A nomogram for predicting hospital mortality of critical ill patients with sepsis and cancer: a retrospective cohort study based on MIMIC-IV and eICU-CRD

Zhen-nan Yuan,1 Yu-juan Xue,2 Hai-jun Wang,1 Shi-ning Qu,1 Chu-lin Huang,1 Hao Wang,1 Hao Zhang,1 Xue-zhong Xing1,3

ABSTRACT

Objective Sepsis remains a high cause of death, particularly in immunocompromised patients with cancer. The study was to develop a model to predict hospital mortality of septic patients with cancer in intensive care unit (ICU).

Design Retrospective observational study.

Setting Medical Information Mart for Intensive Care IV (MIMIC IV) and eICU Collaborative Research Database (eICU-CRD).

Participants A total of 3796 patients in MIMIC IV and 549 patients in eICU-CRD were included.

Primary outcome measures The model was developed based on MIMIC IV. The internal validation and external validation were based on MIMIC IV and eICU-CRD, respectively. Candidate factors were processed with the least absolute shrinkage and selection operator regression and visualised by the nomogram. The model was assessed by the area under the curve (AUC), calibration curve and decision curve analysis curve.

Results The model exhibited favourable discrimination (AUC: 0.726 (95% CI: 0.709 to 0.744) and 0.756 (95% CI: 0.712 to 0.801)) in the internal and external validation sets, respectively, and better calibration capacity than Acute Physiology and Chronic Health Evaluation IV in external validation.

Conclusions Despite that the predicted model was based on a retrospective study, it may also be helpful to predict the hospital mortality of patients with solid cancer and sepsis.

INTRODUCTION

Although the prognosis of patients with cancer has obviously improved with advances in oncological treatment and supportive care,1 patients with cancer were more likely to develop infections or sepsis compared with patients without cancer in previous studies.2,3 Cancer and its treatments (eg, chemotherapy, radiation, surgery) led to an increasing number of immunocompromised patients at high risk of sepsis.1,5 Oncological patients account for up to 20% of intensive care unit (ICU) admissions, and sepsis becomes the leading cause of ICU admission in this group of patients.1,36 Currently, sepsis has become an important reason of hospital mortality in critically ill patients with cancer.7,8 A diagnosis of sepsis for patients with cancer has been shown to increase the risk of hospital mortality up to twofold to threefold, making sepsis a significant threat to cancer survivorship.3,5 In previous studies, hospital mortality of patients with cancer and sepsis was 37%–65%.

There are many critical illness scores that are more applicable to critical patients without cancer,12–15 but the predictive value of these scoring systems in patients with cancer is unclear. Previous scores usually contain too many indicators, making them inconvenient for clinical use.12–14 It is necessary to establish a convenient predictive model to assess the prognosis of septic patients with cancer. Meanwhile, the nomogram is a convenient and reliable tool for predicting certain endpoints, such as disease progression or
We conducted this study to establish a nomogram for assessing the hospital mortality of patients with cancer and sepsis.

**PATIENTS AND METHODS**

**Database**

The model was developed and validated internally based on Medical Information Mart for Intensive Care IV (MIMIC IV, V.2.0) and validated externally based on eICU Collaborative Research Database (eICU-CRD). The MIMIC IV is a publicly available database of patients admitted to the Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA, between 2008 and 2019, which included over 76,000 ICU admissions. It contains detailed information for each admission, including laboratory examination, vital signs, administered medications and status on discharge. The eICU-CRD is a multicentre database of over 200,000 hospitalised ICUs in the USA between 2014 and 2015. The study was based on the suggestions of the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis statement.

**Participants**

A flowchart of the patient selection process is shown in figure 1. The inclusion criteria were as follows: (1) Enrolled patients were diagnosed as malignancy and sepsis according to International Classification of Diseases, Ninth Revision code. The diagnosis of sepsis was based on sepsis definition 3.0. (2) Patients’ age was more than 18 years when admitted by ICU. The exclusion criteria included: (1) Enrolled patients with haematologic tumours. (2) Repeated ICU admissions except for the first time and factors with over 50% missing information.

**Data collection**

Clinical data of septic patients with cancer consisted of the followings (online supplemental table 1): (1) demographic characteristics such as age and gender; (2) cancer stage (the distant metastasis was based on the

---

**Figure 1** Flow diagram of the patient selection in MIMIC IV and eICU-CRD. ICU, intensive care unit; MIMIC-IV, Medical Information Mart for Intensive Care; eICU-CRD, eICU Collaborative Research Database.
definition of the American Joint Committee on Cancer 8); (3) chronic underlying diseases: chronic obstructive pulmonary disease (COPD) and diabetes; (4) organ function based on Sequential Organ Failure Assessment (SOFA) Score; (5) infection sites; (6) laboratory examination; (7) blood cultures. The vital signs and laboratory results within the first 24 hours after ICU admission were collected. The observed endpoints of this study were the hospital death or the safe discharge. R software was used to process the raw data extracted by Navicat for Structure Query Language Server.

Patient and public involvement
Patients and/or the public were not directly involved in this study.

Statistics
The data were described as number (rate) or median (IQR) or means±SD. The most significant factors were selected by the least absolute shrinkage and selection operator (LASSO) in the MIMIC IV dataset. Cross-validation was used to identify appropriate adjusted parameters (λ) for LASSO logistic regression. The selected factors were analysed by multivariate logistic regression and incorporated into nomograms to predict the hospital mortality. The internal validation and external validation were based on MIMIC IV and eICU-CRD, respectively. The internal validation of the enhanced bootstrap technique with 200 resamples from MIMIC IV database in the manuscript was performed. As a result of bootstrapping, performance measures of the derived tool were obtained, namely, the c-statistics and the calibration curve, both corrected for bias. The effectiveness of the nomogram was assessed by the receiver operating characteristic (ROC). The calibration chart was used to further express the consistency between the actual hospital mortality and the predicted hospital mortality. All statistical analyses were performed by R V.4.1. The glmnet package was used for LASSO regression analysis. The Proc and ggplot2 packages were used to compare the performance between two ROC curves.

Feature selection
A total of 19 features were included in this study and considered as potential factors. Factor selection was performed in MIMIC IV using the LASSO regression algorithm. When the partial likelihood binomial deviation reached an acceptable value, the most appropriate tuning parameter λ for LASSO regression was 0.018 (figure 2A). Finally, nine variables with non-zero coefficients were retained in the LASSO analysis (figure 2B, lambda.1se).

RESULT
Patient selection and baseline characteristics
There are 76540 and 200859 ICU stays in MIMIC IV database and eICU-CRD, respectively. A flowchart of the patient selection process and study cohort is shown in figure 1. Finally, 3796 septic patients with cancer from MIMIC IV were designed to training cohort for developing the predicted model. A total of 549 septic patients with cancer from eICU-CRD were designed for validation of the nomogram. The baseline characteristics are described in online supplemental table 1.

Multivariate logistic regression analysis
Potential clinical factors selected by LASSO regression were proceeded with the multivariate logistic regression model (online supplemental table 2, figure 3A). It demonstrated that COPD (OR: 1.407, 95% CI: 1.187 to 1.667, p<0.001), blood infection (OR: 1.373, 95% CI: 1.082 to 1.743, p<0.001), age (OR: 1.015, 95% CI: 1.008 to 1.020, p<0.001), the presence of metastasis (OR: 1.804, 95% CI: 1.533 to 2.122, p=0.011), the value of PH (OR: 0.174, 95% CI: 0.078 to 0.352, p<0.001), the value of anion gap (OR: 1.079, 95% CI: 1.062 to 1.095, p<0.001), ICU admission type (OR: 0.422, 95% CI: 0.230 to 0.775, p<0.001), the level of serum albumin (OR: 0.500, 95% CI: 0.441 to 0.568, p=0.005) and the number of PLT (OR: 1.369, 95% CI: 1.074 to 1.744, p=0.005), (OR: 2.425, 95% CI: 1.866 to 3.118, p=0.005) independently affected hospital mortality of patients with cancer and sepsis.

Development of the nomogram in the MIMIC IV
Nine variables selected by the LASSO regression analysis were used to create a visualisation model via nomogram for hospital mortality. The nomogram was drawn by assigning a weighted score for each selected factor based on the prognostic weights of multiple variables. The higher total scores calculated from the sum of the appointed points for each factor in the nomogram, the higher the hospital mortality rate. Finally, a vertical line is drawn from the total points axis to the probline to obtain the predicted probability of hospital mortality (figure 3B). From the nomogram, we can see that patient with COPD, elder age, late stage of cancer status, blood infection, lower PH of arterial blood gas and high anion gap, lower level of serum albumin and count of platelet, and admission by emergency will have higher scores, indicating a higher hospital mortality.

Internal and external validation of the nomogram
The internal validation and external validation of the nomogram was based on MIMIC IV and eICU-CRD, respectively. The derived model with the nine identified factors had a good discrimination with a bootstrap-corrected c-statistic of 0.726 (95% CI: 0.709 to 0.744). The calibration plot also showed good performance of the predictive model (online supplemental figure 1). The average absolute difference between the predicted and bias-corrected actual probabilities, called the mean absolute error, was low at 0.013.

The area under the curve (AUC) in eICU-CRD was 0.756 (95% CI: 0.712 to 0.801), respectively, showing a good statistical discrimination (figure 4A). The calibration curve demonstrated a good fit both in the MIMIC IV
Figure 2  Selection of tuning parameter ($\lambda$) in the LASSO regression for factors selection. (A) Dotted vertical lines are drawn adopting the minimum rule and the 1 SE of the minimum rule at the suitable values $\log(\lambda)$, where factors are selected. (B) Nine non-zero coefficients are included by LASSO coefficient profiles for clinical factors. LASSO, least absolute shrinkage and selection operator.
and the eICU-CRD based on the observed and predicted hospital mortality (figure 4B). The eICU-CRD consisted of all variables used in the Acute Physiology and Chronic Health Evaluation (APACHE) IV scores and the result of the predictions. So, we future compared the effectiveness of the nomogram with that of APACHE IV in eICU-CRD. The AUC of predicting hospital mortality for APACHE IV in eICU-CRD is 0.678 (95% CI: 0.664 to 0.698), which is lower than that of the nomogram (figure 5A). The nomogram also shows a better calibration compared with APACHE IV (figure 5B). In addition, compared with APACHE IV, the performance of the nomogram has the same discrimination with no statistically significant net reclassification index ((95% CI): 0.017 (−0.073 to 0.107); p=0.710) and integrated discrimination improvement ((95% CI): −0.0036 (−0.047 to 0.040); p=0.871).

Clinical application of the nomogram
The decision curve analysis (DCA) of the nomogram in predicting hospital mortality for septic patients with
Figure 4  The performance of the nomogram in the MIMIC IV and eICU-CRD. (A) AUCs of the nomogram in the MIMIC IV and eICU-CRD. The blue line represented a calibration model of the nomogram between predicted hospital mortality and actual hospital mortality in MIMIC IV. The brown line represented a calibration model of the nomogram between predicted hospital mortality and actual hospital mortality in eICU-CRD. AUC, area under the curve; ICU, intensive care unit; eICU-CRD, eICU Collaborative Research Database; MIMIC IV, Medical Information Mart for Intensive Care IV; ROC, receiver operating characteristic.
Figure 5  The performance of the nomogram and APACHE IV in the eICU-CRD. (A) AUCs of the nomogram and APACHE IV in the eICU-CRD. (B) Calibration curves of the nomogram and the APACHE IV for predicting hospital mortality in the eICU-CRD. The blue line represented a calibration model of the nomogram between predicted hospital mortality and actual hospital mortality in eICU-CRD. The brown line represented a calibration model of the APACHE IV between predicted hospital mortality and actual hospital mortality in eICU-CRD. APACHE IV, Acute Physiology and Chronic Health Evaluation IV; AUC, area under the curve; ICU, intensive care unit; eICU-CRD, eICU Collaborative Research Database; ROC, receiver operating characteristic.
cancer is described in online supplemental figure 2. It demonstrated that the nomogram had a good net benefit both in training and validation cohort, which is also superior to the performance of APACHE IV. Compared with the single variable in the nomogram, the predicted model also shows better using ROC curves (online supplemental figure 3).

DISCUSSION
For critical ill patients, early active intervention response to risk factors is necessary. However, it is difficult to obtain some factors in clinical utility. Besides, there are different degrees of defects in accuracy, sensitivity or specificity among some factors. Many studies showed that SOFA scores had low sensitivity and specificity. The characteristics and outcomes of septic patients with cancer were different from those without cancer. For most of the critical illness scores, cancer-related factors were not assessed. In our research, we created a prediction model presented in the form of nomogram including nine variables. These clinical factors are convenient for clinical use. From the online supplemental table 1, we can see that the baseline characteristics of screened patients in the two databases are quite different, which may be due to hospital admissions. The internal and external validation of the model achieved good results, indicating that the applicability of the model is quite good. However, this is a retrospective study where data were collected from patients’ medical record databases. We were unable to control for data accuracy and bias. In terms of feature selection and classification, there are many other methods that are more widely applied than LASSO, such as random forest, elastic network and deep neural network.

As indicated in this study, the occurrence of distant metastasis is associated with poor prognosis, which may be because of their immunocompromised status. Immunocompromised status indicated poor outcome for septic patients. We should pay more attention to cope with such patients with distant metastasis.

After selecting candidate factors by LASSO regression analysis, the nomogram demonstrated that the decreased platelet count is a high-risk factor for short-term survive of patients with cancer and sepsis. The reduced platelet count indicates a myeloid depression that can lead to coagulopathy. Furthermore, previous studies showed that the presence of coagulopathy was associated with the short-term survive of patients with sepsis. Besides, dynamic changes in coagulopathy referred to be related to the poor prognosis of severe infection. Therefore, clotting status, especially platelet count, should be monitored in these patients.

Chronic disease has been found to be a crucial predictor of short-term outcomes. In our research, COPD increases the hospital death rate in patients with cancer and sepsis. Previous study show that patients with COPD tend to have infections that can occur in other parts of the body, not just the respiratory tract. This increased incidence may be because of the impaired barrier function of the respiratory tract and systemic inflammatory characteristic of COPD. For these patients, we should take active and effective measures for the treatment of COPD.

Many critical illness scores, such as Simplified Acute Physiology Score II and the APACHE II, were created to assess the severity of illness and predict short-term prognosis. APACHE II is commonly used to assess the prognosis of circulatory and respiratory diseases, while SOFA can accurately assess the extent of organ damage, especially for sepsis. However, general prognostic models in the ICU showed good discriminative power but poor calibration, which may underestimate mortality rate in critical ill patients with cancer. The APACHE IV system is also a general severity of illness score to assess critical patients. Compared with APACHE IV, our prediction model has a similar degree of differentiation and slightly better calibration. Besides, we only accepted nine clinical indicators that are very readily available. Compared with APACHE IV, it also demonstrated more net benefits predicted by the nomogram via DCA.

Previous studies have found that it significantly affects the prognosis of patients in the ICU when patients admitted by emergency or elective means. In contrast, the hospital mortality is higher when patients admitted by emergency department. For patients with cancer, emergency admissions also had a worse prognosis than elective admissions. In our study, patients with emergency admission showed similar outcome, which is a high-risk factor for patients with cancer and sepsis. So clinicians should take effective measures to treat it earlier before ICU admission to avoid it admitted by emergency.

In our study, blood infection was also a high-risk factor for poor prognosis of patients. Previous studies showed that the mortality rate could be as high as 18.6% among which staphylococcus aureus had the highest mortality rate. It was reported that blood infection is a risk variable for immunocompromised patients in ICU. So it is important to take blood cultures and clinicians should take effective measures to reduce the rate of blood infection as soon as possible.

Older patients tend to develop sepsis compared with younger patient. There is no doubt that the elder patients become weak when they cope with sepsis. Previous studies have focused on the relationship between the anion gap and the prognosis in many diseases. The anion gap is a traditional indicator to assess the acid–base status, and the abnormal anion gap was considered to be related to acid–base disorders, which have a major impact on the prognosis of critically ill patients. Similarly, we demonstrated that the serum anion gap is a high-risk factor for survive in septic patients with cancer. Besides, our study suggested that decreased level of the serum albumin was associated with poor outcome for septic patients with cancer. Chen et al demonstrated that a high lactic/albunin ratio affected hospital survive rate.
independently for patients with sepsis. For these patients, we should actively correct low protein levels and improve microcirculation.

CONCLUSION
This study creates a prediction model for hospital mortality in septic patients with cancer based on the MIMIC IV database and eICU-CRD. It describes factors associated with having impact on hospital mortality. Despite that the predicted model was based on a retrospective study, it may also be helpful to predict the hospital morality of patients with solid cancer and sepsis.

Author affiliations
1 Department of Intensive Care Unit, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 100021, Beijing, China
2 Department of pediatrics, Peking University People’s Hospital, Beijing, China

Contributors
Conception and design: Z-YZ, X-ZX and Y-JZ. Provision of study materials or patients: S-nQ, C-IH, H-JW and HW. Collection and assembly of data: Y-JX and Z-ny. Data analysis and interpretation: Z-ny and HZ. Final approval of manuscript: all authors. X-Zx is the guarantor of the study.

Funding
The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests
None declared.

Patient and public involvement
Patients and/or the public were not involved in the design, conduct, or reporting, or dissemination plans of this research.

Patient consent for publication
Not applicable.

Ethics approval
This study involves human participants. The name of the Institutional Board is Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College. The model was developed and validated based on public database. After completing the Collaborative Institutional Training Initiative programme, we got permission to access the database (record ID: 36067767). This study used public deidentification databases, so there is no need to obtain the informed consent and approval of the Institutional Review Board. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review
Not commissioned; externally peer reviewed.

Data availability statement
Data are available upon reasonable request. The datasets used and/or analysed during the current study are available from MIMIC-IV and eICU-CRD if they completed CITI programme and got permission to access the database.

Supplemental material
This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD
Xue-zhong Xing http://orcid.org/0000-0001-7546-1379

REFERENCES


36 Guo X, Guo D. A Nomogram based on comorbidities and infection location to predict 30 days mortality of immunocompromised patients in ICU; a retrospective cohort study. *Int J Gen Med* 2021;14:10281–92.


