Establishment of a prognostic model for patients with sepsis based on SOFA: a retrospective cohort study

Hui Liu[1],*, Luming Zhang[1,2],*, Fengshuo Xu[2,3],
Shaojin Li[4], Zichen Wang[5], Didi Han[2,3],
Feng Zhang[1], Jun Lyu[2] (iD) and Haiyan Yin[1] (iD)

Abstract

Objective: To construct a nomogram based on the Sequential Organ Failure Assessment (SOFA) that is more accurate in predicting 30-, 60-, and 90-day mortality risk in patients with sepsis.

Methods: Data from patients with sepsis were retrospectively collected from the Medical Information Mart for Intensive Care (MIMIC) database. Included patients were randomly divided into training and validation cohorts. Variables were selected using a backward stepwise selection method with Cox regression, then used to construct a prognostic nomogram. The nomogram was compared with the SOFA model using the concordance index (C-index), area under the time-dependent receiver operating characteristics curve (AUC), net reclassification improvement (NRI), integrated discrimination improvement (IDI), calibration plotting, and decision-curve analysis (DCA).

Results: A total of 5240 patients were included in the study. Patient's age, SOFA score, metastatic cancer, SpO_2, lactate, body temperature, albumin, and red blood cell distribution width were included in the nomogram. The C-index, AUC, NRI, IDI, and DCA of the nomogram showed that it performs better than the SOFA alone.

[1]Intensive Care Unit, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong Province, China

[2]Department of Clinical Research, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong Province, China

[3]School of Public Health, Xi'an Jiaotong University Health Science Centre, Xi'an, Shaanxi Province, China

[4]Department of Orthopaedics, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong Province, China

[5]Department of Public Health, University of California, Irvine, CA, USA

*These authors contributed equally to this work.

Corresponding authors:
Haiyan Yin and Jun Lyu, Intensive Care Unit, The First Affiliated Hospital of Jinan University, 613 Huangpu Ave W, Guangzhou, Guangdong 510630, China; Department of Clinical Research, The First Affiliated Hospital of Jinan University, 613 Huangpu Ave W, Guangzhou, Guangdong 510630, China.
Email: yinhaiyan1867@126.com; lyujun2020@jnu.edu.cn

Conclusion: A nomogram was established that performed better than the SOFA in predicting 30-, 60-, and 90-day mortality risk in patients with sepsis.

Introduction

According to the third international consensus definitions for sepsis and septic shock (Sepsis-3),[1] sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection. In the clinic, organ dysfunction can be indicated by an increase in the Sequential Organ Failure Assessment (SOFA) score of 2 points or more. Sepsis has a very high morbidity and mortality rate, and more than 30 million people worldwide are estimated to be affected by sepsis each year, which may relate to 6 million deaths annually.[2]

Early sepsis prognostic assessment is as critical as early sepsis diagnosis, as it may lead to greater vigilance among medical workers and the provision of timely and appropriate treatment for the patient. Several scores are widely used for such clinical situations, including the SOFA score, Logistic Organ Dysfunction Score (LODS),[3] Acute Physiology and Chronic Health Evaluation (APACHE) II score, and Simplified Acute Physiology Score (SAPS).[4] The SOFA score, in particular, is demonstrated to be a valuable tool for predicting short-term mortality in patients with sepsis,[3,5] but is not without limitations and cannot replace conventional evaluation indicators (such as procalcitonin and lactate clearance rate) to evaluate the prognosis of patients with sepsis.[6] Lactate displays similar ability to SOFA in discriminating the mortality of patients with sepsis.[7] Furthermore, research by Zhang et al.[8] into the mortality of sepsis has included the development and verification of a new scoring system based on machine learning, that may replace the SOFA score to more effectively predict 30-day mortality of intensive care unit (ICU) patients with sepsis.[9]

A nomogram is a graphical tool based on a statistical prediction model that is used to determine the probability that a single clinical event may occur in a patient.[10] Nomograms combine several risk factors to make an accurate prediction, and are widely used in the clinic.[11] There are numerous predictive models for sepsis, including a nomogram to predict 30-day mortality in patients with septic encephalopathy,[12] and a nomogram to predict the risk of sepsis in patients with cholangitis.[13] The purpose of the present study was to construct a new nomogram based on the SOFA score, that is more suitable for patients with sepsis, and to discuss its value in predicting the risk of mortality in such patients.

Patients and methods

Data source

The Medical Information Mart for Intensive Care (MIMIC)-III database was

established in 2003 at Beth Israel Deaconess Medical Centre and Massachusetts General Hospital and the Massachusetts Institute of Technology (MIT), with funding from the National Institutes of Health (NIH).[14] Version 1.4 of the MIMIC-III database was used in the current study, which covers data obtained between 2001 and 2012 from more than 58 000 patient hospitalizations at Beth Israel Deaconess Medical Centre, including 38 645 adult patients and 7875 neonatal patients.[15] The database provides a large amount of real data that can be utilized in clinical research and comprises information related to patients admitted to critical care units at a large tertiary care hospital. Data include vital signs, medications, laboratory measurements, observations and notes charted by care providers, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, survival data, and more. All data can be extracted in the SQL language for further analysis. Personnel involved in the current research participated in a series of courses provided by the NIH and obtained authorization to access the MIMIC-III database after passing the required assessment (certificate No. 38601114).

The present retrospective study analysed data from a third-party anonymized, publicly available database (MIMIC-III), with pre-existing institutional review board approval. Patient information in the database is anonymous, thus, informed consent for the study was not required. The reporting of this study conforms to STROBE guidelines.[16]

Study population

International Classification of Diseases (ICD)-9 codes 99591, 99592, and 78552 were used to extract data from the MIMIC-III database for patients diagnosed with sepsis, severe sepsis, and septic shock, for subsequent retrospective analyses. Exclusion criteria were as follows: (1) patients aged <18 years; (2) patients with a <24-h stay in the ICU to ensure sufficient data for evaluation; and (3) patients with SOFA scores <2 (as the database comprises data collected between 2001 and 2012, and Sepsis-3 was not updated and used for diagnosis at that time). In patients with ≥ 2 admissions to the ICU, data from first ICU admission only were extracted.

Data extraction

All data were translated into SQL for further analysis. The hadm_id (hospital admission id) variable for each included patient was used to extract the following information from the MIMIC-III database: sex; age; SOFA score; continuous renal replacement therapy use; first care unit (surgical ICU, trauma surgical ICU, medical ICU [MICU], coronary care unit, or cardiac surgery recovery unit); comorbidities, namely, congestive heart failure, cardiac arrhythmia, renal failure, liver disease, metastatic cancer (MC), diabetes, coagulopathy, fluid electrolytes, and blood loss anaemia; laboratory tests, namely, white blood cell count (WBC), neutrophil percentage (NET), red blood cell distribution width (RDW), haematocrit, sodium, potassium, albumin, and lactate levels, and blood pH; and vital signs, comprising heart rate, respiratory rate, body temperature, and SpO_2. All of the above information and data were extracted for the first 24 h of ICU stay.

Statistical analyses and nomogram construction

Categorical variables are presented as frequency and percentage values, and χ^2-test or Fisher's exact test was used to determine differences between cohorts. Continuous variables were assessed for normality of distribution with Shapiro–Wilk test, and are

presented as mean $\pm$ SD or median (interquartile range) depending on normality of distribution. Multivariate Cox regression was used to select variables for plotting the 30-, 60-, and 90-day survival curves of the patients. The survival-probability nomogram was constructed using Cox regression. Data were analysed using SPSS, version 24.0 (IBM, Armonk, NY, USA) and R software, version 4.0.2 (https://www.r-project.org/), and a *P* value <0.05 in a two-sided test was considered statistically significant.

Nomogram validation and performance evaluation

Effectiveness of the nomogram was evaluated by comparing two models using the concordance index (C-index) and the area under the receiver operating characteristic curve (AUC).[17] The predictive accuracy of the model was determined by calculating the integrated discrimination improvement (IDI) and the net reclassification improvement (NRI).[18,19] Calibration plots were used to evaluate consistency between the nomogram prognosis prediction and the actual situation. Finally, the net clinical benefit of the predictive model developed in the present study was assessed using decision-curve analysis (DCA).[20]

Results

Study population characteristics

Data from a total of 7770 patients diagnosed with sepsis, severe sepsis and septic shock between 2001 and 2012 were extracted from the MIMIC-III database. After exclusion of patients who did not meet the study criteria, a total of 5240 patients were included in the study (Figure 1). For laboratory examination results, the indexes with >20% missing values were omitted, and the remaining
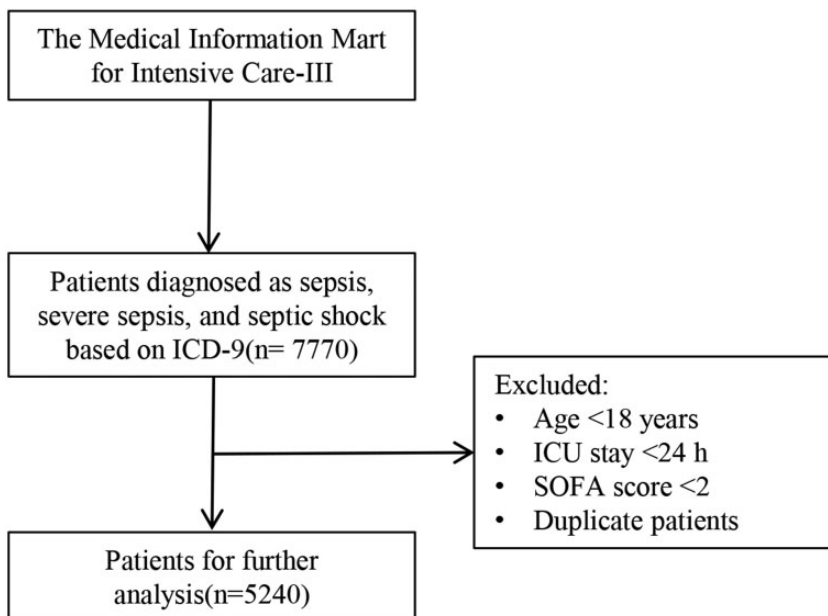

Figure 1. Flowchart of study population selection. ICD, International Classification of Diseases; ICU, intensive care unit; SOFA, Sequential Organ Failure Assessment.

data were filled using the multiple difference complement method. Patients were randomly assigned to the training cohort (70%, $n = 3667$) and validation cohort (30%, $n = 1573$) for constructing and validating the nomogram, respectively (Table 1). The median SOFA score of the entire patient cohort was 6; median ages in the training and validation cohorts were 68 and 67 years, respectively; most patients were male (55.7% and 54.9%); the MICU was the most common first care unit (69.2% and 69.9%); and fluid electrolyte was the most common complication (54.8% and 53.6%). Moreover, most patients in both groups had a WBC count between 10 and 40 k/μl (61.5% and 61.7%); the NET was mainly >70% (78.6% and 78.2%); median SpO_2 values were 97.14 and 97.19%, respectively; and median albumin levels were 2.90 and 2.80 mg/dl, respectively. In both cohorts, the median lactate level was 2.20 mmol/l, median RDW was 15.40%, the most common sodium level was between 130 and 149 k/μl, the most common potassium level was between 3.5 and 5.6 k/μl, the most common $PaCO_2$ value was between 36 and 45 mmHg, the most common heart rate was between 60 and 100 beats/min, the most common respiratory rate was between 20 and 30 breaths/min, and the most common body temperature was between 36 and 37.2°C (all summarised in Table 1).

Multivariate Cox regression analysis and nomogram establishment

Multifactorial Cox regression analysis was performed with data from the training cohort to control for confounding factors (Table 2). After performing a comprehensive evaluation of the variables and then applying Occam's razor (the simplest explanation is preferable to one that is more complex), the variables included in the nomogram for predicting 30-, 60-, and 90-day survival rates in patients with sepsis were: age (hazard ratio [HR] 1.02, 95% confidence interval [CI] 1.02, 1.03), SOFA score (HR 1.07; 95% CI 1.05, 1.08), MC (HR 3.03, 95% CI 2.63, 3.50), SpO_2 (HR 0.97, 95% CI 0.96, 0.98), lactate (HR 1.05, 95% CI 1.03, 1.08), body temperature, albumin (HR 0.81, 95% CI 0.76, 0.87), and RDW (HR 1.12, 95% CI 1.10, 1.14). A graphical illustration of the nomogram is presented in Figure 2, in which age, SOFA score, SpO_2 and RDW are shown to have larger weighting than other included factors.

Nomogram comparison and evaluation

Nomogram performance was evaluated using a variety of metrics, including C-index, AUC, NRI and IDI. The C-index was higher for the nomogram than the single SOFA score, whether in the training cohort (0.717 versus 0.629) or validation cohort (0.713 versus 0.631). The AUCs for 30-, 60-, and 90-day survival probabilities in the nomogram were all higher than for the SOFA score in the training cohort (0.766, 0.771, and 0.772, respectively, versus 0.681, 0.674, and 0.665), and in the validation cohort (0.759, 0.770, and 0.760, respectively, versus 0.700, 0.678, and 0.662; Figure 3). The corresponding median NRI values in the validation cohort were 0.434 (95% CI 0.283, 0.590), 0.520 (95% CI 0.381, 0.653), and 0.530 (95% CI 0.414, 0.681), 30-, 60-, and 90-day survival, respectively. The IDI values for 30-, 60-, and 90-days of follow-up examinations were 0.129, 0.143, and 0.150 in the training cohort and 0.102, 0.131, and 0.140 in the validation cohort, respectively.

As described above, a nomogram for predicting the outcome of patients with sepsis was established by including age, SOFA score, MC, SpO_2, lactate level, body temperature, albumin level, and RDW. Correlations between calibration

Table 1. Demographic and clinical characteristics of the 5240 patients included in the study.

Variable	Validation cohort (n = 1573)	Training cohort (n = 3667)	Statistical significance
Sex			NS
Male	863 (54.9)	2044 (55.7)	
Female	710 (45.1)	1623 (44.3)	
Age, years	67.00 (55.00, 79.00)	68.00 (55.00, 79.00)	NS
SOFA score	6.00 (4.00, 9.00)	6.00 (4.00, 9.00)	NS
CRRT	115 (7.3)	313 (8.5)	NS
First care unit			NS
SICU/TSICU	299 (19.0)	691 (18.8)	
MICU	1099 (69.9)	2536 (69.2)	
CCU	131 (8.3)	313 (8.5)	
CSRU	44 (2.8)	127 (3.5)	
Comorbidity			
Congestive heart failure	586 (37.3)	1353 (36.9)	NS
Cardiac arrhythmias	635 (40.4)	1537 (41.9)	NS
Renal failure	401 (25.5)	883 (24.1)	NS
Liver disease	319 (20.3)	710 (19.4)	NS
Metastatic cancer	113 (7.2)	257 (7.0)	NS
Diabetes	529 (33.6)	1201 (32.8)	NS
Coagulopathy	402 (25.6)	999 (27.2)	NS
Fluid electrolyte	843 (53.6)	2008 (54.8)	NS
Blood loss anaemia	44 (2.8)	84 (2.3)	NS
Laboratory test			
WBC, k/µl			NS
<4	135 (8.6)	301 (8.2)	
4–10	442 (28.1)	1051 (28.7)	
10–40	971 (61.7)	2255 (61.5)	
>40	25 (1.6)	60 (1.6)	
NET, %			NS
<50	118 (7.5)	256 (7.0)	
50–70	225 (14.3)	530 (14.5)	
>70	1230 (78.2)	2881 (78.6)	
RDW, %	15.40 (14.20, 17.20)	15.40 (14.10, 17.10)	NS
HCT, %			NS
<30	421 (26.8)	1068 (29.1)	
30–45.9	1102 (70.1)	2477 (67.5)	
>45.9	50 (3.2)	122 (3.3)	
Sodium, k/µl			$P = 0.002$
<130	150 (9.5)	302 (8.2)	
130–149	1392 (88.5)	3229 (88.1)	
>149	31 (2.0)	136 (3.7)	
Potassium, k/µl			NS
<3.5	268 (17.0)	558 (15.2)	
3.5–5.5	1150 (73.1)	2723 (74.3)	
>5.5	155 (9.9)	386 (10.5)	
Albumin, mg/dl	2.90 (2.50, 3.30)	2.80 (2.40, 3.30)	NS

(continued)

Table 1. Continued.

Variable	Validation cohort (n = 1573)	Training cohort (n = 3667)	Statistical significance
pH			NS
<7.35	667 (42.4)	1552 (42.3)	
7.35–7.45	657 (41.8)	1572 (42.9)	
>7.45	249 (15.8)	543 (14.8)	
Lactate, mmol/l	2.20 (1.50, 3.60)	2.20 (1.50, 3.50)	NS
$PaCO_2$, mmHg			NS
<20	22 (1.4)	38 (1.0)	
21–35	531 (33.8)	1259 (34.3)	
36–45	574 (36.5)	1367 (37.3)	
46–70	379 (24.1)	866 (23.6)	
>70	67 (4.3)	137 (3.7)	
Vital signs			
HR, min^{-1}			NS
<60	35 (2.2)	74 (2.0)	
60–100	1030 (65.5)	2458 (67.0)	
100–120	427 (27.1)	935 (25.5)	
>120	81 (5.1)	200 (5.5)	
RR, min^{-1}			NS
<12	11 (0.7)	23 (0.6)	
12–20	730 (46.4)	1620 (44.2)	
20–30	767 (48.8)	1885 (51.4)	
30–40	65 (4.1)	139 (3.8)	
Temperature, °C			NS
32–35.9	174 (11.1)	425 (11.6)	
36–37.2	927 (58.9)	2095 (57.1)	
37.3–38.5	433 (27.5)	1071 (29.2)	
>38.5	39 (2.5)	76 (2.1)	
SpO_2, %	97.14 (95.77, 98.43)	97.19 (95.76, 98.50)	NS

Data presented as *n* (%) prevalence or median (interquartile range).
SOFA, Sequential Organ Failure Assessment; CRRT, continuous renal replacement therapy; SICU, surgical intensive care unit; TSICU, trauma surgical intensive care unit; MICU, medical intensive care unit; CCU, coronary care unit; CSRU, cardiac surgery recovery unit; WBC, white blood cell; NET, neutrophil percentage; RDW, red cell distribution width; HCT, haematocrit; PaCO2, partial pressure of carbon dioxide in artery; HR, heart rate; RR, respiratory rate; SpO_2, oxyhaemoglobin saturation.
NS, no statistically significant between-group difference ($P > 0.05$).

and standard curves in the nomogram calibration plots for the training and validation cohorts indicated that predicted values for 30-, 60-, and 90-day survival were in good agreement with observed values (Figure 4).

The clinical value of the model and its impact on actual decision-making was verified with DCA. The net benefit of the nomogram was shown to be greater than the SOFA score for any predicted probability, both in the training and validation cohorts (Figure 5), indicating that the nomogram has a substantial net benefit in predicting 30-, 60-, and 90-day survival rates.

Discussion

In this retrospective study of data from the MIMIC-III database, Cox regression

Table 2. Selected variables analysed by multivariable Cox regression in the training cohort.

Variable	HR	95% CI		Statistical significance
Age, years	1.02	1.02	1.03	P <0.001
SOFA score	1.07	1.05	1.08	P <0.001
Renal failure				
No	reference			
Yes	1.15	1.04	1.28	P = 0.007
Liver disease				
No	reference			
Yes	1.28	1.14	1.43	P <0.001
Metastatic cancer				
No	reference			
Yes	3.03	2.63	3.50	P <0.001
Fluid electrolyte				
No				
Yes	1.21	1.11	1.32	P <0.001
First care unit				
SICU/TSICU	reference			
MICU	1.00	0.89	1.12	NS
CCU	1.25	1.05	1.48	P = 0.011
CSRU	1.08	0.85	1.36	NS
SpO$_2$, %	0.97	0.96	0.98	P <0.001
pH				
<7.35	reference			
7.35–7.45	1.15	1.05	1.27	P = 0.005
>7.45	1.39	1.21	1.59	P <0.001
Lactate, mmol/l	1.05	1.03	1.08	P <0.001
Temperature, °C				
32–35.9	reference			
36–37.2	0.74	0.65	0.83	P <0.001
37.3–38.5	0.58	0.50	0.67	P <0.001
>38.5	0.55	0.39	0.79	P = 0.001
Potassium, k/μl				
<3.5	reference			
3.5–5.5	1.13	1.00	1.28	P = 0.050
>5.5	1.29	1.09	1.53	P = 0.003
Sodium, k/μl				
<130	reference			
130–149	0.79	0.68	0.92	P = 0.002
>149	0.95	0.74	1.24	NS
Albumin, mg/dl	0.81	0.76	0.87	P <0.001
RDW, %	1.12	1.10	1.14	P <0.001
NET, %				
<50	reference			
50–70	0.79	0.66	0.95	P = 0.013
>70	0.84	0.71	0.99	P = 0.039

SOFA, Sequential Organ Failure Assessment; SICU, surgical intensive care unit; TSICU, trauma surgical intensive care unit; MICU, medical intensive care unit; CCU, coronary care unit; CSRU, cardiac surgery recovery unit; SpO$_2$, oxyhaemoglobin saturation; RDW, red cell distribution width; NET, neutrophil percentage.
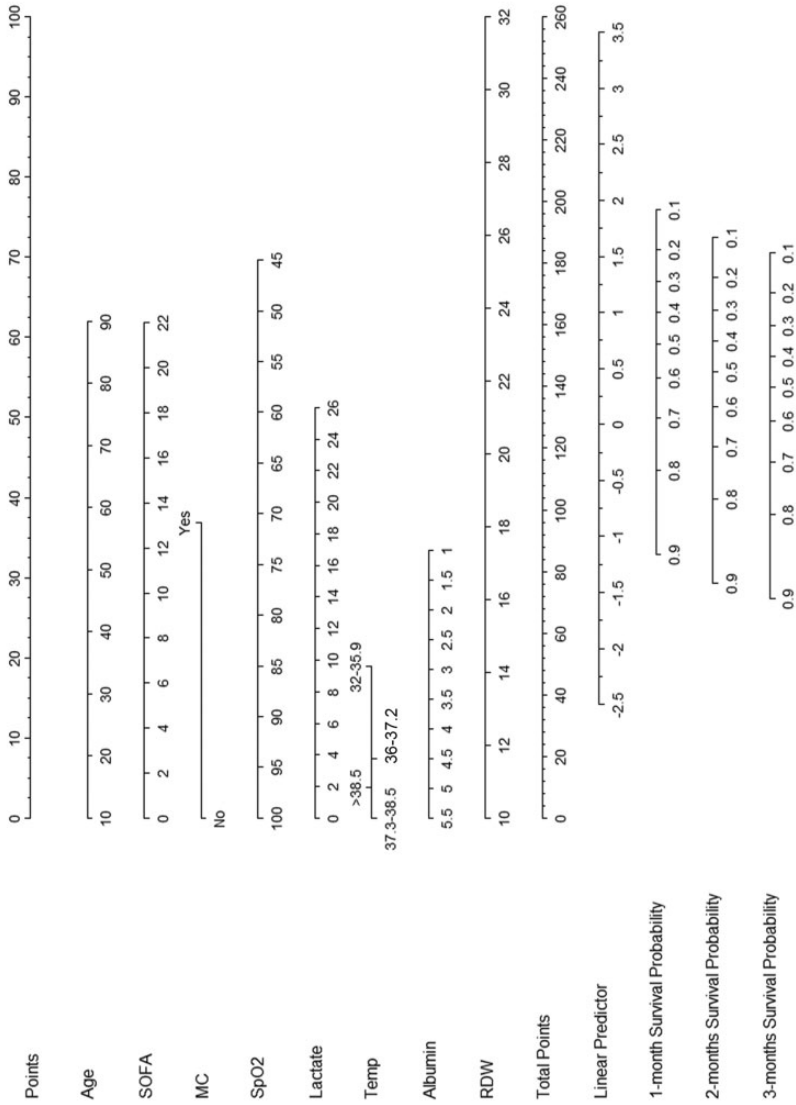NS, no statistically significant association with survival (P >0.05).

Figure 2. Nomogram for predicting 30-, 60-, and 90-day survival. Left column shows the points bar (top) and eight parameters, each to be scored with a vertical line to the points bar, according to the different parameter values. The sum of the points is calculated (total points range, 0–260), and a vertical line is drawn from the total points bar to the 1-month (30-day), 2-month (60-day) and 3-month (90-day) survival probability below, to obtain survival probability of the patient. SOFA, Sequential Organ Failure Assessment; MC, metastatic cancer; Temp, temperature; RDW, red blood cell distribution width.
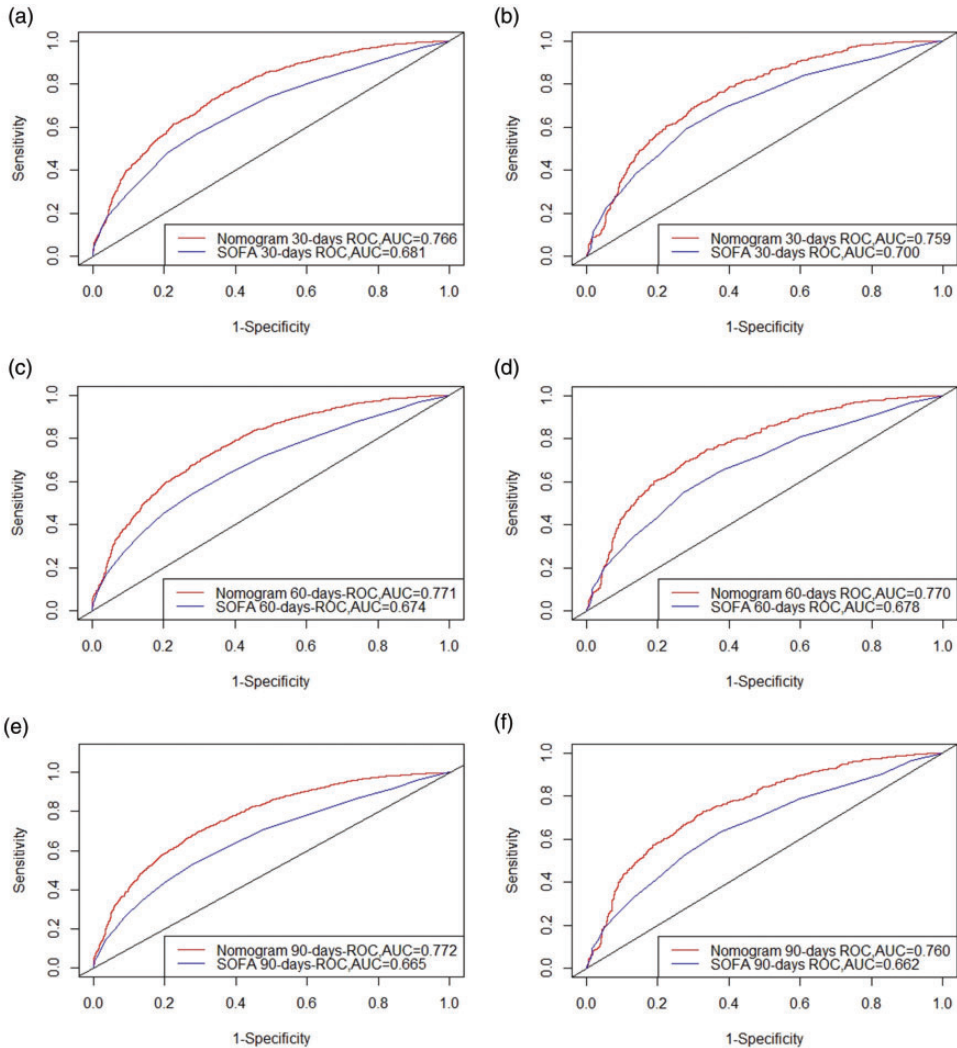
Figure 3. Receiver operating characteristic (ROC) curves for the nomogram and SOFA model, showing AUCs for: 30-day survival (A, training cohort and B, validation cohort); 60-day survival (C, training cohort and D, validation cohort); and 90-day survival (E, training cohort and F, validation cohort). SOFA, Sequential Organ Failure Assessment; AUC, area under the curve.

analysis was used to select variables comprising age, SOFA score, MC, SpO2, lactate, body temperature, albumin, and RDW that were integrated to generate a prediction model for mortality in patients with sepsis, visualized as a nomogram. The C-index and AUC were used to evaluate nomogram performance, and showed

that the model has excellent distinguishability. In addition, the nomogram had a positive improvement effect, shown by NRI and IDI. In the training and validation cohorts, the calibration curve of the new nomogram matched the standard curve very well. In addition, the 30-, 60-, and 90-day DCA curves of the training and
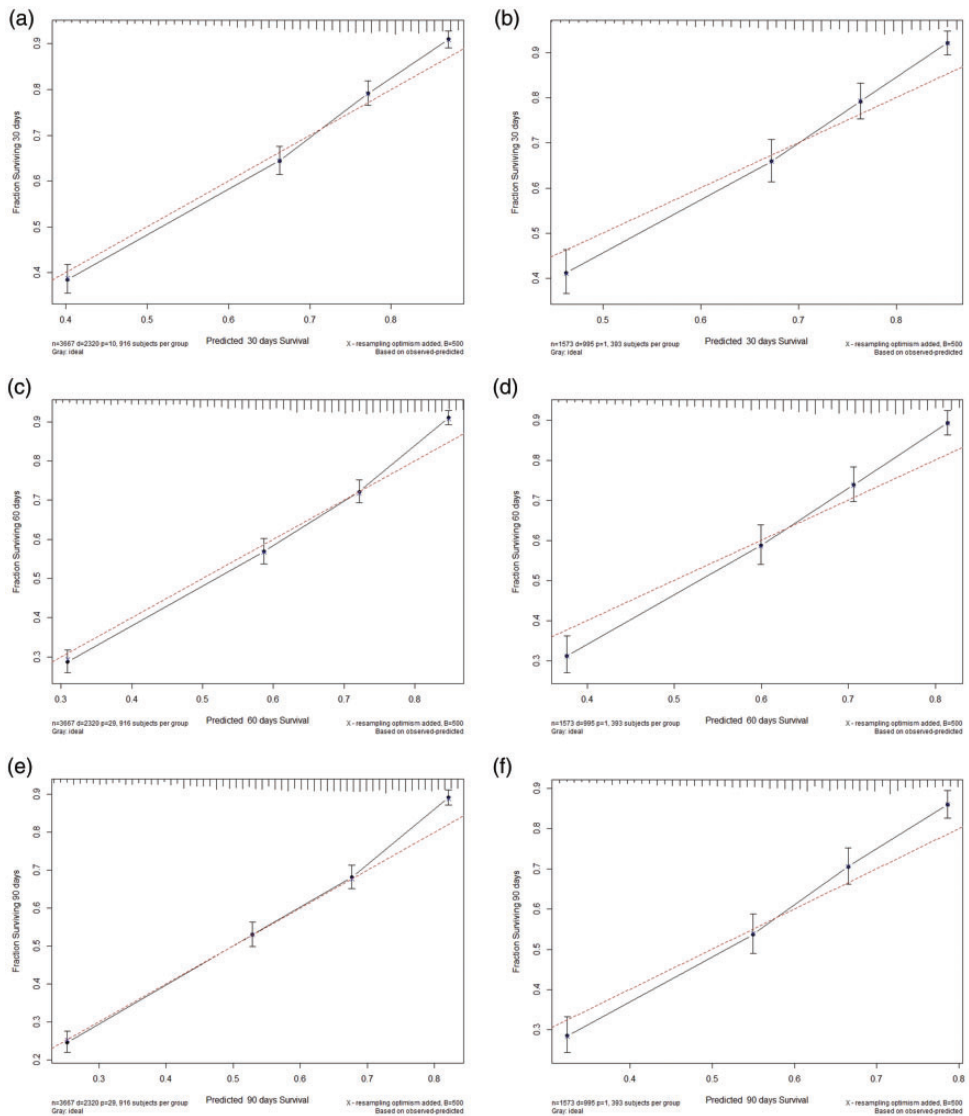
Figure 4. Calibration curves for 30-, 60- and 90-day survival. The abscissa (x-axis) is the predicted survival rate and the ordinate (y-axis) is the actual survival rate. The red dotted line is the reference line (predicted value equals the actual value), the solid black line is the curve fitting line, and the error bars represent 95% confidence intervals. The calibration curves depict the agreement between predicted probabilities and observed outcomes, shown here separately for the training cohort (A, C, and E) and validation cohort (B, D, and F).

validation cohorts revealed that the nomogram generated net benefits. Overall, the results suggest that use of the nomogram may benefit patients with sepsis, as it shows improved performance compared with the SOFA score alone, in predicting 30-, 60- and 90-day mortality in patients with sepsis.
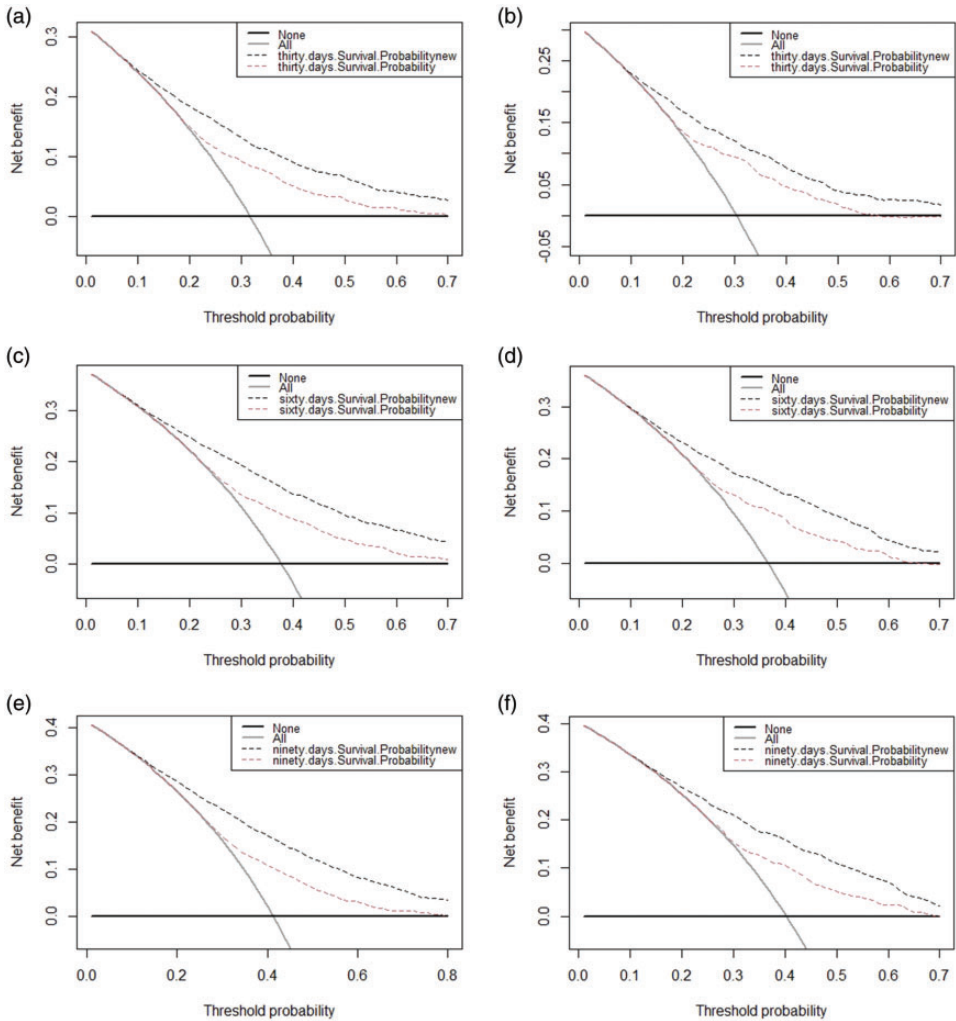
Figure 5. Decision-curve analysis of the training cohort (A, C, E) and validation cohort (B, D, F) for 30-, 60- and 90-day survival. In all figures the black line is above the red line, showing that the area under the curve is larger for the new nomogram model than for the SOFA model. SOFA, Sequential Organ Failure Assessment. Black dotted line, nomogram model; red dotted line, SOFA model.

Age had a relatively strong influence (higher weighting) in the present nomogram. Age has been shown to be an independent risk factor for death in patients with sepsis, and fatality rate is shown to increase linearly with age.[21] The present research revealed the same trend; the nomogram score increased with age, which might be due to elderly patients being more susceptible to infection with gram-negative bacteria, having more comorbidities such as cancer, and having weaker immune function.[22,23] The predictions made using the nomogram regarding whether a patient has MC are also related to the immune system, and the inclusion of albumin in the nomogram reflects the tendency of older patients to have a worse

nutritional status than their younger counterparts.[24,25]

The haemodynamic performance of patients with septic shock is exceptionally complicated.[26,27] In sepsis, endothelial cell dysfunction, the interaction of leukocytes/platelets and endothelial cells, coagulation activation, inflammation, abnormal hemorheology, and functional shunting, together lead to microcirculatory disorders, insufficient tissue perfusion, and hypoxia, ultimately leading to multiple-organ dysfunction or even septic shock.[28] SpO_2 can reflect the body's real-time oxygen supply state and degree of hypoxia, and may therefore serve as a factor associated with sepsis.[29] Additionally, SpO_2 can be measured more conveniently and rapidly than arterial blood gas level, making it an attractive parameter for use in prediction models, and justifying its recruitment and inclusion into the present nomogram.

Patients with sepsis or septic shock will experience anaerobic metabolism due to microcirculatory disturbances and lactate accumulation. Serum lactate level is conventionally regarded as a marker of tissue hypoxia, and used as a clinical indicator of sepsis/septic shock severity and patient prognosis.[30] Lactate also has a high predictive value in the present nomogram, with the score increasing with rising lactate levels.[31]

The indicator with the highest weight in the present nomogram was RDW, a prognostic biomarker for cardiovascular disease, stroke, and metabolic syndrome mortality.[32] A previous retrospective study, based on the MIMIC-III database, found that RDW was useful in predicting long-term all-cause mortality in patients with severe sepsis,[33] and a multicentre observational study showed that RDW is a powerful predictor of risk of all-cause death and blood infection in critically ill patients.[34] Moreover, an increase in the distribution of red blood cells when patients are discharged from the ICU is a powerful predictor of subsequent all-cause mortality.[35] These observations show the predictive value of RDW, however, while this index is commonly measured in clinical workups, it is often ignored and, given its importance in the above studies and the present results, deserves greater attention.

Sepsis is a complex disease, and while guidelines on sepsis are continually updated, the mechanism underlying the pathophysiological changes induced by sepsis remains difficult to characterize. Our understanding of sepsis may be enhanced by summarizing patient data, identifying the degree of risk by patient indicators in the early stages of disease, and adopting reasonable intervention methods, such as improving patient prognosis by improving SpO2 level. The nomogram constructed here may have considerable advantages in predicting the risk of death in patients with sepsis, and is easy to use, making it very suitable for clinical applications.

The results of the present study may be limited by several factors. First, it was a retrospective design based on the MIMIC-III database. Secondly, the nomogram was internally validated using data from only one database. This limitation should be addressed in future studies by using a separate database or data from a separate group or study. Thirdly, further research on the aetiology and antibiotic treatment of sepsis was not conducted. In addition, as the database collected patient data between 2001 and 2012, and Sepsis-3 criteria were not updated at that time, the diagnosis of sepsis was based on Sepsis-1 criteria. Patients with SOFA scores <2 were excluded on this basis, but sepsis diagnoses were still not completely consistent with Sepsis-3. Lastly, sepsis is a heterogeneous disease.[36] In future research, sepsis subclasses should be identified, and reliable predictions should be generated for each

subclass, in order to promote and optimize the clinical management of sepsis.

Conclusions

In the information era, data reusability and data sharing strategies are receiving increasing attention worldwide. Nomograms are a significant component of modern medical decision-making. The present study established a nomogram that performed better than the SOFA alone in predicting 30-, 60-, and 90-day mortality risk in patients with sepsis.

Availability of supporting data

Patient data are available on the MIMIC-III website at https://mimic.physionet.org/, https://doi.org/10.13026/C2HM2Q.

Declaration of conflicting interest

The authors declare that there is no conflict of interest.

ORCID iDs

Jun Lyu https://orcid.org/0000-0002-2237-8771
Haiyan Yin https://orcid.org/0000-0002-9680-4219

References

1. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016; 315: 801–810.

2. Bauer M, Gerlach H, Vogelmann T, et al. Mortality in sepsis and septic shock in Europe, North America and Australia between 2009 and 2019- results from a systematic review and meta-analysis. *Crit Care* 2020; 24: 239.

3. Li Y, Yan C, Gan Z, et al. Prognostic values of SOFA score, qSOFA score, and LODS score for patients with sepsis. *Ann Palliat Med* 2020; 9: 1037–1044.

4. Godinjak A, Iglica A, Rama A, et al. Predictive value of SAPS II and APACHE II scoring systems for patient outcome in a medical intensive care unit. *Acta Med Acad* 2016; 45: 97–103.

5. Karakike E, Kyriazopoulou E, Tsangaris I, et al. The early change of SOFA score as a prognostic marker of 28-day sepsis mortality: analysis through a derivation and a validation cohort. *Crit Care* 2019; 23: 387.

6. Wang Y, Wang D, Fu J, et al. Predictive value of SOFA, qSOFA score and traditional evaluation index on sepsis prognosis. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* 2017; 29: 700–704 [In Chinese, English abstract].

7. Liu Z, Meng Z, Li Y, et al. Prognostic accuracy of the serum lactate level, the SOFA score and the qSOFA score for mortality among adults with sepsis. *Scand J Trauma Resusc Emerg Med* 2019; 27: 51.

8. Zhang Z, Ho KM, Gu H, et al. Defining persistent critical illness based on growth trajectories in patients with sepsis. *Crit Care* 2020; 24: 57.

9. Zhang K, Zhang S, Cui W, et al. Development and validation of a sepsis mortality risk score for Sepsis-3 patients in intensive care unit. *Front Med (Lausanne)* 2021; 7: 609769.

10. Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: more than meets the eye. *Lancet Oncol* 2015; 16: e173–e180.

11. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 2008; 26: 1364–1370.

12. Yang Y, Liang S, Geng J, et al. Development of a nomogram to predict 30-day mortality of patients with sepsis-associated encephalopathy: a retrospective cohort study. *J Intensive Care* 2020; 8: 45.

13. Liu Q, Zhou Q, Song M, et al. A nomogram for predicting the risk of sepsis in patients

with acute cholangitis. *J Int Med Res* 2020; 48: 300060519866100.

14. Yang J, Li Y, Liu Q, et al. Brief introduction of medical database and data mining technology in big data era. *J Evid Based Med* 2020; 13: 57–69.

15. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; 3: 160035.

16. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147: 573–577.

17. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med* 2013; 4: 627–635.

18. Pepe MS, Fan J, Feng Z, et al. The net reclassification index (NRI): a misleading measure of prediction improvement even with independent test data sets. *Stat Biosci* 2015; 7: 282–295.

19. Cook NR, Demler OV and Paynter NP. Clinical risk reclassification at 10 years. *Stat Med* 2017; 36: 4498–4502.

20. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak* 2008; 8: 53.

21. Martin GS, Mannino DM and Moss M. The effect of age on the development and outcome of adult sepsis. *Crit Care Med* 2006; 34: 15–21.

22. Torre LA, Siegel RL, Ward EM, et al. Global cancer incidence and mortality rates and trends–an update. *Cancer Epidemiol Biomarkers Prev* 2016; 25: 16–27.

23. Candeias SM and Gaipl US. The immune system in cancer prevention, development and therapy. *Anticancer Agents Med Chem* 2016; 16: 101–107.

24. Lidoriki I, Frountzas M and Schizas D. Could nutritional and functional status serve as prognostic factors for COVID-19 in the elderly? *Med Hypotheses* 2020; 144: 109946.

25. Lorenzo-López L, Maseda A, de Labra C, et al. Nutritional determinants of frailty in older adults: a systematic review. *BMC Geriatr* 2017; 17: 108.

26. Pilcher J, Ploen L, McKinstry S, et al. A multicentre prospective observational study comparing arterial blood gas values to those obtained by pulse oximeters used in adult patients attending Australian and New Zealand hospitals. *BMC Pulm Med* 2020; 20: 7.

27. Khemani RG, Thomas NJ, Venkatachalam V, et al. Comparison of SpO2 to PaO2 based markers of lung disease severity for children with acute lung injury. *Crit Care Med* 2012; 40: 1309–1316.

28. Lipinska-Gediga M. Sepsis and septic shock-is a microcirculation a main player? *Anaesthesiol Intensive Ther* 2016; 48: 261–265.

29. Hernandez G, Bruhn A and Ince C. Microcirculation in sepsis: new perspectives. *Curr Vasc Pharmacol* 2013; 11: 161–169.

30. Nolt B, Tu F, Wang X, et al. Lactate and immunosuppression in sepsis. *Shock* 2018; 49: 120–125.

31. Ryoo SM, Lee J, Lee YS, et al. Lactate level versus lactate clearance for predicting mortality in patients with septic shock defined by Sepsis-3. *Crit Care Med* 2018; 46: e489–e495.

32. Wang AY, Ma HP, Kao WF, et al. Red blood cell distribution width is associated with mortality in elderly patients with sepsis. *Am J Emerg Med* 2018; 36: 949–953.

33. Han YQ, Zhang L, Yan L, et al. Red blood cell distribution width predicts long-term outcomes in sepsis patients admitted to the intensive care unit. *Clin Chim Acta* 2018; 487: 112–116.

34. Bazick HS, Chang D, Mahadevappa K, et al. Red cell distribution width and all-cause mortality in critically ill patients. *Crit Care Med* 2011; 39: 1913–1921.

35. Purtle SW, Moromizato T, McKane CK, et al. The association of red cell distribution width at hospital discharge and out-of-hospital mortality following critical illness*. *Crit Care Med* 2014; 42: 918–929.

36. Zhang Z, Zhang G, Goyal H, et al. Identification of subclasses of sepsis that showed different clinical outcomes and responses to amount of fluid resuscitation: a latent profile analysis. *Crit Care* 2018; 22: 347.