Evaluation of the Risk Prediction Tools for Patients With Coronavirus Disease 2019 in Wuhan, China: A Single-Centered, Retrospective, Observational Study

Fang-Yan Liu, MD[1,2]; Xue-Lian Sun, BN[1]; Yong Zhang, MD, PhD[3]; Lin Ge, BN[3]; Jing Wang, MD[1]; Xiao Liang, BN[1]; Jun-Fen Li, MD[1]; Chang-Liang Wang, BN[1]; Zheng-Tao Xing, BN[1]; Jagadish K. Chhetri, MD[2,4]; Peng Sun, MD, PhD[5]; Piu Chan, MD, PhD[2,4]

Objectives: To evaluate and compare the efficacy of National Early Warning Score, National Early Warning Score 2, Rapid Emergency Medicine Score, Confusion, Respiratory rate, Blood pressure, Age 65 score, and quick Sepsis-related Organ Failure Assessment on predicting in-hospital death in patients with coronavirus disease 2019.

Design: A retrospective, observational study.

Setting: Single center, West Campus of Wuhan Union hospital-a temporary center to manage critically ill patients with coronavirus disease 2019.

Patients: A total of 673 consecutive adult patients with coronavirus disease 2019 between January 30, 2020, and March 14, 2020.

Interventions: None.

Measurements and Main Results: Data on demography, comorbidities, vital signs, mental status, oxygen saturation, and use of supplemental oxygen at admission to the ward were collected from medical records and used to score National Early Warning Score, National Early Warning Score 2, Rapid Emergency Medicine Score, Confusion, Respiratory rate, Blood pressure, Age 65 score, and quick Sepsis-related Organ Failure Assessment. Total number of patients was 673 (51% male) and median (interquartile range) age was 61 years (50–69 yr). One-hundred twenty-one patients died (18%). For predicting in-hospital death, the area under the receiver operating characteristics (95% CI) for National Early Warning Score, National Early Warning Score 2, Rapid Emergency Medicine Score, Confusion, Respiratory rate, Blood pressure, Age 65 score, and quick Sepsis-related Organ Failure Assessment were 0.882 (0.847–0.916), 0.880 (0.845–0.914), 0.839 (0.800–0.879), 0.766 (0.718–0.814), and 0.694 (0.641–0.746), respectively. Among the parameters of National Early Warning Score, the oxygen saturation score was found to be the most significant predictor of in-hospital death. The area under the receiver operating characteristic (95% CI) for oxygen saturation score was 0.875 (0.834–0.916).

Conclusions: In this single-center study, the discrimination of National Early Warning Score/National Early Warning Score 2 for predicting mortality in patients with coronavirus disease 2019 admitted to the ward was found to be superior to Rapid Emergency Medicine Score, Confusion, Respiratory rate, Blood pressure, Age 65 score, and quick Sepsis-related Organ Failure Assessment. Peripheral oxygen saturation could independently predict in-hospital death in these patients. Further validation of our finding in multiple settings is needed to determine its applicability for coronavirus disease 2019. (*Crit Care Med* 2020; XX:00–00)

Key Words: coronavirus; coronavirus disease 2019; early warning score; pandemic; prognosis; pulse oximetry

The novel coronavirus disease 2019 (COVID-19) initially reported in the city of Wuhan in China and has now spread globally (1–4). As of June 17, 2020, there were 8,061,550 registered cases in 216 countries, and 440,290 people have lost their lives (4).

Based on past reporting, the majority of the patients with COVID-19 had mild symptoms with a good prognosis (5);

however, some patients progressed rapidly to a critical state such as acute respiratory failure state, multiple organ failure, and septic shock (6). For patients infected with COVID-19, early identification of the severity of illness could facilitate appropriate supportive care and prompt access to the ICU if necessary. In an epidemic situation, for patients with mild COVID-19 symptomatic treatment in general isolation is recommended (7). Early intensive care is warranted for patients with rapid worsening of the condition to reduce mortality. However, the appropriate use of intensive care during the epidemic may alleviate the shortage of medical resources (8). Therefore, an easy-to-use risk predictive tool for assessing the possibility of deterioration of patients with COVID-19 is needed. Such a tool could help clinicians to stratify patients into relevant risk categories and facilitate in making appropriate clinical decisions in a state of emergency, where every second counts.

Several scoring systems for detecting potentially critically ill patients in acute settings have been proposed and used in emergency department. Some of the most commonly used scoring systems are the National Early Warning Score (NEWS)/National Early Warning Score 2 (NEWS2) (9–12), Rapid Emergency Medicine Score (REMS) (13, 14), Confusion, Respiratory rate, Blood pressure, Age 65 (CRB65) score (15), and quick Sepsis-related Organ Failure Assessment (qSOFA) (16, 17) which might serve as a risk prediction tool for the patients infected with COVID-19. These tools, including parameters that are easily available in a very basic clinical setting (**Supplementary Table 1**, Supplemental Digital Content 1, http://links.lww.com/CCM/F704; **Supplementary Table 2**, Supplemental Digital Content 2, http://links.lww.com/CCM/F705; **Supplementary Table 3**, Supplemental Digital Content 3, http://links.lww.com/CCM/F706; **Supplementary Table 4**, Supplemental Digital Content 4, http://links.lww.com/CCM/F707; and **Supplementary Table 5**, Supplemental Digital Content 5, http://links.lww.com/CCM/F708) are independent of the laboratory test results and known to predict severe health deterioration. Thus, these scores can be used to objectively assess the severity of the COVID-19 infected patient's clinical condition and predict patient mortality.

We aimed to retrospectively use the clinical data available in a single center and compare these five easy-to-use risk prediction tools for patients with COVID-19. Our study is an attempt to identify the best easy-to-use risk prediction tool to be used as an aid to clinical assessment in a simple clinical setting for the management of COVID-19 patients in a triage situation.

MATERIALS AND METHODS

Study Design and Participants
This was a single-centered retrospective observational study conducted at the West Campus of Wuhan Union Hospital (Wuhan, China), which is a large university hospital, and one of the major designated referral and treatment hospitals for critically ill adult patients (≥ 18 yr old) with COVID-19 (**Supplementary Material 1**, Supplemental Digital Content 6, http://links.lww.com/CCM/F709). We analyzed all data from consecutive patients admitted to the center between January 30, 2020, and March 14, 2020, who had been diagnosed with COVID-19 as per the World Health Organization interim guidance (18). Laboratory confirmation of COVID-19 infection was performed by the local health authority as described elsewhere (1). All these patients had been either discharged or had died by April 12, 2020.

Data Collection
According to the previously published retrospective chart review guidelines (19), the data abstractors were trained clinicians, directly supervised by the principal investigator (PI). Abstractors used data abstraction forms and had a clear definition of abstract variables and were blinded to the study outcome. Elements of electronic medical records (including nursing records and radiological examinations) of all consecutive patients with laboratory-confirmed COVID-19 infection were abstracted. Patient outcomes were abstracted by two nurses. Patients' records missing more than one vital sign were excluded, and for cases where only one vital sign was missing, this value was imputed using the median values. Clinical data, including demographics, comorbidities, vital signs, mental status, oxygen saturation, and use of supplemental oxygen at admission to the ward, were abstracted by research assistants, trained, and supervised by the PI. The collected data were used to score NEWS, NEWS2, REMS, CRB65, and qSOFA (Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/CCM/F704; Supplementary Table 2, Supplemental Digital Content 2, http://links.lww.com/CCM/F705; Supplementary Table 3, Supplemental Digital Content 3, http://links.lww.com/CCM/F706; Supplementary Table 4, Supplemental Digital Content 4, http://links.lww.com/CCM/F707; and Supplementary Table 5, Supplemental Digital Content 5, http://links.lww.com/CCM/F708). The mental status for different tools was based on the hospital recorded level of consciousness (such as conscious, confusion, delirium, drowsiness, sopor, mild coma, medium coma, deep coma). For NEWS/NEWS2, we allocated patients recorded as "confusion, delirium, drowsiness, sopor, mild coma, medium coma, deep coma" a score of 3; for REMS, we allocated patients recorded as "conscious" a score of 0, "confusion, delirium, drowsiness, sopor" a score of 1, "mild coma" a score of 2, "medium coma" a score of 3, "deep coma" a score of 4; and for qSOFA, we allocated patients recorded as "confusion, delirium, drowsiness, sopor, mild coma, medium coma, deep coma" a score of 1, that is, having "altered mental status."

Outcome
The endpoint was in-hospital death.

Statistical Analysis
Categorical variables were described as frequency rates and percentages, and continuous variables were described as median and interquartile range (IQR) values. The discrimination of NEWS, NEWS2, REMS, CRB65, and qSOFA for the prediction of in-hospital death was assessed and compared using the area under the receiver operating characteristic (AUROC) curve. An AUROC of at least 0.70 was defined as "acceptable

discrimination" and an AUROC of at least 0.80 was defined as "excellent discrimination" (20). The optimal cutoff point was defined as the threshold value with the maximal Youden index (21). The AUROC was compared according to the method of DeLong, DeLong, and Clarke-Pearson (22). Calibration was assessed using the Hosmer-Lemeshow test (23) and by visual assessment in loess calibration curve of predicted probabilities with observed risk (24). In order to evaluate the predictive power of each of the constituent elements of the best score, we first undertook univariate analysis, using logistic regression to estimate the association between the variables and death. We then undertook multivariate analysis to determine which individual variables were independent predictors of death. All significant predictor variables (univariate analysis $p < 0.1$) were entered into a logistic regression model, with death as outcome. Then receiver operating characteristics (ROCs) were created to determine the effectiveness of these predictors. In these univariate and multivariate analyses, we used the variables as they are categorized in NEWS (e.g., when a patient's respiration rate was 22 breath/min, his/her respiration rate score was 2. See Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/CCM/F704), rather than the raw data. A p value of less than 0.05 was considered significant. All analyses were performed using SAS Version 9.4 (SAS Institute, Cary, NC).

Ethical Considerations

This study was approved by the Ethics Committee of Wuhan Union Hospital (2020-LSZ-0129). As the study was observational, and the data were anonymized for analysis, individual patient consent was waived by the ethics committees.

RESULTS

Characteristics of Admissions

For this study, medical data of the 673 patients with COVID-19 who fulfilled the inclusion criteria were enrolled (**Supplementary Fig. 1**, Supplemental Digital Content 7, http://links.lww.com/CCM/F710). The median (IQR) age was 61 years (50–69 yr), of which 364 (54%) of them were older than 60 years, 341 patients (51%) were men and 227 patients (34%) had chronic diseases, including hypertension, diabetes, chronic cardiac disease, chronic pulmonary disease, cerebrovascular disease, chronic liver disease, chronic kidney disease and malignancy, and 121 died (18%). The baseline characteristics of patients are presented in **Table 1**.

Performance of NEWS, NEWS2, REMS, CRB65, and qSOFA

Figure 1 shows ROC curves comparing the ability of NEWS, NEWS2, REMS, CRB65, and qSOFA at predicting in-hospital death. AUROC analysis demonstrated the discrimination of the five scoring tools as follows: NEWS, 0.882 (95% CI, 0.847–0.916); NEWS2, 0.880 (95% CI, 0.845–0.914); REMS, 0.839 (95% CI, 0.8–0.879); CRB65, 0.766 (95% CI, 0.718–0.814); and qSOFA, 0.694 (95% CI, 0.641–0.746). NEWS was found to be the best in predicting in-hospital death, and NEWS greater

than or equal to 5 was the optimal threshold, with a sensitivity of 84.3% and specificity of 76.8% (**Table 2**). DeLong test found AUROC of NEWS to be significantly larger than that of REMS ($p = 0.048$), CRB65 ($p < 0.001$), and qSOFA ($p < 0.001$) (**Supplementary Table 6**, Supplemental Digital Content 8, http://links.lww.com/CCM/F711). Calibration plots are shown in **Supplementary Figure 2** (Supplemental Digital Content 9, http://links.lww.com/CCM/F712), which showed NEWS, NEWS2, REMS, CRB65, and qSOFA to be well calibrated.

Physiologic Parameters and Mortality

Univariate analysis (**Table 3**) showed that of the seven parameters in the NEWS score, respiration rate score, oxygen saturation score, supplemental oxygen, temperature score, heart rate score, and AVPU score were associated with death in patients with COVID-19 (at the level of $p < 0.1$).

Multivariate analysis showed that respiration rate score, oxygen saturations score, temperature score, and AVPU score were independent predictors of death in patients with COVID-19. After adjusting for other variables, supplemental oxygen and heart rate score did not predict death.

Performance of Each Parameter of National Early Warning Score

Supplementary Figure 3 (Supplemental Digital Content 10, http://links.lww.com/CCM/F713) shows ROC curves comparing the ability of each parameter of NEWS at predicting in-hospital death. AUROC analysis demonstrated the discrimination of the each parameter as follows: respiration rate score, 0.687 (95% CI, 0.636–0.738); oxygen saturation score, 0.875 (95% CI, 0.838–0.913); systolic blood pressure score, 0.511 (95% CI, 0.476–0.546); temperature score, 0.544 (95% CI, 0.503–0.585); heart rate score, 0.534 (95% CI, 0.479–0.589); and AVPU score, 0.566 (95% CI, 0.534–0.597). Oxygen saturation score was found to be the best in predicting in-hospital death, and oxygen saturation score greater than or equal to 2 (i.e., oxygen saturation ≤ 93%) was the optimal threshold, with a sensitivity of 77.7% and specificity of 89.7%. DeLong test found AUROC of oxygen saturation score to be significantly larger than that of respiration rate score ($p < 0.001$); systolic blood pressure score ($p < 0.001$); temperature score ($p < 0.001$); heart rate score ($p < 0.001$); and AVPU score ($p < 0.001$) (**Supplementary Table 7**, Supplemental Digital Content 11, http://links.lww.com/CCM/F714).

Therefore, we chose to compare the oxygen saturation score with the other prediction scoring systems (Fig. 1). The oxygen saturation score was found to be better than CRB65 ($p < 0.001$) and qSOFA ($p < 0.001$) and not worse than NEWS (p = 0.63), NEWS2 (p = 0.74), and REMS (p = 0.1) in predicting in-hospital death (**Supplementary Table 8**, Supplemental Digital Content 12, http://links.lww.com/CCM/F715). Calibration plots showed that oxygen saturation score was well calibrated (Supplementary Fig. 2, Supplemental Digital Content 9, http://links.lww.com/CCM/F712). The oxygen saturation score appears to give a better specificity but a lower sensitivity than NEWS (Table 2).

TABLE 1. Characteristics of the Study Cohort at Admission

Variables	Whole Cohort	Nonsurvivors	Survivors
n	673	121	552
Age, yr, median (IQR)	61 (50–69)	69 (61–78)	58 (48–67)
Male, n (%)	341 (50.67)	79 (65.29)	262 (47.46)
Chronic medical illness, n (%)	227 (33.73)	35 (28.93)	192 (34.78)
Hypertension	143 (21.25)	37 (30.58)	106 (19.2)
Diabetes	94 (13.97)	22 (18.18)	72 (13.04)
Chronic cardiac disease	65 (9.66)	20 (16.53)	45 (8.15)
Chronic pulmonary disease	25 (3.71)	5 (4.13)	20 (3.62)
Cerebrovascular disease	23 (3.42)	10 (8.26)	13 (2.36)
Chronic kidney disease	13 (1.93)	5 (4.13)	8 (1.45)
Chronic liver disease	14 (2.08)	2 (1.65)	12 (2.17)
Malignancy	39 (5.79)	11 (9.09)	28 (5.07)
Vital signs			
Respiratory rate, breaths/min, median (IQR)	20 (20–24)	24 (20–30)	20 (20–23)
Oxygen saturation %, median (IQR)	96 (94–98)	86 (73.5–93)	96 (96–98)
Supplemental oxygen, n (%)	273 (40.56)	108 (89.26)	165 (29.89)
Systolic blood pressure, mm Hg, median (IQR)	130 (120–140.5)	135 (120–150)	130 (119.25–139)
Mean arterial pressure, mm Hg, median (IQR)	97.33 (90–105.67)	98 (88–106.83)	96.83 (90–105.25)
Temperature, °C, median (IQR)	36.7 (36.3–37.2)	36.9 (36.35–37.8)	36.7 (36.3–37.1)
Heart rate, beats/min, median (IQR)	90 (80–102)	90 (80–106)	90 (80–101)
Alert, Verbal, Pain, Unresponsiveness, score, median (IQR)	0 (0–0)	0 (0–0)	0 (0–0)
Illness severity, median (IQR)			
National Early Warning Score	3 (1–6)	8 (6–10)	3 (1–4)
National Early Warning Score 2	3 (1–6)	8 (5.5–10)	3 (1–4)
Rapid Emergency Medicine Score	5 (3–7)	8 (6–11)	4 (2–6)
Confusion, Respiratory rate, Blood pressure, Age 65 score	0 (0–1)	1 (1–2)	0 (0–1)
Quick Sepsis-related Organ Failure Assessment	0 (0–1)	1 (0–1)	0 (0–1)
Length of stay, d, median (IQR)	17 (10–30)	6 (3–12)	20 (13–34)

IQR = interquartile range.

DISCUSSION

It is vital to determine as quickly as possible which patients with COVID-19 infection are at high risk of death, especially in poor healthcare resource settings so as to make proper use of all available resources. Thus, the risk prediction tools employed for aiding triage decisions should be based on rapidly obtainable and direct prognosis-related parameters. We have compared five well established such tools: NEWS, NEWS2, REMS, qSOFA, and CRB65 for predicting death risk in a critical situation.

NEWS is a validated clinical assessment tool developed by the Royal College of Physicians in the United Kingdom (9–11). NEWS comprises respiratory rate, oxygen saturation, temperature, systolic blood pressure, heart rate, and level of consciousness (Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/CCM/F704). In December 2017, NEWS2 was released superseding the original (2012) Version, which includes two oxygen saturation scales; the second adjusts target saturations to 88–92% for those with hypercapnic respiratory failure (12) (Supplementary Table 2, Supplemental Digital Content 2, http://links.lww.com/CCM/F705). We have previously shown that the discrimination of NEWS (AUROC, 0.85; 95% CI, 0.79–0.90) for the prediction of death within 24 hours of acutely ill patients in Beijing was excellent (11). But the performance of NEWS2 in Chinese population is still
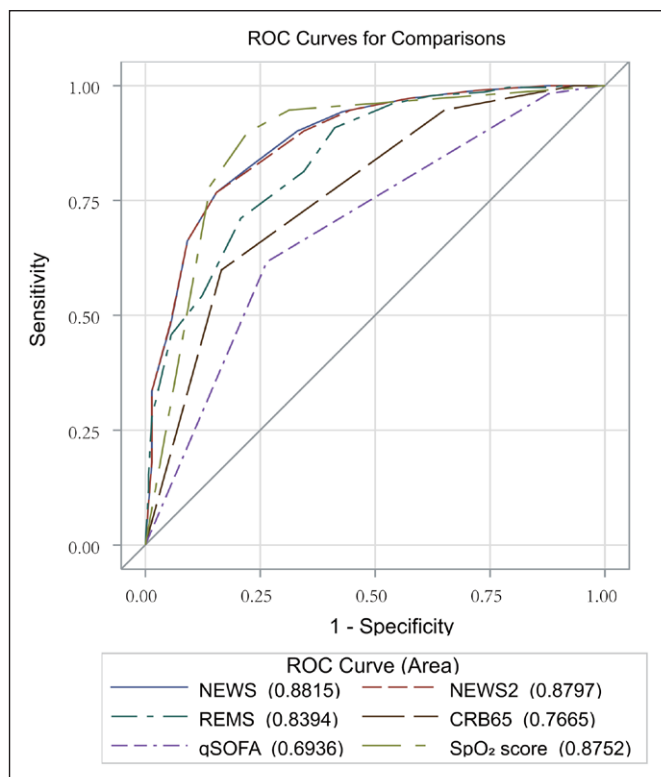
Figure 1. Receiver operating characteristic (ROC) curves comparing the ability of National Early Warning Score (NEWS), National Early Warning Score 2 (NEWS2), Rapid Emergency Medicine Score (REMS), Confusion, Respiratory rate, Blood pressure, Age 65 (CRB65) score, quick Sepsis-related Organ Failure Assessment (qSOFA), and peripheral oxygen saturation (Spo$_2$) at predicting in-hospital death.

unvalidated. In our current study, we found that the discrimination of NEWS and NEWS2 for the prediction of in-hospital death in patients with COVID-19 infection in Wuhan was both equally excellent. There was no significant difference between the performances of NEWS and NEWS2, which could be due to the small size of patients with hypercapnic respiratory failure. Therefore, for ease of use, NEWS might be more practical.

Redfern et al (25) found that NEWS had better discrimination over qSOFA for identifying high-risk non-ICU patients with/without infection. In this study, we found that the efficiency of NEWS/NEWS2 were the best compared with the other three risk prediction score systems as a predictor of in-hospital death in patients with COVID-19. A previous study has shown that the NEWS was good at predicting acute mortality (10). In this study, we found that the NEWS was also good at predicting in-hospital mortality in patients with COVID-19. As a predictor of in-hospital death, the optimal cut point for NEWS was 5. With a sensitivity of 84.3% and specificity of 76.8%, it was more accurate and had higher negative predictive value (NPV) (95.7%) and positive predictive value (PPV) (44.4%) than REMS, CRB65, and qSOFA (Table 2). In addition, our results show that it is more appropriate to admit those with scores of 5 and above to ICU because the clinical situation of patients with COVID-19 could rapidly deteriorate. Using NEWS as a triage assistant tool would not only simplify the decision-making process but also improve the quality of medical care and save the

medical resources for needed patients. Based on the risk prediction score, the critical care resource could be reserved for the identified high-risk patients. During the COVID-19 pandemic, critical care is a scarce resource facing soaring demand; thus, preplanned resource allocation and prioritization based on NEWS could largely improve care efficiency and ultimately lead to better outcomes on one hand, but also raise the alarm of severity to promote early intervention and reduce fatality.

A recent study suggested to apply the Multilobular infiltration, hypo-Lymphocytosis, Bacterial coinfection, Smoking history, hyper-Tension and Age (MuLBSTA) score in predicting the risk of mortality in COVID-19 infection (5). The MuLBSTA score requires six indexes, which are age, smoking history, hypertension, bacterial coinfection, multilobular infiltration, and lymphopenia (26). Since this score takes longer time to assess and, more importantly, includes laboratory tests, we did not include the MuLBSTA score in this study. In addition, it should be noted that many countries with poor medical resources or other limited resources may be unable to perform such complicated tests in an emergency situation or on site. Thus, simple tools such as NEWS could be more useful for rapid assessment and management of the COVID-19 pandemic. NEWS requires only vital signs, oxygen saturation, use of supplemental oxygen, and mental status that nurses or even volunteers can use the tool without much training.

Baker et al (27) found that even a single deranged physiologic parameter at admission was associated with mortality in a critically ill population. Our study showed admission median oxygen saturation of survivors and nonsurvivors with COVID-19 was 86% (IQR, 73.5–93%) and 96% (IQR, 96–98%), respectively. The value of oxygen saturation was confirmed by multivariate analysis. Furthermore, our study demonstrated that oxygen saturation score had excellent discrimination and good calibration. The discrimination of oxygen saturation score was better than CRB65 and qSOFA and not worse than NEWS, NEWS2, and REMS in predicting in-hospital death. In other words, we may also use the oxygen saturation level alone to predict death in patients with COVID-19 infection. Peripheral oxygen saturation can be quickly and easily obtained by any health professional almost anywhere. It may be used as an inexpensive reliable tool for assessing the severity of COVID-19 and detecting patients at high risk in poor resource settings. However, our results suggest that while oxygen saturation score gives a higher specificity and PPV, it has a lower sensitivity as compared with NEWS. The lower sensitivities demonstrate that oxygen saturation score is relatively poor at correctly identifying those patients who will die subsequently in hospital, based at admission data. Smith et al (28) found that early warning scores, such as NEWS, provide better detection of adverse outcomes at a lower trigger rate than a single vital sign variable. Therefore, if resources are unlimited, NEWS would be a better choice.

It should also be noted that during the initial phase of the COVID-19 outbreak, due to the rapid spread and limited medical resources, patients admitted to hospital were more likely to have a higher risk of death. Therefore, more than 30% of patients included in this study were severe and critically ill patients. Thus, these patients had a higher mortality

TABLE 2. The Performance for National Early Warning Score, National Early Warning Score 2, Rapid Emergency Medicine Score, Confusion, Respiratory Rate, Blood Pressure, Age 65 Score, Quick Sepsis-Related Organ Failure Assessment, and Peripheral Oxygen Saturation in Predicting In-Hospital Death

Score	Area Under the Receiver Operating Characteristic (95% CI)	p	Optimal Cutoff	Youden Indices	Sensitivity, %	Specificity, %	Positive Predictive Value, %	Negative Predictive Value, %	False Positive Rate, %	False Negative Rate, %	Accuracy, %	Hosmer-Lemeshow Test p
National Early Warning Score	0.882 (0.847–0.916)	< 0.001	5	0.61	84.30	76.81	44.35	95.71	23.19	15.70	78.16	0.597
National Early Warning Score 2	0.880 (0.845–0.914)	< 0.001	5	0.61	84.30	76.63	44.16	95.70	23.37	15.70	78.01	0.458
Rapid Emergency Medicine Score	0.839 (0.800–0.879)	< 0.001	6	0.51	79.34	71.20	37.65	94.02	28.80	20.66	72.66	0.083
Confusion, Respiratory rate, Blood pressure, Age 65 score	0.766 (0.718–0.814)	< 0.001	1	0.44	83.47	59.96	31.37	94.30	40.04	16.53	64.19	0.601
Quick Sepsis-related Organ Failure Assessment	0.694 (0.641–0.746)	< 0.001	1	0.35	73.55	61.59	29.57	91.40	38.41	26.45	63.74	–a
Peripheral oxygen saturation score	0.875 (0.834–0.916)	< 0.001	2	0.67	77.69	89.67	62.25	94.83	10.33	22.31	87.52	0.442

aQuick Sepsis-related Organ Failure Assessment has only four categories, and only 24 patients scored 2 or 3 in this study, so when we were performing Hosmer-Lemeshow test, χ^2 value was 0.014 and degrees of freedom was 0.

TABLE 3. Logistic Regression Analyses to Determine the Association of Each Parameter of National Early Warning Score With Death

Physiologic Variables	Univariate Analysis		Multivariate Analysis	
	OR (95% CI)	p	OR 95% CI	p
Respiration rate	1.73 (1.48–2.04)	< 0.001	1.27 (1.03–1.58)	0.03
Oxygen saturation	4.07 (3.31–5.02)	< 0.001	3.46 (2.5–4.78)	< 0.001
Supplemental oxygen	19.49 (10.66–35.63)	< 0.001	1.28 (0.8–2.03)	0.30
Systolic blood pressure	1.22 (0.89–1.66)	0.22	—	—
Temperature	1.75 (1.17–2.64)	0.007	2.36 (1.25–4.45)	0.008
Heart rate	1.32 (1.03–1.70)	0.03	0.79 (0.55–1.13)	0.20
Alert, Verbal, Pain, Unresponsiveness score	2.62 (1.86–3.67)	< 0.001	14.46 (3.71–56.26)	< 0.001

OR = odds ratio.
Dashes indicate systolic blood pressure was not included in multivariate analysis because it had a p value of ≥0.1 in univariate analysis.

rate (17.9%) than the previously reported case fatality rate in China (5.6%) (29).

The results of this study must be interpreted considering some limitations. First, as a retrospective study, all data were collected as a part of usual care rather than for research, and thus, some of the clinically meaningful data were not available (e.g., the mode of oxygen delivery for all patients). Second, this is a single-center study, and our findings may not be generalizable to all hospitals. Third, a further external validation of oxygen saturation score as a prediction model may be needed. Fourth, our study was not intended to create a new triage system but rather validation

and comparison of existing scoring systems widely used in emergency situation for applying to COVID-19 based on the data from our large samples in an initial phase of the pandemic so that different triage systems can be applied in different situation for further care of COVID-19 patients. Based on the different situations of available medical resources, balance between PPV and NPV of each scoring system should be weighted. Last, data of the suspected but undiagnosed cases were not included in our study. It is warranted to conduct a multicenter study including as many patients as possible in future to have a more comprehensive understanding of using several scoring systems' performance in detecting the potentially critically ill patients with COVID-19 infection.

CONCLUSIONS

In this single center study, the discrimination of NEWS/NEWS2 for predicting in-hospital death in patients with COVID-19 admitted to the ward was found to be superior to REMS, CRB65, and qSOFA. In addition, peripheral oxygen saturation could independently predict in-hospital death in these patients. Further validation in different multiple settings may be needed to determine the widespread applicability of these tools.

ACKNOWLEDGMENTS

REFERENCES

1. Wang D, Hu B, Hu C, et al: Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 2020; 323:1061–1069

2. Zhu N, Zhang D, Wang W, et al: A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020; 382:727–733

3. Xu XW, Wu XX, Jiang XG, et al: Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: Retrospective case series. *BMJ* 2020; 368:m606

4. World Health Organization: Novel Coronavirus (COVID-19) Situation Report-149. 2020. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200617-covid-19-sitrep-149.pdf?sfvrsn=3b3137b0_4. Accessed June 17, 2020

5. Wu Z, McGoogan JM: Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 2020; 323:1239–1242

6. Chen N, Zhou M, Dong X, et al: Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet* 2020; 395:507–513

7. World Health Organization: Novel Coronavirus (COVID-19) Situation Report-46. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf_4. Accessed June 17, 2020

8. Yang X, Yu Y, Xu J, et al: Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: A single-centered, retrospective, observational study. *Lancet Respir Med* 2020; 8:475–481

9. Royal College of Physicians: National Early Warning Score (NEWS): Standardising the Assessment of Acute-Illness Severity in the NHS. London, United Kingdom, Royal College of Physicians, 2012. Available at: https://www.rcplondon.ac.uk/file/9559/download Accessed July 26, 2020

10. Smith GB, Prytherch DR, Meredith P, et al: The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation* 2013; 84:465–470

11. Liu FY, Qin J, Wang RX, et al: A prospective validation of National Early Warning Score in emergency intensive care unit patients at Beijing. *Hong Kong J Emerg Med* 2015; 22:137–144

12. Royal College of Physicians: National Early Warning Score (NEWS) 2: Standardising the Assessment of Acute-Illness Severity in the NHS. Updated Report of a Working Party. London, United Kingdom, Royal College of Physicians, 2017. Available at: https://www.rcplondon.ac.uk/file/8636/download Accessed July 26, 2020

13. Olsson T, Lind L: Comparison of the Rapid Emergency Medicine Score and APACHE II in nonsurgical emergency department patients. *Acad Emerg Med* 2003; 10:1040–1048

14. Olsson T, Terent A, Lind L: Rapid Emergency Medicine score: A new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. *J Intern Med* 2004; 255:579–587

15. Lim WS, van der Eerden MM, Laing R, et al: Defining community acquired pneumonia severity on presentation to hospital: An international derivation and validation study. *Thorax* 2003; 58:377–382

16. Singer M, Deutschman CS, Seymour CW, et al: The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:801–810

17. Seymour CW, Liu VX, Iwashyna TJ, et al: Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:762–774

18. World Health Organization: Clinical Management of Severe Acute Respiratory Infection When Novel Coronavirus (nCoV) Infection Is Suspected: Interim Guidance. 2020. Available at: https://apps.who.int/iris/handle/10665/330854. Accessed June 17, 2020

19. Worster A, Bledsoe RD, Cleve P, et al: Reassessing the methods of medical record review studies in emergency medicine research. *Ann Emerg Med* 2005; 45:448–451

20. Hosmer DW, Lemeshow S, Sturdivant RX: Applied Logistic Regression. Third Edition. New York, NY, Wiley, 2013

21. Youden WJ: Index for rating diagnostic tests. *Cancer* 1950; 3:32–35

22. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 1988; 44:837–845

23. Hosmer D: A goodness-of-fit test of the multiple logistic regression model. *Commun Stat Theory Methods* 1980; 9:1043–1069

24. Van Calster B, Nieboer D, Vergouwe Y, et al: A calibration hierarchy for risk models was defined: From utopia to empirical data. *J Clin Epidemiol* 2016; 74:167–176

25. Redfern OC, Smith GB, Prytherch DR, et al: A comparison of the Quick Sequential (Sepsis-Related) Organ Failure Assessment Score

and the National Early Warning Score in non-ICU patients with/without infection. *Crit Care Med* 2018; 46:1923–1933

26. Guo L, Wei D, Zhang X, et al: Clinical features predicting mortality risk in patients with viral pneumonia: The MuLBSTA score. *Front Microbiol* 2019; 10:2752

27. Baker T, Blixt J, Lugazia E, et al: Single deranged physiologic parameters are associated with mortality in a low-income country. *Crit Care Med* 2015; 43:2171–2179

28. Smith GB, Prytherch DR, Jarvis S, et al: A comparison of the ability of the physiologic components of Medical Emergency Team Criteria and the U.K. National Early Warning Score to discriminate patients at risk of a range of adverse clinical outcomes. *Crit Care Med* 2016; 44:2171–2181

29. National Health Commission of the People's Republic of China: National Health Commission of the People's Republic of China Home Page. 2020. Available at: http://www.nhc.gov.cn. Accessed June 17, 2020