



# OPEN Predictive model of risk factors for 28-day mortality in patients with sepsis or sepsis-associated delirium based on the MIMIC-IV database

Li Zhang<sup>1,2,3</sup>, Xiang Li<sup>4</sup>, Jinyong Huang<sup>5</sup>, Yanjie Yang<sup>1,2</sup>, Hu Peng<sup>1,2</sup>, Ling Yang<sup>1,2</sup> & Xiangyou Yu<sup>4</sup>✉

Research on the severity and prognosis of sepsis with or without progressive delirium is relatively insufficient. We constructed a prediction model of the risk factors for 28-day mortality in patients who developed sepsis or sepsis-associated delirium. The modeling group of patients diagnosed with Sepsis-3 and patients with progressive delirium of related indicators were selected from the MIMIC-IV database. Relevant independent risk factors were determined and integrated into the prediction model. Receiver operating characteristic (ROC) curves and the Hosmer–Lemeshow (HL) test were used to evaluate the prediction accuracy and goodness-of-fit of the model. Relevant indicators of patients with sepsis or progressive delirium admitted to the intensive care unit (ICU) of a 3A hospital in Xinjiang were collected and included in the verification group for comparative analysis and clinical validation of the prediction model. The total length of stay in the ICU, hemoglobin levels, albumin levels, activated partial thrombin time, and total bilirubin level were the five independent risk factors in constructing a prediction model. The area under the ROC curve of the predictive model (0.904) and the HL test result ( $\chi^2 = 8.518$ ) indicate a good fit. This model is valuable for clinical diagnosis and treatment and auxiliary clinical decision-making.

**Keywords** Sepsis with delirium, Risk factors, Prediction model, MIMIC- IV database

Sepsis is a life-threatening organ dysfunction caused by the dysregulated response of an organism to infection<sup>1</sup>. Sepsis and septic shock impact millions of people worldwide yearly, killing between one in three and one in six of those it affects. Sepsis-associated delirium (SAD) is a cerebral dysfunction that occurs during sepsis. SAD is among the most serious complications of sepsis and is an independent risk factor for death in patients with sepsis. Its pathophysiological mechanism is complex, with no standard, unified diagnosis or treatment method at present<sup>2</sup>. Early and accurate identification of high-risk patients and timely adoption of effective therapeutic measures are of great importance to improve patient prognosis. Currently, the acute physiology and chronic health evaluation II (APACHE II) and sequential organ failure assessment (SOFA) scores are most often applied in clinical practice to assess the condition of patients with sepsis and to evaluate the physiological function of organs, respectively<sup>3</sup>. The most recent clinical practice guidelines on delirium from the Society of Critical Care Medicine recommend regular assessment of delirium using validated tools, such as the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) or the Intensive Care Delirium Screening Checklist<sup>2</sup>. However, early assessment of SAD progress and prediction model sensitivity in patients with sepsis require further research. Risk prediction models have been widely applied in medical research to predict future onset. Although not an integral part of treatment, risk prediction is essential to improve clinical decision-making<sup>4</sup>. Moreover, risk prediction models are used to predict the likelihood or risk of the onset of a particular outcome

<sup>1</sup>Xinjiang Medical University, Urumqi 830000, China. <sup>2</sup>School of Nursing, Xinjiang Medical University, Urumqi 830000, China. <sup>3</sup>Department of Nursing, the First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China. <sup>4</sup>Centre for Critical Care Medicine, The First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, Xinjiang, China. <sup>5</sup>Department of Traumatology and Orthopaedics, the First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China. ✉email: yu2796@163.com

or event in an individual and are often established through statistical methods based on disease etiology or related risk factors. The quantitative relationships between risk factors in a model are analyzed to predict the probability of a particular outcome event as well as the risk level of an event in a population with certain specific risk factors<sup>5,6</sup>. Therefore, the rapid identification of risk factors for sepsis onset and its progression to SAD may facilitate early medical prevention and interventions, which are important for improving patient prognosis and clinical outcomes. The Multiparameter Intelligent Monitoring in Intensive Care-II (MIMIC-II) database is the largest public clinical database<sup>7</sup> in the field of emergency and critical care medicine. It was established in a joint effort of emergency medicine physicians, intensive care unit (ICU) physicians, and computer science experts from the Massachusetts Institute of Technology, Massachusetts General Hospital, and Beth Israel Deaconess Medical Center. In September 2016, the MIMIC-II database was upgraded to the Medical Information Mart for Intensive Care-III (MIMIC-III), and the latest version is the Medical Information Mart for Intensive Care-IV (MIMIC-IV)<sup>8</sup>.

In recent years, the MIMIC databases and various machine learning methods have gradually been applied in medical research. Compared with commonly used clinical indicators, predictive models established based on different clinical requirements and data characteristics that make full use of these repositories can better predict the onset of adverse outcomes<sup>9–11</sup>. Rahmatinejad et al. used disease scores combined with clinical variables to establish and effectively verify prediction models, which played an important role in early clinical prediction<sup>12</sup>.

The present study was designed to screen adult patients who met the diagnostic criteria for Sepsis 3 through the MIMIC-IV database, based on the third International consensus definition of sepsis and septic shock, or SAD, for key features of sepsis onset or progression to SAD. Based on these characteristics, a prediction model of risk factors for the onset or development of sepsis into SAD was constructed to facilitate early identification and intervention and improve the prognosis of patients.

## Methods

### Data source

Patient-related information was screened from the MIMIC-IV database, including demographic information, vital signs, laboratory tests, treatment, diagnosis, assessment, and prognosis. The data were verified by several studies to be of high quality. Access to the database was obtained for the present study (database authorization certificate ID: 57,771,514), and the database is available to all investigators who comply with the requirements for its use (<https://mimic.physionet.org>)<sup>13</sup>. Data were extracted from the database for all adult patients (aged  $\geq 18$  years) admitted to the ICU who met the inclusion criteria. Patients with significant missing information were excluded.

### Study participants

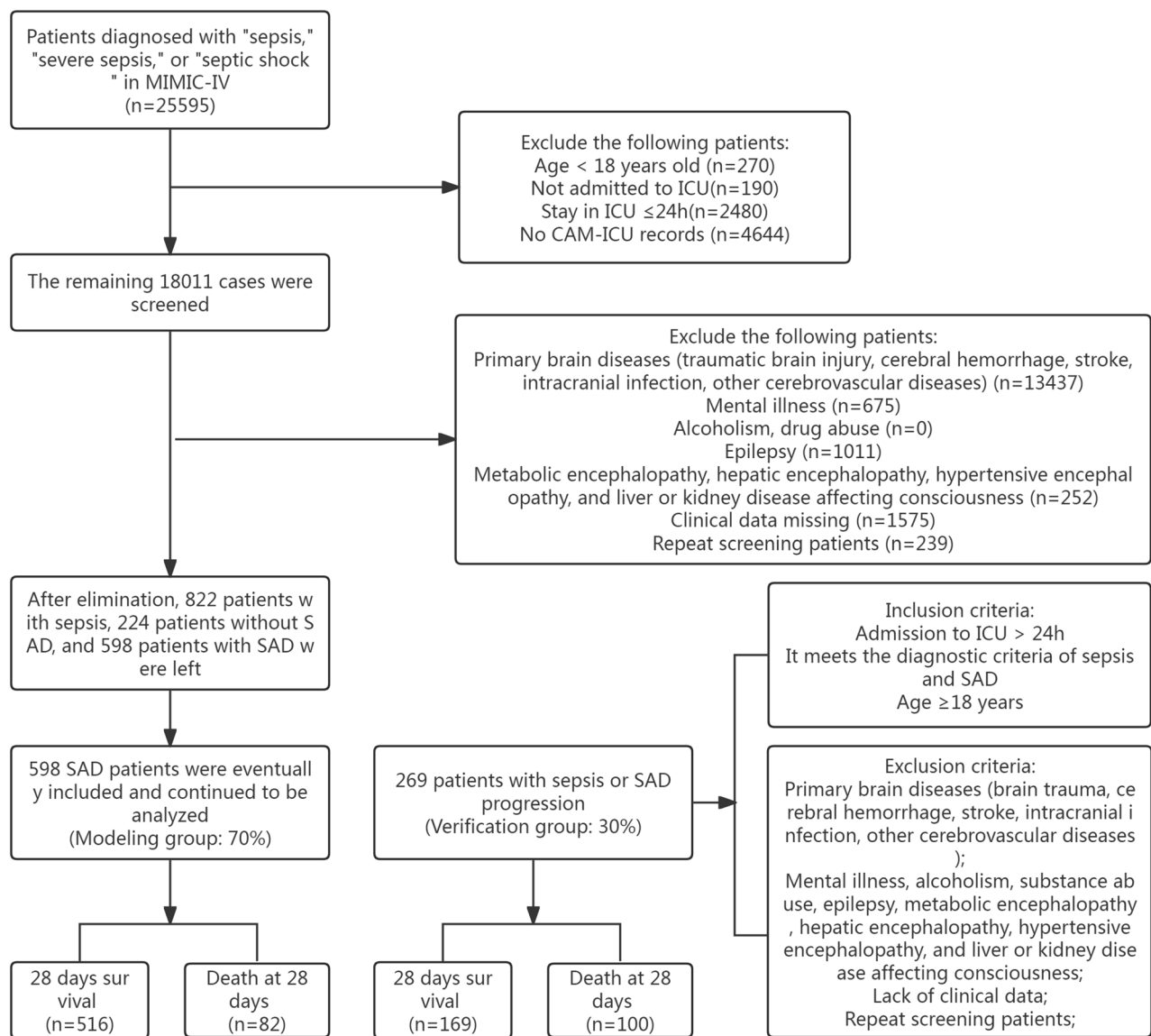
In this study, Structured Query Language (SQL) via Navicat Premium (version 15.0) was used to extract meaningful baseline clinical data based on the “subject\_id” of each patient in the ICU. If a patient’s data was measured multiple times, only the first measurement was used<sup>14</sup>. Patient information was extracted from the MIMIC-IV database according to the following inclusion criteria: (1) first admission to the ICU; (2) age of  $\geq 18$  years; (3) diagnosis of sepsis; (4) length of stay in the ICU of  $\geq 24$  h; and (5) completion of delirium assessment. The diagnosis of sepsis was based on the definition of Sepsis-3, including patients with documented or suspected infection, SOFA scores of  $\geq 2$ , and microbial culture results that clearly indicate infection<sup>15,16</sup>. The CAM-ICU was used to screen and identify delirium. Patients with sepsis who had a positive delirium assessment were defined as having SAD, and those with delirium associated with drug or alcohol abuse, neurological disorders, mental disorders, and other conditions were excluded<sup>17,18</sup>. Finally, the modeling group in this study included 598 patients (358 men and 240 women). The validation group included 269 patients with sepsis or SAD who were admitted to the Department of Intensive Care Medicine of a grade A tertiary hospital in Xinjiang, China, between February 2018 and December 2022. This study complied with the medical ethics standards of the Declaration of Helsinki and was approved by the Ethics Committee of the First Affiliated Hospital of Xinjiang Medical University (ethics approval no.: K202301-43). The patient enrollment process of the study is illustrated in Fig. 1

### Data extraction

According to the key words “sepsis,” “severe sepsis,” and “septic shock,” the original data of patients diagnosed with “sepsis,” “severe sepsis,” or “septic shock” were searched in the MIMIC-IV database, and the patients were further screened according to the inclusion and exclusion criteria. The detailed process of patient screening is shown in Fig. 1. The requirements for multivariate logistic regression analysis indicate a sample size of at least 10 times the number of variables<sup>19</sup>. In the present study, 598 patients who met the criteria were screened from the database. According to the allocation principle, the sample sizes of the modeling and verification groups accounted for 70% and 30% of the total sample size, respectively. The sample size of the verification group was calculated to be 269 patients for external verification.

### Data collection

This study included the clinical data of all patients in the MIMIC-IV database who met the diagnostic criteria. These data included: (1) general demographic information (sex, age, etc.); (2) underlying comorbidities (myocardial infarction, heart failure, cerebrovascular disease, diabetes mellitus with chronic comorbidities, malignant neoplasm, and severe liver disease); (3) laboratory results within 24 h of admission to the ICU, including routine blood indicators (white blood cell count, neutrophil count, hemoglobin, platelet count, and albumin), coagulation indicators (international normalized ratio, prothrombin time, and activated partial thromboplastin time [aPTT]), and blood gas analysis (pH value, partial pressure of oxygen, partial pressure of



**Figure 1.** Flow chart of patient enrollment.

carbon dioxide, blood potassium, blood sodium, and lactate); (4) biochemistry indices (creatinine, total bilirubin, alanine transaminase, alkaline phosphatase, aspartate transaminase, etc.); (5) prognostic scores assessed within 24 h of admission to the ICU that indicate patient prognosis (acute physiology scores, APACHE II scores, SOFA scores, and total length of stay in the ICU).

### Data extraction and processing

Access to the data was granted after registering with the database website, obtaining a certificate following ethics training and an examination, signing a data usage agreement outlining appropriate data usage and safety standards, and obtaining permission to conduct the corresponding research. No further ethical approval was required<sup>20</sup>. Patients in the MIMIC-IV database between 2008 and 2019 were screened according to the inclusion criteria. SQL and PostgreSQL were used to extract basic patient data, underlying comorbidities, laboratory indices, and prognostic outcomes. The data were further filtered and processed using SQL based on the exclusion and rejection criteria (missing data ratio of > 30%). As the proportion of missing data in all extracted data was < 5%, these data were retained and subjected to multiple imputations<sup>21</sup>, based on which the data were further analyzed. Clinical data collection for this study was reviewed and approved by the institutional ethics committee (K202301-43).

### Model construction

Uni- and multivariate logistic regression analyses were performed on the retrieved clinical data to screen for factors influencing in-hospital mortality. Based on the results of multivariate logistic regression, regression models for predicting mortality outcomes were established using the influencing factors as the main predictive

indices and subject survival as the predictive outcome. The regression equation was visualized as a nomograph using the “rms” package in R, version 4.3.1. The values corresponding to each factor in the nomograph were identified, and the total score was calculated by summing all the values. The total score corresponded to the probability of 28-day mortality among patients with sepsis or SAD, in which a higher total score was associated with a higher probability of death<sup>22</sup>.

### Statistical analysis

Data were extracted using SQL and PostgreSQL. Data entry and analyses were performed using Excel 2019 and IBM SPSS Statistics for Windows, version 26.0, respectively. R was used to construct the models and assess model performance. The Shapiro–Wilk test was performed to test for normality, and measurement data conforming to a normal distribution are expressed as  $\bar{x} \pm s$ . Comparisons were made using the two-sample independent t-test. Measures with skewed distributions are expressed as medians (quartiles), and intergroup comparisons were made using the rank-sum test. Count data are expressed as the percentages of cases (%), and intergroup comparisons were made using the chi-square test.  $P < 0.05$  indicated statistical significance. Statistically significant variables were included in the multivariate analyses. Binary logistic regression analyses were used to screen independent risk factors and construct predictive models. Nomograms were made to visualize the predictive models. The predictive accuracy of the model was assessed by plotting the receiver operating characteristic (ROC) and calibration curves. The goodness-of-fit of the predictive model was determined using the Hosmer–Lemeshow (HL) test. Data from the validation group were substituted. ROC and calibration curves were plotted to determine the predictive accuracy of the model.

## Results

### Comparison of baseline data

Patients in the modeling and validation groups were categorized into survival and mortality groups based on patient outcomes on day 28 after admission. Overall, 598 adult patients (aged  $\geq 18$  years) diagnosed with sepsis combined with SAD were screened in the modeling group, of which 82 were in the mortality group. Nine indicators showed statistically significant differences between the groups ( $P < 0.05$ ), including age, comorbidity of heart failure, acute physiology score, SOFA score, and total length of stay in the ICU. The validation group comprised 269 adult patients (aged  $\geq 18$  years) with sepsis combined with SAD or who progressed to SAD (survival group: 169 patients; mortality group: 100 patients). Table 1 shows the results of the comparisons of data of patients with sepsis and SAD in the modelling and validation groups.

### Analysis of factors associated with sepsis onset or progression to SAD in patients in the modeling group

The risk factors with statistically significant differences in the univariate analyses ( $P < 0.05$ ) screened in the modeling group were further subjected to multivariate logistic regression analysis. The results showed that five indicators (total length of stay in the ICU, hemoglobin, albumin, aPTT, and total bilirubin) were independent risk factors for 28-day mortality in patients with sepsis or SAD ( $P < 0.05$ ) (Table 2).

### Prediction model development and validation

Based on the results of multivariate logistic regression analysis, the five independent risk factors were assigned values from X1 to X5. The following regression equation was then established based on the above results:

$$\text{Logit}(P) = 3.561 + 0.196X1 - 0.24X2 - 1.236X3 - 0.033X4 + 0.061X5$$

where P is the predicted probability of death within 28 days of admission for patients with sepsis complicated with delirium.

The R software package was used to construct a 28-day mortality prediction model for patients with sepsis or SAD (Fig. 2). The ROC curve of the predictive model indicated a sensitivity of 91.5%, a specificity of 76.4%, and an area under the ROC curve (AUC) of 90.4%. The predictive model was then applied to the validation group and showed a model sensitivity of 75%, a specificity of 81.1%, and an AUC of 83.3% (Figs. 3 and 4). The calibration curves for both the modeling and validation groups aligned with the actual curve, indicating that the model had a favorable prediction accuracy (Figs. 5 and 6). The results of the HL goodness-of-fit test showed  $\chi^2 = 8.518$  and  $P = 0.385$  ( $> 0.05$ ) for the modeling group, suggesting that the predictive model had a high goodness-of-fit in both datasets.

## Discussion

SAD is a systemic organ and tissue injury caused by inflammation. Ely et al. found a high incidence of sepsis-associated encephalopathy in the ICU, with approximately 20–50% of patients with sepsis having delirium or disturbance of consciousness<sup>23</sup>. However, different cohort studies have also confirmed that sepsis-associated encephalopathy is closely related to increased short-term mortality in patients with sepsis<sup>24</sup>. Existing clinical scores, such as the APACHE-II, Simplified Acute Physiology Score-II, SOFA, quick SOFA (qSOFA), and systemic inflammatory response syndrome, are widely used to assess the risk degree of critically ill patients and predict poor prognosis. The use of objective and accurate prognostic scoring systems can help distinguish and manage complex patients, improve baseline indicators, and quantify disease severity<sup>25–27</sup>. However, studies on the severity and prognosis of sepsis with or without progressive delirium are relatively insufficient. Therefore, we constructed a prediction model of risk factors of 28-day mortality in sepsis with or without progressive delirium based on the MIMIC-IV database.

Item	Modeling group				Validation group			
	Survival group (516 patients)	Mortality group (82 patients)	Z/ $\chi^2$ value	P-value	Survival group (169 patients)	Mortality group (100 patients)	t/Z/ $\chi^2$ two-value	P-value
Sex								
Male	304 (84.9%)	54 (5.1%)			122 (61.3%)	77 (38.7%)		
Female	212 (88.3%)	28 (11.7%)	1.418	0.234	47 (67.1%)	23 (32.9%)	0.755	0.385
Age (years)	69 (56, 81)	60 (44, 70)	-4.491	<0.001	54.71 ± 14.17	58.81 ± 16.24	-2.170	0.031
Complication								
Myocardial infarction	98 (89.9%)	11 (10.1%)	1.447	0.224	9 (64.3%)	4 (35.7%)	0.240	0.624
Heart failure	162 (91.5%)	15 (8.5%)	5.829	0.016	11 (50%)	11 (50%)	1.687	0.194
Cerebrovascular disease	75 (83.3%)	15 (16.7%)	0.781	0.377	14 (66.7%)	7 (33.3%)	0.144	0.704
Diabetes with chronic comorbidities	50 (84.7%)	9 (15.3%)	0.132	0.717	19 (42.2%)	26 (57.8%)	9.822	0.002
Malignant tumor	78 (84.8%)	14 (15.2%)	0.208	0.648	14 (43.8%)	18 (56.2%)	5.659	0.017
Severe liver disease	31 (77.5%)	9 (22.5%)	2.798	0.094	21 (65.5%)	11 (34.4%)	0.122	0.727
Scoring system								
Acute physiology score (points)	49 (38, 66)	60 (48, 83)	-4.627	<0.001	31 (18, 49)	59 (30, 75)	-6.200	<0.001
SAPS II (points)	37 (29, 47)	39 (28, 48)	-0.158	0.875	28.69 ± 17.01	37.39 ± 18.23	-3.947	<0.001
SOFA score (points)	6 (4, 8)	7 (5, 10)	-3.869	<0.001	11 (8, 13)	13 (9, 15)	-2.505	0.012
Co-morbidity index (points)	6 (4, 8)	6 (4, 8)	-1.199	0.231	1 (1, 2)	2 (1, 3)	-1.845	0.065
Laboratory index								
pH value	7.37 (7.33, 7.42)	7.36 (7.33, 7.42)	-0.896	0.370	7.40 (7.33, 7.45)	7.40 (7.34, 7.46)	-0.315	0.753
Total ICU stay (days)	3 (2, 6)	15 (5, 27)	-9.128	<0.001	4 (2, 7)	15 (10, 20)	-3.717	<0.001
Lactic acid (mmol/L)	2.17 (1.60, 2.17)	2.17 (1.50, 2.34)	-0.041	0.967	1.70 (1.20, 2.55)	1.90 (1.30, 3.40)	-1.248	0.212
PaO <sub>2</sub> (mmHg)	170 (129, 177)	170 (121, 193)	-0.553	0.580	112 (84.5, 143.5)	101 (78, 144)	-1.284	0.199
PaCO <sub>2</sub> (mmHg)	42 (39, 42)	42 (37, 44)	-0.464	0.643	31 (27.5, 37.9)	32 (25, 36.9)	-0.115	0.908
Creatinine (μmol/L)	24.8 (23.5, 25.0)	24.8 (22.6, 25.9)	-0.119	0.905	93 (63, 199)	101 (78, 253)	-1.198	0.231
Sodium (mmol/L)	138.5 (136.0, 141.0)	138.8 (136.5, 141.4)	-0.150	0.880	136 (131, 142)	138 (132, 143)	-0.896	0.370
Potassium (mmol/L)	4.15 (3.85, 4.55)	4.08 (3.66, 4.40)	-1.688	0.091	3.8 (3.40, 4.20)	4 (3.58, 4.40)	-2.441	0.015
Platelets (10 <sup>9</sup> /L)	190 (133, 254)	187 (110, 239)	-1.227	0.220	168 (95, 258)	176 (74, 242)	-0.493	0.622
WBC (10 <sup>9</sup> /L)	11.0 (8.0, 14.6)	12.4 (8.6, 17.2)	-1.656	0.098	12 (7.9, 16.8)	12.48 (8.7, 18.4)	-1.235	0.217
Albumin (g/dL)	3.18 (3.18, 3.30)	3.18 (2.70, 3.18)	-4.030	<0.001	2.89 (2.56, 3.55)	2.93 (2.48, 3.37)	-0.377	0.706
Neutrophils (10 <sup>9</sup> /L)	10.3 (6.8, 11.5)	10.3 (8.1, 10.5)	-0.782	0.434	9.5 (5.3, 14.6)	10.6 (5.8, 10.6)	-1.596	0.111
INR	1.30 (1.10, 1.48)	1.30 (1.11, 1.50)	-0.186	0.852	1.17 (1.03, 1.36)	1.24 (1.08, 1.57)	-1.899	0.058
PT (s)	14.1 (12.7, 16.2)	14.4 (12.7, 16.7)	-0.013	0.990	13.5 (12, 15.9)	15 (12.6, 18.4)	-2.644	0.008
aPTT (s)	35 (29, 46)	29 (26, 37)	-3.894	<0.001	32 (29, 37)	34 (30, 38)	-1.425	0.154
ALT (U/L)	76 (22, 106)	57 (21, 106)	-0.627	0.531	32 (17.6, 101)	33.4 (20, 59)	-0.367	0.714
Alkaline phosphatase (U/L)	104 (73, 104)	104 (72, 105)	-0.491	0.623	61.4 (53.4, 87.7)	65.1 (54.4, 93.6)	-1.261	0.207
AST (U/L)	90 (32, 182)	131 (30, 182)	-0.647	0.518	52.5 (28.4, 152)	42.8 (28, 115.3)	-0.463	0.643
Total bilirubin (mg/dL)	1.52 (0.50, 1.66)	1.66 (0.70, 1.66)	-2.102	0.036	1.13 (0.82, 2.09)	1.6 (0.96, 3.03)	-0.022	0.008
Hemoglobin (g/dL)	9.60 (8.20, 11.50)	8.90 (7.70, 10.63)	-2.443	0.015	10.8 (8.2, 13.3)	9.7 (7.6, 13.3)	-2.024	0.252

**Table 1.** Comparison of general data, related scores, and laboratory indexes of patients with sepsis complicated with SAD between the modeling and verification groups. Abbreviations: ALT, alanine transaminase; aPTT, activated partial thromboplastin time; AST, aspartate transaminase; ICU, intensive care unit; INR, international normalized ratio; SAPS, simplified acute physiology score; SOFA, sequential organ failure assessment; WBC, white blood cell.

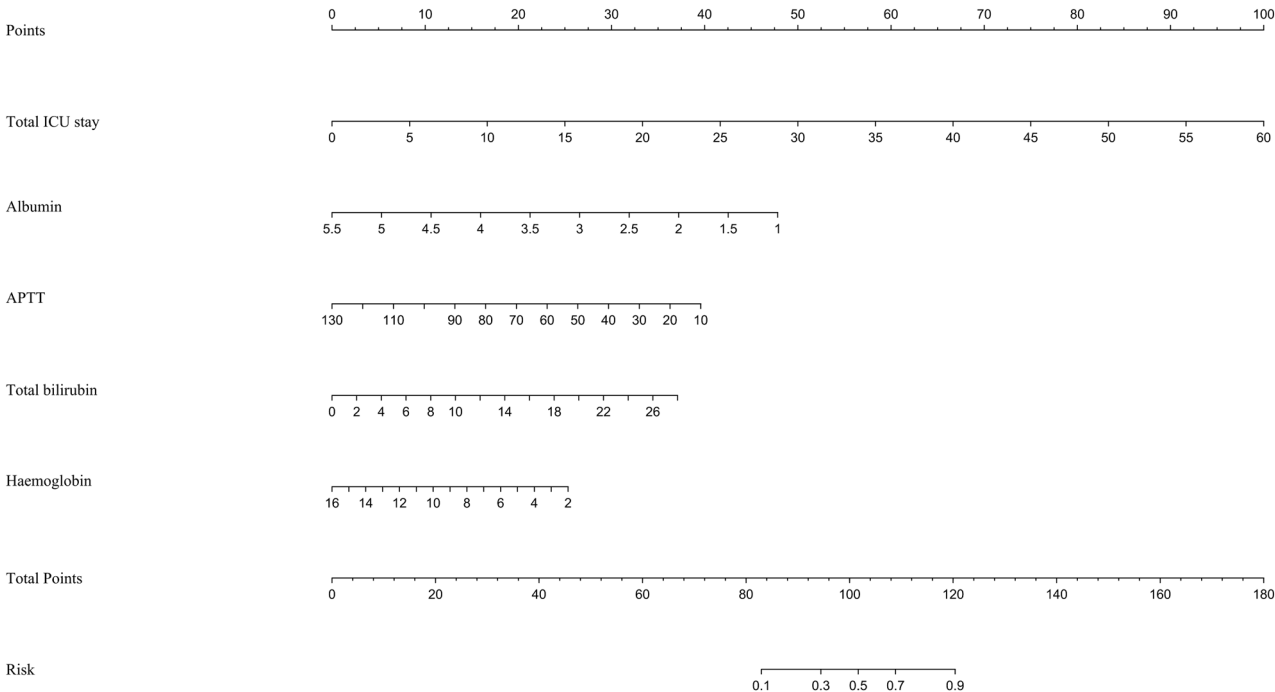
### Independent risk factors in the modeling group to rapidly screen patients for sepsis onset or progression to SAD

The results of this study identified five indicators (total length of stay in the ICU, hemoglobin level, albumin level, aPTT, and total bilirubin level) as independent risk factors for 28-day mortality in patients with sepsis or those who progressed to SAD ( $P < 0.05$ ). Multivariate logistic regression showed that the total length of stay in the ICU was positively correlated with the 28-day survival rate of patients with sepsis or advanced delirium (odds ratio, 1.217; 95% confidence interval, 1.165–1.271). The most direct harm caused by the reduction of hemoglobin levels is the weakening of oxygen transport capacity and hypoxia of tissue cells. This study demonstrated that the



Impact factors	$\beta$	SE	OR (95% CI)	P-value
Total ICU stay (days)	0.196	0.022	1.217 (1.165–1.271)	< 0.001
Hemoglobin (g/L)	-0.240	0.072	0.787 (0.683–0.907)	0.001
Albumin (g/L)	-1.236	0.325	0.291 (0.154–0.549)	< 0.001
aPTT (s)	-0.033	0.013	0.967 (0.943–0.992)	0.011
Total bilirubin ( $\mu$ mol/L)	0.061	0.026	1.063 (1.010–1.118)	0.018

**Table 2.** Results of multivariate logistic regression in the modeling group. Abbreviations: aPTT, activated partial thromboplastin time; CI, confidence interval; OR, odds ratio.

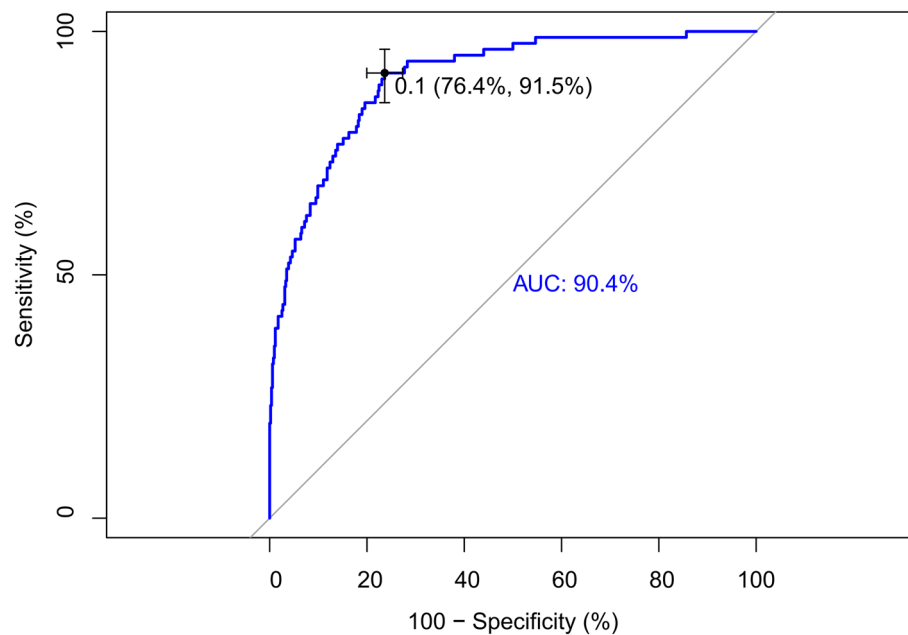


**Figure 2.** Nomogram of the risk prediction model for 28-day mortality among patients in the modelling group with sepsis or who progressed to SAD.

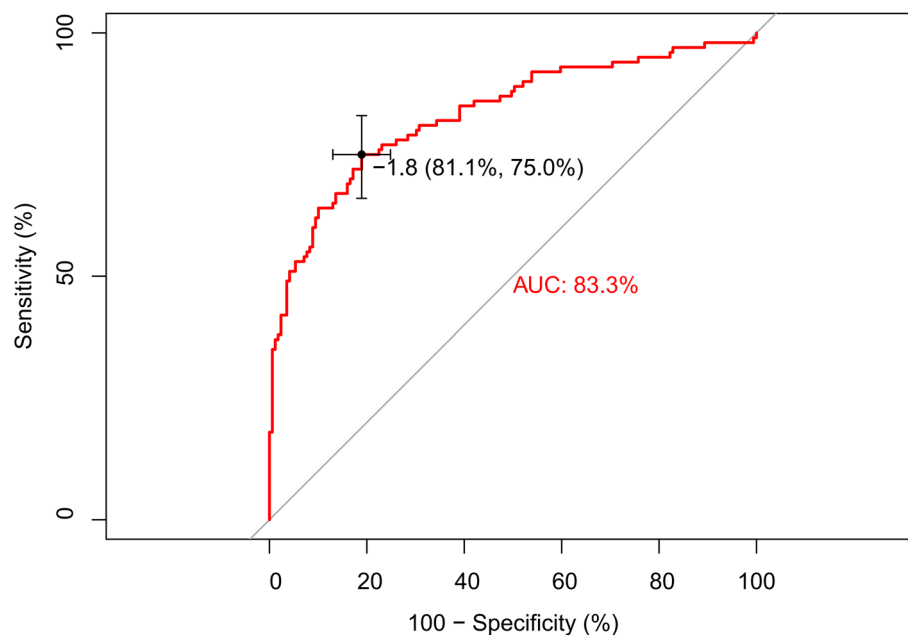
hemoglobin level of patients in the survival group was significantly higher than that in the mortality group, and the histogram indicated that the hemoglobin level was a protective factor for prognosis. These findings underscore the importance of maintaining appropriate hemoglobin levels in patients with sepsis or those who progressed to SAD. In addition, decreased hemoglobin and albumin levels are clear features of malnutrition, indicating an increased nutritional risk, which is highly likely to affect sepsis prognosis<sup>28</sup>. In recent years, several studies have demonstrated that inflammation and coagulation disorder in patients with sepsis promote and interact with each other, thus amplifying the cascade reaction. Coagulation disorder in sepsis is also a key factor that accelerates the progression of the disease and affects patient prognosis<sup>29</sup>, which is consistent with the results of the present study. Moreover, serum bilirubin levels were associated with mortality in patients with severe sepsis; in particular, elevated serum bilirubin levels within 72 h of admission were associated with increased mortality in patients with sepsis<sup>30,31</sup>. Bilirubin impairs the bactericidal properties of neutrophils, inhibits inducible nitric oxide synthase, and inhibits platelet aggregation by suppressing platelet activation<sup>32</sup>. These findings are consistent with the results of the present study, which indicate that total bilirubin was an independent risk factor for the prognosis of patients with sepsis. Furthermore, Xiao et al.<sup>33</sup> demonstrated that the combination of bilirubin, creatinine, and qSOFA score had a high sensitivity and specificity in predicting sepsis progression and prognosis.

**Construction of a risk prediction model for patients with sepsis and those who progress to SAD based on the MIMIC-IV database**

We constructed a model to predict 28-day mortality in patients with sepsis or those who progressed to SAD, employing the following formula:  $\text{Logit}(P) = 3.561 + 0.196X_1 - 0.24X_2 - 1.236X_3 - 0.033X_4 + 0.061X_5$ . The left side of the nomogram diagram displays the variables included in the prediction model. The corresponding line segment of each variable is marked with a scale, representing the value range of the variable, while the length of the line segment reflects the contribution of the factor to the outcome event. The total point represents the total

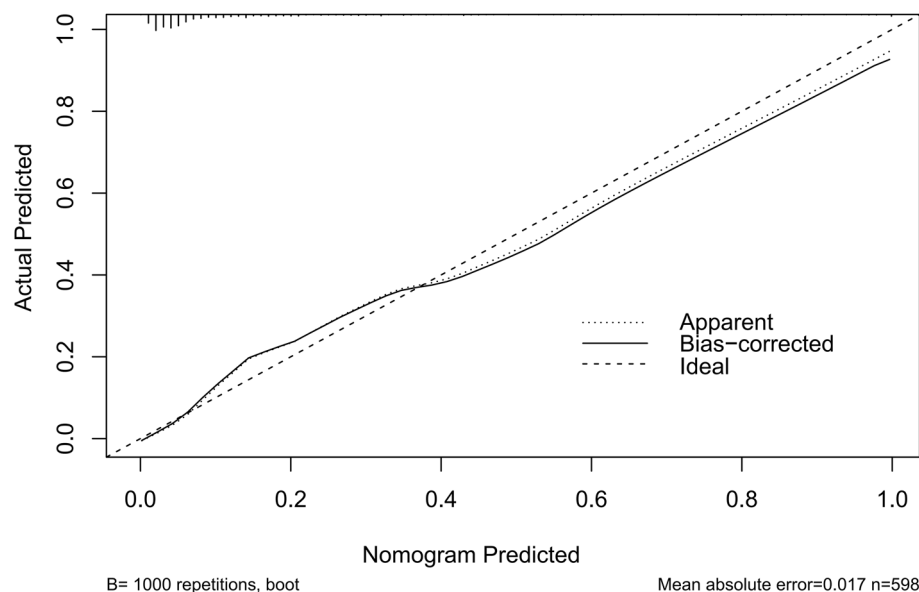


**Figure 3.** Receiver operating characteristic (ROC) curve for predicting risk factors for 28-day mortality among patients in the modelling group.

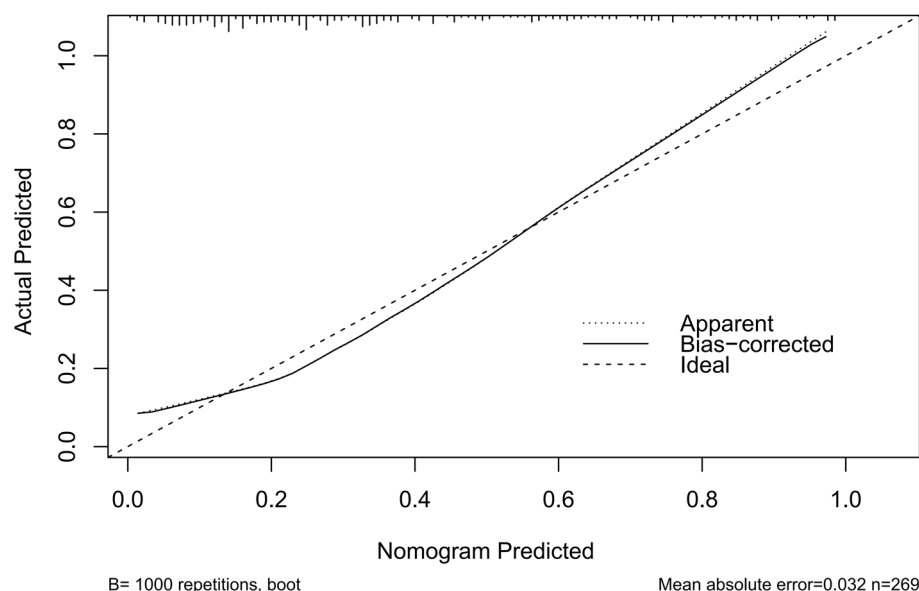


**Figure 4.** Receiver operating characteristic (ROC) curve for predicting the risk factors for 28-day mortality among patients in the validation group.

score of the single scores corresponding to the values of all variables. The predicted probability of death within 28 days in patients with sepsis or those who progressed to SAD can be evaluated according to the total point of the column chart. The model was visualized as a nomogram to help clinicians assess the prognosis of the patients and provide targeted treatment at an early stage, offering valuable references for clinical practice. Regarding model performance evaluation, the results of the HL test showed a likelihood ratio chi-square value of 8.518 ( $df=8$ ,  $P>0.05$ ), indicating that the prediction model was well-fitted<sup>33,34</sup>. The differentiation of prediction models was measured by the AUC. An AUC of  $<0.7$  indicated poor evaluation efficiency of the model, an AUC of  $0.7-0.9$  indicated medium evaluation efficiency, and an AUC of  $>0.9$  indicated good evaluation efficiency<sup>35,36</sup>. The AUC values of the predictive model in the modeling and validation cohorts were 90.4% and 83.3%, respectively. These



**Figure 5.** Receiver operating characteristic (ROC) validation curve for predicting risk factors 28-day mortality among patients in the modelling group.



**Figure 6.** Receiver operating characteristic (ROC) validation curve for predicting the risk factors 28-day mortality among patients in the validation group.

values indicate that the constructed model has a good degree of differentiation. The calibration curve reflects the consistency of the model. In the cartesian coordinate system in Figs. 3 and 5, the calibration curve of the model coincides with the height of the 45° straight line, indicating that the 28-day mortality risk predicted by the model was in good agreement with the actual mortality rate.

#### **Risk prediction model for patients with sepsis or those who progressed to SAD may serve as a valuable guide in clinical practice**

In this study, based on the existing challenges in clinical practice and with the help of a large clinical database, a prognostic model for patients with sepsis and those with occurring or progressing delirium was constructed for rapid screening and identification of sensitive early warning indices. Moreover, the relationship between prognosis and various blood pressure indices in the early stage of sepsis was explored. Based on five clinical variables for the occurrence of sepsis or SAD, the probability of mortality within 28 days in this patient group can be predicted. This tool may help clinicians evaluate disease severity, guide the stratification of disease risk,



quickly adjust prevention and treatment plans, and improve the patient prognosis. In addition, the reliability and generalizability of the model were improved by scientifically screening the risk factors of 28-day mortality in patients with sepsis and applying clinical data for verification. The results of the present study may be used to identify risk factors and poor prognosis of high-risk groups early in clinical practice, guide personalized treatment and prevention programs to improve risk factor indicators, and reduce risk scores, which are important for early warning and improving the clinical outcomes of patients.

## Strengths and limitations

In summary, a model was constructed to rapidly predict mortality in patients with sepsis or those who progressed to SAD within 28 days after ICU admission based on the MIMIC-IV database. Moreover, evaluation tools, such as the SOFA and APACHE-II scores, may be combined for a greater clinical reference that can assist in clinical decision-making, adjusting therapeutic plans, and preventing the onset of adverse outcomes<sup>37</sup>. However, some limitations must be addressed. The data retrieved from the MIMIC-IV database may be subject to variation in disease severity, treatment strategies, and risk factors. The extraction of some validation indicators was incomplete, the data of the validation group were mostly from retrospective studies, and the research results were for reference and application in clinical practice. Additionally, in the process of external verification, verification based on a limited number of medical records in a single center may have certain limitations. In the future, multicenter studies with increased sample sizes are required to further verify the validity of the model.

## Data availability

The MIMIC-IV database is available from <https://mimic.physionet.org>. The raw data were extracted using structure query language (SQL) and PostgreSQL, as well as using Excel 2019 and IBM SPSS Statistics for Windows (version 26.0) for data entry and analysis, respectively.

Received: 28 October 2023; Accepted: 2 August 2024

Published online: 13 August 2024

## References

1. Singer, M. *et al.* The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* **315**, 801–810 (2016).
2. Atterton, B., Paulino, M. C., Povoia, P. & Martin-Loeches, I. Sepsis associated delirium. *Medicina (Kaunas)* **56**, 240 (2020).
3. Bahtouee, M., Eghbali, S. S., Maleki, N., Rastgou, V. & Motamed, N. Acute Physiology and Chronic Health Evaluation II score for the assessment of mortality prediction in the intensive care unit: a single-centre study from Iran. *Nurs. Crit. Care* **24**, 375–380 (2019).
4. Rahmatinejad, Z. *et al.* Comparing in-hospital mortality prediction by senior emergency resident's judgment and prognostic models in the emergency department. *BioMed Res. Int.* **2023**, 6042762 (2023).
5. Zhang, R., Zheng, L. & Pan, G. Application and establishment of disease incidence risk prediction models. *Chin. J. Health Stat.* **32**, 724–726 (2015).
6. Chen, L. *Construction and Application of Risk Prediction Model for Stroke-Associated Pneumonia in Elderly Stroke Patients [D]* (Changchun Univ. of Traditional Chinese Medicine, 2022).
7. Zhao, J. *et al.* Construction of an nomogram for predicting the prognosis of patients with sepsis-associated acute kidney injury. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue.* **35**, 1255–1261 (2023).
8. Zhou, M. *et al.* Analysis of Characteristics of the MIMIC-III Database and Implications for Constructing a Shared Traditional Chinese Medicine Dataset. *Chin. J. Libr. Inf. Sci. Tradit. Chin. Med.* **43**, 1–5 (2019).
9. Vellido, A., Ribas, V., Morales, C., Ruiz Sanmartín, A. & Ruiz Rodríguez, J. C. Machine learning in critical care: state-of-the-art and a sepsis case study. *Biomed. Eng. OnLine* **17**(S1), 135 (2018).
10. Awad, A., Bader-El-Den, M., McNicholas, J., Briggs, J. & El-Sonbaty, Y. Predicting hospital mortality for intensive care unit patients: time-series analysis. *Health Inform. J.* **26**, 1043–1059 (2020).
11. Ocampo-Quintero, N. *et al.* Enhancing sepsis management through machine learning techniques: a review. *Med. Intensiva (Engl. Ed.)* (2020).
12. Rahmatinejad, Z. *et al.* Internal validation of the predictive performance of models based on three ED and ICU scoring systems to predict in-hospital mortality for intensive care patients referred from the emergency department. *Biomed. Res. Int.* **2022**, 3964063 (2022).
13. Tang, J. *et al.* The relationship between potassium levels and 28-day mortality in sepsis patients: Secondary data analysis using the MIMIC-IV database. *Heliyon*. **10**, e31753 (2024).
14. Huang, X. *et al.* The hemoglobin-to-red cell distribution width ratio to predict all-cause mortality in patients with sepsis-associated encephalopathy in the MIMIC-IV database. *Int. J. Clin. Pract.* **2022**, 7141216 (2022).
15. Johnson, A. E. *et al.* MIMIC-III, a freely accessible critical care database. *Sci. Data* **3**, 160035 (2016).
16. Fernando, S. M., Rochwerf, B. & Seely, A. J. E. Clinical implications of the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *CMAJ* **190**, E1058–E1059 (2018).
17. Seymour, C. W. *et al.* Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* **315**, 762–774. doi (2016). Erratum in: *JAMA* **315**, 2237 (2016).
18. Gao, Y. & Zhang, J. Sample size determination for logistic regression analysis. *J. Evid. Based Med.* **18**, 122–124 (2018).
19. Qi, S. & Zhou, F. A review of screening methods for sepsis cases from electronic medical record databases. *Acad. J. Chin. PLA Med. Sch.* **41**, 918–929 (2020).
20. Jia, D. *Development of a 28-Day Mortality Risk Prediction Model for Patients with Sepsis-Induced Coagulopathy Based on the MIMIC-III Database [D]* (Anhui Medical Univ, 2022).
21. Li, Y. M. *et al.* Simulation study on missing data imputation methods for longitudinal data in cohort studies. *Chin. J. Epidemiol.* **42**, 1889–1894 (2021).
22. Wang, Z. *et al.* Construction of a 28-day Mortality Prediction Model for Patients with Pulmonary Infection Complicated by Sepsis. *Shandong Med. J.* **63**, 37–43 (2023).
23. Ely, E. W. *et al.* Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. *JAMA* **291**, 1753–1762 (2004).
24. Sonnevile, R. *et al.* Potentially modifiable factors contributing to sepsis-associated encephalopathy. *Intensive Care Med.* **43**, 1075–1084 (2017).

25. Rahmatinejad, Z. *et al.* Prognostic utilization of models based on the APACHE II, APACHE IV, and SAPS II scores for predicting in-hospital mortality in emergency department. *Am. J. Emerg. Med.* **38**, 1841–1846 (2020).
26. Rahmatinejad, Z. *et al.* Predictive performance of the SOFA and mSOFA scoring systems for predicting in-hospital mortality in the emergency department. *Am. J. Emerg. Med.* **37**, 1237–1241 (2019).
27. Rahmatinejad, Z. *et al.* Comparison of six scoring systems for predicting in-hospital mortality among patients with SARS-COV2 presenting to the emergency department. *Indian J. Crit. Care Med.* **27**, 416–425 (2023).
28. Chen, R., Zhou, X., Rui, Q. & Wang, X. Combined predictive value of the risk factors influencing the short-term prognosis of sepsis. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* **32** (2020).
29. Ishikawa, M. *et al.* Neutropenic enterocolitis-induced sepsis and disseminated intravascular coagulation after chemotherapy: a case report. *BMC Womens Health* **21**, 187 (2021).
30. Snow, G. L. *et al.* Comparative evaluation of the clinical laboratory-based Intermountain risk score with the Charlson and Elixhauser comorbidity indices for mortality prediction. *PLOS ONE* **15**, e0233495 (2020).
31. Chou, H. C., Huang, C. T. & Sheng, W. H. Differential roles of comorbidity burden and functional status in elderly and non-elderly patients with infections in general wards. *J. Formos. Med. Assoc.* **119**, 821–828 (2020).
32. Yang, Z. X., Lv, X. L. & Yan, J. Serum total bilirubin level is associated with hospital mortality rate in adult critically ill patients: a retrospective study. *Front. Med. (Lausanne)* **8**, 697027 (2021).
33. Patel, J. J. *et al.* The association of serum bilirubin levels on the outcomes of severe sepsis. *J. Intensive Care Med.* **30**, 23–29 (2015).
34. Chen, J. *Establishment and Validation of a Delirium Risk Prediction Model for Neurosurgical ICU Patients [D]* (Nanjing Univ, 2020).
35. Fan, H. *et al.* Development and validation of a dynamic delirium prediction rule in patients admitted to the intensive care units (DYNAMIC-ICU): a prospective cohort study. *Int. J. Nurs. Stud.* **93**, 64–73 (2019).
36. Zhu, X. *Analysis of Risk Factors for Delirium in ICU and Construction of Risk Prediction Model [D]*. People's Liberation Army of China Army Medical University, (2017).
37. Li, Y. *et al.* Research Progress on Risk Prediction Models for Perioperative Hypothermia. *Nurs. Res.* **35**, 3107–3110 (2021).

## Author contributions

L.Z. designed, researched and wrote the manuscript, X.L. retrieved and collated the data, J.H. assisted in translating and collating the literature, Y.Y., H.P. and L.Y. analyzed the statistical data, and X.Y. the supervisor, guided the overall research plan and process.

## Funding

This study was funded by the 2022 Critical Care Medicine Runze Fund of the Wu Jieping Medical Foundation (320.6750.2023–02–3).

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to X.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024