Clin. Lab. 2024;70:XXX-XXX


## ORIGINAL ARTICLE

# A Prediction Model for Prolonged Hospital Length of Stay (ProLOS) in Coronavirus Disease 2019 (COVID-19) Patients

Ling Zhang [1], Guo C. Mi [2], Xue X. Yao [1], Si Y. Sun [1], Ai S. Fu [1], Yan L. Ge [1]

[1] *Department of Respiratory Medicine, North China University of Science and Technology Affiliated Hospital, Tangshan, Hebei, China*
[2] *School of Clinical Medicine, North China University of Science and Technology, Tangshan, Hebei, China*

### SUMMARY

***Background:*** **Coronavirus disease 2019 (COVID-19) is an acute respiratory infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). With the normalization of COVID-19 globally, it is crucial to construct a prediction model that enables clinicians to identify patients at risk for ProLOS based on demographics and serum inflammatory biomarkers.**
***Methods:*** **The study included hospitalized patients with a confirmed diagnosis of COVID-19. These patients were randomly grouped into a training (80%) and a test (20%) cohort. The LASSO regression and ten-fold cross-validation method were applied to filter variables. The training cohort utilized multifactorial logistic regression analyses to identify the independent factors of ProLOS in COVID-19 patients. A 4-variable nomogram was created for clinical use. ROC curves were plotted, and the area under the curve (AUC) was calculated to evaluate the model's discrimination; calibration analysis was planned to assess the validity of the nomogram, and decision curve analysis (DCA) was used to evaluate the clinical usefulness of the model.**
***Results:*** **The results showed that among 310 patients with COVID-19, 80 had extended hospitalization (80/310). Four independent risk factors for COVID-19 patients were identified: age, coexisting chronic respiratory diseases, white blood cell count (WBC), and serum albumin (ALB). A nomogram based on these variables was created. The AUC in the training cohort was 0.808 (95% CI: 0.75 - 0.8671), and the AUC in the test cohort was 0.815 (95% CI: 0.7031 - 0.9282). The model demonstrates good calibration and can be used with threshold probabilities ranging from 0% to 100% to obtain clinical net benefits.**
***Conclusions:*** **A predictive model has been created to accurately predict whether the hospitalization duration of COVID-19 patients will be prolonged. This model incorporates serum WBC, ALB levels, age, and the presence of chronic respiratory system diseases.**
**(Clin. Lab. 2024;70:xx-xx. DOI: 10.7754/Clin.Lab.2023.231203)**

**Correspondence:**
Yan L. Ge
North China University of Science and Technology Affiliated Hospital
Tangshan
Hebei
China
Phone: + 86 15932081296
Email: 495732196@qq.com

Manuscript accepted December 13, 2023

**KEYWORDS**

COVID-19, prediction model, Prolonged Hospital Length of Stay

## INTRODUCTION

Over the past year, COVID-19, caused by SARS-CoV-2, has become a new normal. According to the Chinese Center for Disease Control and Prevention's report in November 2023 on the nationwide COVID-19 infection, from Oct 1 to Oct 31, there were 209 new severe cases and 24 deaths, all attributed to the Omicron vari-

ant. These cases covered 73 evolutionary branches, with the predominant circulating strains being the XBB lineage variants. The top three variants in terms of proportion were XBB.1.9 and its sub-branches, XBB.1.22 and its sub-branches, and XBB.1.16 and its sub-branches. Although when assessing the prognosis and severity of COVID-19, some scoring systems previously used for CAP, such as the CURB-65 and Pneumonia Severity Index (PSI) [1-3], or newly developed scoring systems based on mechanical ventilation or mortality outcomes [4], already exist. However, each scoring system has some limitations in predicting the need for prolonged hospitalization in patients with COVID-19. Since the Omicron variant became prevalent, the mortality rate has been significantly reduced. At the same time, due to the highly contagious nature of the Omicron variant, it leads to a concentrated outbreak, resulting in a significant increase in the demand for hospital beds. Since the normalization of COVID-19, a study has been conducted to predict the length of hospital stay for COVID-19 patients [5]. They finally selected five variables: immunotherapy, familial cluster, heparin, runny nose, and APTT. The main focus of their study is to identify the factors influencing the length of hospital stay for COVID-19 patients. Our primary objective in developing this model is to effectively predict the probability of prolonged hospitalization for COVID-19 patients based on simple variables. This serves two purposes: facilitating communication between clinicians and patients and identifying preventable and solvable factors contributing to prolonged hospital stays. By doing so, we aim to reduce the risk of worsening patient conditions and alleviate the strain on healthcare resources. To account for significant regional differences in economic levels, we have chosen binary variables to discuss the risk factors associated with prolonged hospitalization and establish a predictive model.

## MATERIALS AND METHODS

### Patients

We conducted a retrospective study. The study was conducted at a provincial teaching hospital in China with more than 1,300 beds. We enrolled 310 confirmed cases of COVID-19 using positive RT-PCR tests for SARS-CoV-2 at North China University of Science and Technology Affiliated Hospital from January to September 2023. Those with a history of malignant hematological diseases, uremia, advanced lung cancer, refractory heart failure, rheumatic diseases, radiotherapy, immuno-suppressive agents, transplantation, medical history, and incomplete history of auxiliary examination were excluded. The following information was collected from the patient: age, gender, smoking, alcohol use, coexisting hypertension, coexisting diabetes, coexisting coronary heart disease, coexisting chronic respiratory diseases (active pulmonary tuberculosis, COPD, bronchial asthma, chronic bronchitis, bronchiectasis, interstitial lung disease). Laboratory test results upon admission: C-reactive protein (CRP), white blood cell count (WBC), neutrophil (NEU), lymphocyte (LYM), platelet (PLT), hemoglobin (HGB), serum albumin (ALB), serum creatinine (Scr), blood urea nitrogen (BUN), myoglobin (MYO), creatine kinase (CK), creatine kinase isoenzymes (CK-MB), lactate dehydrogenase (LDH), alanine aminotransferase (ALT), aspartate transaminase (AST), total bilirubin (TBIL), thrombin time (TT), activated partial thromboplastin time (APTT), international normalized ratio (INR), fibrinogen (FIB), D-dimer. The length of hospital stay for enrolled patients, defined as the $\geq$ 75th percentile (14 days), was considered prolonged hospitalization. According to this definition, a length of stay of $\geq$ 14 days was defined as the ProLOS group (80 patients), and a length of stay of $<$ 14 days was determined as the non-ProLOS group (230 patients). These patients were randomly grouped into a training (80%) and a validation (20%) cohort. ProLOS was used as a predictor of outcome in this study.

### Measurement of serum inflammatory marker biomarkers

According to our research design, blood routine was measured by Abbott's five-category automatic blood cell analyzer as part of routine testing. Coagulation indicators are determined with France's STAGO automatic coagulation analyzer. Serum CRP level was measured by Prunus Medical's PA8800V unique protein analyzer. Bayer's 1650 automatic biochemistry analyzer measured blood biochemistry indicators.

## RESULTS

### Statistical analyses

Continuous variables are expressed as median and interquartile range (IQR), while categorical variables are presented as frequencies and percentages (%). We used the chi-squared test for categorical variables to compare the difference between groups and Wilcoxon rank-sum tests or *t*-tests for continuous variables. The LASSO regression and ten-fold cross-validation method were applied to filter variables. The training cohort utilized multifactorial logistic regression analyses to identify the independent factors of prolonged hospitalization in COVID-19 patients. The COVID-19 patient ProLOS prediction model was constructed, and the nomogram was plotted. ROC curves were plotted, and the area under the curve (AUC) was calculated to evaluate the model's discrimination; calibration analysis was planned to assess the validity of the nomogram, and decision curve analysis (DCA) to evaluate the clinical usefulness of the model. Statistical analysis was conducted using R software (version 4.3.2; R Foundation for Statistical Computing), all tests were two-tailed, and a p-value $< 0.05$ was considered statistically significant.

**Table 1. Baseline characteristics of patients with COVID-19.**

| Variables | Total (n = 310) | Non-ProLOS (n = 230) | ProLOS (n = 80) | p-value |
|---|---|---|---|---|
| Gender, n (%) | | | | 0.314 |
| Male | 173 (56) | 106 (46) | 31 (39) | |
| Female | 137 (44) | 124 (54) | 49 (61) | |
| Age (years) * | 68 (59, 76) | 66 (57, 73.75) | 74 (67, 82) | < 0.001 |
| Smoking, n (%) | | | | 0.642 |
| Yes | 66 (21) | 47 (20) | 19 (24) | |
| No | 244 (79) | 183 (80) | 61 (76) | |
| Alcohol use, n (%) | | | | 0.468 |
| Yes | 36 (12) | 29 (13) | 7 (9) | |
| No | 274 (88) | 201 (87) | 73 (91) | |
| Coexisting hypertension, n (%) | | | | 0.87 |
| Yes | 139 (45) | 102 (44) | 37 (46) | |
| No | 171 (55) | 128 (56) | 43 (54) | |
| Coexisting diabetes, n (%) | | | | 0.555 |
| Yes | 72 (23) | 51 (22) | 21 (26) | |
| No | 238 (77) | 179 (78) | 59 (74) | |
| Coexisting coronary heart disease, n (%) | | | | 0.128 |
| Yes | 79 (25) | 53 (23) | 26 (32) | |
| No | 231 (75) | 177 (77) | 54 (68) | |
| Chronic respiratory diseases, n (%) | | | | < 0.001 |
| Yes | 40 (13) | 15 (7) | 25 (31) | |
| No | 270 (87) | 215 (93) | 55 (69) | |
| CRP (mg/L) * | 31.15 (10.53, 60.48) | 25.05 (9.8, 53.95) | 40.55 (15.85, 87.68) | 0.001 |
| WBC (x $10^9$) * | 5.8 (4.3, 8.1) | 5.4 (4.12, 7.5) | 7.3 (5.6, 9.3) | < 0.001 |
| NEU (x $10^9$) * | 4.07 (2.69, 6.01) | 3.76 (2.5, 5.36) | 5.29 (3.38, 7.3) | < 0.001 |
| LYM (x $10^9$) * | 1 (0.65, 1.49) | 1.02 (0.66, 1.51) | 0.88 (0.55, 1.37) | 0.337 |
| PLT (x $10^9$) * | 188.5 (134, 245.75) | 184 (130.25, 241.75) | 195.5 (139.75, 258) | 0.489 |
| HGB (g/L) # | 127.18 ± 18.93 | 127.72 ± 17.98 | 125.65 ± 21.47 | 0.442 |
| ALB (g/L) # | 38.49 ± 4.52 | 39.31 ± 4.21 | 36.14 ± 4.6 | < 0.001 |
| Scr (μmol/L) * | 68 (56, 87) | 69 (58, 87.75) | 64.5 (54, 81.25) | 0.09 |
| BUN (mmol/L) * | 5.53 (4.27, 6.86) | 5.47 (4.18, 6.78) | 5.96 (4.72, 7.5) | 0.02 |
| MYO (μg/L) * | 22 (13, 39.75) | 20 (13, 40.5) | 24 (16, 39.25) | 0.246 |
| CK (U/L) * | 61 (43, 96.5) | 65 (47, 101) | 51 (32, 78) | 0.001 |
| CK-MB (U/L) * | 11 (9, 15) | 11 (9, 14) | 12 (9, 17) | 0.179 |
| LDH (U/L) * | 247.5 (202, 313.5) | 244 (198.5, 292.5) | 268.5 (216, 366.25) | 0.012 |
| ALT (U/L) * | 23 (15, 35) | 22 (15, 34.75) | 24 (14, 36.25) | 0.92 |
| AST (U/L) * | 25 (20, 33.75) | 25 (21, 32) | 25.5 (18, 35.5) | 0.488 |
| TBIL (μmol/L) * | 11.3 (8.9, 15.3) | 11.45 (9.2, 15.7) | 11 (8.5, 14) | 0.272 |
| TT (s) * | 17.5 (16.7, 18.4) | 17.4 (16.7, 18.2) | 17.75 (16.9, 18.83) | 0.049 |
| APTT (s) * | 37.4 (34.12, 40.88) | 37.5 (34.6, 40.38) | 37.05 (33.6, 41.85) | 0.959 |
| INR * | 0.99 (0.94, 1.06) | 0.99 (0.93, 1.05) | 1 (0.94, 1.08) | 0.142 |
| FIB (g/L) * | 4.94 (4.12, 5.83) | 4.84 (4.1, 5.74) | 5.2 (4.27, 5.96) | 0.106 |
| D-dimer (ng/mL) * | 511 (316, 1,127.75) | 469 (271.25, 949.75) | 655.5 (397.75, 1,572) | 0.001 |

*** - Data presented as median (interquartile range).**
**# - Data presented as mean ± SD.**

**Table 2. Results of multivariate logistic regression analysis.**

| Item | p | OR | 95% CI | |
|---|---|---|---|---|
| | | | Lower limit | Ceiling |
| WBC (x $10^9$) | 0.0023 | 1.146 | 1.0498 | 1.2503 |
| ALB (g/L) | 0.0315 | 0.911 | 0.8375 | 0.9917 |
| Age (years) | 0.0090 | 1.041 | 1.0102 | 1.0739 |
| Chronic respiratory diseases | < 0.0001 | 7.146 | 2.9426 | 17.3552 |

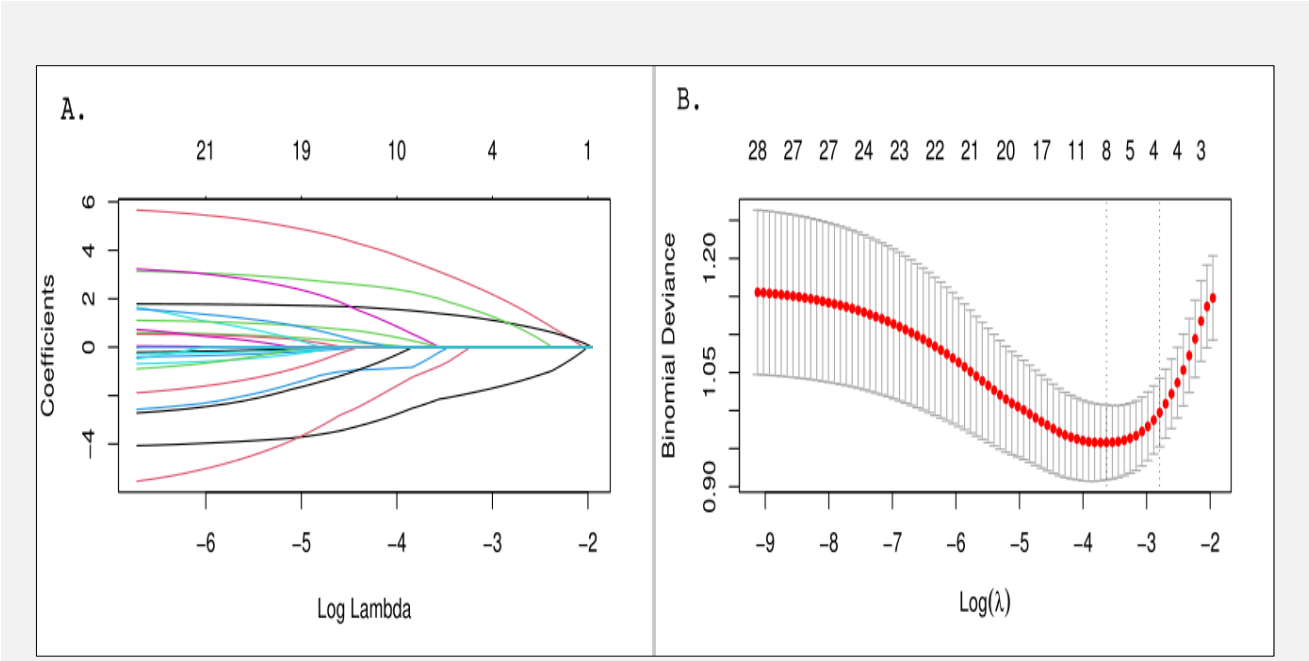


**Figure 1. Predictor selection using the LASSO regression analysis with ten-fold cross-validation.**

**A - Variable coefficient penalty plot. B - Tuning parameter (lambda) selection of deviance in the LASSO regression based on the minimum criteria (left dotted line) and the 1-SE criteria (right dotted line) (the present study selected).**

## Patient characteristics

A total of 310 eligible patients were included in this study, including 137 males and 173 females, with a median age of 68 (range 59 - 76 years). Among them, 80 patients (80/310) had prolonged hospitalization for further treatment. The results showed differences in age, comorbid chronic respiratory diseases, CRP, WBC, NEU, ALB, BUN, CK, LDH, TT, and D-dimer between the two groups were significant ($p < 0.05$). In contrast, LYM and coexisting hypertension were not ($p > 0.05$). Baseline characteristics of patients with COVID-19 are shown in Table l.

## LASSO regression and ten-fold cross-validation

Four of the 29 variables collected from patients were selected based on non-zero coefficients calculated by LASSO regression analysis (Figure 1). These variables included age, WBC, ALB, and coexisting chronic respiratory diseases.

The training cohort utilized binary logistic regression analyses to explore the independent factors of prolonged hospitalization in COVID-19 patients.

In the training set, the binary logistic regression analysis results showed that all four variables were independent risk factors for prolonged hospitalization in COVID-19 patients, as shown in Table 2.

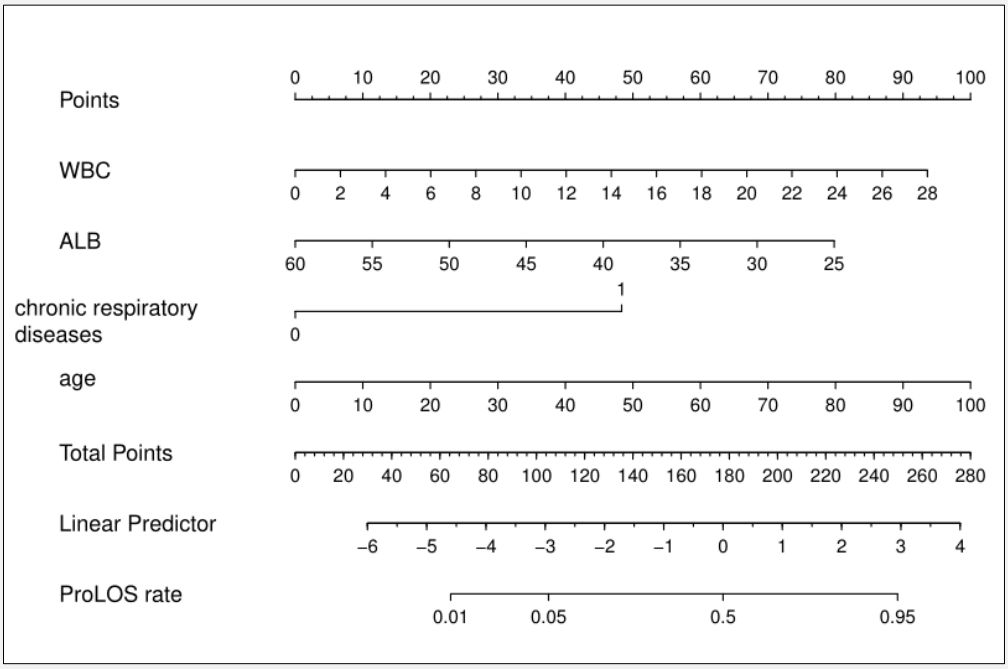


**Figure 2. Construction of nomogram model for predicting ProLOS of COVID-19 patients. Each variable for COVID-19 patients corresponds to a point, and then all points are added together to correspond to the predicted probability of ProLOS based on the total points.**

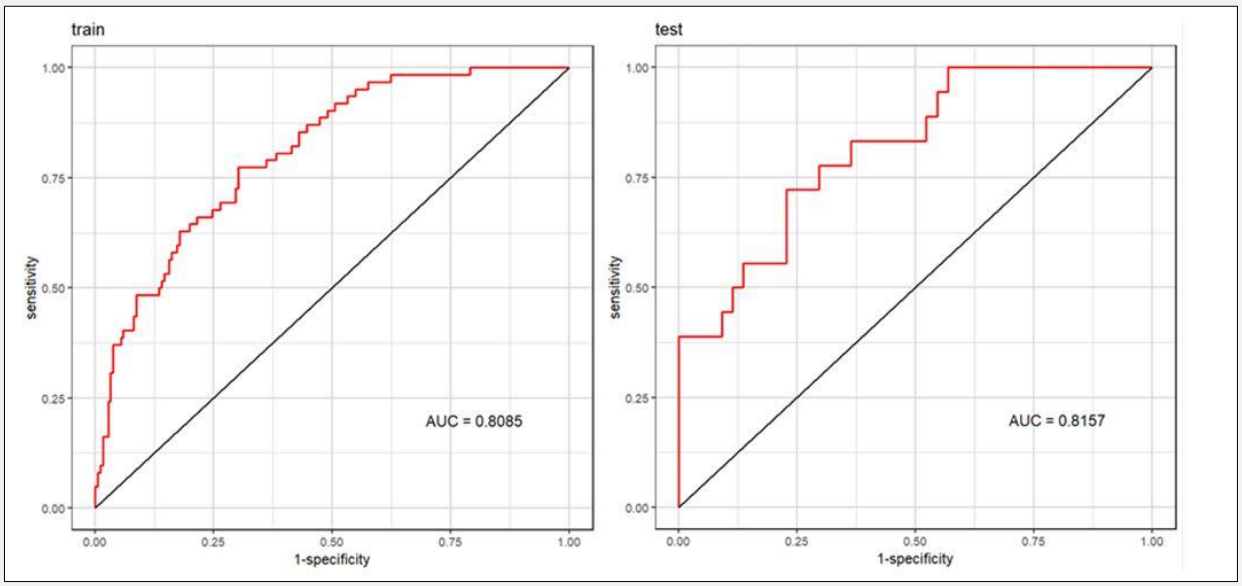


**Figure 3. ROC curve analysis of the prediction model.**

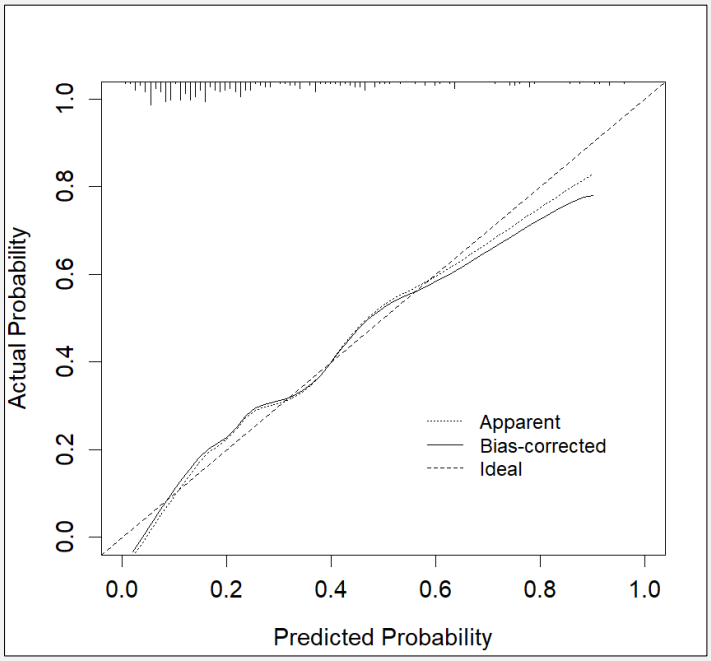


**Figure 4. Validation plot of the calibration curve for the model.**

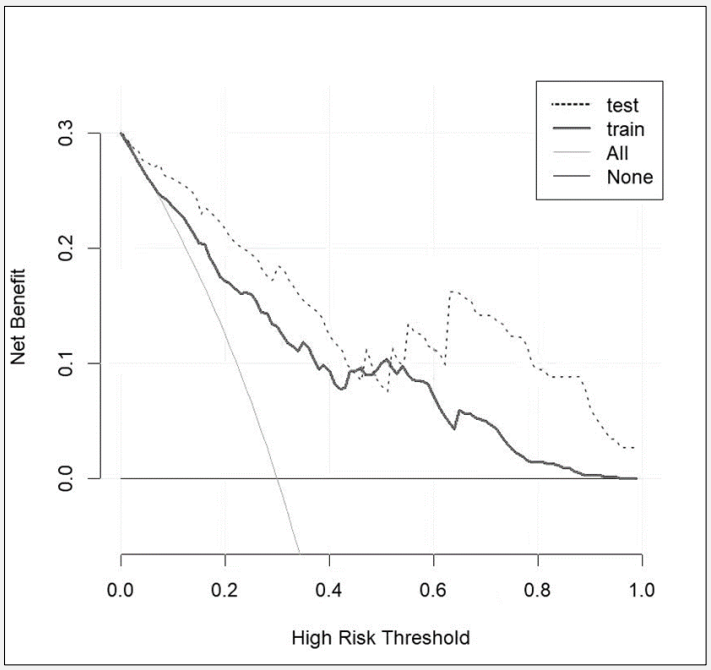


**Figure 5. DCA of the nomogram. The dotted line represents the test set. The solid line represents the training set. Clinical net benefit is achieved when the threshold probability of ProLOS is between 0% and 100%. DCA, decision curve analysis; ProLOS, prolonged length of stay in hospital.**

**Constructing predictive models and plotting the nomogram**

The prediction model was built based on these four variables, and a nomogram was plotted (Figure 2).

**The new model for evaluation and validation**

The ROC curve was plotted (Figure 3), and the calculated area under the curve (AUC) was 0.805 (CI: 0.75 - 0.8671) with an internal validation AUC of 0.815 (95% CI: 0.7031 - 0.9282). The model was well calibrated (Figure 4), and a net clinical benefit could be obtained at threshold probabilities in the range of 0% - 100% using the submodel (Figure 5).

## DISCUSSION

In the COVID-19 outbreak caused by SARS-CoV-2, multiple risk factors have been identified as having a potential impact on SARS-CoV-2 infection and prognosis, including older age, male gender, pre-existing comorbidities and racial/ethnic disparities, variations in laboratory markers and pro-inflammatory cytokines, and possible complications [6]. However, good nutritional status and vaccination may be a protective factor for SARS-CoV-2 infection [6]. The retrospective study found that the 75th percentile LOS was 14 days. Age, WBC, ALB, and coexisting chronic respiratory diseases in COVID-19 patients were independently associated with ProLos and were included in the final nomogram. Based on these four simple variables, it predicts whether patients will stay longer, optimizes patient management and alleviates healthcare resource shortages. Advanced age is an independent risk factor for COVID-19 patients, consistent with previous studies [7]. Due to the ageing of the immune system in elderly patients, the tolerance to infectious diseases is significantly reduced [8]. However, ageing also has a sustained pro-inflammatory environment of low-grade innate immune activation, which can increase infection-induced tissue damage in older individuals [9]. At the same time, older people usually have more comorbidities and comorbid underlying conditions are significantly associated with susceptibility and poorer clinical outcomes in COVID-19 patients [10-12]. However, the relationship between asthma and COVID-19 susceptibility and severity is controversial. Only 1 of the 310 patients collected in this study had comorbid asthma, which may be related to the fact that the therapeutic use of inhaled glucocorticoids in asthmatic patients helps to reduce the risk of hospitalization for COVID-19. A study has also indicated reduced expression of ACE2 and TMPRSS2 in sputum samples from asthma patients treated with ICSs [13]. However, the univariate analysis of this study did not show any difference in hypertension, diabetes, and coronary heart disease. COVID-19 patients with older combined underlying chronic respiratory diseases need careful management of underlying lung disease [14]. The data showed a statistically significant effect of WBC on prolonged hospitalization in COVID-19, with higher WBC levels associated with more extended hospital stays in patients with COVID-19. This may be related to combined bacterial infections, and mixed infections of the lungs are challenging to diagnose. Patients with COVID-19 tend to exhibit reduced lymphocyte counts, so the neutrophil-to-lymphocyte ratio (NLR) is associated with disease prognosis [15]. Even though the median lymphocyte count was lower in the ProLos group than in the Non-ProLos group in this study, it was not statistically significant at $p > 0.05$. COVID-19 patients have longer hospital stays once co-infected with bacterial infections, and timely and appropriate use of antibiotics is essential to improve patient prognosis [16]. In general, the lower the ALB, the worse the nutritional status of the patient, and the accompanying loss of fluids, coupled with viral infections and fever symptoms, are common and may even lead to more serious hypovolemic shock. This is consistent with a multicenter retrospective study in which patients with COVID-19 hypoalbuminemia tended to have worse clinical outcomes, and the administration of nutritional supportive therapy improved clinical outcomes in these patients [17]. These four predictors are easily accessible in the clinic. The nomogram has good discriminatory and corrective power, and DCA evaluation shows its value for clinical applications. The study was conducted in a single center, and the results may not be extrapolated to other regions. However, as a retrospective single-center study, we also found the valuable prediction model, and this gives us into the next phase of a multicenter, large sample of prospective study provides the research foundation.

## CONCLUSION

A predictive model has been created to accurately predict whether the hospitalization duration of COVID-19 patients will be prolonged. This model incorporates serum WBC, ALB levels, age, and the presence of chronic respiratory system diseases.

**Acknowledgment:**

We would like to thank other members of the Department of Respiratory Medicine, Affiliated Hospital of North China University of Technology, for the constructive criticism.

**Source of Support:**

This work was supported by the Hebei Province Science Development Program [20201246] and the 2020 Doctoral Research Startup Project of the Affiliated Hospital of North China University of Science and Technology [bs2110].

**Ethical Approval:**
This study was approved by the ethics committee of North China University of Science and Technology Affiliated Hospital. All procedures performed in studies were in accordance with the ethical standards. Informed consent was obtained.

**Declaration of Interest:**
No conflicts of interest.

**References:**


1. Arora P, Shankar T, Joshi S, et al. Prognostication of COVID-19 patients using ROX index and CURB-65 score - A retrospective observational study. J Family Med Prim Care 2022 Oct;11(10): 6006-14. (PMID: 36618245)

2. Elmoheen A, Abdelhafez I, Salem W, et al. External Validation and Recalibration of the CURB-65 and PSI for Predicting 30-Day Mortality and Critical Care Intervention in Multiethnic Patients with COVID-19. Int J Infect Dis 2021 Oct;111:108-16. (PMID: 34416403)

3. Guo J, Zhou B, Zhu M, et al. CURB-65 may serve as a useful prognostic marker in COVID-19 patients within Wuhan, China: a retrospective cohort study. Epidemiol Infect 2020 Oct;148:e241. (PMID: 32998791)

4. Liang W, Liang H, Ou L, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med 2020 Aug;180(8):1081-9. (PMID: 32396163)

5. Zhang J, Li L, Hu X, et al. Development of a model by LASSO to predict hospital length of stay (LOS) in patients with the SARS-Cov-2 omicron variant. Virulence 2023 Dec;14(1):2196177. (PMID: 37041726)

6. Zhang JJ, Dong X, Liu GH, Gao YD. Risk and Protective Factors for COVID-19 Morbidity, Severity, and Mortality. Clin Rev Allergy Immunol 2022 Feb;64(1):90-107. (PMID: 35044620)

7. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar;395(10229):1054-62. (PMID: 32171076)

8. Flook M, Jackson C, Vasileiou E, et al. Informing the public health response to COVID-19: a systematic review of risk factors for disease, severity, and mortality. BMC Infect Dis 2021 Apr; 21(1):342. (PMID: 33845766)

9. Shaw AC, Joshi S, Greenwood H, Panda A, Lord JM. Curr Opin Immunol 2010 Aug;22(4):507-13. (PMID: 20667703)

10. Stokes EK, Zambrano LD, Anderson KN, et al. Coronavirus Disease 2019 Case Surveillance - United States, January 22-May 30, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun;69(24):759-65. (PMID: 32555134)

11. Chan JF, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020 Feb; 395(10223):514-23. (PMID: 31986261)

12. Guan WJ, Liang WH, Zhao Y, et al. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. Eur Respir J 2020 May;55(5):2000547. (PMID: 32217650)

13. Peters MC, Sajuthi S, Deford P, et al. COVID-19–related Genes in Sputum Cells in Asthma. Relationship to Demographic Features and Corticosteroids. Am J Respir Crit Care Med 2020 Jul; 202(1):83-90. (PMID: 32348692)

14. Awatade N, Wark P, Chan A, et al. The Complex Association between COPD and COVID-19. J Clin Med 2023 May;12(11): 3791. (PMID: 37297985)

15. Asperges E, Albi G, Zuccaro V, et al. Dynamic NLR and PLR in Predicting COVID-19 Severity: A Retrospective Cohort Study. Infect Dis Ther 2023 Jun;12(6):1625-40. (PMID: 37198387)

16. Wu HY, Chang PH, Chen KY, et al. Coronavirus disease 2019 (COVID-19) associated bacterial coinfection: Incidence, diagnosis and treatment. J Microbiol Immunol Infect 2022 Dec; 55(6):985-92. (PMID: 36243668)

17. Chen C, Zhang Y, Zhao X, Tao M, Yan W, Fu Y. Hypoalbuminemia - An Indicator of the Severity and Prognosis of COVID-19 Patients: A Multicentre Retrospective Analysis. Infect Drug Resist 2021 Sep;14:3699-710 (PMID: 34526790)