Comparison of Six Scoring Systems for Predicting In-hospital Mortality among Patients with SARS-COV2 Presenting to the Emergency Department

Zahra Rahmatinejad[1], Benyamin Hoseini[2], Hamidreza Reihani[3], Ameen Abu Hanna[4], Ali Pourmand[5], Seyyed Mohammad Tabatabaei[6], Fatemeh Rahmatinejad[7], Saeid Eslami[8]

ABSTRACT

Background: The study aimed to compare the prognostic accuracy of six different severity-of-illness scoring systems for predicting in-hospital mortality among patients with confirmed SARS-COV2 who presented to the emergency department (ED). The scoring systems assessed were worthing physiological score (WPS), early warning score (EWS), rapid acute physiology score (RAPS), rapid emergency medicine score (REMS), national early warning score (NEWS), and quick sequential organ failure assessment (qSOFA).

Materials and methods: A cohort study was conducted using data obtained from electronic medical records of 6,429 confirmed SARS-COV2 patients presenting to the ED. Logistic regression models were fitted on the original severity-of-illness scores to assess the models' performance using the Area Under the Curve for ROC (AUC-ROC) and Precision-Recall curves (AUC-PR), Brier Score (BS), and calibration plots were used to assess the models' performance. Bootstrap samples with multiple imputations were used for internal validation.

Results: The mean age of the patients was 64 years (IQR:50–76) and 57.5% were male. The WPS, REMS, and NEWS models had AUROC of 0.714, 0.705, and 0.701, respectively. The poorest performance was observed in the RAPS model, with an AUROC of 0.601. The BS for the NEWS, qSOFA, EWS, WPS, RAPS, and REMS was 0.18, 0.09, 0.03, 0.14, 0.15, and 0.11 respectively. Excellent calibration was obtained for the NEWS, while the other models had proper calibration.

Conclusion: The WPS, REMS, and NEWS have a fair discriminatory performance and may assist in risk stratification for SARS-COV2 patients presenting to the ED. Generally, underlying diseases and most vital signs are positively associated with mortality and were different between the survivors and non-survivors.

Keywords: COVID-19, Emergency department, Mortality prediction, Performance measures, Scoring system.

Indian Journal of Critical Care Medicine (2023): 10.5005/jp-journals-10071-24463

HIGHLIGHTS

- Study compared six different severity-of-illness scores for predicting in-hospital mortality among confirmed COVID-19 patients presenting to the emergency department.
- The WPS, REMS, and NEWS models had fair discriminatory performance, with AUROC values ranging from 0.701 to 0.714.
- The NEWS model had excellent calibration, while the other models had proper calibration, indicating that these scores may support risk stratification for COVID-19 patients presenting to the emergency department.

INTRODUCTION

Since March 11, 2020, the coronavirus disease 2019 (COVID-19) outbreak has been declared a global pandemic by the World Health Organization (WHO) due to its rapid spread across the globe.[1–3] This pandemic is caused by a novel virus, SARS-CoV-2, which shares similarities with the virus responsible for the severe acute respiratory syndrome (SARS) outbreak in 2003.[1] SARS-CoV-2 belongs to the Coronaviridae family and possesses a single-stranded positive-sense RNA genome.[1,2] The general symptoms of COVID-19 infection include fever, cough, and fatigue.[3–6] However, patients can develop anorexia and/or diarrhea,[3,6–8] dyspnea, chest

[1,6]Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

[2]Pharmaceutical Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran

[3]Department of Emergency Medicine, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

[4]Department of Medical Informatics, Amsterdam UMC – Location AMC, University of Amsterdam, the Netherlands

[5]Department of Emergency Medicine, The George Washington University, School of Medicine and Health Sciences, Washington DC, United States

[7]Department of Health Information Technology, Faculty of Paramedical Sciences, Mashhad University of Medical Sciences, Mashhad, Iran

[8]Department of Medical Informatics, Faculty of Medicine; Pharmaceutical Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran; Department of Medical Informatics, Amsterdam UMC – Location AMC, University of Amsterdam, the Netherlands

Corresponding Author: Hamidreza Reihani, Department of Emergency Medicine, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran, Phone: +98 5138525312, e-mail: Reihanihr@mums.ac.ir

pain, and cardiovascular involvement such as acute dysfunction of the left ventricle of the heart, arrhythmia, myocardial inflammation, microvascular injury, and thrombosis.[9]

The COVID-19 pandemic has led to an unprecedented volume of hospital visits and admissions, including those with suspected COVID-19, resulting in elevated levels of overcrowding in hospitals and healthcare centers.[1,10,11] This circumstance has contributed to a delay in care and poorer outcomes, such as increased morbidity and mortality, especially in emergency departments (EDs) where resources are limited.[3,12,13] Critically ill patients are particularly vulnerable to encountering these adverse events.[14–16] Given the susceptibility of critically ill patients to experiencing adverse events in hospitals and healthcare centers, it is crucial to implement an efficient triage system capable of identifying and prioritizing those at greater risk of adverse outcomes.[14,15,17–19]

Various scoring systems have been developed and implemented in the emergency department that can be useful for prioritizing critically ill patients and predicting mortality for triage.[18–21] These models primarily rely on patients' vital signs and level of consciousness.[21–23] The present study incorporates six such models: Study: National Early Warning Score (NEWS), quick Sequential Organ Failure Assessment (qSOFA), Early Warning Score (EWS), Worthing Physiological Score (WPS), Rapid Acute Physiology Score (RAPS), and Rapid Emergency Medicine Score (REMS).[24–29]

These models are characterized by their ability to rapidly identify critically ill patients in need of urgent intervention through the use of easily obtainable bedside parameters. However, their ability in predicting mortality for patients with COVID-19 in the ED setting remains unclear. Therefore, the primary objective of this study was to evaluate and compare the performance of six commonly used models, namely the NEWS, qSOFA, EWS, WPS, RAPS, and REMS in predicting mortality among patients with COVID-19. Additionally, the study aimed to investigate differences in demographic and clinical characteristics between survivors and

non-survivors. We performed the study on a cohort of patients from Iran, a country with a high number of confirmed COVID-19 cases and related deaths.[3,6,11,13,30]

MATERIALS AND METHODS

Study Design

This single-center, cohort study was conducted at Emam Reza Hospital, which is located in Mashhad, Northeast of Iran. This center is known as a referral university hospital and one of the main centers for SARS-COV2 patients. The annual volume of ED visits in this center is more than 200,000 visits. The study duration for this report was from 29 February to 30 July 2021. We obtained university ethics committee approval *(Number: IR.MUMS.REC.1400.141)*.

Data Collection

The data of all confirmed SARS-COV2 patients who were admitted to ED, was extracted from the electronic medical records of a university hospital in Mashhad, Iran. The patients with respiratory symptoms underwent assessment by both RT-PCR tests using throat and nose swab specimens and lung imaging incredibly high-resolution Computed Tomography (CT) scans. A radiologist who was blind to clinical diagnosis, management, and outcome reported the

Flowchart 1: Flowchart of patient selection process and reasons for exclusion
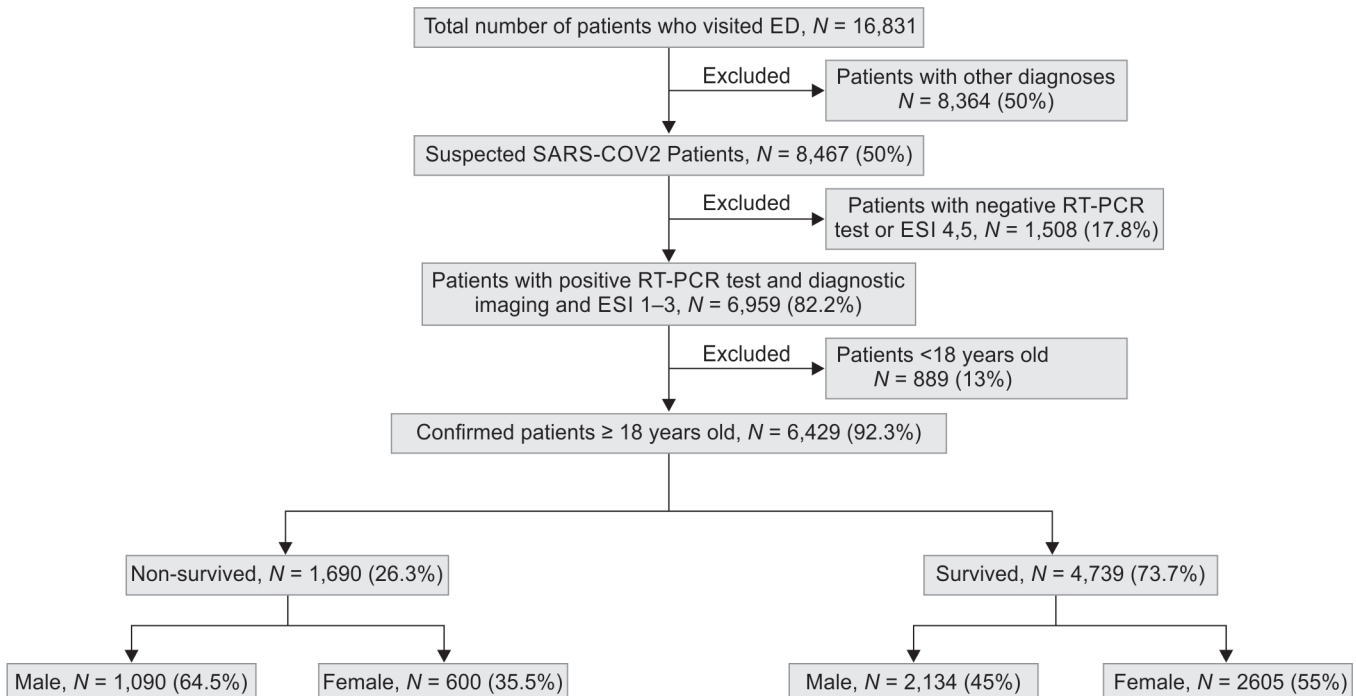
Table 1: The point assignment scheme of each scoring system

Model (Min–Max)	Temperature (°C)	SBP (mm Hg)	MAP (mm Hg)	RR (breaths/min)	Pulse (beats/min)	GCS	AVPU	O_2 sat (%)	O_2 therapy	Age (year)
				Variables						
NEWS (0–19)	≤35→3 35.1–36→1 36.1–38→0 38.1–39→1 ≥39.1→2	≤90→3 91–100→2 101–110→1 111–219→0 ≥220→3	NA	≤8→3 9–11→1 12–20→0 21–24→2 ≥25→3	≤40→3 41–50→1 51–90→0 91–110→1 111–130→2 ≥131→3	NA	A→0 Other→3	≤91→3 92–93→2 94–95→1 ≥96→0	Yes→2 No→0	NA
qSOFA (0–3)	NA	≤100→1	NA	≥22→1	NA	≤14→1	NA	NA	NA	NA
EWS (0–18)	<35→2 35.1–38→0 38–39.5→2 >39.5→3	<80→3 81–90→2 91–100→1 101–199→0 >200→2	NA	<8→3 9–19→0 20–25→1 25–29→2 >30→3	<40→3 41–50→2 51–100→0 101–110→1 111–129→2 >130→3	NA	A→0 V→1 P→2 U→3	<85→3 85–89→2 90–94→1 >95→0	NA	NA
WPS (0–14)	≥35.3→0 <35.3→3	≥100→0 ≤99→2	NA	≤19→0 20–21→1 ≥22→2	≤101→0 ≥102→1	NA	A→0 Other→3	96–100→0 94–95→1 92–93→2 <92→3		NA
RAPS (0–16)	NA	NA	70–109→0 50–69→2 110–129→2 130–159→3 ≤49→4 ≥160→4	12–24→0 10–11→1 25–34→1 6–9→2 35–49→3 ≥5→4 ≤50→4	70–109→0 50–69→2 110–139→2 40–54→3 140–179→3 ≤39→4 ≥180→4	≥14→0 11–13→1 8–10→2 5–7→3 ≤4→4	NA	NA	NA	NA
REMS (0–26)	NA	NA	70–109→0 50–69→2 110–129→2 130–159→3 ≤49→4 ≥160→4	12–24→0 10–11→1 25–34→1 6–9→2 35–49→3 ≥5→4 ≤50→4	70–109→0 50–69→2 110–139→2 40–54→3 140–179→3 ≤39→4 ≥180→4	≥14→0 11–13→1 8–10→2 5–7→3 ≤4→4	NA	>89→0 86–89→1 75–85→3 <75→4	NA	<45→0 45–54→2 55–64→3 65–73→5 ≥74→6

images. Then, the patients were confirmed to be COVID-19 infected based on positive real-time reverse transcription-polymerase chain (RT-PCR) testing and were included in the study. During the study period, 16,831 suspected SARS-COV2 patients were presented in this center (Flowchart 1). The database contains demographic and clinical data including age, gender, diagnosis, level of triage, vital signs (i.e., f9obody temperature, respiratory rate, systolic blood pressure, heart rate, and mental status, cyanosis, distress), comorbidities (i.e., diabetes, hypertension, cancer, coronary heart disease (CHD), asthma, cerebrovascular disease (CVD), chronic kidney disease (CKD), and living status (alive or deceased). The total scores of NEWS, qSOFA, EWS, WPS, RAPS, and REMS were calculated according to the variables included in each score (Table 1).

Eligibility Criteria

All adult patients (≥18 years) with an emergency severity index (ESI) of 1, 2, and 3 (resuscitation, emergent, and urgent) were included.

Statistical Analysis

When reporting descriptive statistics of the sample, the Kolmogorov-Smirnov (KS) test was used to test the normality of continuous variables. Normally distributed variables are reported as means and standard deviations, and; non-normally distributed variables as median and interquartile ranges (IQRs). Categorical variables were presented as frequencies and percentages. We used non-parametric tests, as appropriate, to compare groups: The Mann-Whitney U-test, Fisher's exact test, or the Chi-square test. p-values less than 0.05 were considered to indicate statistical significance.

We handled missing data by applying multiple imputations within bootstrap samples. Specifically, we used the boot MI method in which we applied 200 bootstrap samples that included the missing values and then generated three imputed datasets for each sample.[31] We fitted a logistic regression model on each imputed dataset and calculated the average of the model's coefficients of each of the three imputed datasets belonging to a bootstrap sample. These 200 (averaged) results were then used to calculate the means and confidence intervals of the results. The logistic regression model for each of the NEWS, EWS, REMS, WPS, RAPS, and qSOFA scores included the score as the independent variable and hospital mortality as the response. The probability of mortality is calculated by using the logit formula: where X is the score, and β_0 and β_1 are the model's coefficients.

We assessed the predictive performance of the models by the area under the receiver operating curve (AUC) for discrimination; the area under the precision-recall curve (AUPRC) for the balance between the positive predictive value and sensitivity; and the Brier score (BS) for the accuracy of probabilistic prediction. We used calibration graphs to inspect the agreement between

the predictions and the true proportion of mortality (using one imputation of the original dataset with 1,000 bootstraps). The De-long test was used for AUC pairwise comparison. The Youden Index was applied to find the best cutoff for the specificity, sensitivity, positive predictive value (PPV), negative predictive value (NPV), and classification accuracy. All analyses were performed in R using pROC, rms, PRROC, and the bootImpute packages.

Results

Table 2 presents the baseline characteristics of the study cohort, highlighting the key finding that the hospital mortality rate was 26.3 and 57.5% of participants were male.

The study found significant differences in age, gender Glasgow coma scale (GCS), systolic blood pressure (SBP), respiratory rate (RR), oxygen saturation (SpO$_2$), mean arterial pressure (MAP), pulse, distress, cyanosis, pain, alert, verbal, pain, unresponsive (AVPU) score, and comorbidities such as diabetes, hypertension, cancer, coronary heart disease (CHD), and cardiovascular disease (CVD) between non-survivors and survivors. Triage level in the non-survivors was predominantly emergent (level 2) compared to survivors (74.9%). Approximately 44.3% of patients who died were transported by ambulance. The median age of non-survivors was higher (70, IQR: 60–80) than survivors (62, IQR: 47–74), and non-survivors were predominantly elderly men with comorbidities, especially cancer and cardiovascular disease. The non-survivor group had slightly lower GCS. The median SpO$_2$ of survivors was 90 (85–93) compared with 85 (77–90) for non-survivors. At the presentation to ED, there was a large proportion of patients with lower MAP in non-survivors as compared to the group of survivors 90 (79–100) vs 93 (83.3–100).

Based on the AVPU scale, 6,151patients (96.3%) were graded as "alert", 154 patients (2.4%) as "vocally responsive", 68 patients (6.1%) as "painfully responsive", and 56 patients as "unresponsive" (0.2%). All median scores were significantly lower in the non-survivors than the survivors.

There were on average 28.5% missing values in the data. After applying multiple imputations within bootstrap samples, the estimates of the intercept and slope (and their 95% confidence intervals) in the linear predictors of the logistic regression models were:

- For NEWS: –2.65 (–2.73, –2.63) and 0.31 (0.30, 0.33)
- For qSOFA: –1.54 (–1.58, –1.5) and 0.97 (0.94, 1.04)
- For EWS: –2.92 (–2.96, –2.7) and 0.28 (0.26, 0.31)
- For WPS: –2.86 (–2.99, –2.83) and 0.46 (0.45, 0.48)
- For RAPS: –1.67 (–1.7, –1.6) and 0.20 (0.19, 0.22)
- For REMS: –3.53 (–3.57, –3.5) and 0.22 (0.21, 0.24)

This means for example that the linear predictor of the NEWS model is –2.65 + 0.31 *NEWS and that the probability of mortality is

$$P = \frac{1}{1 + e^{-(-2.65 + 0.31*NEWS)}}$$

The ROC and PR plots are shown in Figure 1. Table 3 displays the corresponding bootstrap-validated predictive performance

**Table 2:** Baseline characteristics of the patients with SARS-COV2

Characteristics	Alive (N = 4,739)	Dead (N = 1,690)	Total (N = 6,429)	p-value
Demographic				
Age (year)	62 (47–74)	70 (60–80)	64 (50–76)	<0.001[a]
Gender				
Male	2,605 (55%)	1,090 (64.5%)	3,695 (57.5%)	<0.001[b]
Female	2,134 (45%)	600 (35.5%)	2,734 (42.5%)	
Clinical parameters				
GCS	15 (15–15)	15 (14–15)	15 (15–15)	<0.001[a]
SBP (mm Hg)	120 (110–140)	120 (108–140)	120 (110–140)	0.002[a]
TEMPR (°C)	37 (36.5–37)	37 (37–37.5)	37 (37–37)	0.1[a]
RR (breaths/min)	18 (17–20)	19 (18–22)	18 (18–20)	<0.001[a]
SpO$_2$ (%)	90 (85–93)	85 (77–90)	89 (84–92)	<0.001[a]
MAP (mm Hg)	93 (83.3–100)	90 (79–100)	93.3 (83–100)	0.003[a]
Pulse (beats/min)	95 (85–108)	98 (87–111)	95 (86–110)	<0.001[a]
Distress	52 (1.1%)	71 (4.2%)	123 (1.9%)	<0.001[b]
Cyanoses	6 (0.1%)	14 (0.8%)	20 (0.3%)	<0.001[b]
Pain	214 (4.5%)	144 (8.5%)	358 (5.6%)	<0.001[b]
AVPU	4,596 (97%)	1,555 (93%)	6,151 (96.3%)	0.001[c]
Alert	104 (2.2%)	50 (3%)	154 (2.4%)	
Voice	19 (0.4%)	49 (3%)	68 (1%)	
Pain	20 (0.1%)	36 (1%)	56 (0.2%)	
Unresponsive				
Comorbidities				
Diabetes	736 (15.5%)	325 (19.2%)	1,061 (16.5%)	0.001[b]
Hypertension	759 (16%)	352 (19.2%)	1,084 (16.9%)	0.001[b]

(Contd..)

Table 2: *(Contd..)*

Characteristics	Alive (N = 4,739)	Dead (N = 1,690)	Total (N = 6,429)	p-value
Cancer	99 (2.1%)	53 (3.1%)	152 (2.4%)	0.015
CKD	199 (4.2%)	99 (5.9%)	298 (4.6%)	0.005[b]
CHD	399 (8.4%)	169 (10%)	568 (8.8%)	0.049[b]
Asthma	122 (2.6%)	41 (2.4%)	163 (2.5%)	0.7[b]
CVD	73 (1.5%)	51 (3%)	124 (1.9%)	<0.001[b]
ATM	1,481 (31.3%)	748 (44.3%)	2,229 (34.7)	<0.001[b]
Triage level				
Level 1: Resuscitation	63 (1.3%)	91 (5.4%)	154 (2.4%)	<0.001[c]
Level 2: Emergent	2,509 (52.9%)	1,265 (74.9%)	3,774 (58.7%)	
Level 3: Urgent	2,167 (45.7%)	334 (19.8%)	2,501 (38.9%)	
Scoring systems				
NEWS	4 (3–6)	6 (4–8)	4 (3–6)	<0.001[a]
qSOFA	1 (0–1)	1 (0–1)	0 (0–1)	<0.001[a]
EWS	3 (1–4)	4 (2–5)	3 (2–4)	<0.001[a]
WPS	3 (3–4)	4 (3–6)	3 (3–5)	<0.001[a]
RAPS	0 (0–2)	1 (0–3)	0 (0–2)	<0.001[a]
REMS	6 (3–8)	8 (6–10)	6 (3–9)	<0.001[a]

Values are presented as Median (IQR) or N (%). ATM, ambulance transferring mode; CHD, coronary heart disease; CKD, chronic kidney disease; CVD, cerebrovascular disease; ESI, emergency severity index; EWS, early warning score; GCS, Glasgow coma scale; MAP, mean arterial pressure; SPO_2, peripheral capillary oxygen saturation; NEWS, national early warning score; qSOFA, quick sequential organ failure assessment; RAPS, rapid acute physiology score; REMS, rapid emergency medicine score; WPS, worthing physiological scoring system. [a]Analysis by Mann-Whitney U test. [b]Analysis by Fisher's exact test. [c]Analysis by Chi-square test
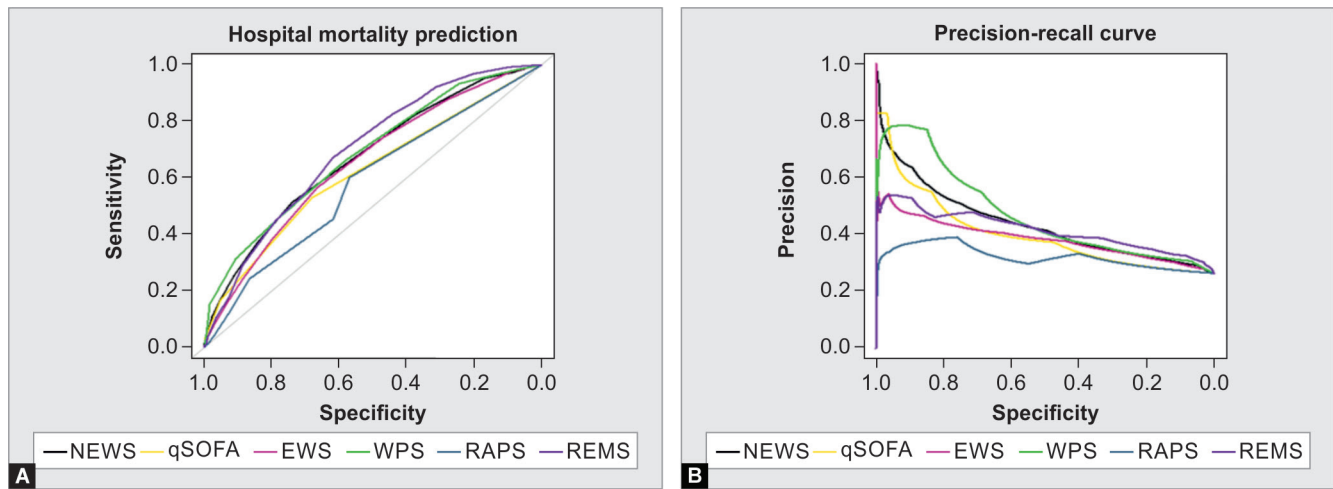

Figs 1A and B: ROC and PR plots. Receiver operating characteristic curves for NEWS (0.701), qSOFA (0.651), EWS (0.669), WPS (0.714), RAPS (0.601), and REMS (0.705) in the emergency department (left panel) and Precision-Recall curves (right panel) for the six evaluated models

Table 3: Performance measures of the NEWS, qSOFA, EWS, WPS, RAPS, and REMS models for predicting in-hospital mortality at presentation to the emergency department

Models	AUROC (95% CI)	AUCPR (95% CI)	BS	Threshold	SE	SP	PPV	NPV	Acc
NEWS	0.701 (0.676,0.723)	0.437 (0.424–0.449)	0.03	5.5	0.61	0.70	0.41	0.81	0.68
qSOFA	0.651 (0.629,0.676)	0.40 (0.388–0.411)	0.14	0.5	0.52	0.68	0.37	0.80	0.64
EWS	0.669 (0.645,0.689)	0.384 (0.372–0.395)	0.15	3.5	0.56	0.67	0.38	0.81	0.64
WPS	0.714 (0.688,0.742)	0.471 (0.458–0.483)	0.09	3.5	0.66	0.58	0.36	0.82	0.60
RAPS	0.601 (0.677,0.728)	0.319 (0.308–0.329)	0.18	2.5	0.60	0.57	0.33	0.80	0.58
REMS	0.705 (0.679,0.731)	0.415 (0.402–0.427)	0.11	6.5	0.67	0.62	0.39	0.84	0.63

Acc, classification accuracy; AUCPR, area under the precision-recall curve; AUROC, area under the receiver operating characteristic curve; BS, brier score; CI, confidence interval; EWS, early warning score; NEWS, national early warning score; NPV, negative predictive value; PPV, positive predictive value; qSOFA, quick sequential organ failure assessment; RAPS, rapid acute physiology score; REMS, rapid emergency medicine score; SE, sensitivity; SP, specificity; WPS, worthing physiological scoring system
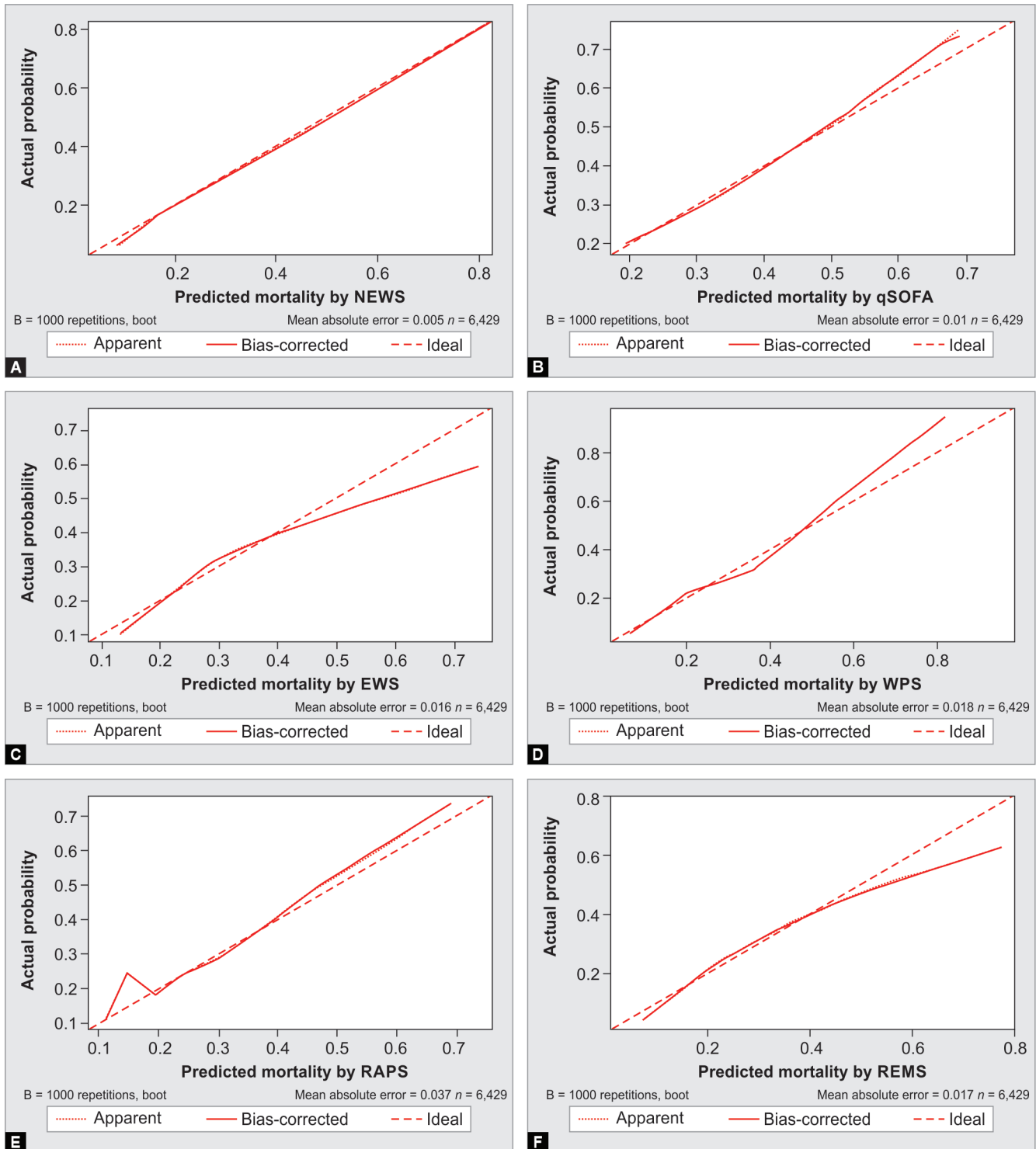
Figs 2A to F: Calibration plots of the NEWS, qSOFA, EWS, WPS, RAPS, and REMS models in the emergency department

of the models. The highest AUCs and AUCPRs belonged to the WPS, REMS, and NEWS models, and the lowest to the RAPS model. A pairwise comparison of AUROCs showed no statistically significant differences between the WPS, REMS, and NEWS (>0.05).

The AUROC of the WPS, REMS, and NEWS were significantly higher than the RAPS, qSOFA, and EWS. Additionally,

no significant differences were observed between the AUROC of EWS and the qSOFA. Regarding prediction accuracy, the Brier score of NEWS was the lowest (0.03), indicating the best performance, whereas RAPS had the highest Brier score (0.18), indicating the worst performance. Figure 2 displays the calibration graphs. The NEWS showed excellent calibration. The rest of the

models showed a reasonable calibration but with over or under-prediction at the high end of the predicted probabilities. Among the models, the NEWS showed the largest range of predicted probabilities.

Discussion

The variation in severity of the SARS-COV2 has made it a challenging task to assess patients' outcomes. This difficulty is far more pronounced in ED settings. In this circumstance using a proper scoring system can help prioritize and manage these complicated patients.

In this single-center but large cohort study, we investigated the predictive performance of the NEWS, qSOFA, EWS, WPS, RAPS, and REMS scoring systems which rely mostly on demographics and vital signs to predict in-hospital mortality. To the best of our knowledge, this is the first study to evaluate and compare different models to predict in-hospital mortality in ED settings in patients with SARS-COV2.

Main Finding

All scores demonstrated a clear positive association with the outcome of in-hospital mortality. The WPS, REMS, and NEWS models exhibited the best and nearly identical performance as evidenced by their AUC, AUCPR, and Brier scores. Their AUC values, ranging from 0.7 to 0.8, are considered "fair". The RAPS, qSOFA, and EWS models had poor discriminatory power. There were no significant differences between the qSOFA and the EWS models. The RAPS model had the worst AUROC, showing poor discrimination between survivors and non-survivors.

The NEWS showed excellent calibration, indicating good agreement between predictions and observed outcomes. EWS and REMS models showed reasonable calibration but with some overestimation, and qSOFA, WPS, and RAPS showed reasonable calibration but with or underestimation at the higher range of the predicted probabilities. The other models showed worse calibration.

Comparison to Other Related Studies

Recently, numerous studies have been conducted to evaluate the predictive performance of rapid scoring systems such as REMS, MEWS, and qSOFA on SARS-COV2 patients worldwide as shown in Table 4. Our findings are consistent with a Chinese study that reported better performance of the REMS model in predicting mortality among SARS-CoV-2 patients as compared to the RAPS and MEWS models.[32]

Other evidence also supports the superior discriminatory power of REMS in comparison to MEWS.[33] However, in our study, the qSOFA model performed poorly, which is in contrast to the fair performance observed in the Chinese study (AUCROC: 0.74, CI 95%: 0.66–0.82).[34] There is also a similar finding (indicating poor discrimination ability) in predicting the outcome by the qSOFA in another study. The only difference with the present study was that respiratory failure within 24 hours of admission was considered as an outcome of the model rather than mortality.[35]

In our study, comorbidities and different underlying medical conditions such as hypertension, cancer, diabetes, CHD, CKD, and CVD (except asthma) had a significant relationship with the outcome. This is consistent with the recent reports that found a significant relationship between hypertension, CKD, CHD, and also diabetes with the outcome.[3,36] In contrast, it has been reported

Table 4: Published evaluation studies of the rapid scoring system in SARS-COV2 patients

Study	Year	Country	Patients (N)	Male gender (%)	Age[a]	Mortality rate (%)	Models	Model specifications			Mean ± SD or Median (IQR) Score		Calibration
								AUROC (95% CI)	SE	SP	Survivors	Non-survivors	
Sabetian G et al.[30]	2020	China	79	41.6%	Survivors: 56.52 ± 16.9 Non-survivors: 75.05 ±12.94	24%	MEWS REMS RAPS	NA	NA	NA	1.48 ± 0.87 3.90 ± 2.98 0.80 ± 1.31	2.37 ± 1.53 7.90 ± 2.8 1.58 ± 1.71	NA
Schomaker M et al.[31]	2020	China	105	50.94%	60.82 ± 16.32	18%	REMS MEWS	0.84 (0.76–0.91) 0.68 (0.58–0.77)	89.5% 68.4%	69.8% 65.1%	5 (2, 6) 1 (1, 2)	7 (6, 10) 2 (1, 3)	NA
Hu H et al.[32]	2020	China	140	70.09%	61.25 ± 15.53	15.7%	SOFA qSOFA	0.89 (0.83–0.96) 0.74 (0.66–0.82)	90% 70%	83.2% 80.4%	1 (0, 2) 0 (0, 0)	4 (3, 5) 1 (0, 1)	NA
Current study	2020	Iran	6,429	57.5%	Survivors: 59 ± 17.9 Non-survivors: 68 ± 15.9 Or 62.3 ± 17.71	26.3%	NEWS qSOFA EWS WPS RAPS REMS	0.701 (0.676,0.723) 0.651 (0.629,0.676) 0.669 (0.645,0.689) 0.714 (0.688,0.742) 0.601 (0.677,0.728) 0.705 (0.679,0.731)	0.61 0.52 0.56 0.66 0.60 0.67	0.70 0.68 0.67 0.58 0.57 0.62	4 (3–6) 1 (0–1) 3 (1–4) 3 (3–4) 0 (0–2) 6 (3–8)	6 (4–8) 1 (0–1) 4 (2–5) 4 (3–6) 1 (0–3) 8 (6–10)	Excellent Reasonable Reasonable Reasonable Reasonable Reasonable

[a]Age is represented as mean ± SD or median (IQR). AUC, area under the receiver operating characteristic curve; CI, confidence interval; EWS, early warning score; NA, not available; MEWS, modified early warning score; NEWS, national early warning score; qSOFA, quick sequential organ failure assessment; RAPS, rapid acute physiology score; REMS, rapid emergency medicine score; WPS, worthing physiological scoring system

in two Chinese studies that, there was no significant relationship between Carcinoma and outcome while our study showed a positive association.[34,36] In another study, diabetes, hypotension, cardiovascular disease, chronic liver, and kidney disease, was associated with an increased mortality but cerebrovascular disease, and malignancy was not.[37]

Strengths and Limitations

This study has several important strengths. Firstly, it involves a large number of patients, approximately 46 times more than the largest related study in Table 4. Secondly, it utilizes various evaluation methods to assess model performance, including discrimination, calibration, Brier score, and the area under the precision-recall curve. Furthermore, this study compares the performance of six models and statistical tests for differences among them. Despite its strengths, the study also has some limitations. One limitation is the presence of missing values for some predictors, which accounted for 28.5% of the overall data. However, we addressed this issue by using multiple imputations within bootstrap samples. Another limitation is that the generalizability of our findings may be limited since the study was conducted in a single center. Nonetheless, the hospital where the study was conducted is one of the main centers for SARS-COV2 patients.

Implications

This study provides important insights into the use of prediction models for mortality in COVID-19 patients. The results can be valuable for policy makers who need to prioritize the care of critically ill patients and allocate resources efficiently during the pandemic.[38,39] Resource can be challenging, particularly in crowded emergency departments during the pandemic. So, more accurate models could help both physicians and decision makers to determine the appropriate care for their patients. Additionally, these models can prevent early patient discharge or serve as a precise tool for ICU admissions. The methodology employed in this study can also be beneficial for future researchers. Multicentric studies are necessary to further investigate the performance of these models in the SARS-COV2 patient population, taking laboratory results into account. The application of machine learning techniques can enhance the accuracy of these models and is strongly recommended for future investigations.[30,40–43]

Conclusion

This study evaluates the internal validity of six prediction models based on prominent scores and compares them. It demonstrates the potential utility of the models based on REMS and EWS as risk-stratification tools for SARS-COV2 patients at presentation to ED.

Ethics Approval and Consent to Participate

The study was approved by Mashhad University of Medical Sciences (Number: IR.MUMS.REC.1400.141) and conformed to the Declaration of Helsinki principles.

Availability of Data and Materials

Data analyzed in this study will be available upon reasonable request from corresponding author.

Authorship Contribution Statement

Zahra Rahmatinejad: Conceptualization, Methodology, Investigation, Formal analysis, Writing – review and editing. **Saeid Eslami:** Conceptualization, Methodology, Writing – review & editing. **Benyamin Hoseini:** Conceptualization, Methodology, Writing – review and editing. **Seyyed Mohammad Tabatabaei:** Conceptualization, Methodology, Investigation, Writing – review & editing. **Fatemeh Rahmatinejad:** Conceptualization, Methodology, Formal analysis, Writing – review & editing. Ali Pourmand: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Ameen Abu-Hanna:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Hamidreza Reihani:** Conceptualization, Methodology, Investigation, Writing – review and editing.

Orcid

Zahra Rahmatinejad https://orcid.org/0000-0003-1168-7234
Benyamin Hoseini https://orcid.org/0000-0002-0355-6181
Hamidreza Reihani https://orcid.org/0000-0003-0617-9374
Ameen Abu Hanna https://orcid.org/0000-0003-4324-7954
Ali Pourmand https://orcid.org/0000-0002-8440-8454
Seyyed Mohammad Tabatabaei https://orcid.org/0000-0002-3153-8968
Fatemeh Rahmatinejad https://orcid.org/0000-0002-6565-9493
Saeid Eslami https://orcid.org/0000-0003-3755-1212

References

1. Ghale-Noie ZN, Salmaninejad A, Bergquist R, Mollazadeh S, Hoseini B, Sahebkar A. Genetic aspects and immune responses in COVID-19: Important organ involvement. Adv Exp Med Biol 2021;1327:3–22. DOI: 10.1007/978-3-030-71697-4_1.
2. Bhatraju PK, Ghassemieh BJ, Nichols M, Kim R, Jerome KR, Nalla AK, et al. COVID-19 in critically ill patients in the seattle region - Case series. N Engl J Med 2020;382(21):2012–2022. DOI: 10.1056/NEJMoa2004500.
3. Goshayeshi L, Akbari Rad M, Bergquist R, Allahyari A, Hashemzadeh K, Hoseini B. Demographic and clinical characteristics of severe COVID-19 infections: a cross-sectional study from Mashhad University of Medical Sciences, Iran. BMC Infect Dis 2021;21(1):656. DOI: 10.1186/s12879-021-06363-6
4. Carlos WG, Dela Cruz CS, Cao B, Pasnick S, Jamil S. Novel wuhan (2019-nCoV) coronavirus. Am J Respir Crit Care Med 2020;201(4):P7–P8. DOI: 10.1164/rccm.2014P7.
5. Paules CI, Marston HD, Fauci AS. Coronavirus infections—more than just the common cold. JAMA 2020;323(8):707–708. DOI: 10.1001/jama.2020.0757.
6. Goshayeshi L, Milani N, Bergquist R, Sadrzadeh SM, Rajabzadeh F, Hoseini B. Covid-19 presented only with gastrointestinal symptoms: a case report of a 14-year-old patient. Govaresh. 2021;25(4): 300–304. http://govaresh.org/index.php/dd/article/view/2341.
7. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the united states. N Engl J Med2020;382(10):929–936. DOI: 10.1056/NEJMoa2001191.
8. Redd WD, Zhou JC, Hathorn KE, McCarty TR, Bazarbashi AN, Thompson CC, et al. Prevalence and characteristics of gastrointestinal symptoms in patients with sars-cov-2 infection in the United States: a multicenter cohort study. Gastroenterology 2020;159(2):765–767. DOI: https://doi.org/10.1053/j.gastro.2020.04.045.
9. Long B, Brady WJ, Koyfman A, Gottlieb M. Cardiovascular complications in covid-19. Emerg Med 2020;38(7):1504–1507. DOI: 10.1016/j.ajem.2020.04.048.

10. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. Ann Emerg Med 2018;71(5):565–574. e2. DOI: 10.1016/j.annemergmed.2017.08.005.

11. Khodashahi R, Naderi H, Bojdy A, Heydari AA, Tavanaee SA, Javad MG, et al. Comparison the effect of arbidol plus hydroxychloroquine vs hydroxychloroquine alone in treatment of covid-19 disease: A randomized clinical trial. Current Respiratory Medicine Reviews 2020;16(4):252–262. DOI: 10.2174/1573398X17666210129125703.

12. Mareiniss DP. The impending storm: COVID-19, pandemics and our overwhelmed emergency departments. Am J Emerg Med 2020;38(6):1293–1294. DOI: 10.1016/j.ajem.2020.03.033.

13. Khodashahi R, Naderi HR, Sedaghat A, Allahyari A, Sarjamee AS, Eshaghi AS, et al. Intravenous immunoglobulin for treatment of patients with covid-19: A case-control study. research article. Arch Clin Infect Dis 2021;16(1):e108068. DOI: 10.5812/archcid.108068.

14. Rahmatinejad Z, Hoseini B, Rahmatinejad F, Abu-Hanna A, Bergquist R, Pourmand A, et al. Internal validation of the predictive performance of models based on three ed and icu scoring systems to predict inhospital mortality for intensive care patients referred from the emergency department. Biomed Res Int 2022;2022:3964063. DOI: 10.1155/2022/3964063.

15. Rahmatinejad Z, Rahmatinejad F, Sezavar M, Tohidinezhad F, Abu-Hanna A, Eslami S. Internal validation and evaluation of the predictive performance of models based on the PRISM-3 (Pediatric Risk of Mortality) and PIM-3 (Pediatric Index of Mortality) scoring systems for predicting mortality in Pediatric Intensive Care Units (PICUs). BMC Pediatr 2022;22(1):199. DOI: 10.1186/s12887-022-03228-y.

16. Alireza A, Leila A, Zahra R, Mirmohammad M, Najmeh N, Saeid E. Development of a national core dataset for the Iranian ICU patients outcome prediction: a comprehensive approach. J Innov Health Inform 2018;25(2):71–76. DOI: 10.14236/jhi.v25i2.953.

17. Rahmatinejad Z, Tohidinezhad F, Reihani H, Rahmatinejad F, Pourmand A, Abu-Hanna A, et al. Prognostic utilization of models based on the APACHE II, APACHE IV, and SAPS II scores for predicting in-hospital mortality in emergency department. The American Journal of Emergency Medicine 2020/09/01/ 2020;38(9):1841–1846. DOI: https://doi.org/10.1016/j.ajem.2020.05.053.

18. Rahmatinejad Z, Tohidinezhad F, Rahmatinejad F, Eslami S, Pourmand A, Abu-Hanna A, et al. Internal validation and comparison of the prognostic performance of models based on six emergency scoring systems to predict in-hospital mortality in the emergency department. BMC Emerg Med 2021;21(1):68. DOI: 10.1186/s12873-021-00459-7.

19. Rahmatinejad Z, Reihani H, Tohidinezhad F, Rahmatinejad F, Peyravi S, Pourmand A, et al. Predictive performance of the SOFA and mSOFA scoring systems for predicting in-hospital mortality in the emergency department. The American Journal of Emergency Medicine 2019;37(7):1237–1241. DOI: https://doi.org/10.1016/j.ajem.2018.09.011.

20. Rahmatinejad F, Rahmatinejad Z, Kimiafar K, Eslami S, Hoseini B. Performance of pediatric risk of mortality and pediatric index of mortality in pediatric intensive care units: A case study of patients with digestive diseases. GOVARESH. 2022;26(3):132–142.

21. Brabrand M, Folkestad L, Clausen NG, Knudsen T, Hallas J. Risk scoring systems for adults admitted to the emergency department: a systematic review. Scand J Trauma Resusc Emerg Med 2010;18(1):8. DOI: 10.1186/1757-7241-18-8.

22. Olivia D, Nayak A, Balachandra M, John J. A classification model for prediction of clinical severity level using qSOFA medical score. Information Discovery and Delivery 2020; DOI: 10.1108/idd-02-2019-0013.

23. Song C-Y, Xu J, He J-Q, Lu Y-Q. COVID-19 early warning score: A multi-parameter screening tool to identify highly suspected patients. medRxiv. 2020; DOI: https://doi.org/10.1101/2020.03.05.20031906.

24. Kim I, Song H, Kim HJ, Park KN, Kim SH, Oh SH, et al. Use of the National Early Warning Score for predicting in-hospital mortality in older adults admitted to the emergency department. Clin Exp Emerg Med 2020;7(1):61–66. DOI: 10.15441/ceem.19.036.

25. Guo Y, Wang Y, Ma C, Li R, Li T. Performance of quick sequential organ failure assessment (qSOFA) score for prognosis of heat-related hospitalized patients. Heart Lung 2020;49(4):415–419. DOI: 10.1016/j.hrtlng.2020.02.040.

26. Groarke J, Gallagher J, Stack J, Aftab A, Dwyer C, McGovern R, et al. Use of an admission early warning score to predict patient morbidity and mortality and treatment success. Emerg Med J 2008;25(12):803–806. DOI: 10.1136/emj.2007.051425.

27. Duckitt R, Buxton-Thomas R, Walker J, Cheek E, Bewick V, Venn R, et al. Worthing physiological scoring system: derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based single-centre study. Br J Anaesth 2007;98(6):769–774. DOI: 10.1093/bja/aem097.

28. Rhee KJ, Fisher Jr CJ, Willitis NH. The rapid acute physiology score. Am J Emerg Med 1987;5(4):278–282. DOI: 10.1016/0735-6757(87)90350-0.

29. Olsson T, Terént A, Lind L. Rapid emergency medicine score: A new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. J Intern Med 2004;255(5):579–587. DOI: 10.1111/j.1365-2796.2004.01321.x.

30. Sabetian G, Azimi A, Kazemi A, Hoseini B, Asmarian N, Khaloo V, et al. Prediction of patients with covid-19 requiring intensive care: a cross-sectional study based on machine-learning approach from iran. Indian J Crit Care Med 2022;26(6):688–695. DOI: 10.5005/jp-journals-10071-24226.

31. Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. Stat med 2018;37(14):2252–2266. DOI: 10.1002/sim.7654.

32. Hu H, Kong W, Yao N, Qiu Y, Gu H, Xu W. Prognostic value of three rapid scoring scales and combined score for the assessment of patients with coronavirus disease 2019. Nurs Open 2022;9(3):1865–1872. DOI: 10.1002/nop2.934.

33. Hu H, Yao N, Qiu Y. Comparing rapid scoring systems in mortality prediction of critically ill patients with novel coronavirus disease. Acad Emerg Med 2020;27(6):461–468. DOI: 10.1111/acem.13992.

34. Liu S, Yao N, Qiu Y, He C. Predictive performance of SOFA and qSOFA for in-hospital mortality in severe novel coronavirus disease. Am J Emerg Med 2020;38(10):2074–2080. DOI: 10.1016/j.ajem.2020.07.019.

35. Haimovich AD, Ravindra NG, Stoytchev S, Young HP, Wilson FP, van Dijk D, et al. Development and validation of the quick COVID-19 severity index: A prognostic tool for early clinical decompensation. Ann Emerg Med 2020;76(4):442–453. DOI: 10.1016/j.annemergmed.2020.07.022.

36. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020;395(10229):1054–1062. DOI: 10.1016/S0140-6736(20)30566-3.

37. Zhou Y, He Y, Yang H, Yu H, Wang T, Chen Z, et al. Development and validation a nomogram for predicting the risk of severe COVID-19: A multi-center study in Sichuan, China. Plos one 2020;15(5):e0233328. DOI: 10.1371/journal.pone.0233328.

38. Ehmann MR, Zink EK, Levin AB, Suarez JI, Belcher HME, Daugherty Biddison EL, et al. Operational recommendations for scarce resource allocation in a public health crisis. Chest 2020;159:10231. DOI: 10.1016/j.chest.2020.09.246.

39. Maves RC, Downar J, Dichter JR, Hick JL, Devereaux A, Geiling JA, et al. Triage of scarce critical care resources in COVID-19: An implementation guide for regional allocation an expert panel report of the task force for mass critical care and the American College of Chest Physicians. Chest 2020; 158(1):212–225. DOI: 10.1016/j.chest.2020.03.063.

40. Cheng FY, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, et al. Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. J Clin Med 2020;9(6):1668. DOI: 10.3390/jcm9061668.

41. Heo J, Han D, Kim H-J, Kim D, Lee YK, Lim D, et al. Prediction of patients requiring intensive care for COVID-19: development and validation of an integer-based score using data from Centers for Disease

Control and Prevention of South Korea. J Intensive Care 2021;9(1):16. DOI: 10.1186/s40560-021-00527-x.

42. Saba T, Abunadi I, Shahzad MN, Khan AR. Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types. Microsc Res Tech;84(7):1462–1474. DOI: 10.1002/jemt.23702.

43. Surme S, Buyukyazgan A, Bayramlar OF, Cinar AK, Copur B, Zerdali E, et al. Predictors of intensive care unit admission or mortality in patients with coronavirus disease 2019 pneumonia in Istanbul, Turkey. Jpn J Infect Dis 2021;74(5):458–464. DOI: 10.7883/yoken.JJID.2020.1065.