

Original Article

Development and validation of a clinical prediction model for *Clostridioides difficile* associated diarrheaRuiying Zheng^{1,2,3#}, Hwei Luan^{4#}, Jun Zhou^{1,2}, Zhixin Shi^{1,2}, Xin Hong^{1,2}, Lei Huang^{1,2}, Genyan Liu^{1,2}¹ Department of Laboratory Medicine, The First Affiliated Hospital with Nanjing Medical University, National Key Clinical Department of Laboratory Medicine, Nanjing 210029, PR China² Branch of National Clinical Research Center for Laboratory Medicine, Nanjing 210029, PR China³ Department of Clinical Laboratory, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, PR China⁴ Department of Laboratory Medicine, Jurong People's Hospital Affiliated to Jiangsu University, 212400, PR China.

Authors contributed equally to this work.

Abstract

Introduction: The aim of this study was to develop and validate a clinical prediction model for *Clostridioides difficile* associated diarrhea (CDAD) based on routine laboratory tests.

Methodology: Data from 121 CDAD patients and 123 patients with non-CDAD who presented at the First Affiliated Hospital of Nanjing Medical University between May 2017 and January 2022 were used to create a nomogram based on logistic regression. In addition, 109 stool samples from diarrhea patients in Jurong People's Hospital were collected to detect *Clostridioides difficile* toxin genes. The performance of the prediction model was assessed by the area under the curve (AUC), Hosmer-Lemeshow goodness of fit, and decision curve analysis (DCA).

Results: The following variables were included in the new multivariate regression model: white blood cell (WBC), lymphocyte (LY), hemoglobin (HGB), mean corpuscular volume (MCV), activated partial thromboplastin time (APTT), D-dimer, urea, creatinine (Cr), and uric acid (UA). The AUC of the prediction model was 0.793 (95% CI = 0.737–0.849) for the derivation sets and 0.708 (95% CI = 0.506–0.910) for the validation set. The calibrated values were 0.874 and 0.543, respectively. The nomogram showed better net benefit when prediction probability values were above 0.1 in the DCA curve.

Conclusions: A new diagnostic prediction model for CDAD was established. Clinicians can use the nomogram to initially assess the likelihood of CDAD when the patient suffers diarrhea, to ensure timely specific laboratory tests, and appropriate diagnostic and treatment measures.

Key words: *Clostridioides difficile*; diarrhea; diagnosis; nomogram.

J Infect Dev Ctries 2025; 19(5):766-775. doi:10.3855/jidc.20006

(Received 18 February 2024 – Accepted 03 October 2024)

Copyright © 2025 Zheng *et al.* This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Clostridioides difficile (*C. difficile*) is an anaerobic Gram-positive bacterium [1,2]. There is a potential for fatal consequences when *C. difficile* infection (CDI) causes nosocomial diarrhea and pseudomembranous colitis [2,3]. Over the past years, an increasing trend in morbidity and mortality caused by CDI has been observed globally [4,5]. The French healthcare-associated infection (HAI) early warning response system (HAI-EWRS, 2012–17) and the National Reference Laboratory Data (2012–17), which track statistics on the incidence of severe or fulminant CDI showed that in 2016, the incidence of CDI in the emergency department was 3.6 cases per 10,000 patient days (PD), and in 2010–2016, the incidence of CDI increased yearly at a rate of approximately 14% per year [6]. A study reported from Hong Kong in 2019 show

that the 30-day mortality rate was up to 16.8%, and the 60-day recurrence rate remained at 11% [7]. Furthermore, nearly 500,000 cases of CDI occur in the United States each year, causing 30,000 deaths and over \$5 billion in damages [8]. Some patients with other diseases are more likely to get CDI, and this adds a serious burden to the cost of treatment [9–12]. Hence, it is crucial to determine if a patient with diarrhea has *C. difficile* associated diarrhea (CDAD) and needs timely treatment.

Most CDAD patients who suffer from toxigenic *C. difficile* get diarrhea [2]. The nucleic acid amplification test (NAAT) for the toxin gene can be conducted directly from stool samples and is used for testing for *C. difficile*. Studies have shown that the NAAT has high specificity and sensitivity compared to toxigenic *C. difficile* culture [13] and is used as the predominant

laboratory diagnostic tool [2]. Although the NAAT method is relatively quick, it is expensive [13]. In China, nearly 14% of patients with diarrhea are infected with toxigenic *C. difficile*. Many hospitals in China lack the necessary diagnostic capabilities, including microbiology, immunology and molecular technologies, to identify the presence of CDAD [14]. Similar situation also exists in many developing countries [15], which could lead to a missed or incorrect diagnosis, and even serious or fatal consequences. Therefore, it is necessary to establish a prediction model to recognize CDAD in primary hospitals when specific CDI tests are not available.

The purpose of this study was to develop a nomogram of the incidence of CDAD based on routine laboratory test indicators which can assist in the early identification of CDAD in patients with diarrhea when specific CDI tests are not available.

Methodology

Design and setting

A retrospective case-control study was performed at the First Affiliated Hospital of Nanjing Medical University to develop a CDAD clinical diagnosis prediction model. This study included inpatients who were admitted between May 2017 and January 2022. The study population met the diagnostic criteria based on the guidelines of the American Thoracic Society/Infectious Diseases Society of America (ATS/IDSA). The inpatients who developed symptoms with suspicion of infectious diarrhea, hospitalization for at least 48 hours, and positive results by the GeneXpert *C. difficile* polymerase chain reaction (PCR) test were selected for the case group [2]. The control group population included those who showed clinical signs of diarrhea with a negative result by the GeneXpert *C. difficile* PCR test. We aimed for a case-control ratio of 1:1, and the controls were randomly selected according to age, gender, and department.

Data collection

Clinical data were extracted from the electronic patient record system, and demographic characteristics such as age, gender, and inpatient department were collected. Clinical laboratory results were recorded in a laboratory information system, and included routine blood parameters, biochemical analyses, and coagulation tests. The entirety of the laboratory data was derived from the patients’ initial test results at the onset of the diarrhea.

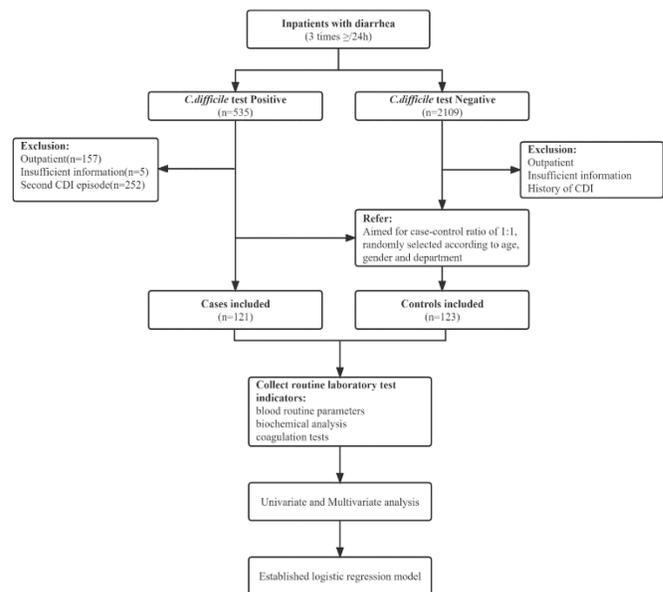
The variables had missing data in < 5% of patients, except for fibrinogen (FIB) and D-dimer, which were

incomplete in > 6% (Supplementary Figure 1). The values were entered using multiple imputations to account for missing data in multivariable analysis. This method is appropriate when values are missing at random (MAR), which seemed reasonable to apply in this study (Supplementary Figure 1). All potential predictors and the outcome variables were included in the imputation procedure.

Statistical analysis

All laboratory results for the derivation samples were classified into two groups. The best cut-off value for each variable was based on the receiver operating characteristic (ROC) curve. Multiple logistic regression was performed, starting with all potential predictors. Univariate analysis was used to select variables to determine the best prediction model. The occurrence of *C. difficile* infection probabilities was estimated using the nomogram. The discrimination capacity of the model in the derivation sample was assessed using the area under the curve (AUC), which plotted the sensitivity against the 1-specificity for all possible cut-off values to predict two endings AUC values range from no discrimination (0.5) to perfect discrimination (1.0), and AUC values greater than 0.70 suggested a reasonable estimation [16]. The calibration of the model was assessed by Hosmer-Lemeshow goodness of fit and visualized by a calibration curve. Internal validation of the model was performed using bootstrapping

Figure 1. *Clostridioides difficile* associated diarrhea (CDAD) prediction model building flow chart.



CDI: *Clostridioides difficile* infection.

techniques (n = 500), and bias-corrected estimates of the calibration curve.

A total of 109 stool samples were collected from diarrhea patients in Jurong People's Hospital affiliated to Jiangsu University, and the clinical information and laboratory results of these patients were recorded.

CDI results of the case and control groups were defined in the same way as the derivation set. This dataset was used as a validation set for external validation of the model. The decision curve analyses (DCA) were used to assess the usefulness and clinical application of the model.

The occurrence of *C. difficile* infection was the endpoint of interest in the present study. The variance inflation factor (VIF) > 4.0 in the nomogram was interpreted as indicating multicollinearity, and the use of those variables was avoided in the final model analysis. Influential cases > 0.1 were defined as strong influential cases, which were the parameters affecting the model scale imbalance and the final prediction model avoided use these parameters.

All statistical analyses were performed using the R programming language and environment (<http://www.r-project.org/>) and IBM SPSS 26 (IBM Corp, Armonk, NY, USA). Moreover, the R code used in this research is available in GitHub (<https://github.com/Wingerlin/R>).

Results

Data distribution

A total of 121 CDAD inpatients and 123 controls were included in the derivation study. The exclusion criteria are summarized in Figure 1. The demographic and variable distribution characteristics after imputation are summarized in Table 1. The majority of the patients included in the study was ≤ 50 years of age (42.98%, n = 121), and the main department of distribution was gastroenterology (66.12%, n = 121).

The stool samples collected from Jurong People's Hospital were tested for the toxin gene, along with routine tests (Table 1). Based on the results of the toxin gene tests, there were 9 cases in the case group and 100 cases in the control group. All the patients of the case group were > 50 years old; and the majority of the patients' diagnoses was complicated with presence of cardiovascular, liver and kidney, and respiratory diseases (88.90%, n = 9).

Variables screening and nomogram construction

Based on the results of the univariate analysis and after avoiding missing important variables, we finally adopted *p* < 0.1 as the significance level for the multivariate logistic regression analysis. Basophil (BA), hematocrit (HCT), platelet count (PLT) and calcium (Ca) were the criteria used for inclusion of the patients in the CDAD group, but not in the control group.

Table 1. Characteristics of patients with *Clostridioides difficile* infection (CDI) diarrhea and the control group.

Variable	Derivation		Validation	
	Control (n = 123)	Case (n = 121)	Control (n = 100)	Case (n = 9)
Age, n (%) years				
≤ 50	57 (46.34)	52 (42.98)	15 (15.00)	0 (0.00)
50–59	23 (18.70)	23 (19.01)	26 (26.00)	1 (11.11)
60–69	20 (16.26)	18 (14.88)	17 (17.00)	5 (55.56)
70–79	15 (12.20)	14 (11.57)	29 (29.00)	2 (22.22)
≥ 80	8 (6.50)	14 (11.57)	13 (13.00)	1 (11.11)
Gender, n (%)				
Male	78 (63.41)	74 (61.16)	54 (54.00)	6 (66.67)
Female	45 (36.59)	47 (38.84)	46 (46.00)	3 (33.33)
Department, n (%)				
Gastroenterology	82 (66.67)	80 (66.12)	15 (15.00)	1 (11.11)
Other	41 (33.33)	41 (33.88)	85 (85.00)	8 (88.89)
Diagnosis, n (%)				
Inflammatory bowel disease	56 (45.53)	59 (48.76)	15 (15.00)	1 (11.11)
Cardiovascular disease	14 (11.38)	11 (9.09)	9 (9.00)	2 (22.22)
Liver and Kidney Diseases	6 (4.88)	9 (7.44)	15 (15.00)	3 (33.33)
Respiratory diseases	10 (8.13)	9 (7.44)	20 (20.00)	3 (33.33)
Leukemia or malignancy	11 (8.94)	10 (8.26)	6 (6.00)	0 (0.00)
No or other diseases	26 (21.14)	23 (19.01)	35 (35.00)	0 (0.00)
WBC, n (%)				
≤ 8.740	101 (82.11)	79 (65.29)	78 (78.00)	7 (77.78)
> 8.740	22 (17.89)	42 (34.71)	22 (22.00)	2 (22.22)
LY, n (%)				
≤ 1.045	45 (36.59)	29 (23.97)	42 (42.00)	6 (66.67)
> 1.045	78 (63.41)	92 (76.03)	58 (58.00)	3 (33.33)

Table 1 (continued). Characteristics of patients with *Clostridioides difficile* infection (CDI) diarrhea and the control group.

Variable	Derivation		Validation	
	Control (n = 123)	Case (n = 121)	Control (n = 100)	Case (n = 9)
MO, n (%)				
≤ 0.435	74 (60.16)	84 (69.42)	47 (47.00)	3 (33.33)
> 0.435	49 (39.84)	37 (30.58)	53 (53.00)	6 (66.67)
EO, n (%)				
≤ 0.095	65 (52.85)	58 (47.93)	57 (57.00)	6 (66.67)
> 0.095	58 (47.15)	63 (52.07)	43 (43.00)	3 (33.33)
BA, n (%)				
≤ 0.025	64 (52.03)	45 (37.19)	71 (71.00)	8 (88.89)
> 0.025	59 (47.97)	76 (62.81)	29 (29.00)	1 (11.11)
HGB, n (%)				
≤ 104.500	60 (48.78)	33 (27.27)	23 (23.00)	4 (44.44)
> 104.500	63 (51.22)	88 (72.73)	77 (77.00)	5 (55.56)
HCT, n (%)				
≤ 35.550	80 (65.04)	55 (45.45)	39 (39.00)	6 (66.67)
> 35.550	43 (34.96)	66 (54.55)	61 (61.00)	3 (33.33)
MCV, n (%)				
≤ 85.050	43 (34.96)	18 (14.88)	48 (48.00)	1 (11.11)
> 85.050	80 (65.04)	103 (85.12)	52 (52.00)	8 (88.89)
PLT, n (%)				
≤ 349.500	100 (81.30)	109 (90.08)	97 (97.00)	8 (88.89)
> 349.500	23 (18.70)	12 (9.92)	3 (3.00)	1 (11.11)
MPV, n (%)				
≤ 9.130	33 (26.83)	24 (19.83)	5 (5.00)	0 (0.00)
> 9.130	90 (73.17)	97 (80.17)	95 (95.00)	9 (100.00)
PDW, n (%)				
≤ 11.650	50 (40.65)	37 (30.58)	25 (25.00)	3 (33.33)
> 11.650	73 (59.35)	84 (69.42)	75 (75.00)	6 (66.67)
PT, n (%)				
≤ 13.050	81 (65.85)	92 (76.03)	83 (83.00)	6 (66.67)
> 13.050	42 (34.15)	29 (23.97)	17 (17.00)	3 (33.33)
APTT, n (%)				
≤ 28.650	65 (52.85)	41 (33.88)	94 (94.00)	1 (11.11)
> 28.650	58 (47.15)	80 (66.12)	6 (6.00)	8 (88.89)
FIB, n (%)				
≤ 2.180	19 (15.45)	29 (23.97)	30 (30.00)	2 (22.22)
> 2.180	104 (84.55)	92 (76.03)	70 (70.00)	7 (77.78)
D-dimer, n (%)				
≤ 2.160	93 (75.61)	104 (85.95)	83 (83.00)	7 (77.78)
> 2.160	30 (24.39)	17 (14.05)	17 (17.00)	2 (22.22)
ALT, n (%)				
≤ 17.650	83 (67.48)	70 (57.85)	51 (51.00)	5 (55.56)
> 17.650	40 (32.52)	51 (42.15)	49 (49.00)	4 (44.44)
AST, n (%)				
≤ 20.150	73 (59.35)	58 (47.93)	55 (55.00)	4 (44.44)
> 20.150	50 (40.65)	63 (52.07)	45 (45.00)	5 (55.56)
LDH, n (%)				
≤ 161.500	34 (27.64)	23 (19.01)	14 (14.00)	3 (33.33)
> 161.500	89 (72.36)	98 (80.99)	86 (86.00)	6 (66.67)
ALB, n (%)				
≤ 35.450	70 (56.91)	55 (45.45)	30 (30.00)	6 (66.67)
> 35.450	53 (43.09)	66 (54.55)	70 (70.00)	3 (33.33)
Urea, n (%)				
≤ 6.450	93 (75.61)	78 (64.46)	83 (83.00)	5 (55.56)
> 6.450	30 (24.39)	43 (35.54)	17 (17.00)	4 (44.44)
Cr, n (%)				
≤ 40.050	11 (8.94)	2 (1.65)	44 (44.00)	3 (33.33)
> 40.050	112 (91.06)	119 (98.35)	56 (56.00)	6 (66.67)
UA, n (%)				
≤ 393.500	88 (71.54)	99 (81.82)	89 (89.00)	7 (77.78)
> 393.500	35 (28.46)	22 (18.18)	11 (11.00)	2 (22.22)
Ca, n (%)				
≤ 2.205	83 (67.48)	64 (52.89)	44 (44.00)	5 (55.56)
> 2.205	40 (32.52)	57 (47.11)	56 (56.00)	4 (44.44)

WBC: white blood cell count; LY: lymphocyte; MO: monocyte; EO: eosinophil; BA: basophil; HGB: hemoglobin; HCT: hematocrit; MCV: mean corpuscular volume; PLT: platelet count; MPV: mean platelet volume; PDW: platelet volume distribution width; PT: prothrombin time; APTT: activated partial thromboplastin time; FIB: fibrinogen; ALT: alanine aminotransferase; AST: aspartate aminotransferase; LDH: lactate dehydrogenase; ALB: albumin; Cr: creatinine; UA: uric acid; Ca: calcium.

However, the predictive performance of the final clinical prediction model was not improved by adding these metrics; so, a more concise prediction model was chosen. The univariate analysis results are shown in Table 2. The other indicators were not significantly correlated with the occurrence of CDAD. We generated a predicted model for CDI according to the variables screened, as shown in multivariate analysis (Table 2); and the nomogram (Figure 2) was generated with the final inclusion of 9 predictors, including white blood cell (WBC), lymphocyte (LY), hemoglobin (HGB), mean corpuscular volume (MCV), activated partial thromboplastin time (APTT), D-dimer, urea, creatinine (Cr), and uric acid (UA).

Figure 2. Nomogram to predict the incidence of *Clostridioides difficile* associated diarrhea (CDAD).

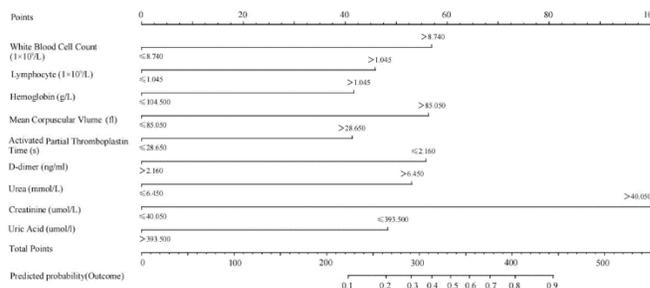


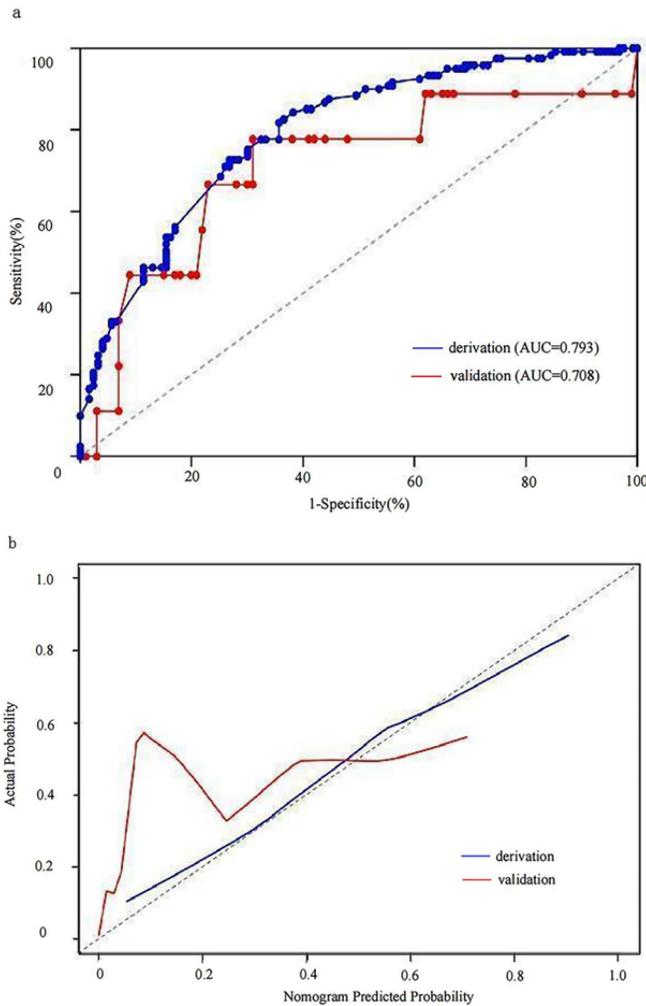
Table 2. Results of univariate and multivariate analyses.

Variables	Univariate analysis				Multivariate analysis			
	β coefficient	OR	95% CI	p value	β coefficient	OR	95% CI	p value
WBC								
≤ 8.740					refer			
> 8.740	0.303	2.441	(1.348–4.420)	0.003	1.132	3.101	(1.528–6.294)	0.002
LY								
≤ 1.045					refer			
> 1.045	0.604	1.83	(1.050–3.190)	0.033	0.912	2.488	(1.229–5.036)	0.011
BA								
≤ 0.025					refer			
> 0.025	0.605	1.832	(1.099–3.054)	0.020	0.294	1.342	(0.697–2.585)	0.379
HGB								
≤ 104.500					refer			
> 104.500	0.932	2.54	(1.489–4.332)	0.001	0.828	2.289	(1.161–4.512)	0.017
HCT								
≤ 35.550					refer			
> 35.550	-0.803	0.448	(0.268–0.750)	0.002	0.285	1.33	(0.568–3.114)	0.512
MCV								
≤ 85.050					refer			
> 85.050	-1.124	0.325	(0.174–0.606)	0.000	1.12	3.064	(1.471–6.380)	0.003
PLT								
≤ 349.500					refer			
> 349.500	0.737	2.089	(0.988–4.418)	0.054	-0.562	0.57	(0.206–1.574)	0.278
APTT								
≤ 28.650					refer			
> 28.650	0.782	2.187	(1.304–3.666)	0.003	0.822	2.275	(1.240–4.173)	0.008
D-dimer								
≤ 2.160					refer			
> 2.160	-0.68	0.507	(0.263–0.978)	0.043	-1.109	0.330	(0.139–0.782)	0.012
Urea								
≤ 6.450					refer			
> 6.450	0.536	1.709	(0.981–2.977)	0.058	1.054	2.869	(1.310–6.282)	0.008
Cr								
≤ 40.050					refer			
> 40.050	-1.765	0.171	(0.037–0.789)	0.024	1.988	7.299	(1.198–44.490)	0.031
UA								
≤ 393.500					refer			
> 393.500	0.582	1.79	(0.977–3.280)	0.060	-0.96	0.383	(0.190–0.772)	0.007
Ca								
≤ 2.205					refer			
> 2.205	0.614	1.848	(1.099–3.106)	0.020	0.758	2.135	(0.937–4.864)	0.071

WBC: white blood cell count; LY: lymphocyte; BA: basophil; HGB: hemoglobin; HCT: hematocrit; MCV: mean corpuscular volume; PLT: platelet count; APTT: activated partial thromboplastin time; Cr: creatinine; UA: uric acid; Ca: calcium; refer: reference; OR: odds ratio; CI: confidence interval. p values that were < 0.1 in univariate analysis: and < 0.05 in multivariate analysis are in bold.

The corresponding scores for the above factors were summed to a total score corresponding to the predicted value at the bottom. The variance inflation factor (VIF) values were all < 4 (Supplementary Table 1), which meant no collinearity existed between screened variables. Influential cases of the model are shown Supplementary Figure 2 and Cook's distance was less than 1 for all variables.

Figure 3. Receiver operating characteristic (ROC) curves and calibration curves of nomogram.



A. The area under the subject operating characteristic curve (AUC) of the prediction model. The blue line represents the receiver operating characteristic (ROC) curve of the derivation and the AUC is 0.793 [95% CI (0.737–0.849)]; The red line is the ROC curve after external validation, and the AUC of the internal validation data is 0.708 [95% CI (0.506–0.910)]. **B.** Calibration of the nomogram to predict the outcome of diarrhea in an inpatient in the training set and validation set. The blue line represents the calibration of the derivation sets, the red line is the performance of the validation set.

Nomogram internal validation

After plotting the ROC for the nomogram, the AUC value was 0.793 (95% confidence interval (CI) = 0.737–0.849), and after internal validation by bootstrapping techniques, AUC was 0.761 (95% CI = 0.702–0.821), which showed good discrimination. The AUC > 0.70 for prediction in both training and validation data (Figure 3A), indicated favorable discrimination by the nomogram. The nomogram calibration curves showed good agreement of probabilities between predicted and observed values, and after Hosmer-Lemeshow goodness of fit analysis, the *p* value was 0.874 (Figure 3B). Thus, the nomogram had high confidence in discrimination and calibration.

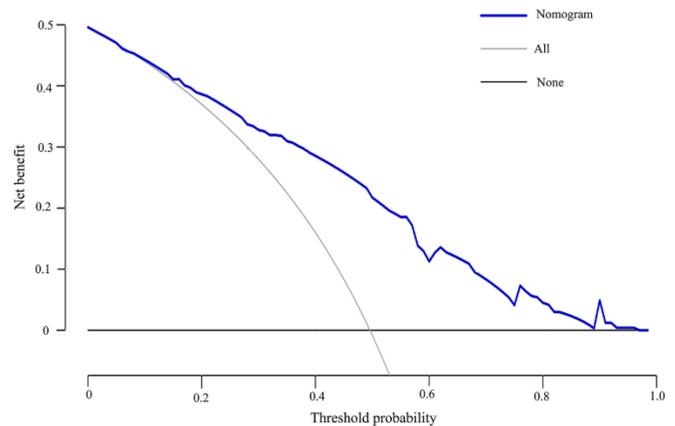
Nomogram external validation

The performance of the model was evaluated using the validation set again, and the ROC curve was plotted as shown with the red curve in Figure 3A. The AUC was 0.708 (95% CI = 0.506–0.910), and the prediction model had a modest performance among the external validation set. The calibration curves were plotted as shown with the red curve in Figure 3B and the *p* value was 0.543 after Hosmer-Lemeshow goodness of fit analysis. It was observed that the prediction model underestimated the occurrence of the events when the predicted probability was less than 0.5, and overestimated the occurrence of events when the predicted probability was greater than 0.5.

Clinical effectiveness of the nomogram

The DCA demonstrated the clinical utility of the nomogram (Figure 4). The grey curve (all variables) in Figure 4 represents the clinical benefits associated with changes in thresholds after all indicators were included

Figure 4. Decision curve analysis of nomogram.



The y-axis represents the net benefit, the x-axis represents the risk threshold of infection in diarrhea.

in the multivariable logistic regression, and the black curve (no variables) represents the clinical benefits resulting from the exclusion of any indicator. This was able to better predict the disease occurrence, as it showed better net benefit when the threshold probability was > 0.1 , and this could alert clinicians to the possibility of CDAD in patients with diarrhea to avoid missing diagnosis.

Discussion

CDAD is an important clinical condition worldwide [17]. Its clinical manifestations include severe or bloody diarrhea, pseudomembranous enterocolitis, and toxic megacolon; which are often life-threatening and fatal [18]. Colonization with toxigenic *C. difficile* significantly increases the risk of clinical infection, and it is important to develop infection prevention strategies [19]. However, in many primary care settings, specific *C. difficile* tests are not available. It is important to develop a useful prediction model to identify CDAD to reduce missed or incorrect CDAD diagnoses, especially in China and other developing countries [14,15].

Previously developed clinical prediction rules were most concerned with the possible consequences of CDI or recurrent CDI [15,20]. Tilton *et al.* developed a prediction model for the risk of nosocomial *C. difficile* infection in patients receiving systemic antibiotic therapy, which included 200 subjects (100 cases and 100 controls); and the ultimately identified meaningful predictors were patients receiving systemic antibiotics, age > 70 years, recently admitted (within the last 90 days), and an AUC value was 0.7 [21]. Cobo *et al.* reported a recurrence risk model for *C. difficile* infection and completed external validation in 183 patients with an AUC of 0.72 [22].

So far, no clinical rules have been developed to address the possibility of CDI in patients with diarrhea. To our knowledge, the present study is the first diagnosis prediction model for CDAD. This study successfully developed a nomogram to transform complex regression equations into a simple and visual graph, which has the potential to be a wieldy CDAD decision-making method. The study further evaluated the discrimination and calibration of the nomogram. The AUC was 0.793 in derivation and 0.761 after bootstrap resampling technique ($n = 500$); and the p value was 0.874 after Hosmer-Leneshow goodness-of-fit test. The AUC of the model was 0.708 and $p = 0.543$ after Hosmer-Leneshow goodness-of-fit. The prediction model showed good performance in the derivation set, internal validation, and external validation to identify CDI; which will hopefully

provide a new diagnostic pathway for healthcare facilities that are not yet able to perform specific diagnosis of CDAD.

The performance of the model with the external validation data was slightly lower compared to the training set. This was due to the fact that the distribution of clinical characteristics was different between the derivation set and the validation set.

The CDI severity was defined using serum WBC counts and Cr levels by the Infection Diseases Society of America (IDSA)/Society for Healthcare Epidemiology of America (SHEA) criteria [2]. One of the pathogenic mechanisms of *C. difficile* is the induction of intestinal inflammation through the release of toxins A and toxin B [23], with glycosylate host Rho and Rac GTPases, leading to disruption of the actin cytoskeleton and loss of epithelial barrier integrity, with subsequent apoptosis and tissue damage leading to inflammation [24,25]. Moreover, Pope *et al.* found that *C. difficile* toxin B can activate group 3 innate lymphocytes [26]. In the present study, the odds ratio (OR) of WBC and LY were 3.101 and 2.488 respectively; thus, acting as an important indicator of inflammation or infection, exhibiting similar behavior in the model.

Choi *et al.* demonstrated that low hemoglobin levels were significantly more specific in the recurrent CDI group compared to the control group [27]. Lee *et al.* also found that anemia was associated with poor outcomes during community onset of CDI [28]. CDI often occurs in patients with underlying diseases, such as inflammatory bowel disease [29,30], liver cirrhosis [31], malignancies [32], and hematologic disorders [9]. These diseases are often accompanied by hemorrhagic anemia, abnormal hematopoiesis, or lack of hematopoietic material. Moreover, some patients with severe CDI may also present with blood in the stool [2]. In the present study, the OR of HGB was 2.289, which was inconsistent with the findings of the previous study. This could be due to several reasons. First, patients with mild or common CDI were less likely to have bloody stools; Second, without a combination of other bleeding causes, the amount of bleeding caused by CDI was limited. In addition, we also found that OR of MCV was 3.064; and there were few existing literatures on MCV and CDI. However, we concluded that the bone marrow of CDAD was normal for hematopoiesis and that the hematopoietic material was sufficient, and could cause microcytic anemia.

Aronsson *et al.* measured antithrombin III, protein C, and free protein S; and found that CDI may lead to loss of coagulation inhibitors, causing a risk of

thromboembolic complication [33]. Our findings were inconsistent with this previous study. Considering the normal reference range (< 0.55mg/L) for laboratory testing of D-dimers, we inferred that it may have been caused by unreasonable statistical grouping. Interestingly, however, we found that the OR of APTT was 2.275, which suggested a statistical correlation between prolonged APTT and CDI. Previously, Alba Isasi *et al.* reported that venous sinus thrombosis could form secondary to disseminated intravascular coagulation after CDI [34], and the elevated fibrin degradation products (FDP) may have been a cause of the prolonged APTT. In addition, the lack of production of endogenous coagulation factors in disease states could have been another important reason for prolonged APTT.

Crivaro *et al.* showed that levels of urea, Cr, and WBC were positively correlated when CDAD was diagnosed [35]. Nomura *et al.* found that blood urea was significantly higher in the thoroughbred racehorse CDAD group than in the control group [36]. In this study, the OR of urea was 2.869, which suggested that elevated blood urea was associated with CDI. Moreover, we also found the OR of Cr was 7.299. Previously, the IDSA/SHEA criteria defined the CDI severity using serum WBC counts and Cr levels ($\geq 1.5\text{mg/dL}$). Some patients with severe CDI had renal failure, which was closely related to elevated urea and Cr.

There were few reports about the relationship between UA and CDI. In this study, low UA levels were associated with higher incidence of CDI (OR = 0.383). We concluded that this could have been due to the following reasons. These patients were mostly associated with underlying diseases or long-term antibiotic use resulting in intestinal dysfunction and malnutrition. In addition, some patients with renal tubular acidosis caused by renal failure may have had low UA.

Despite the model being a promising CDAD predictive tool in patients with diarrhea, there were several limitations in this study. The difference of distribution characteristics between the derivation set and the validation set, and the small volume of data in the validation set, may have influenced our results. It is therefore necessary to extend the generalizability of the model to multi-regional and multi-center data in order to validate it.

Conclusions

This study established a prediction model for CDAD through routine laboratory tests. Clinicians

should consider the possibility of CDAD when patients develop diarrhea, and they can use the nomogram to make a primary assessment of the probability of CDAD in patients to prescribe specific diagnostic tests and corresponding treatment measures in a timely manner.

Ethical considerations

This study was approved by the Ethics Office of the First Affiliated Hospital of Nanjing Medical University with approval number 2022-SR-306, and the hospital Ethics Committee waived the requirement for informed consent from patients since the study was retrospective and all procedures were routine. We certify that the study was performed in accordance with the relevant guidelines and regulations.

Funding

This research was supported by the Project of Jiangsu Provincial Medical Key Discipline (grant number ZDXK202239).

Corresponding author

Genyan Liu, MD, PhD.

Department of Laboratory Medicine, The First Affiliated Hospital with Nanjing Medical University, Nanjing 210029, PR China.

Tel: (86 25) 68308025

Fax: (86 25) 83724440;

Email: liugenyang@njmu.edu.cn.

Conflict of interests

No conflict of interests is declared.

References

1. Lin CY, Cheng HT, Kuo CJ, Lee YS, Sung CM, Keidan M, Rao K, Kao JY, Hsieh SY (2022) Proton pump inhibitor-induced gut dysbiosis increases mortality rates for patients with *Clostridioides difficile* infection. *Microbiol Spectr* 10: e0048622. doi: 10.1128/spectrum.00486-22.
2. McDonald LC, Gerding DN, Johnson S, Bakken JS, Carroll KC, Coffin SE, Dubberke ER, Garey KW, Gould CV, Kelly C, Loo V, Shaklee Sammons J, Sandora TJ, Wilcox MH (2018) Clinical practice guidelines for *Clostridium difficile* infection in adults and children: 2017 update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin Infect Dis* 66: e1–e48. doi: 10.1093/cid/cix1085.
3. Magnusson C, Mernelius S, Bengner M, Noren T, Serrander L, Forsell S, Matussek A (2022) Characterization of a *Clostridioides difficile* outbreak caused by PCR ribotype 046, associated with increased mortality. *Emerg Microbes Infect* 11: 850–859. doi: 10.1080/22221751.2022.2049981.
4. Dudzicz S, Kujawa-Szewieczek A, Kwiecien K, Wiecek A, Adamczak M (2018) *Lactobacillus plantarum* 299v reduces the incidence of *Clostridium difficile* infection in nephrology

- and transplantation ward — results of one year extended study. *Nutrients* 10: 1–11. doi: 10.3390/nu10111574.
5. Manthey CF, Dranova D, Christner M, Drolz A, Kluge S, Lohse AW, Fuhrmann V (2019) Initial therapy affects duration of diarrhoea in critically ill patients with *Clostridioides difficile* infection (CDI). *Crit Care* 23: 399. doi: 10.1186/s13054-019-2648-6.
 6. Colomb-Cotinat M, Assouvie L, Durand J, Daniau C, Leon L, Maugat S, Soing-Altrach S, Gateau C, Couturier J, Arnaud I, Astagneau P, Berger-Carbonne A, Barbut F (2019) Epidemiology of *Clostridioides difficile* infections, France, 2010 to 2017. *Euro Surveill* 24: 7–16. doi: 10.2807/1560-7917.ES.2019.24.35.1800638.
 7. Guo CLT, Kwong TNY, Mak JWY, Zhang L, Lui GCY, Wong GLH, Ip M, Yu J, Sung JY, Wu WKK, Wong SH (2021) Trends in incidence and clinical outcomes of *Clostridioides difficile* infection, Hong Kong. *Emerg Infect Dis* 27: 3036–3044. doi: 10.3201/eid2712.203769.
 8. Scaria E, Powell WR, Birstler J, Alagoz O, Shirley D, Kind AJH, Safdar N (2020) Neighborhood disadvantage and 30-day readmission risk following *Clostridioides difficile* infection hospitalization. *BMC Infect Dis* 20: 762. doi: 10.1186/s12879-020-05481-x.
 9. Duhalde L, Lurienne L, Wingen-Heimann SM, Guillou L, Buffet R, Bandinelli PA (2020) Excess burden associated with *Clostridioides difficile* infection in haematological patients occurring during hospitalization with induction chemotherapy in the USA. *J Hosp Infect* 104: 560–566. doi: 10.1016/j.jhin.2019.12.017.
 10. Braae UC, Moller FT, Ibsen R, Ethelberg S, Kjellberg J, Molbak K (2020) The economic burden of *Clostridioides difficile* in Denmark: a retrospective cohort study. *Front Public Health* 8: 562957. doi: 10.3389/fpubh.2020.562957.
 11. Granata G, Bartoloni A, Codeluppi M, Contadini I, Cristini F, Fantoni M, Ferraresi A, Fornabaio C, Grasselli S, Lagi F, Masucci L, Puoti M, Raimondi A, Taddei E, Trapani FF, Viale P, Johnson S, Petrosillo N, On Behalf of the CloVid Study Group (2020) The burden of *Clostridioides difficile* infection during the COVID-19 pandemic: a retrospective case-control study in Italian hospitals (CloVid). *J Clin Med* 9: 1–11. doi: 10.3390/jcm9123855.
 12. Wijarnprecha K, Aby ES, Kim D, Ungprasert P, Cheungpasitporn W, Thongprayoon C, Lukens FJ, Harnois DM, Kroner PT (2021) The burden of *Clostridioides difficile* infection in patients with history of liver transplant and during index admission. *Eur J Gastroenterol Hepatol* 33: 894–898. doi: 10.1097/MEG.0000000000001812.
 13. Kim N, Lee SY, Park J, Lee J (2022) Comparative evaluation of three immunoassays for the simultaneous detection of *Clostridioides difficile* glutamate dehydrogenase and toxin A/B. *Microorganisms* 10: 1–11. doi: 10.3390/microorganisms10050947.
 14. Tang C, Cui L, Xu Y, Xie L, Sun P, Liu C, Xia W, Liu G (2016) The incidence and drug resistance of *Clostridium difficile* infection in Mainland China: a systematic review and meta-analysis. *Sci Rep* 6: 37865. doi: 10.1038/srep37865.
 15. Curcio D, Cane A, Fernandez FA, Correa J (2019) *Clostridium difficile*-associated diarrhea in developing countries: a systematic review and meta-analysis. *Infect Dis Ther* 8: 87–103. doi: 10.1007/s40121-019-0231-8.
 16. Linden A (2006) Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. *J Eval Clin Pract* 12: 132–139. doi: 10.1111/j.1365-2753.2005.00598.x.
 17. Dingle KE, Didelot X, Quan TP, Eyre DW, Stoesser N, Golubchik T, Harding RM, Wilson DJ, Griffiths D, Vaughan A, Finney JM, Wyllie DH, Oakley SJ, Fawley WN, Freeman J, Morris K, Martin J, Howard P, Gorbach S, Goldstein EJC, Citron DM, Hopkins S, Hope R, Johnson AP, Wilcox MH, Peto TEA, Walker AS, Crook DW, Modernising Medical Microbiology Informatics Group (2017) Effects of control interventions on *Clostridium difficile* infection in England: an observational study. *Lancet Infect Dis* 17: 411–421. doi: 10.1016/S1473-3099(16)30514-X.
 18. Chen YH, Li TJ, Tsai BY, Chen LK, Lai YH, Li MJ, Tsai CY, Tsai PJ, Shieh DB (2019) Vancomycin-loaded nanoparticles enhance sporicidal and antibacterial efficacy for *Clostridium difficile* infection. *Front Microbiol* 10: 1141. doi: 10.3389/fmicb.2019.01141.
 19. Arriola V, Tischendorf J, Musuuza J, Barker A, Rozelle JW, Safdar N (2016) Assessing the risk of hospital-acquired *Clostridium difficile* infection with proton pump inhibitor use: a meta-analysis. *Infect Control Hosp Epidemiol* 37: 1408–1417. doi: 10.1017/ice.2016.194.
 20. Hu MY, Katchar K, Kyne L, Maroo S, Tummala S, Dreisbach V, Xu H, Leffler DA, Kelly CP (2009) Prospective derivation and validation of a clinical prediction rule for recurrent *Clostridium difficile* infection. *Gastroenterology* 136: 1206–1214. doi: 10.1053/j.gastro.2008.12.038.
 21. Tilton CS, Johnson SW (2019) Development of a risk prediction model for hospital-onset *Clostridium difficile* infection in patients receiving systemic antibiotics. *Am J Infect Control* 47: 280–284. doi: 10.1016/j.ajic.2018.08.021.
 22. Cobo J, Merino E, Martinez C, Cozar-Llisto A, Shaw E, Marrodan T, Calbo E, Bereciartua E, Sanchez-Munoz LA, Salavert M, Perez-Rodriguez MT, Garcia-Rosado D, Bravo-Ferrer JM, Galvez-Acebal J, Henriquez-Camacho C, Cuquet J, Pino-Calm B, Torres L, Sanchez-Porto A, Fernandez-Felix BM, Nosocomial Infection Study Group (2018) Prediction of recurrent *Clostridium difficile* infection at the bedside: the GEIH-CDI score. *Int J Antimicrob Agents* 51: 393–398. doi: 10.1016/j.ijantimicag.2017.09.010.
 23. Wang J, Ghali S, Xu C, Mussatto CC, Ortiz C, Lee EC, Tran DH, Jacobs JP, Lagishetty V, Faull KF, Moller T, Rossetti M, Chen X, Koon HW (2018) Ceragenin CSA13 reduces *Clostridium difficile* infection in mice by modulating the intestinal microbiome and metabolites. *Gastroenterology* 154: 1737–1750. doi: 10.1053/j.gastro.2018.01.026.
 24. Fletcher JR, Pike CM, Parsons RJ, Rivera AJ, Foley MH, McLaren MR, Montgomery SA, Theriot CM (2021) *Clostridioides difficile* exploits toxin-mediated inflammation to alter the host nutritional landscape and exclude competitors from the gut microbiota. *Nat Commun* 12: 462. doi: 10.1038/s41467-020-20746-4.
 25. Anjuwon-Foster BR, Tamayo R (2018) Phase variation of *Clostridium difficile* virulence factors. *Gut Microbes* 9: 76–83. doi: 10.1080/19490976.2017.1362526.
 26. Pope RL, Chitrakar A, Sah P, Shadid T, Ballard JD, Zenewicz LA (2022) *Clostridioides difficile* toxin B activates group 3 innate lymphocytes. *Infect Immun* 90: e0007322. doi: 10.1128/iai.00073-22.
 27. Choi HK, Kim KH, Lee SH, Lee SJ (2011) Risk factors for recurrence of *Clostridium difficile* infection: effect of vancomycin-resistant enterococci colonization. *J Korean Med Sci* 26: 859–864. doi: 10.3346/jkms.2011.26.7.859.

28. Lee E, Song KH, Bae JY, Yoon D, Hwang JH, Choe PG, Park WB, Bang JH, Kim ES, Park SW, Kim NJ, Oh MD, Kim HB (2018) Risk factors for poor outcome in community-onset *Clostridium difficile* infection. *Antimicrob Resist Infect Control* 7: 75. doi: 10.1186/s13756-018-0365-6.
29. Gros B, Soto P, Causse M, Marin S, Iglesias E, Benitez JM (2022) Impact of *Clostridioides difficile* infection in patients admitted with ulcerative colitis. *Scand J Gastroenterol* 58: 232–239. doi: 10.1080/00365521.2022.2121175.
30. Boeriu A, Roman A, Fofiu C, Dobru D (2022) The current knowledge on *Clostridioides difficile* infection in patients with inflammatory bowel diseases. *Pathogens* 11: 1–16. doi: 10.3390/pathogens11070819.
31. Liu Y, Chen M (2022) *Clostridioides difficile* infection in liver cirrhosis: a concise review. *Can J Gastroenterol Hepatol* 2022: 4209442. doi: 10.1155/2022/4209442.
32. Sahu KK, Mishra AK, Jindal V, Siddiqui AD, George SV (2021) To study the contributing factors and outcomes of *Clostridioides difficile* infection in patients with solid tumors. *Heliyon* 7: e08450. doi: 10.1016/j.heliyon.2021.e08450.
33. Aronsson B, Blomback M, Eriksson S, Egberg N (1992) Low levels of coagulation inhibitors in patients with *Clostridium difficile* infection. *Infection* 20: 58-60. doi: 10.1007/BF01711063.
34. Alba Isasi MT, Garcia Nunez DF, Navarro Garcia JC, Valero Lopez G (2022) Venous sinus thrombosis secondary to disseminated intravascular coagulation after *Clostridioides difficile* infection. *Neurologia (Engl Ed)* 37: 499–501. doi: 10.1016/j.nrleng.2022.03.003.
35. Crivaro AN, Carasi P, Salto I, Hugo A, Soldavini Pelichotti PC, Bengoa A, Fragomeno M, Serradell MA, Minnaard J, Rolny I, Alul E, Arregui L, Fabra Martinez ME, Moreno Valero OJ, Facente A, Magarinos F, Jewtuchowicz V, Perez PF, Trejo FM (2022) *Clostridioides difficile*: characterization of the circulating toxinotypes in an Argentinean public hospital. *Rev Argent Microbiol* 55: 73–82. doi: 10.1016/j.ram.2022.05.010.
36. Nomura M, Kuroda T, Tamura N, Muranaka M, Niwa H (2020) Mortality, clinical findings, predisposing factors and treatment of *Clostridioides difficile* colitis in Japanese thoroughbred racehorses. *Vet Rec* 187: e14. doi: 10.1136/vr.105605.

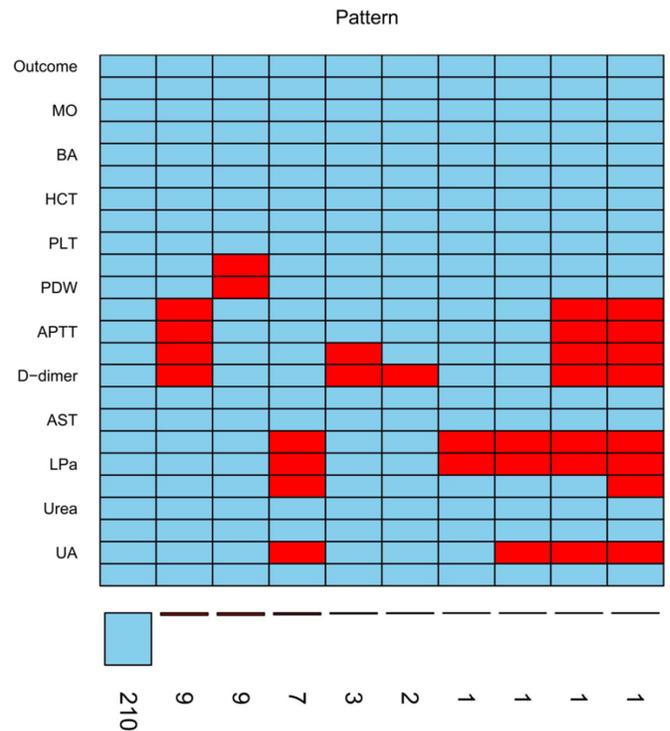
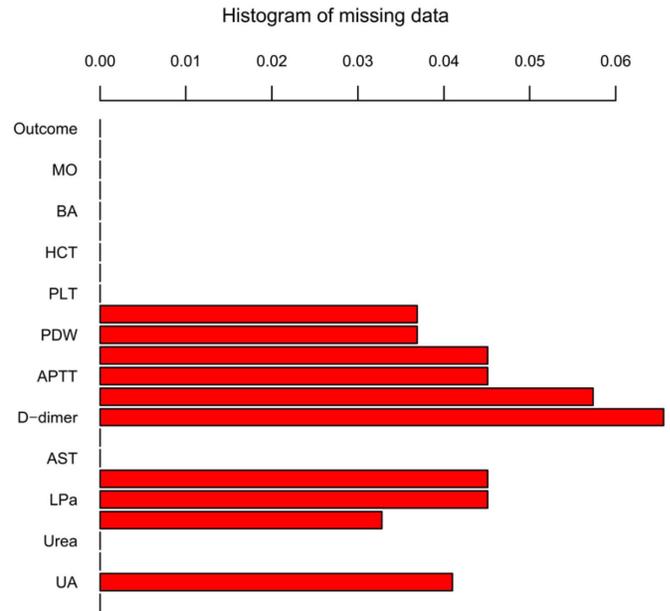
Annex – Supplementary Items

Supplementary Table 1. Multicollinearity test of included variables.

Variables	Multicollinearity
WBC	1.105375
LY	1.173308
HGB	1.233631
MCV	1.138659
APTT	1.068625
D-dimer	1.339326
Urea	1.461708
Cr	1.038089
UA	1.076766

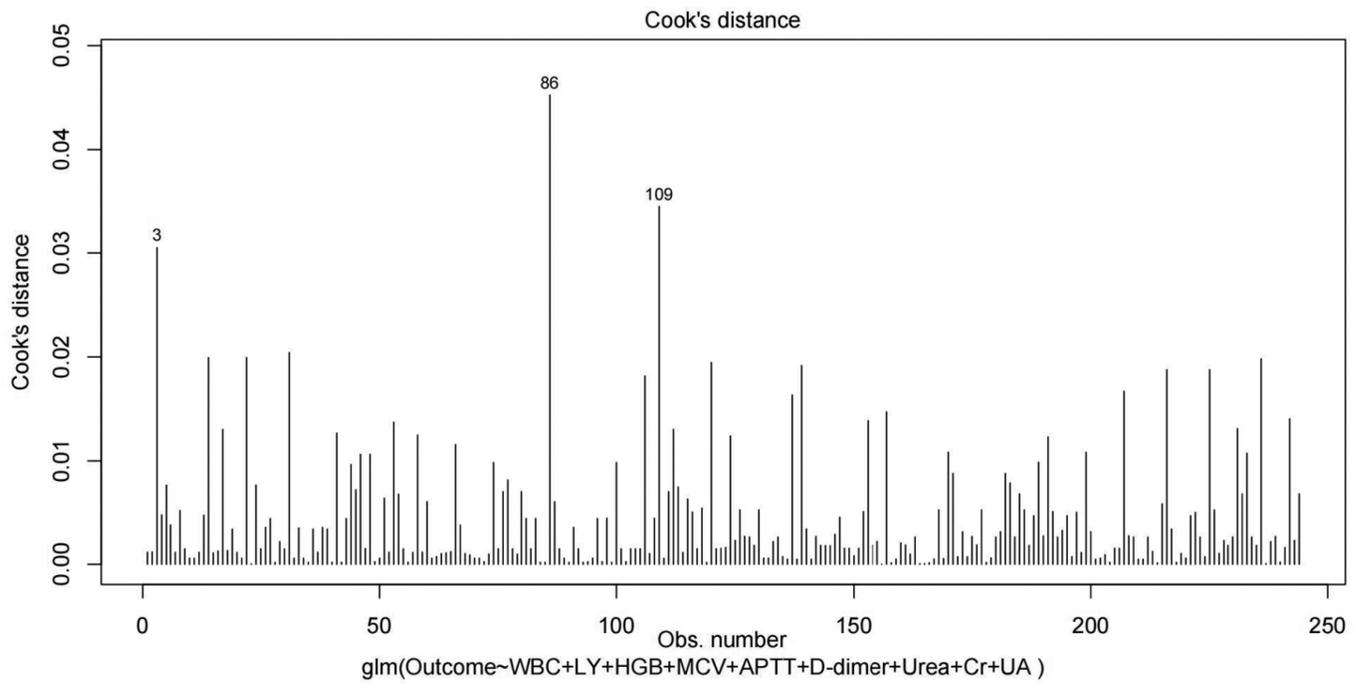
WBC: white blood cell count; LY: lymphocyte; HGB: hemoglobin; MCV: mean corpuscular volume; APTT: activated partial thromboplastin time; Cr: creatinine; UA: uric acid.

Supplementary Figure 1. Missing data of all candidate variables.



MO: monocyte; BA: basophil; HCT: hematocrit; PLT: platelet count; PDW: platelet volume distribution width; APTT: activated partial thromboplastin time; AST: aspartate aminotransferase; LP(a): lipoprotein(a); UA: uric acid.

Supplementary Figure 2. Influential case of nomogram.



WBC: white blood cell count; LY: lymphocyte; HGB: hemoglobin; MCV: mean corpuscular volume; APTT: activated partial thromboplastin time; Cr: creatinine; UA: uric acid.