral shedding (continuous variable) and whether the time of viral shedding was more than 14 days (binary variable) were used as outcome variables. Through LASSO regression, influential factors were screened, and the regression equation between viral shedding time and influential factors was established. The factors impacting the outcome variable, namely whether the viral shedding time was longer than 14 days, were screened using LASSO regression. The most significant risk factors were then picked using multivariate logistic regression analysis, and a nomogram was drawn. The assessment of the prediction model using the consistency index (C-index). The R package "corrplot" was used for correlation analysis, "glmnet" was used for LASSO regression, "rms" and "survival" were used for logistic regression, and a nomogram was drawn.

Results

General Data Included in Analysis

This retrospective study comprised 768 cases admitted to this hospital between April 1, 2022, and May 31, 2022, with 631 cases included in the data analysis after 137 cases with incomplete observational variable data were excluded. The 631 cases included 321 males (50.4%) and 310 females (49.6%) with an average age of 62.1 years; most of these cases had underlying diseases (57.5%), and most were not vaccinated (51.3%); COVID-19 carriers were primarily clinically manifested by fever (50.9%) and cough (44.2%), and the virus was shed within 14 days in the majority of patients (68.8%) (Table 1).

Observation of Potential Factors Based on the Correlation Analysis Between Clinical Data and Viral Shedding Time

Based on the variable features, Spearman correlation analysis was used to determine the relationship between the data acquired for this investigation and the period of viral shedding. The viral shedding time in males was significantly higher than in females; the frequency of vaccination and lymphocyte percentage were significantly negatively correlated with viral shedding time, while Ct values of N and ORF1ab genes and the time from diagnosis to admission were significantly positively correlated (Figure 1).

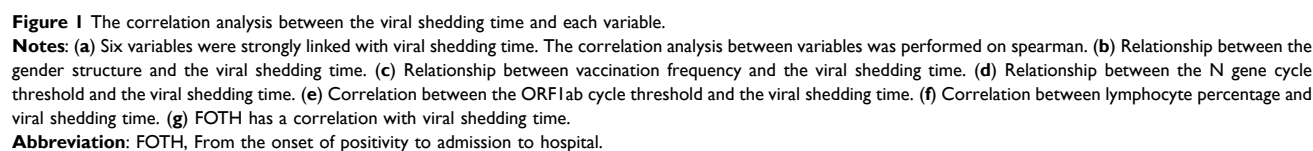### LASSO Regression Screening of Factors Influencing Viral Shedding Time

In this study, 30 independent variables were used to build the LASSO regression model, with viral shedding time being the dependent variable to prevent overfitting; 631 cases were divided into the training set and the validation set by a ratio of 8:2, with 504 cases in the training set and 127 cases in the validation set. There was no significant difference between the training set and the validation set in terms of gender, age, smoking, underlying diseases, frequency of vaccination, viral shedding time, and the admission time (P > 0.05). On the training set, LASSO regression was conducted, and the

Table 1 Demographic and Clinical Features of All Participants' Enrollment

Characteristics	Estimates	Percentage
Gender		
Male	321	50.9%
Female	310	49.10%
Age (mean(SD))(year)	62.2(18.7)	
18–45	128	20.30%
46–69	265	42.00%
>69	238	37.70%
Smoking		
Yes	13	2.10%
No	618	97.90%
Comorbidities		
Yes	363	57.50%
No	268	42.50%
None of Comorbidities		
1	247	39.10%
2	103	16.30%
≥3	13	2.10%
Vaccination (time)		
0	324	51.30%
1	36	5.70%
2	125	19.80%
3	146	23.1
Time of viral shedding (Median, Quartile)(day)	12	
≤14	434	68.80%
>14	197	31.20%
From the onset of positive to the admission to hospital (Median, Quartile)(day)	5	
Fever		
Yes	427	50.90%
No	204	49.10%
Cough		
Yes	279	44.20%
No	352	55.80%
Pharyngalgia		
Yes	122	19.30%
No	509	80.70%
Dyspnea		
Yes	20	3.20%
No	611	96.80%
Constipation		
Yes	92	14.60%
No	539	85.40%
Insomnia		
Yes	63	10.00%
No	568	90.00%

Abbreviation: SD, standard deviation.

complexity of the LASSO regression model was adjusted with parameter λ; after cross validation, the model corresponding to the minimum target parameter mean (lambda.min) and that corresponding to the fewest variables included in a variance range of lambda.min (lambda.1se) were obtained based on the λ value, and the optimal model was chosen based on the testing set data. Of the two models, the model corresponding to lambda.min included 13 variables, of which

Figure 1 The correlation analysis between the viral shedding time and each variable.

Notes: (**a**) Six variables were strongly linked with viral shedding time. The correlation analysis between variables was performed on spearman. (**b**) Relationship between the gender structure and the viral shedding time. (**c**) Relationship between vaccination frequency and the viral shedding time. (**d**) Relationship between the N gene cycle threshold and the viral shedding time. (**e**) Correlation between the ORF1ab cycle threshold and the viral shedding time. (**f**) Correlation between lymphocyte percentage and viral shedding time. (**g**) FOTH has a correlation with viral shedding time.

Abbreviation: FOTH, From the onset of positivity to admission to hospital.

the time from diagnosis to admission, smoking, and hypersensitive C-reactive protein value were positively correlated with viral shedding time, and strongly positive IgG antibody, Ct value of ORF1ab, CD4 T cell count, IL-8, IL-1b, CD4 T cell count percentage, lymphocyte percentage, cough, and three shots of COVID-19 vaccine were negatively correlated with viral shedding time. The model corresponding to Lambda.1se contained four variables, of which the period between diagnosis and admission was positively correlated with viral shedding time, while strongly positive IgG antibody, Ct value of ORF1ab, and CD4 T cell count were negatively correlated with viral shedding time. The LASSO model corresponding to lambda.min was used to predict the dependent variable of the validation set data, and the goodness-of-fit was evaluated based on the predicted value and the actual value, with a coefficient of determination of 0.751; the LASSO model corresponding to lambda.1se was used to predict the dependent variable of the validation set data, and the goodness-of-fit was evaluated based on the predicted value and the actual value, with a coefficient of determination of 0.692. The mean absolute error (MAE) and root-mean-square error (RMSE) of the LASSO model corresponding to lambda.min in predicting the dependent variable and in comparing with the actual value were 2.331 and 3.169, respectively, which were lower than those (2.630 and 3.570, respectively) of the model corresponding to lambda.1se, indicating that the 13-variable LASSO regression model corresponding to lambda.min was more accurate (Figure 2).

Viral shedding time = 13.424+(5.596*FOTH+0.166*Smoking+0.097*HsCRP+0.414*Coprostasis-0.006*vaccination3-1.526*IgG2-0.106*IL8-0.085*IL1β-0.024*Lymphocyte-1.248*ORF1abCt-0.054*CD4T%-0.507*CD4T-0.007*Cough)

Analysis of Case Characteristics with 14 Days as the Viral Shedding Time Node

434 and 197 cases reached the viral-shedding criteria within and beyond 14 days, respectively, among the 631 cases included in this study, with differences in gender distribution between the two groups (P<0.01); the two groups showed differences in vaccination (P<0.05), with a higher proportion of vaccination in the group <14 days than in the group >14 days. The incidence rates of age distribution, underlying diseases, smoking, fever, cough, pharyngalgia, insomnia, and dyspnea did not differ significantly between the two groups; however, the incidence rate of constipation was higher in the group >14 days than in the group <14 days; There were differences in the positive rate of IgG antibody between the two groups, with that in the group >14 days being higher; there were significant differences in the Ct values of nucleic acid N and ORF1ab genes between the two groups (P<0.01), with the Ct value in the group >14 days being higher than that in the group <14 days; there was no significant difference in the viral shedding time after admission between the two groups (P > 0.05), the time from diagnosis to admission in the group <14 days was significantly lower than that in the group >14 days (P < 0.01); The lymphocyte percentage in the group <14 days was significantly higher than that in the group >14 days (P<0.05); there were no significant differences between the two groups in terms of hemoglobin, hypersensitive C-reactive protein, D-dimer, CD4 cell count, CD4 cell percentage, TNF-α, IL-1β, IL-2, IL-6, IL-8, IL-12p-70, IL-17, and IFN-γ (Table 2).

Screening of Factors Influencing Whether the Viral Shedding Time > 14 Days and Drawing of Nomogram

In this study, 30 factors were chosen as independent variables, and if the viral shedding time was longer than 14 days it was the outcome variable. LASSO and logistic regression was used to screen the influential factors with the most significant differences. The training set data were derived from the above data set, the λ value was determined by cross validation following LASSO regression, and then lambda.min and lambda. 1se models were created based on the λ value, from which the optimal model was chosen based on the data from the testing set. The model corresponding to lambda. min contained 17 variables, whereas the model relating to lambda. 1se contained four variables. On the basis of the two models using the training set and testing set, dependent variables were predicted, and the ROC curve was created to determine the AUCs. The AUCs corresponding to lambda.min were 0.865 and 0.823, while those relating to lambda.1se were 0.829% and 0.779. On the basis of the model corresponding to lambda.min, logistic regression was performed on the training set data, and five variables evaluated, including the time from diagnosis to admission, CD4 cell count, Ct value of ORF1ab gene, IL-8, and constipation, were significantly different. Five variables were used to draw a nomogram, and the model was developed and evaluated; the discriminating index C-index was 0.856, and the
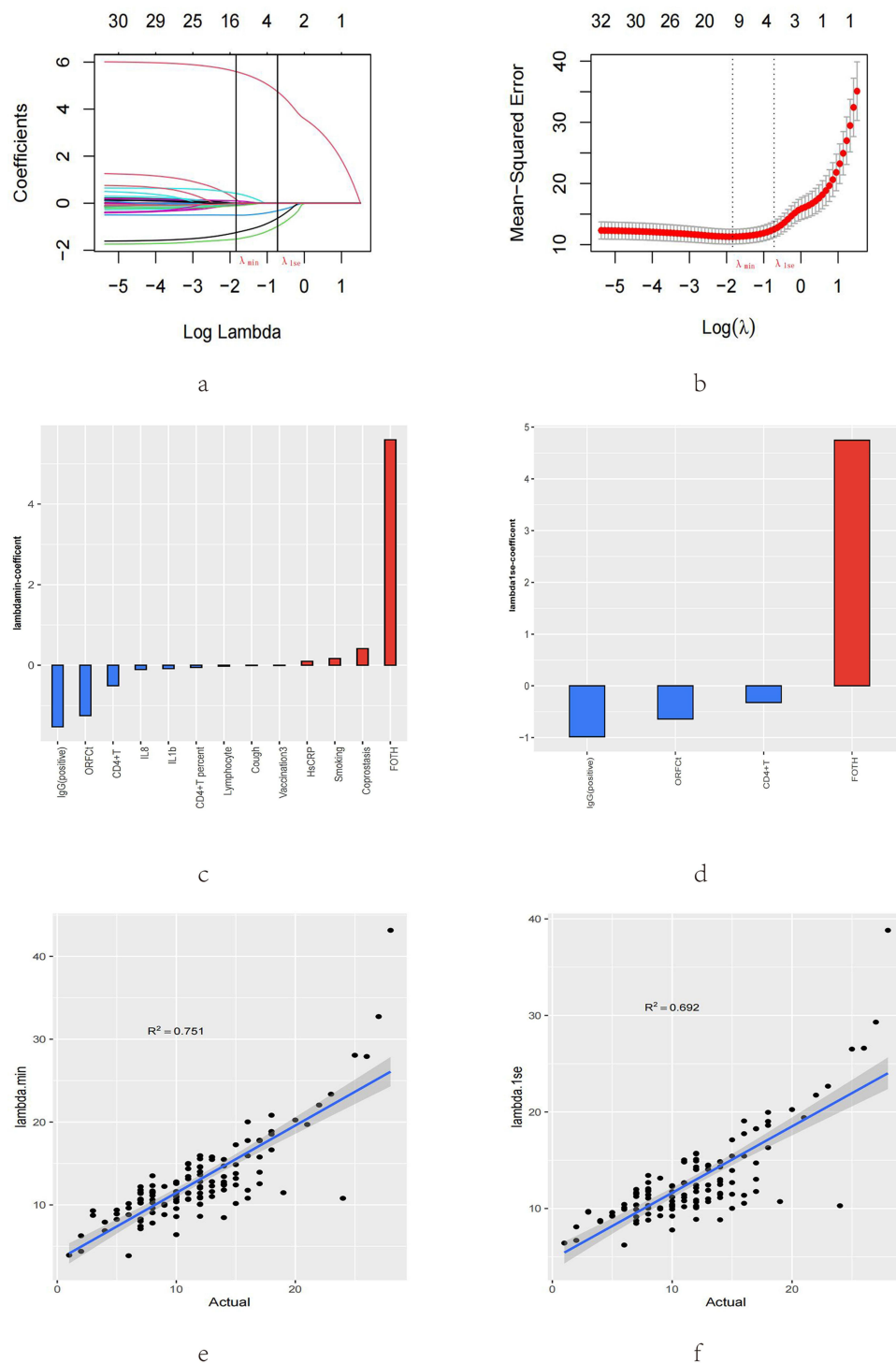
Figure 2 LASSO regression analysis based on viral shedding time.
Notes: (**a**) The map of coefficient distribution was drawn based on LASSO. (**b**) LASSO was utilized to create parametric maps. (**c**) The coefficients of the 13-parameter model that correspond to lambda.min. (**d**) Coefficients of the 13-parameter model corresponding to lambda.1se. (**e**) The model corresponding to lambda.min was assessed using data from the testing set. (**f**) The model corresponding to lambda.1se was assessed using the data from the testing set.
Abbreviation: LASSO, Least Absolute Shrinkage and Selection Operator.

calibration index indicated that the model was common (P<0.05). The model was tested using the testing set data, for which the discrimination index C-index was 0.888 and the calibration index demonstrated that the model fit well (P > 0.05). The AUC estimated using the data from the testing set was 0.847; the sensitivity of the model was 0.744 and its specificity was 0.830 (Table 3, Figures 3 and 4).

Table 2 The Logistic Regression Analysis for the Binary Outcome Indicates

Feature	Coefficients	se	OR (95% CI)	p
FOTH	0.384	0.036	1.458(1.372,1.584)	<0.01
CD4+T Number	−0.002	0.000	0.998(0.997,0.999)	<0.01
ORFCt	−0.058	0.017	0.944(0.912,0.975)	<0.01
Coprostasis	1.033	0.295	2.811(1.573,5.007)	<0.01
IL-8	−0.006	0.003	0.993(0.987,0.998)	0.036

Abbreviations: FOTH, From the onset of positive to the admission to hospital; CD4+T, CD4 cell count; ORFCt, ORF gene cycle threshold; IL, Interleukin.

Table 3 Demographic, Clinical Features, and Laboratory Results of Participant Distinguished by the 14 Days

Characteristics	Viral Shedding (Days≤14)	Viral Shedding (Days>14)	P value
Patients (n=631)	434(69.1%)	197(30.9%)	
Gender			
Female	230(52.9%)	80(40.6%)	<0.01
Male	204(46.1%)	117(59.4%)	
Age (years)			
18–45	83(19.1%)	45(22.8%)	0.427
46–69	187(43.1%)	78(39.6%)	
>69	164(37.8%)	74(37.6%)	
Smoking			
Yes	10(2.3%)	3(1.5%)	0.537
No	424(97.7%)	194(98.5%)	
Comorbidities			
Yes	255(58.8%)	108(54.8%)	0.367
No	179(41.2%)	89(45.2%)	
None of Comorbidities			
1	172(39.6%)	75(38.1%)	0.833
2	75(17.3%)	28(14.2%)	
≥3	8(1.8%)	5(2.5%)	
Vaccination (time)			
0	208(47.9%)	116(58.9%)	<0.05
1	24(5.5%)	12(6.1%)	
2	90(20.7%)	35(17.8%)	
3	112(25.8%)	34(17.3%)	
Fever			
Yes	139(32.0%)	65(33.0%)	0.860
No	295(68.0%)	132(67.0%)	
Cough			
Yes	190(43.8%)	89(45.2%)	0.758
No	244(56.2%)	108(54.8%)	
Pharyngalgia			
Yes	91(20.9%)	31(15.7%)	0.144
No	343(79.1%)	166(84.3%)	
Dyspnea			
Yes	11(2.5%)	9(4.6%)	0.166
No	423(97.5%)	188(95.4%)	
Constipation			
Yes	53(12.2%)	39(19.8%)	<0.05
No	381(87.8%)	158(80.2%)	

(Continued)

Table 3 (Continued).

Characteristics	Viral Shedding (Days≤14)	Viral Shedding (Days>14)	P value
Insomnia			
Yes	49(11.3%)	14(7.1%)	0.116
No	385(88.7%)	183(92.9%)	
IgG			
Negative	292(67.3%)	100(50.8%)	<0.01
Weakly positive	3(0.7%)	5(2.5%)	
Positive	139(32.0%)	92(46.7)	
IgM			
Negative	419(96.5%)	188(95.4%)	0.514
Weakly positive	6(1.4%)	2(1.0%)	
Positive	9(2.1%)	7(3.6%)	
NCt (Median, Quartile)	26.3(21.1,34.8)	36.2(22.7,45.5)	<0.01
ORF1abCt (Median, Quartile)	28.4(23.4,36.9)	37.7(25.1,50.0)	<0.01
From the onset of positive to the admission to hospital (Median, Quartile)	3(2,7)	14(5,19)	<0.01
From hospitalization to viral shedding (Median, Quartile)	5(2,8)	4(1,12)	0.984
Lymphocyte (95% confidence interval)(percent)	29.9(28.9,30.9)	27.0(25.4,28.6)	<0.05
Hb (Median, Quartile)	129(119,141)	127(115,142)	0.111
hsCRP (Median, Quartile)	2.49(0.25,6.47)	1.64(0.25,6.35)	0.117
D dimer	0.35(0.21,0.69)	0.38(0.21,0.82)	0.435
CD4+T Number	565(354,821)	578(409,804)	0.582
CD4+T Number (95% confidence interval)(percent)	41.1(40.3,41.9)	40.1(38.6,41.5)	0.186
TNF-α (Median, Quartile)	0.34(0.06,0.96)	0.41(0.13,1.11)	0.296
IL1-β (Median, Quartile)	3.04(1.14,6.96)	3.24(1.29,7.49)	0.608
IL-2 (Median, Quartile)	0.66(0.35,1.42)	0.7(0.41,1.27)	0.803
IL-6 (Median, Quartile)	1.77(0.91,3.54)	1.84(0.85,3.81)	0.872
IL-8 (Median, Quartile)	0.01(0.00,12.6)	0.00(0.00,6.29)	0.212
IL12p70 (Median, Quartile)	0.95(0.76,1.21)	0.95(0.76,1.20)	0.786
IL17 (Median, Quartile)	2.55(1.58,3.99)	2.55(1.73,3.83)	0.942
IFN-γ (Median, Quartile)	1.93(1.09,3.39)	1.59(0.92,3.06)	0.103

Abbreviations: NCt, N gene cycle threshold; ORF1abCt, ORF1ab gene cycle threshold; Hb, Hemoglobin; hsCRP, hypersensitive C-reactive protein; TNF-α, tumor necrosis factor-α; IL, Interleukin; IFN-γ, Interleukin.

Discussion

COVID-19 is a fulminating infectious disease; nevertheless, the majority of patients in China are mild and consume a large number of medical resources for diagnosis and treatment after admission.[15] In this study, LASSO regression models were used to predict the period of viral shedding in 631 mild cases based on their general circumstances and hematological examination results.

In this study, the analysis of clinical characteristics revealed that mild COVID-19 carriers were, on average, 62.1 years old; the majority had underlying diseases; more than 50% were not vaccinated; fever and cough were the most common symptoms; and the nucleic acid test became negative within 14 days in most COVID-19 carriers. The older the patients, the longer the hospital stays, which is correlated with age-related immune remodeling.[16] Underlying diseases may increase the expression of ACE2 receptors, thereby affecting viral infectivity; COPD, diabetes mellitus, malignancy, cardiovascular disease, and so on are common diseases that affect viral invasion.[17,18] Vaccines are effective at preventing infection, shortening hospital stays, and reducing mortality.[19] Asymptomatic carriers are quarantined first and treated in the hospital if their health changes during quarantine; therefore, tracking the time between when a patient is diagnosed and being hospitalized can help forecast the duration of viral shedding.[20–22] As the most common clinical symptoms of COVID-19 infection, fever, cough, and dyspnea are also significantly correlated with disease severity;[23,24] gender is
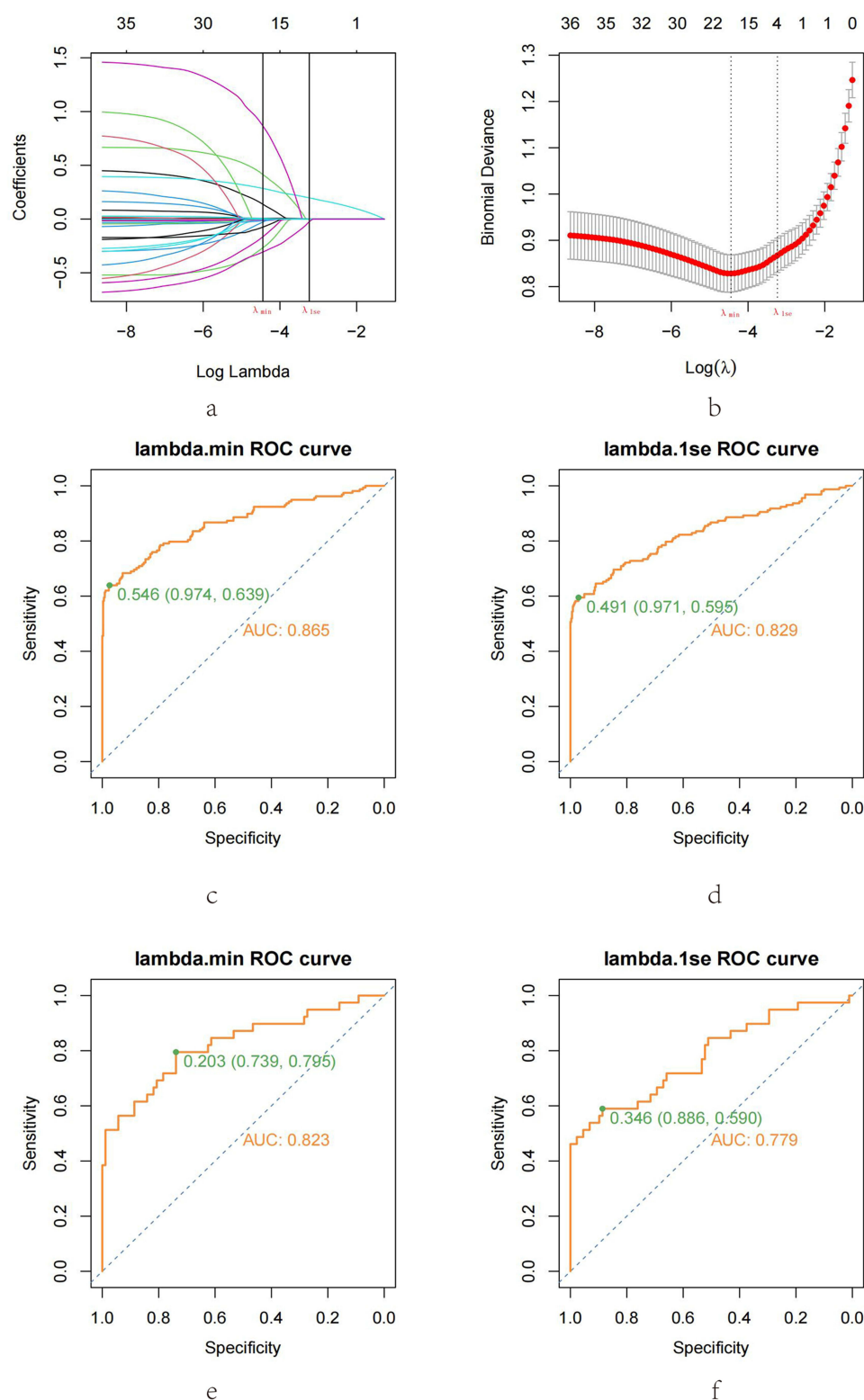
Figure 3 LASSO regression analysis was performed depending on whether the viral shedding duration was longer than 14 days, and models were evaluated using the ROC curve.

Notes: (**a**) The map of coefficient distribution was created using LASSO. (**b**) LASSO was utilized to create parametric maps. (**c**) Based on the model corresponding to lambda.min, the ROC curve of the training set was drawn. (**d**) The model matching to lambda.1se was utilized to generate the ROC curve of the training set. (**e**) The ROC curve of the testing set was generated using the model corresponding to lambda.min. (**f**) The ROC curve of the testing set was created using a model corresponding to lambda.1se.

Abbreviations: LASSO, Least Absolute Shrinkage and Selection Operator; ROC, receiver operator characteristic curve.

...

a

b

c

d
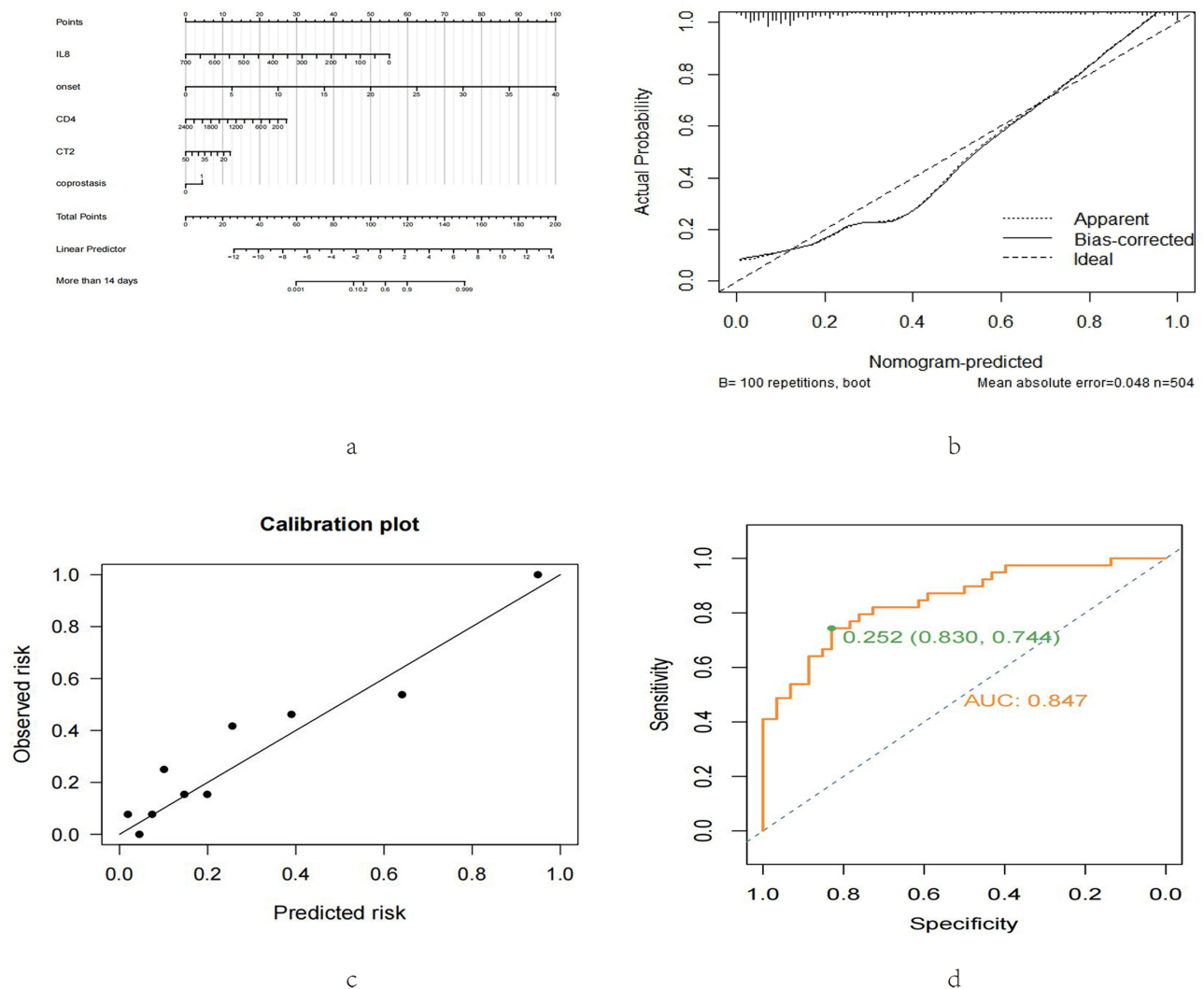
Figure 4 A nomogram was drawn using the LASSO-selected variables for logistic modeling.
Notes: (**a**) A nomogram was created using five factors that were significantly associated. (**b**) Based on the logistic model, the calibration curve for the training set was derived. (**c**) The calibration curve of the testing set was drawn using the logistic model. (**d**) The ROC curve of the testing set was drawn using the logistic model.

clearly associated with SRAS-COV2 susceptibility and prognosis, and females, for example, have a lower hospitalization rate and mortality.[25]

Several variables may be associated with viral clearance time. As proven by a number of studies,[26] delayed admission following a diagnosis may result in a prolonged viral clearance time. In trials where the median length of hospitalization was 3 days, the median duration of complete viral clearance was 13 days,[27] which is comparable to our finding. Due to the fact that this was a retrospective study in which the time of the first positive nucleic acid result was derived from the medical history and not confirmed by clear nucleic acid test reports, the data may be unreliable. Gastrointestinal symptoms may interfere with viral clearance.[28,29] Constipation, a common gastrointestinal symptom in patients with COVID-19, occurred in only 92 cases in this study, but it was less observed in patients with viral shedding within 14 days, indicating that constipation may influence viral clearance;[30] changes in the value of hypersensitive C-reactive protein are highly associated with the degree of lung injury, and are frequently used biomarkers to predict the severity of COVID-19 cases.[31,32] IgG antibodies produced from humoral immunity are detected in serum within 7 to 14 days of disease diagnosis and can last for several weeks.[33] As a result, these antibodies are frequently utilized for early disease detection.[34] Contrary to our

findings, Masiá et al found that IgG antibodies are positively correlated with viral exposure time. The reasons may be as follows: First, the prior outcome factors included mortality and severe instances, and because some patients were severely affected, the results varied.[35] Second, the participants in our study were mild cases, some of whom were asymptomatic for an extended period of time and were found to have passed through the acute phase prior to admission and thus had high IgG antibody titers;[36] the higher the antibody positive titers after admission, the shorter the time for viral clearance, and IgG antibodies produced by vaccines may also affect viral shedding.[37] The COVID-19 viral shedding time can be accurately predicted by monitoring the ORF1ab gene cycle threshold. Similar to our model, Aranha et al found that the ORF1ab gene cycle threshold was adversely correlated with viral clearance time.[38] It was difficult to explain why the ORF1ab gene cycle threshold of patients with viral shedding within 14 days of admission was much lower than that of patients with viral shedding beyond 14 days. Zhang et al discovered that the ORF1ab gene cycle threshold peaked on day 7 in patients with early viral shedding and on day 14 in patients with delayed viral shedding,[39] whereas we discovered that the median hospital stay was 3 days in patients with early viral shedding and 14 days in patients with delayed viral shedding. This may be the cause of our observations. The drop in lymphocyte percentage increases the likelihood and severity of COVID-19 infection and the role of CD4 T cells in SARS-COV2 infection, while a decrease in CD4 T cell count may prolong the time required for viral clearance.[40,41] The suppression of pulmonary immunity by smoking is a risk factor for the progression of COVID-19 infection, and damage to airway epithelial cells and cilia impacts viral clearance;[42] IL-8 and IL-1β as pro-inflammatory factors are associated with the severity of SARS-COV2 infection.[43,44] Nonetheless, we noticed that the elevation of IL-8 and IL-1β decreased viral shedding time. Huang et al similarly identified the significance of pro-inflammatory factors inhibition in extending viral shedding time, but no further explanation was provided.[45] This may be because the release of pro-inflammatory molecules stimulates macrophages, hence accelerating viral clearance.[46,47] In our study, we used data to screen for risk factors that affect viral shedding >14 days in order to predict the length of hospital stay, the time from diagnosis to admission, CD4 cell count, Ct value of ORF1ab gene, IL-8, and constipation were found to be the most significant risk factors. The above variables are explained in detail above. The results of the data show that the proportion of male patients in the group of >14 days is large. Although there is no relevant report on the effect of gender differences on the time of viral shedding, the susceptibility and prognosis of SRAS-COV2 virus have been very clear. Due to gender differences in the expression of the angiotensin-converting enzyme receptor ACE2 and the effect of sex hormones on immune regulation, female patients are associated with lower hospitalization and mortality rates.[25]

The outbreak of COVID-19 in Shanghai in 2022 was mainly caused by Omicron variants BA.2 and BA.2.2. Due to the viral variants and the widespread vaccination in Shanghai, the mortality and severity rate were low. A significant negative correlation between vaccination and shedding time was also found in our study. However, as Omicron continues to mutate, with increasing infectivity but decreasing pathogenicity, the application of the model needs to be updated to adapt to different variants by collecting more clinical data.[3,48] Studies have found that from BA.1 to XBB, the immune evasion ability of the virus is increasing, which also means that the vaccines that have been inoculated are not sufficient to prevent virus infection. Studies have also found that with the continuous mutation of the virus, from Delta, BA.2.12 to XBB.1, the pathogenic ability is gradually decreased, and many symptoms may become mild,[49,50] which also promote the continuous update of the model by collecting more data.

Most COVID-19 carriers exhibit mild symptoms; therefore, it is crucial to dedicate adequate medical resources for the possible expansion of the epidemic. Building a mathematical model utilizing general data and test results to forecast viral shedding time is of considerable benefit. However, in this retrospective study conducted at a single center, the assessment of the model may be affected by the varying testing procedures of other institutions and only cases in April and May 2022 were collected. Our study focused only on the Omicron variant in Shanghai, it is not appropriate to predict other variants and more data is needed if it is applied in other regions and other variants. Despite these limitations, we were able to create a highly accurate prediction model for viral shedding time, which we want to use to manage medical resources.

Conclusion

After acquiring observation indexes, we were able to determine the viral shedding time in mild COVID-19 cases using a model that accurately predicted the viral shedding time in patients with mild COVID-19.

Data Sharing Statement

All data generated or analysed during this study are included in this article. Further enquiries can be directed to the corresponding author.

Acknowledgments

We would like to acknowledge the hard and dedicated work of all the staff that implemented the intervention and evaluation components of the study.

Funding

Disclosure

The authors report no conflicts of interest in this work.

References

1. Diaz Badial P, Bothorel H, Kherad O, Dussoix P, Tallonneau Bory F, Ramlawi M. A new screening tool for SARS-CoV-2 infection based on self-reported patient clinical characteristics: the COV19-ID score. *BMC Infect Dis*. 2022;22(1):187. doi:10.1186/s12879-022-07164-1
2. Safiabadi Tali SH, LeBlanc JJ, Sadiq Z, et al. Tools and techniques for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)/COVID-19 detection. *Clin Microbiol Rev*. 2021;34(3):e00228–20. doi:10.1128/CMR.00228-20
3. Fan Y, Li X, Zhang L, Wan S, Zhang L, Zhou F. SARS-CoV-2 Omicron variant: recent progress and future perspectives. *Signal Transduct Target Ther*. 2022;7(1):141. doi:10.1038/s41392-022-00997-x
4. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369:m1328. doi:10.1136/bmj.m1328
5. Shankar S, Mohakuda SS, Kumar A, et al. Systematic review of predictive mathematical models of COVID-19 epidemic. *Med J Armed Forces India*. 2021;77(Suppl 2):S385–S392. doi:10.1016/j.mjafi.2021.05.005
6. Brodin P. Immune determinants of COVID-19 disease presentation and severity. *Nat Med*. 2021;27(1):28–33. doi:10.1038/s41591-020-01202-8
7. Huang J, Xu Y, Wang B, et al. Risk stratification scores for hospitalization duration and disease progression in moderate and severe patients with COVID-19. *BMC Pulm Med*. 2021;21(1):120. doi:10.1186/s12890-021-01487-6
8. Al Dhamen MA, Alhashim AF, Alqattan HH, Pottoo FH. COVID-19: an update on pathogenesis and treatment. *Curr Pharm Des*. 2021;27(32):3454–3461. doi:10.2174/1381612826666201222161312
9. Rees EM, Nightingale ES, Jafari Y, et al. COVID-19 length of hospital stay: a systematic review and data synthesis. *BMC Med*. 2020;18(1):270. doi:10.1186/s12916-020-01726-3
10. Chen B, Geng P, Shen J, et al. Traditional Chinese medicine Jingyingubiao formula therapy improves the negative conversion rate of SARS-CoV2 in patients with mild COVID-19. *Int J Biol Sci*. 2022;18(15):5641–5652. doi:10.7150/ijbs.76699
11. Wang J, Bao A, Gu J, et al. A retrospective analysis of 12,400 SARS-CoV-2 RNA tests in patients with COVID-19 in Wuhan. *Medicine*. 2021;100(20):e25916. doi:10.1097/MD.0000000000025916
12. Ma C, Sun W, Tang T, et al. Effectiveness of adenovirus type 5 vectored and inactivated COVID-19 vaccines against symptomatic COVID-19, COVID-19 pneumonia, and severe COVID-19 caused by the B.1.617.2 (Delta) variant: evidence from an outbreak in Yunnan, China, 2021. *Vaccine*. 2022;40(20):2869–2874. doi:10.1016/j.vaccine.2022.03.067
13. Han J, Shi LX, Xie Y, et al. Analysis of factors affecting the prognosis of COVID-19 patients and viral shedding duration. *Epidemiol Infect*. 2020;148:e125. doi:10.1017/S0950268820001399
14. Li K, Zhang C, Qin L, et al. A nomogram prediction of length of hospital stay in patients with covid-19 pneumonia: a retrospective cohort study. *Dis Markers*. 2021;2021:5598824. doi:10.1155/2021/5598824
15. Liang W, Liang H, Ou L, et al., China Medical Treatment Expert Group for COVID-19. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med*. 2020;180(8):1081–1089. doi:10.1001/jamainternmed.2020.2033
16. Chen Y, Klein SL, Garibaldi BT, et al. Aging in COVID-19: vulnerability, immunity and intervention. *Ageing Res Rev*. 2021;65:101205. doi:10.1016/j.arr.2020.101205
17. Ejaz H, Alsrhani A, Zafar A, et al. COVID-19 and comorbidities: deleterious impact on infected patients. *J Infect Public Health*. 2020;13(12):1833–1839. doi:10.1016/j.jiph.2020.07.014
18. Fang X, Li S, Yu H, et al. Epidemiological, comorbidity factors with severity and prognosis of COVID-19: a systematic review and meta-analysis. *Aging*. 2020;12(13):12493–12503. doi:10.18632/aging.103579

19. Tregoning JS, Flight KE, Higham SL, Wang Z, Pierce BF. Progress of the COVID-19 vaccine effort: viruses, vaccines and variants versus efficacy, effectiveness and escape. *Nat Rev Immunol*. 2021;21(10):626–636. doi:10.1038/s41577-021-00592-1

20. Chiam T, Subedi K, Chen D, et al. Hospital length of stay among COVID-19-positive patients. *J Clin Transl Res*. 2021;7(3):377–385.

21. Liu H, Shang X, Chen S, Li T, Zhang J. Cautions on the laboratory indicators of COVID-19 patients on and during admission. *J Clin Lab Anal*. 2021;35(5):e23767. doi:10.1002/jcla.23767

22. Salian VS, Wright JA, Vedell PT, et al. COVID-19 transmission, current treatment, and future therapeutic strategies. *Mol Pharm*. 2021;18 (3):754–771. doi:10.1021/acs.molpharmaceut.0c00608

23. Zheng B, Cai Y, Zeng F, et al. An interpretable model-based prediction of severity and crucial factors in patients with COVID-19. *Biomed Res Int*. 2021;2021:8840835. doi:10.1155/2021/8840835

24. Ma X, Li A, Jiao M, et al. Characteristic of 523 COVID-19 in Henan province and a death prediction model. *Front Public Health*. 2020;8:475. doi:10.3389/fpubh.2020.00475

25. Raza HA, Sen P, Bhatti OA, Gupta L. Sex hormones, autoimmunity and gender disparity in COVID-19. *Rheumatol Int*. 2021;41(8):1375–1386. doi:10.1007/s00296-021-04873-9

26. Shu HM, He S, Sun Y, et al. Factors influencing viral clearance in mild COVID-19 and clinical characteristics of asymptomatic patients. *Biomed Res Int*. 2021;2021:5909612. doi:10.1155/2021/5909612

27. Samrah SM, Al-Mistarehi AH, Kewan T, et al. Viral clearance course of COVID-19 outbreaks. *J Multidiscip Healthc*. 2021;14:555–565. doi:10.2147/JMDH.S302891

28. Han C, Duan C, Zhang S, et al. Digestive symptoms in COVID-19 patients with mild disease severity: clinical presentation, stool viral RNA testing, and outcomes. *Am J Gastroenterol*. 2020;115(6):916–923. doi:10.14309/ajg.0000000000000664

29. Chen H, Tong Z, Ma Z, et al. Gastrointestinal bleeding, but not other gastrointestinal symptoms, is associated with worse outcomes in COVID-19 patients. *Front Med*. 2021;8:759152. doi:10.3389/fmed.2021.759152

30. Allali I, Bakri Y, Amzazi S, Ghazal H, Marquart ME. Gut-lung axis in COVID-19. *Interdiscip Perspect Infect Dis*. 2021;2021:6655380. doi:10.1155/2021/6655380

31. Jaeken J, Wadman SK, Duran M, et al. Adenylosuccinase deficiency: an inborn error of purine nucleotide synthesis. *Eur J Pediatr*. 1988;148 (2):126–131. doi:10.1007/BF00445919

32. Tjendra Y, Al Mana AF, Espejo AP, et al. Predicting disease severity and outcome in COVID-19 patients: a review of multiple biomarkers. *Arch Pathol Lab Med*. 2020;144(12):1465–1474. doi:10.5858/arpa.2020-0471-SA

33. Anka AU, Tahir MI, Abubakar SD, et al. Coronavirus disease 2019 (COVID-19): an overview of the immunopathology, serological diagnosis and management. *Scand J Immunol*. 2021;93(4):e12998. doi:10.1111/sji.12998

34. Xie J, Ding C, Li J, et al. Characteristics of patients with coronavirus disease (COVID-19) confirmed using an IgM-IgG antibody test. *J Med Virol*. 2020;92(10):2004–2010. doi:10.1002/jmv.25930

35. Masiá M, Telenti G, Fernández M, et al. SARS-CoV-2 seroconversion and viral clearance in patients hospitalized with COVID-19: viral load predicts antibody response. *Open Forum Infect Dis*. 2021;8(2):ofab005. doi:10.1093/ofid/ofab005

36. Xiao T, Wang Y, Yuan J, et al. Early viral clearance and antibody kinetics of COVID-19 among asymptomatic carriers. *Front Med*. 2021;8:595773. doi:10.3389/fmed.2021.595773

37. Wisnewski AV, Campillo Luna J, Redlich CA, Ansari AA. Human IgG and IgA responses to COVID-19 mRNA vaccines. *PLoS One*. 2021;16(6): e0249499. doi:10.1371/journal.pone.0249499

38. Aranha C, Patel V, Bhor V, Gogoi D. Cycle threshold values in RT-PCR to determine dynamics of SARS-CoV-2 viral load: an approach to reduce the isolation period for COVID-19 patients. *J Med Virol*. 2021;93(12):6794–6797. doi:10.1002/jmv.27206

39. Zhang W, Zhou S, Wang G, et al. Clinical predictors and RT-PCR profile of prolonged viral shedding in patients with SARS-CoV-2 Omicron variant in Shanghai: a retrospective observational study. *Front Public Health*. 2022;10:1015811. doi:10.3389/fpubh.2022.1015811

40. Sette A, Crotty S. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell*. 2021;184(4):861–880. doi:10.1016/j.cell.2021.01.007

41. Jordan SC. Innate and adaptive immune responses to SARS-CoV-2 in humans: relevance to acquired immunity and vaccine responses. *Clin Exp Immunol*. 2021;204(3):310–320. doi:10.1111/cei.13582

42. Patanavanich R, Glantz SA. Smoking is associated with COVID-19 progression: a meta-analysis. *Nicotine Tob Res*. 2020;22(9):1653–1656. doi:10.1093/ntr/ntaa082

43. Ghazavi A, Ganji A, Keshavarzian N, Rabiemajd S, Mosayebi G. Cytokine profile and disease severity in patients with COVID-19. *Cytokine*. 2021;137:155323. doi:10.1016/j.cyto.2020.155323

44. Gao YD, Ding M, Dong X, et al. Risk factors for severe and critically ill COVID-19 patients: a review. *Allergy*. 2021;76(2):428–455. doi:10.1111/all.14657

45. Huang L, Chen Z, Ni L, et al. Impact of angiotensin-converting enzyme inhibitors and angiotensin receptor blockers on the inflammatory response and viral clearance in COVID-19 patients. *Front Cardiovasc Med*. 2021;8:710946. doi:10.3389/fcvm.2021.710946

46. Panigrahy D, Gilligan MM, Huang S, et al. Inflammation resolution: a dual-pronged approach to averting cytokine storms in COVID-19? *Cancer Metastasis Rev*. 2020;39(2):337–340. doi:10.1007/s10555-020-09889-4

47. Regidor PA, De La Rosa X, Santos FG, Rizo JM, Gracia Banzo R, Silva RS. Acute severe SARS COVID-19 patients produce pro-resolving lipids mediators and eicosanoids. *Eur Rev Med Pharmacol Sci*. 2021;25(21):6782–6796. doi:10.26355/eurrev_202111_27123

48. Chen X, Yan X, Sun K, et al. Estimation of disease burden and clinical severity of COVID-19 caused by Omicron BA.2 in Shanghai, February-June 2022. *Emerg Microbes Infect*. 2022;11(1):2800–2807. doi:10.1080/22221751.2022.2128435

49. Shrestha LB, Foster C, Rawlinson W, Tedla N, Bull RA. Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: implications for immune escape and transmission. *Rev Med Virol*. 2022;32(5):e2381. doi:10.1002/rmv.2381

50. Mohandas S, Shete A, Kumar A, et al. Comparative pathogenicity of BA.2.12, BA.5.2 and XBB.1 with the delta variant in Syrian hamsters. *Front Microbiol*. 2023;14:1183763. doi:10.3389/fmicb.2023.1183763