BMC Emergency Medicine

RESEARCH

Open Access

Development of a severity score based on the International Classification of Disease-10 for general patients visiting emergency centers

Ji Eun Kim[1], Jinwoo Jeong[1*], Yuri Choi[1] and Sung Woo Lee[2]

Abstract

Background When comparing mortality, the severity of illness or injury should be considered; therefore, scoring systems that represent severity have been developed and used. Given that diagnosis codes in the International Classification of Disease (ICD) and vital signs are part of routine data used in medical care, a severity scoring system based on these routine data would allow for the comparison of severity-adjusted treatment outcomes without substantial additional efforts.

Methods This study was based on the National Emergency Department Information System database of the Republic of Korea. Patients aged 15 years or older were included. Data from between 2016 and 2018 were used to develop the scoring system, and data from 2019 were used for testing. We calculated the products of the number of disease-specific survival probabilities (DSPs) to reflect the severity of the patients with multiple diagnoses. A logistic regression model was developed using DSPs, age, and physiological parameters to develop a more accurate mortality prediction model.

Results The newly developed model showed predictive ability, as indicated by an area under the receiver-operating characteristic curve of 0.975 (95% CI: 0.974–0.977). When a threshold value of -5.869 was used for determining mortality, the overall accuracy was 0.958 (0.958–0.958).

Conclusion We developed a scoring system based on ICD codes, age, and vital signs to predict the in-hospital mortality of emergency patients, and it achieved good performance. The scoring system would be useful for standardizing the severity of emergency patients and comparing treatment results.

Keywords Patient acuity, Emergency service, hospital, Emergency room visits, Hospital mortality, International Classification of Diseases

*Correspondence:
Jinwoo Jeong
jinwoo@dau.ac.kr
[1]Department of Emergency Medicine, Dong-A University College of Medicine, 26 Daesingongwon-Ro, Seo-gu, Busan 49201, Republic of Korea
[2]Department of Emergency Medicine, Korea University College of Medicine, Seoul 02841, Republic of Korea

Background

To improve the quality of emergency care, measurement of treatment outcomes is essential [1]. In-hospital mortality is commonly used as an outcome indicator for evaluating treatment outcomes in the emergency department (ED) [2]. When comparing mortality, the severity of illness or injury should be considered; therefore, scoring systems that represent severity have been developed and used [2, 3].

For patients with infections, severity scores derived from intensive care units, such as the Sequential Organ Failure Assessment (SOFA) or Acute Physiology and Chronic Health Evaluation (APACHE) scores, have been evaluated in the ED [4, 5]. However, these scores have limitations for general ED patients because not all ED patients have their blood tested for calculation of the scores.

In trauma care, the Trauma and Injury Severity Score (TRISS) has long been used to compare severity-adjusted outcomes [6]. The TRISS incorporates the Injury Severity Score (ISS) to represent the anatomical severity of the injury and the Revised Trauma Score (RTS) to represent the physiological parameters and age. The ISS has limitations in wide adoption because it requires that all injuries be described in the Abbreviated Injury Scale (AIS) lexicon, which is an expensive step that is not easily applicable in hospitals other than specialized trauma centers [7]. Therefore, the International Classification of Disease (ICD)-based ISS (ICISS) was developed as an alternative to the ISS that allows representation of anatomical injury severity without requiring separate coding other than the widely used ICD system [7]. The performance of the ICISS has been reported to be comparable to that of the ISS or TRISS [8, 9].

Efforts have been made to utilize ICD codes to predict treatment outcomes in patients other than those with injuries. In one study, a morbidity and comorbidity score was developed on the basis of ICD-10 codes and was used to predict hospital mortality among patients admitted to acute care hospitals. The area under the receiver operating characteristic curve (AUROC) was reported to be 0.910 (95% confidence interval [CI]: 0.907–0.913) [10].

Given that diagnosis codes in the ICD and vital signs are part of routine data used in medical care, a severity scoring system based on these routine data would allow for the comparison of severity-adjusted treatment outcomes without substantial additional efforts. This practical approach ensures that the system can be easily implemented in various healthcare settings [10]. To date, few severity scoring systems based on such data are applicable to emergency patients in general.

The goal of this study was to develop and validate a severity scoring system that predicts hospital mortality on the basis of parameters routinely collected from the national database of patients who visit emergency centers nationwide.

Methods

Study design and setting

The study was an observational study based on the National ED Information System (NEDIS) database maintained by the National Emergency Medical Center (NEMC) of the Republic of Korea (dataset identification number N20211820511). The NEDIS is a nationwide registry of demographic and clinical data regarding ED visits from all emergency medical facilities in Korea [2].

The NEMC continuously maintains the quality of the data recorded in the NEDIS. When the registered values are too extreme to be true or out of context, for example, when the time of ED disposition is earlier than patient arrival at the ED, the NEMC requests that the data manager in the ED correct the transmitted data. The recorded ICD codes of the sampled cases are compared to the corresponding medical records for accuracy per the annual ED evaluation program. Clinical trial number: not applicable.

Selection of participants

From the NEDIS dataset, cases submitted by the designated regional and local emergency medical centers were included for analysis, whereas cases from the smaller community hospitals were excluded because fewer fields were provided by the hospitals. Only patients aged 15 years or older were included. Data collected between 2016 and 2018 were used for the development of the scoring system, and data collected during 2019 were used for validation.

Measurements

Age (coded at 5-year intervals), sex, consciousness (coded using the alert, voice, pain, or unresponsive (AVPU) scale), vital signs at presentation, and diagnoses at the time of disposition from the ED, either at admission or at discharge, were coded according to the Korean adaptation of the ICD, 10th Revision (ICD-10). Mortality in the ED or after admission was recorded.

Creation of a diagnosis-specific survival probability (DSP, formerly referred to as the survival risk ratio (SRR)) set [11]

All of the ICD-10 codes were listed, except for the V, W, X, Y, and Z code groups, which describe the circumstances of the events rather than the diagnoses. The first four digits were taken from the codes to build the DSP table. DSP was calculated as the ratio of survivors within the cases with each corresponding ICD code as one of their diagnoses at the time of disposition from the ED on the basis of the derivation dataset between 2016 and 2018. For those codes with fewer than 20 cases, DSPs

were calculated with cases matching the first three digits of the corresponding ICD codes [7, 11]. After the above process, some code had zero cases matched with the first three digits. Those codes were assigned a DSP of 1 so that the code would not influence the final calculation.

Derivation of the ICD-based emergency severity score

Because we attempted to apply the ICISS methodology to all ED cases, including injuries and diseases, we calculated the products of the lowest DSPs to reflect the severity of the patients with multiple diagnoses and named the value "ICD-based Emergency Severity Score (ICESS)." [7].

$$\mathrm{ICESS} = \mathrm{DSP}_1 \times \mathrm{DSP}_2 \ldots \times \mathrm{DSP}_n$$

We evaluated the performance of the ICESS in predicting mortality by changing the maximal number of DSPs incorporated into the calculation, thereby determining the optimal number of DSPs for the derivation of the severity score. In cases where the number of ICD codes corresponding to the case is fewer than the maximum DSP number, the entire DSP corresponding to the ICD was used.

Expanding the prediction model with additional variables

To develop a more accurate mortality prediction model, a logistic regression model was developed using the ICESS, age and physiological parameters. The model was inspired by the TRISS used for trauma victims, which is the logit for mortality calculated from age, anatomical injury severity represented by the AIS, and physiological parameters represented by the RTS [6, 12]. Age was converted into age scores using univariate associations with mortality before being incorporated into the regression (Table 1). The modified early warning score (MEWS) was used to represent severity derived from physiological parameters. Logistic regression analysis was performed using the three variables to predict in-hospital mortality. Therefore, the expanded prediction model can be calculated as follows:

Table 1 Scores derived from the relationship between age and mortality

Age (y)	Score
15–54	0
55–74	1
75–84	2
85–89	3
90–94	6
95–99	11
100 or above	17

$$\mathrm{Y} = \mathrm{b}_0 + \mathrm{b}_1 \times \mathrm{MEWS} + \mathrm{b}_2 \times \mathrm{ICESS} + \mathrm{b}_3 \times \mathrm{AGE\,SCORE}$$

Here, Y represents the logit for mortality in the model, and the probability of survival is calculated as follows:

$$\text{probability for survival} = \frac{1}{1 + e^{\mathrm{Y}}}$$

To mitigate issues regarding possible overfitting, four-fold cross-validation was applied. The derivation dataset was randomly assigned to one of four groups; three were used for logistic regression, and the other was used for validation.

Outcome

On the basis of the test dataset from the NEDIS in 2019, the AUROCs and the area under the precision–recall curves (AUPRCs) and their corresponding 95% CIs were calculated and used as measures of the performance of the ICESS and the expanded logistic regression model in predicting hospital mortality. The cutoff point for determining in-hospital mortality was derived from the ROC curve via Youden's method. Model accuracy, sensitivity and specificity and their corresponding 95% confidence intervals were calculated.

A calibration plot and Brier score were used to determine the calibration of the proposed prediction model. It is defined by the mean squared difference between the observed value of a binary outcome and its predicted probability. A value of 0 represents complete agreement, whereas a value of 1 represents complete disagreement [13]. We calculated the Brier score to determine the agreement between the actual survival and the calculated probability for survival.

Statistical analysis

R version 4.4.1 (R Foundation for Statistical Computing, Vienna, Austria, 2024) and the package 'tidyverse' version 2.0 were used for the statistical analysis. The package 'precrec' version 0.14.4 was used to derive the ROC and the PRC curves and calculate the AUCs [14]. The 95% CIs of the AUCs were derived by bootstrapping for 2,000 iterations. The package "pROC" version 1.18.5 was used to calculate Youden's index for cutoff values of the score derived by the logistic regression model [15]. The package "scoring" version 0.6 was used to calculate the Brier score. Categorical data are summarized as frequencies and percentages, whereas continuous data are presented as medians and interquartile ranges. The lack of overlap between the corresponding 95% CIs indicated statistical significance.

Results

Characteristics of the study subjects

A total of 12,889,082 patients were eligible for the derivation of the ICESS (Fig. 1). The general characteristics of the derivation set are summarized in Table 2.

Disease-specific survival probabilities

The DSPs for each ICD are presented in the supplemental table (Table S1).

The optimal number of DSPs for the calculation of ICESS

The performance of ICESS was evaluated by varying the maximum number of DSPs incorporated into the equation. Increasing the number of DSPs to more than one in each case did not improve the prediction performance (Table 3). Therefore, ICESS was defined as the minimum single DSP for further analysis.

$$\text{ICESS} = \text{the lowest DSP}$$

Prediction model

The results of model derivation and validation for each fold are presented in Table 4. The fourfold cross-validation resulted in very similar results with overlapping confidence intervals. We proceeded with coefficients derived from the third fold and derived the following formula:

$$\begin{aligned}\text{logit for mortality} = &-2.565 - 6.413 \times \text{ICESS} \\ &+ 0.608 \times \text{MEWS} \\ &+ 0.308 \times (\text{age score})\end{aligned}$$

We defined the Emergency Severity Score (ESS) as the logit for mortality derived from the 2016–2018 dataset with fourfold cross-validation. The threshold for predicting mortality was determined to be -5.869 using Youden's method to optimize sensitivity and specificity.

Test

The predictive performance of the ESS was verified with AUROC and AUPRC analyses using the 2019 dataset. The ESS showed excellent predictive capability, with an AUROC of 0.975 (95% CI: 0.974—0.977) and an AUPRC of 0.724 (0.717—0.731) (Fig. 2). When the threshold value of -5.869 was used for determining mortality, the overall accuracy was 0.958 (0.958–0.958), the sensitivity

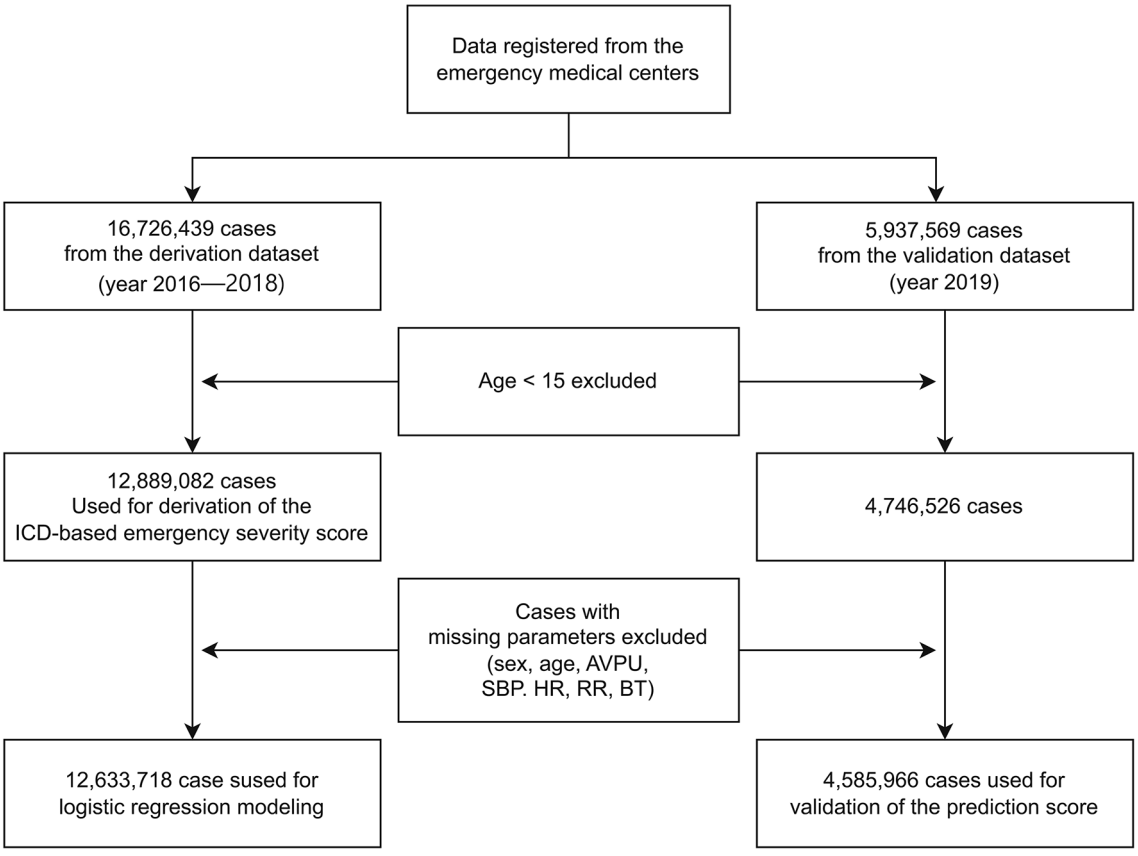

Fig. 1 Flow diagram of the study population in the present study. The included and excluded cases analyzed in the present study are presented. AVPU: consciousness recorded on the scale of alert, response to voice, response to pain and unresponsive; SBP: systolic blood pressure; HR: heart rate; RR: respiratory rate; BT: body temperature

Table 2 General characteristics of the derivation and validation sets. The values are presented as numbers (%) or medians (interquartile ranges)

	Derivation set (*n* = 12,633,718)	Validation set (*n* = 4,585,966)
Sex		
Male	6,258,347 (49.5)	2,261,689 (49.3)
Age (y)		
15–54	7,179,310 (56.8)	2,466,580 (53.8)
55–74	3,627,721 (28.7)	1,376,552 (30.0)
75–84	1,382,447 (10.9)	545,000 (11.9)
85–89	314,750 (2.5)	137,749 (3.0)
90–94	103,103 (0.8)	48,026 (1.0)
95–99	23,017 (0.2)	10,738 (0.2)
100 or above	3,370 (0.0)	1,321 (0.0)
Event		
Disease	9,159,256 (72.5)	3,346,492 (73.0)
Other	3,474,462 (27.5)	1,239,474 (27.0)
Consciousness		
Alert	12,208,188 (96.6)	4,435,800 (96.7)
Response to voice	145,115 (1.1)	81,215 (1.8)
Response to pain	55,027 (0.4)	51,628 (1.1)
Unresponsive	225,388 (1.8)	17,323 (0.4)
Systolic blood pressure (mmHg)	130 (118–148)	130 (119–150)
Heart rate (min^{-1})	82 (74–94)	83 (74–95)
Respiratory rate (min^{-1})	20 (18–20)	20 (18–20)
Body temperature (°C)	36.6 (36.4–37.0)	36.7 (36.4–37.0)
Pulse oxygen saturation (%)	98 (97–99)	98 (97–99)
Medical Early Warning Score	1 (1–2)	1 (1–2)
Admission	3,060,583 (24.2)	1,107,268 (24.1)
ED[a] mortality	50,947 (0.4)	15,245 (0.3)
Mortality	50,722 (0.4)	15,268 (0.3)

[a]ED: Emergency department

Table 3 Number of DSPs incorporated into the calculation and the corresponding prediction capabilities

Maximum Number of DSPs[a]	AUROC[b]	AUPRC[c]
1	0.969 (0.968–0.970)	0.671 (0.664–0.679)
2	0.969 (0.968–0.970)	0.670 (0.662–0.677)
3	0.969 (0.968–0.970)	0.670 (0.663–0.678)
4	0.969 (0.968–0.970)	0.671 (0.663–0.678)
5	0.969 (0.968–0.970)	0.671 (0.663–0.678)

[a]DSP: Diagnosis-specific survival probability; [b]AUROC: Area under the receiver operating characteristic curve; [c]AUPRC: Area under the precision–recall curve

was 0.880 (0.875–0.885), and the specificity was 0.958 (0.958–0.959). The Brier score was calculated as 0.001 using the validation dataset. The calibration plot revealed that the ESS tends to underestimate survival in patients with severe disease, with an expected survival probability of less than 50% (Fig. 3).

Table 5 summarizes the performance of individual parameters and combinations of parameters. Table 6 presents the accuracy, sensitivity and specificity for different thresholds.

Discussion

Scoring systems designed to determine disease severity are necessary not only for predicting adverse outcomes in patients but also for quality improvement and preventive measures [16]. We developed a model to predict the mortality of emergency patients on the basis of the initial clinical variables and the diagnosis coded in the ICD with good accuracy. The newly developed severity score performs better than previously reported severity scores, such as the SOFA or APACHE II scores [5, 16, 17]. Our severity score has an AUROC of 0.975 for the prediction of mortality, outperforming that of the SOFA and APACHE II with values between 0.7 and 0.8 when applied to patients in the ED [4, 5, 16]. Our score has additional advantages in that it does not require laboratory values and is easy to calculate from large datasets collected from ED patients at the national level [2].

In the field of traumatology, scoring systems such as the ISS, TRISS and ICISS have been used to compare the performance of trauma care between systems or for quality improvement [3, 18]. W-scores calculated from the survival ratio predicted by the TRISS have been used as a method of comparing trauma care results between trauma care systems [12, 19, 20]. The TRISS is calculated by combining the patient's age, type of injury, RTS, and ISS to estimate the probability of survival. The RTS represents patients' physiological responses to injury, and the ISS represents anatomical injury [9]. The TRISS has been considered the standard tool for measuring trauma severity for decades. However, ISS, a representation of anatomical injury severity, has been criticized for relying on a consensus rather than an empirically derived scale, and it requires personnel extensively trained in AIS coding [7, 9, 21]. The ICISS was proposed as an alternative to the ISS and has advantages, including simpler calculations because it is based on the ICD code system, which is an official system of disease and mortality statistics [21]. In a study based on seven countries, the ICISS was combined with age and sex in a logistic regression model to predict in-hospital mortality, with an AUROC of 0.87, and the differences from the use of country-specific or pooled DSPs were minor [11]. The use of our model is not limited to those with trauma. Instead, the concept of the ICDSS is extended to all ED patients, and the model can be used to predict mortality while also considering age and physiological parameters. Considering that the prediction model based on the TRISS has been reported to have an AUROC of 0.98 for penetrating trauma and an AUROC of 0.84 for blunt trauma [22, 23], our AUROC of 0.975 in all ED patients, including both trauma patients and disease patients, could be regarded as successful work without necessitating additional coding efforts. Moreover, with the aging population, underlying medical conditions and medical complications are becoming

Table 4 Derivation and validation of logistic regression models of the fourfold cross-validation process

	Fold 1	Fold 2	Fold 3	Fold 4
Derivation set size (n)	10,272,021	10,269,926	10,271,579	10,272,824
Validation set size (n)	3,423,429	3,425,524	3,423,871	3,422,626
Regression coefficients				
Intercept	-2.593	-2.532	-2.565	-2.541
Age score	0.307	0.308	0.308	0.308
MEWS[a]	0.608	0.606	0.608	0.609
ICESS[b]	-6.385	-6.447	-6.413	-6.461
Youden's threshold	-5.900	-5.874	-5.869	-5.916
AUROC[c]	0.975 (0.974—0.976)	0.974 (0.973—0.976)	0.975 (0.973—0.976)	0.973 (0.972—0.975)
AUPRC[d]	0.741 (0.739—0.753)	0.738 (0.734—0.749)	0.746 (0.742—0.757)	0.735 (0.732—0.746)
True positive (n)	11,351	11,237	11,294	11,551
False positive (n)	201,366	192,619	192,770	200,229
True negative (n)	3,209,396	3,220,223	3,218,434	3,209,427
False negative (n)	1316	1445	1373	1419
Accuracy	0.941 (0.940—0.944)	0.943 (0.940—0.945)	0.943 (0.941—0.944)	0.941 (0.940—0.943)
Sensitivity	0.896 (0.890—0.090)	0.886059 (0.882—0.893)	0.8916081 (0.887—0.899)	0.8905937 (0.885—0.896)
Specificity	0.942 (0.941—0.944)	0.944 (0.941—0.945)	0.943 (0.941—0.944)	0.941 (0.940–0.944)

The numbers in parentheses represent 95% confidence intervals

[a]MEWS: Medical early warning score; [b]ICESS: International Classification of Diseases-based emergency severity score; [c]AUROC: Area under the receiver operating characteristic curve; [d]AUPRC: Area under the precision–recall curve

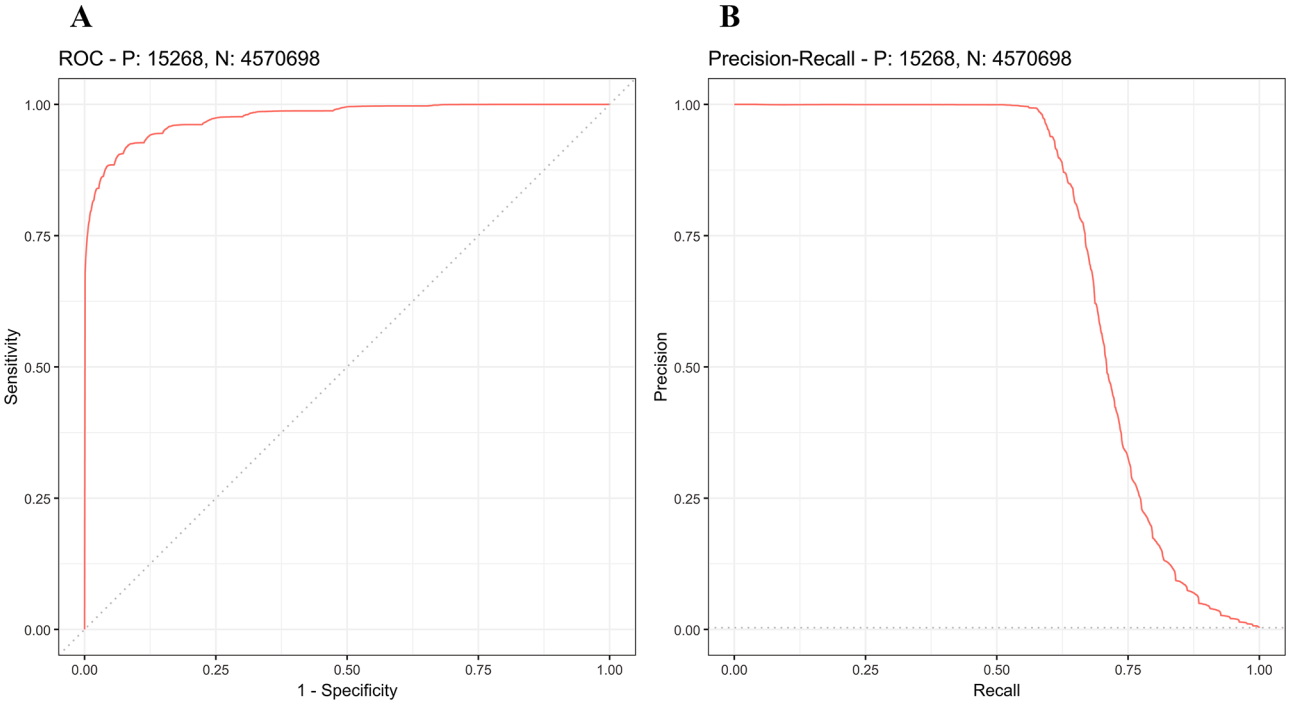

Fig. 2 The performance of the Emergency Severity Score (ESS) developed in the present study in predicting in-hospital mortality in patients visiting emergency centers in the validation dataset. (**A**) The receiver operating characteristic curve yielded an area under the curve of 0.975 (95% CI: 0.974–0.977). (**B**) The precision–recall curve yielded an area under the curve of 0.724 (95% CI: 0.717–0.731)

increasingly important for injured patients. Compared with the ICISS, our model has an advantage: we included DSPs of medical diagnoses with injured patients.

Our model achieved an AUPRC of 0.724 in the test dataset. In highly imbalanced datasets, such as ours (where survival is significantly more frequent than mortality), the PRC is considered a more suitable performance metric than the ROC curve [24]. Unlike the AUROC, which is always above 0.5, the baseline AUPRC is equal to the prevalence of positive cases. Additionally, there is an unreachable area in the precision–recall space [25]. In our dataset, the baseline AUPRC is 0.003, making

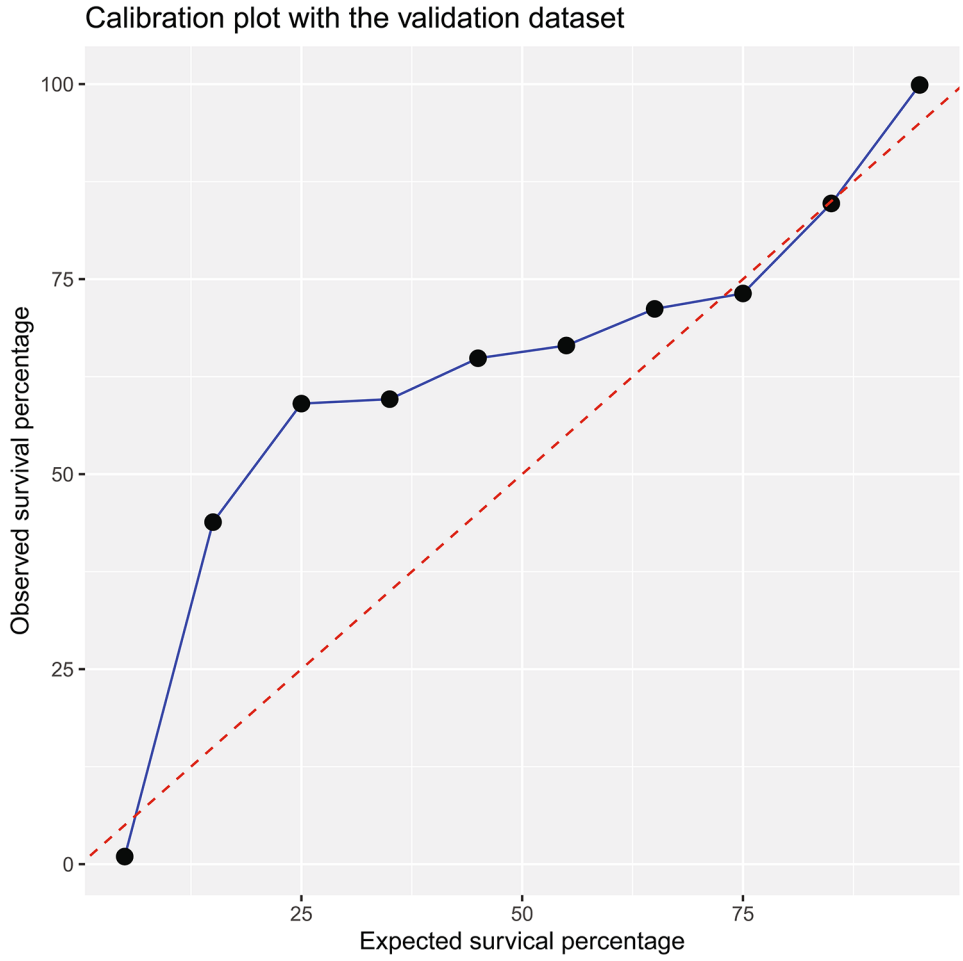

Fig. 3 The calibration plot presents the relationship between the expected survival based on the Emergency Severity Score (ESS) developed in the present study and the observed survival in the test dataset. The observed survival was greater than the range of predicted survival (less than 0.75), indicating that the ESS overestimates severity in the corresponding range

Table 5 The performance of the prediction models on the basis of individual parameters and a combination of parameters in the test dataset

Model parameters	AUROC	AUPRC	Accuracy	Sensitivity	Specificity
ICESS	0.969 (0.968—0.970)	0.6715 (0.664—679)	0.868 (0.868—0.890)	0.925 (0.905—0.929)	0.868 (0.868—0.890)
MEWS	0.956 (0.953—0.957)	0.653 (0.646—0.661)	0.916 (0.916—0.916)	0.883 (0.878—0.888)	0.916 (0.916—0.916)
AGE	0.775 (0.771—0.778)	0.020 (0.019—0.021)	0.839 (0.838—0.839)	0.586 (0.578—0.594)	0.839 (0.839—0.840)
ICESS + AGE	0.956 (0.955—0.958)	0.663 (0.656—0.671)	0.913 (0.913—0.914)	0.857 (0.851—0.862)	0.913 (0.913—0.915)
ICESS + MEWS	0.973 (0.971—0.974)	0.723 (0.716—0.730)	0.945 (0.941—0.946)	0.890 (0.885—0.896)	0.945 (0.941—0.945)
MEWS + AGE	0.965 (0.964—0.967)	0.671 (0.664—0.678)	0.934 (0.934—0.935)	0.881 (0.876—0.886)	0.935 (0.934—0.935)
ICESS + MEWS + AGE	0.975 (0.974—0.977)	0.724 (0.717–0.731)	0.958 (0.958—0.958)	0.880 (0.875–0.885)	0.958 (0.958—0.989)

The numbers in parentheses represent 95% confidence intervals

[a]MEWS: medical early warning score; [b]ICESS: International Classification of Diseases-based emergency severity score; [c]AUROC: area under the receiver operating characteristic curve; [d]AUPRC: area under the precision–recall curve

0.724 a sufficiently high value. However, as the AUPRC is not commonly reported in the medical literature, direct comparisons with other studies are limited [24].

The calibration plot revealed that actual survival was greater than predicted survival in the higher severity group, with an expected survival probability of less than 0.75. This uneven calibration causes the case-mix problem when comparing the treatment results between groups with different severity distributions by employing the W statistic or observed-to-expected ratio. The case-mix problem with the W statistic using TRISS was acknowledged as early as 1995 [12]. SAPS 2 and APACHE II scores also suffer from uneven calibration [26, 27]. To overcome the case-mix problem, a method of

Table 6 Accuracy, sensitivity and specificity of the proposed prediction model in the test dataset at different thresholds

Survival probability	Threshold	True positive	False positive	True negative	False negative	Accuracy	Sensitivity	Specificity
0.1	2.197	8824	87	4,570,611	6444	0.999 (0.999—0.999)	0.578 (0.570—0.586)	1.000 (1.000—1.000)
0.2	1.386	8829	169	4,570,529	6339	0.999 (0.999—0.999)	0.585 (0.577—0.593)	1.000 (1.000—1.000)
0.3	0.847	9049	342	4,570,356	6219	0.999 (0.999—0.999)	0.593 (0.585—0.600)	1.000 (1.000—1.000)
0.4	0.405	9154	497	4,570,201	6114	0.999 (0.999—0.999)	0.600 (0.592—0.607)	1.000 (1.000—1.000)
0.5	0	9329	820	4,569,878	5939	0.999 (0.998—0.999)	0.611 (0.603—0.619)	1.000 (1.000—1.000)
0.6	-0.405	9457	1074	4,569,624	5811	0.998 (0.998—0.999)	0.619 (0.612—0.627)	1.000 (1.000—1.000)
0.7	-0.847	9695	1662	4,569,036	5573	0.998 (0.998—0.998)	0.635 (0.627—0.643)	1.000 (1.000—1.000)
0.8	-1.386	9951	2360	4,568,338	5317	0.998 (0.998—0.998)	0.652 (0.644—0.659)	0.999 (0.999—1.000)
0.9	-2.197	10,332	4470	4,566,228	4936	0.998 (0.998—0.998)	0.677 (0.669—0.684)	0.999 (0.999—0.999)
0.99	-4.595	11,961	42,963	4,527,735	3307	0.990 (0.990—0.990)	0.783 (0.777—0.790)	0.991 (0.991—0.991)
0.999	-6.907	14,434	680,757	3,889,941	834	0.851 (0.851—0.852)	0.945 (0.942—0.949)	0.851 (0.851—0.851)
Youden	-5.869	13,439	190,555	4,380,143	1829	0.958 (0.958—0.958)	0.880 (0.875—0.885)	0.958 (0.958—0.958)

The numbers in parentheses represent 95% confidence intervals

standardizing the W statistic to the benchmark severity distribution has been suggested [12].

The ICISS was defined as the product of DSPs, and the performance was reported to increase as the number of DSPs incorporated into the calculation increased to five [7]. Unlike the ICISS, the performance of our model did not increase with the number of diagnoses considered. This may be because diagnoses of medical conditions are more closely related to each other than are diagnoses of injuries. For example, when the diagnoses "R572, septic shock" and "J181, lobar pneumonia" are present in a patient, the patient's condition should be interpreted as "septic shock caused by lobar pneumonia", and a more severe diagnostic code would represent the patient's condition better than the product of the two DSPs. On the other hand, "S064, epidural hematoma" and "S7231, fracture of shaft of femur" were more independent.

Our model was affected by the limitations inherent in the ICD-10 coding system. The ICD-10 cannot be used to differentiate between small liver lacerations and more severe lacerations. Separate codes are not available for bilateral pneumothorax or tension pneumothorax. Therefore, the precision of our model was limited by the use of the ICD coding system.

Our scoring system did not include laboratory results or pulse oximetry data. These variables might have improved the predictive performance of the model. However, we found that pulse oximetry was not routinely measured in all patients in every emergency center. Therefore, including pulse oximetry in the calculation would suffer from substantial selection bias, so we decided to exclude this parameter. We could not include laboratory results because the database did not contain those data. However, we believe that laboratory values would have been affected by the same selection bias problems.

The treatment results of emergency patients are heavily dependent on the population and health system. Our results cannot be generalized to countries other than Korea. However, our model and the supplemental DSP table could be regarded as benchmarks when applied to other countries.

Conclusions

In conclusion, we developed a scoring system to predict the in-hospital mortality of emergency patients on the basis of ICD codes, age and vital signs and achieved good performance. This scoring system would be useful for standardizing the severity of emergency patients and comparing treatment results.

Abbreviations

ED Emergency department
SOFA Sequential Organ Failure Assessment
APACHE Acute Physiology and Chronic Health Assessment

TRISS	Trauma and Injury Severity Score
ISS	Injury Severity Score
RTS	Revised Trauma Score
AIS	Abbreviated Injury Scale
ICD	International Classification of Diseases
ICISS	International Classification of Diseases-based Injury Severity Score
AUROC	Area under the receiver operating characteristic curve
CI	Confidence interval
NEDIS	National Emergency Department Information System
NEMC	National Emergency Medical Center
AVPU	Alert, voice, pain, unresponsive
DSP	Diagnosis-specific survival probability
SRR	Survival risk ratio
ICESS	ICD-based Emergency Severity Score
MEWS	Modified Early Warning Score
AUPRC	Area under the precision–recall curve
ESS	Emergency Severity Score

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12873-025-01214-y.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
JEK drafted the manuscript. JJ designed the study and performed the statistical analysis. SWL conceptualized the study and participated in interpreting the results. YC contributed to the acquisition and analysis of the data. All the authors were involved in the revision of the manuscript and provided intellectual content of critical importance. All the authors have read and gave final approval of the version to be published.

Funding
The study was supported by the Dong-A University Research Fund.

Data availability
The data that support the findings of this study are available from the National Emergency Medical Center of the Republic of Korea, but restrictions apply to the availability of these data, which were used under license for the current study and therefore are not publicly available.

Declarations

Ethics approval and consent to participate
The institutional review board of Dong-A University Hospital determined that the study was exempt from formal approval because it involved a deidentified version of the preexisting national dataset. The need for informed consent was deemed exempt because the study involved a deidentified version of the preexisting national dataset. (DAUHIRB-EXP-21-065).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 11 November 2024 / Accepted: 28 March 2025
Published online: 05 April 2025

References

1. Pines JM, Fee C, Fermann GJ, Ferroggiaro AA, Irvin CB, Mazer M, Frank Peacock W, Schuur JD, Weber EJ, Pollack CV. The role of the society for academic emergency medicine in the development of guidelines and performance measures. Acad Emerg Med. 2010;17(11):e130–140.
2. Jeong J, Lee SW, Kim WY, Han KS, Kim SJ, Kang H. Development and validation of a scoring system for mortality prediction and application of standardized W statistics to assess the performance of emergency departments. BMC Emerg Med. 2021;21(1).
3. Mock C, Juillard C, Brundage S, Goosen J, M J. Guidelines for trauma quality improvement programmes. Geneva, Switzerland: WHO; 2009.
4. Rahmatinejad Z, Tohidinezhad F, Reihani H, Rahmatinejad F, Pourmand A, Abu-Hanna A, Eslami S. 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. 2020;38(9):1841–6.
5. Abdullah SOB, Sørensen RH, Nielsen FE. Prognostic accuracy of SOFA, qSOFA, and SIRS for mortality among emergency department patients with infections. Infect Drug Resist. 2021;14:2763–75.
6. Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. Trauma score and the injury severity score. J Trauma. 1987;27(4):370–8.
7. Osler T, Rutledge R, Deis J, Bedrick E. ICISS: an international classification of disease-9 based injury severity score. J Trauma. 1996;41(3):380–6. discussion 386–388.
8. Tohira H, Jacobs I, Mountain D, Gibson N, Yeo A. Systematic review of predictive performance of injury severity scoring tools. Scand J Trauma Resusc Emerg Med. 2012;20:63.
9. Wong SSN, Leung GKK. Injury severity score (ISS) vs. ICD-derived injury severity score (ICISS) in a patient population treated in a designated Hong Kong trauma centre. Mcgill J Med. 2008;11(1):9–13.
10. Stausberg J, Hagn S. New morbidity and comorbidity scores based on the structure of the ICD-10. PLoS ONE. 2015;10(12):e0143365.
11. Gedeborg R, Warner M, Chen LH, Gulliver P, Cryer C, Robitaille Y, Bauer R, Ubeda C, Lauritsen J, Harrison J, et al. Internationally comparable diagnosis-specific survival probabilities for calculation of the ICD-10-based injury severity score. J Trauma Acute Care Surg. 2014;76(2):358–65.
12. Hollis S, Yates DW, Woodford M, Foster P. Standardized comparison of performance indicators in trauma: a new approach to case-mix variation. J Trauma. 1995;38(5):763–6.
13. Yang W, Jiang J, Schnellinger EM, Kimmel SE, Guo W. Modified Brier score for evaluating prediction accuracy for binary outcomes. Stat Methods Med Res. 2022;31(12):2287–96.
14. Saito T, Rehmsmeier M. Precrec: fast and accurate precision–recall and ROC curve calculations in R. Bioinformatics. 2017;33(1):145–7.
15. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S + to analyze and compare ROC curves. BMC Bioinformatics. 2011;12(1):77.
16. Kammar-García A, Castillo-Martínez L, Mancilla-Galindo J, Villanueva-Juárez JL, Pérez-Pérez A, Rocha-González HI, Arrieta-Valencia J, Remolina-Schlig M, Hernández-Gilsoul T. SOFA score plus impedance ratio predicts mortality in critically ill patients admitted to the emergency department: retrospective observational study. Healthcare. 2022;10(5):810.
17. Thakur R, Naga Rohith V, Arora JK. Mean SOFA Score in Comparison With APACHE II Score in Predicting Mortality in Surgical Patients With Sepsis. Cureus. 2023.
18. Mock C, Lormand JD, Goosen J, Joshipura M, M P. Guidelines for essential trauma care. Geneva: World Health Organization; 2004.
19. Jung K, Huh Y, Lee JC, Kim Y, Moon J, Youn SH, Kim J, Kim TY, Kim J, Kim H. Reduced mortality by Physician-Staffed HEMS dispatch for adult blunt trauma patients in Korea. J Korean Med Sci. 2016;31(10):1656–61.
20. Kim OH, Roh YI, Kim HI, Cha YS, Cha KC, Kim H, Hwang SO, Lee KH. Reduced mortality in severely injured patients using Hospital-based helicopter emergency medical services in interhospital transport. J Korean Med Sci. 2017;32(7):1187–94.
21. Kim Y, Jung KY, Kim CY, Kim YI, Shin Y. Validation of the international classification of diseases 10th Edition-based injury severity score (ICISS). J Trauma. 2000;48(2):280–5.
22. de Munter L, Polinder S, Lansink KW, Cnossen MC, Steyerberg EW, de Jongh MA. Mortality prediction models in the general trauma population: A systematic review. Injury. 2017;48(2):221–9.
23. Millham FH, LaMorte WW. Factors associated with mortality in trauma: re-evaluation of the TRISS method using the National trauma data bank. J Trauma. 2004;56(5):1090–6.
24. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE. 2015;10(3):e0118432–0118432.

25. Boyd K, Santos Costa V, Davis J, Page CD. Unachievable region in Precision-Recall space and its effect on empirical evaluation. Proc Int Conf Mach Learn. 2012;2012:349.
26. Jahn M, Rekowski J, Gerken G, Kribben A, Canbay A, Katsounas A. The predictive performance of SAPS 2 and SAPS 3 in an intermediate care unit for internal medicine at a German university transplant center; A retrospective analysis. PLoS ONE. 2019;14(9):e0222164.
27. Vandenbrande J, Verbrugge L, Bruckers L, Geebelen L, Geerts E, Callebaut I, Gruyters I, Heremans L, Dubois J, Stessel B. Validation of the acute physiology and chronic health evaluation (APACHE) II and IV score in COVID-19 patients. Crit Care Res Pract. 2021:1–9.

Publisher's note