

OPEN

Performance Analysis of the National Early Warning Score and Modified Early Warning Score in the Adaptive COVID-19 Treatment Trial Cohort

OBJECTIVES: We sought to validate prognostic scores in coronavirus disease 2019 including National Early Warning Score, Modified Early Warning Score, and age-based modifications, and define their performance characteristics.

DESIGN: We analyzed prospectively collected data from the Adaptive COVID-19 Treatment Trial. National Early Warning Score was collected daily during the trial, Modified Early Warning Score was calculated, and age applied to both scores. We assessed prognostic value for the end points of recovery, mechanical ventilation, and death for score at enrollment, average, and slope of score over the first 48 hours.

SETTING: A multisite international inpatient trial.

PATIENTS: A total of 1,062 adult nonpregnant inpatients with severe coronavirus disease 2019 pneumonia.

INTERVENTIONS: Adaptive COVID-19 Treatment Trial 1 randomized participants to receive remdesivir or placebo. The prognostic value of predictive scores was evaluated in both groups separately to assess for differential performance in the setting of remdesivir treatment.

MEASUREMENTS AND MAIN RESULTS: For mortality, baseline National Early Warning Score and Modified Early Warning Score were weakly to moderately prognostic (c-index, 0.60–0.68), and improved with addition of age (c-index, 0.66–0.74). For recovery, baseline National Early Warning Score and Modified Early Warning Score demonstrated somewhat better prognostic ability (c-index, 0.65–0.69); however, National Early Warning Score+age and Modified Early Warning Score+age further improved performance (c-index, 0.68–0.71). For deterioration, baseline National Early Warning Score and Modified Early Warning Score were weakly to moderately prognostic (c-index, 0.59–0.69) and improved with addition of age (c-index, 0.63–0.70). All prognostic performance improvements due to addition of age were significant ($p < 0.05$).

CONCLUSIONS: In the Adaptive COVID-19 Treatment Trial 1 cohort, National Early Warning Score and Modified Early Warning Score demonstrated moderate prognostic performance in patients with severe coronavirus disease 2019, with improvement in predictive ability for National Early Warning Score+age and Modified Early Warning Score+age. Area under receiver operating curve for National Early Warning Score and Modified Early Warning Score improved in patients receiving remdesivir versus placebo early in the pandemic for recovery and mortality. Although these scores are simple and readily obtainable in myriad settings, in our data set, they were insufficiently predictive to completely replace clinical judgment in coronavirus disease 2019 and may serve best as an adjunct to triage, disposition, and resourcing decisions.

Christopher J. Colombo, MD, MA,
FACP, FCCM^{1,2}

Rhonda E. Colombo, MD, MHS,
FACP, FIDSA¹⁻³

Ryan C. Maves, MD, FCCM, FCCP,
FIDSA^{2,4}

Angela R. Branche, MD⁵

Stuart H. Cohen, MD⁶

Marie-Carmelle Elie, MD⁷

Sarah L. George, MD⁸

Hannah J. Jang, PhD, RN, CNL, PHN⁹

Andre C. Kalil, MD, MPH¹⁰

David A. Lindholm, MD, FACP^{2,11}

Richard A. Mularski, MD, MSHS,
MCR, ATSF, FCCP, FACP¹²

Justin R. Ortiz, MD, MS, FACP,
FCCP¹³

Victor Tapson, MD¹⁴

C. Jason Liang, PhD¹⁵

On behalf of the ACTT-1 Study
Group

Written work prepared by employees of the Federal Government as part of their official duties is, under the U.S. Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the United States Code is not available. As such, copyright does not extend to the contributions of employees of the Federal Government.

DOI: 10.1097/CCE.0000000000000474

KEY WORDS: age; coronavirus disease 2019; Modified Early Warning Score; National Early Warning Score; prognostic scores; severe acute respiratory syndrome coronavirus 2

Throughout the coronavirus disease 2019 (COVID-19) pandemic (1), disease incidence, the need for critical care, and mortality rates have varied widely among industrialized nations (2–4). In addition to geographical variation and temporal trends in the pandemic, older age, male sex, and non-White race have been associated with an increased risk for ICU admission and death (5). Baseline comorbidities, such as diabetes mellitus, hypertension, and obesity, are also associated with increased risk of severe COVID-19 (6, 7). Nevertheless, factors that predict clinical outcomes in COVID-19 remain incompletely characterized (8), pointing to the need for evaluating the ability of clinical scores to predict outcomes in this novel clinical entity. Prognostic scores have been used to predict outcomes in multiple infections (9–11), guiding clinicians on management decisions, including hospitalization and the need for critical care.

Distinctions among mild disease not requiring hospitalization, severe disease requiring hospitalization, and critical illness requiring potentially scarce resources (e.g., ventilators and ICU beds) have become magnified in the COVID-19 pandemic. Hospital beds have become limited resources in many instances during the pandemic, giving new urgency to determining the utility of these scores for resource-limited settings.

Evaluating prognostic scores post hoc using data from clinical studies can identify previously unappreciated risk factors and combinations of risk factors. The COVID-19 pandemic is neither resolved nor likely to be the last pandemic of its kind; thus, there is a need to assess and optimize accuracy of prognostic models for this disease. Multiple established prognostic scores have been evaluated in COVID-19 (12–14), but many of these are small and/or single-center investigations limiting generalizability. Additionally, many novel prognostic scores have been developed (15–20) but are limited by reporting limitations, potential biases (21), and requirements for nonreadily available data elements.

The National Early Warning Score (NEWS), initially developed to detect clinical worsening in adult patients, was widely used early in the pandemic. NEWS

is calculated using the patient's respiratory rate, oxygen saturation, use of supplemental oxygen, temperature, systolic blood pressure, heart rate, and level of consciousness (22). A graded scoring system applied to each parameter generates an overall score. A NEWS greater than or equal to 7 represents a high clinical risk of deterioration, and a score of 5 or 6 indicates medium risk (22). The predictive value of NEWS and NEWS2 (a modification of NEWS that incorporates an additional oxygenation scale for patients at risk of hypercapnic respiratory failure) (23) for worse clinical outcomes in COVID-19, including ICU admission and death, varies depending on the study population, definition of clinical deterioration, and thresholds used (2, 3, 12–14, 24–26). Both NEWS (14) and NEWS2 (27) were shown to outperform quick Sequential Organ Failure Assessment (qSOFA) in predicting outcomes for severe COVID-19 but neither offered an advantage over the other (3, 12). Acknowledging age as a strong risk factor in COVID-19 (6, 28), a modified NEWS incorporating age greater than or equal to 65 years was reported in the literature early in the pandemic (29). Despite the utility of age as a risk modifier specific to COVID-19, this score (NEWS+age) has not yet been validated. Another single-center study concluded the addition of age to NEWS offered no additional prognostic value (13). However, a separate multicenter study found that adding age along with various laboratory markers to NEWS2 improved its predictive accuracy in COVID-19 (30).

The Modified Early Warning Score (MEWS), another score aimed at early detection of clinical deterioration, uses the same variables as NEWS, minus oxygen saturation, and use of supplemental oxygen (31). However, the weighting of the variables differs. In two small studies, MEWS showed worse predictive performance compared with NEWS, NEWS2, and a machine learning algorithm (12, 32). However, MEWS's ease of application makes it an attractive candidate and warrants further systematic investigation in a well-defined COVID-19 cohort.

NEWS and MEWS were originally evaluated for their ability to predict eventual deterioration in the emergency department prior to admission. Despite this original intended use, these scores have also been used during the COVID-19 pandemic to predict deterioration in inpatients and, home care patients, both as a single measurement to support disposition decisions,

and with ongoing serial measurements to determine trends for early warning of deterioration.

Here, we evaluate NEWS, MEWS, and age-based modifications of each in the Adaptive COVID-19 Treatment Trial (ACTT)-1 study population (33), enabling evaluation of these prognostic scores to predict outcomes (mortality, time to recovery, and time to deterioration) in a strictly defined inpatient cohort using prospectively obtained data and accounting for the impact of remdesivir as treatment.

MATERIALS AND METHODS

Design

We analyzed prospectively collected data from ACTT-1, an international, multisite, double-blind, randomized, placebo-controlled trial of IV remdesivir in 1,062 adults hospitalized for COVID-19. The ACTT-1 trial protocol was approved by a central institutional review board (IRB; National Institute of Health Division of Microbiology and Infectious Disease, protocol approval number 20-0006) or local institutional review boards when required (see **Supplemental Material**, <http://links.lww.com/CCX/A698>, for complete list). It was overseen by an independent data and safety monitoring board. Subsequent analyses by the investigators of the deidentified data collected in the initial trial were deemed exempt from additional IRB review.

Inclusion and exclusion criteria, study population, and design of ACTT-1 have been previously described (33). Briefly, ACTT-1 enrolled subjects with severe COVID-19, defined as hospitalization for COVID-19 with evidence of lower respiratory tract infection and with an estimated glomerular filtration rate greater than or equal to 30 mL/min. Subjects were randomized 1:1 to receive remdesivir or placebo and followed for 29 days with daily assessments while hospitalized. Individual clinical status was assessed using an 8-category ordinal scale that differentiates patients by their level of required respiratory support; the highest ordinal category in the preceding 24 hours was captured for a given day. NEWS at baseline and daily was captured just prior to study product administration on dosing days and every subsequent inpatient day at approximately the same time every day. We analyzed the results separately by treatment arm, in acknowledgment that prognostic score performance may differ between the two populations.

Validating Existing Risk Scores

Our primary objective was to validate previously published risk scores in a hospitalized COVID-19 population accounting for remdesivir treatment. We selected the risk scores based on review of previously published predictive models (21, 25) and variable availability in the dataset (e.g., blood gas analysis was not collected in ACTT-1, and we were unable to calculate NEWS2 as a comparator). We chose NEWS, because it was prospectively collected as part of ACTT-1, has shown promise in previously published studies, uses readily available physiologic parameters, and is easy to implement. MEWS was chosen as a comparator as it shares many of these positive features of NEWS, and is also widely used in assessing risk in inpatient settings. We calculated MEWS from the data points comprising NEWS.

Evaluating Modifications to Existing Risk Scores

The secondary analysis assessed how modifications to NEWS and MEWS may improve prognostic performance. Advanced age has been consistently identified as a risk for worse outcomes in COVID-19 (25). An age-based modification of NEWS (NEWS+age), wherein patients 65 years or older were assigned three additional points, has previously been reported in the literature (29) but has not yet been validated. We defined an analogous age-based modification of MEWS (MEWS+age). MEWS and NEWS have the potential for short-term variability, which could either enhance sensitivity or introduce unhelpful volatility. As an approach to minimize the potential volatility of a single measurement, we evaluated the 48-hour average of NEWS, MEWS, NEWS+age, and MEWS+age from the first 2 days following study enrollment. In order to evaluate the value of score trend versus single measurements, we evaluated the 1-day change (slope) from baseline for all scores.

End Points

The risk scores were evaluated for their ability to predict time to mortality, time to recovery, and time to deterioration. Time to mortality was defined as days from randomization to death; similar to the ACTT-1 primary analysis, death was censored at 29 days. Time to recovery was defined as days from

randomization to the first day the patient met criteria for category 1, 2, or 3 on the eight-category ordinal scale used in ACTT-1 (**Supplemental Table 1**, <http://links.lww.com/CCX/A699>) (33). To determine the risk of utilizing specific critical care resources (e.g., noninvasive and invasive ventilators), the time to deterioration was defined as days from randomization to the first day a participant met criteria for ordinal category 6, 7, or 8. For each end point, participants were counted only once based on when the outcome was first achieved. Additionally, the risk scores were evaluated for their ability to predict the three binary end points of 14-day mortality, recovery, and deterioration, as well as the three binary end points of 28-day mortality, recovery, and deterioration.

Statistical Analysis

We registered an a priori analysis plan, available at <https://aspredicted.org/kq798.pdf>. The study was conceived with knowledge of the primary results from ACTT-1 (33) but prior to gaining access to the patient-level data. A description of deviations from the analysis plan is detailed in the **Supplement** (<http://links.lww.com/CCX/A699>).

For the time-to-event end points, the c-index (34) was used to evaluate prognostic value. For the binary end points of 14- and 28-day events, the area under the receiver operating curve (AUC) was used to evaluate prognostic value. Specifically, the cumulative/dynamic AUC (35) was used for estimation in order to account for censored observations. The positive predictive value (PPV) and negative predictive value (NPV) were also estimated for meaningful score thresholds. Graphics of the receiver operating characteristic (ROC) curves for both 14- and 28-day events are presented in the figures.

Analysis of the average NEWS/MEWS and slope NEWS/MEWS was restricted to participants still in the study at day 2. For analysis of time to recovery, patients who died were censored at day 29, similar to the ACTT-1 primary analysis. For analyses using time to deterioration, participants categorized as ordinal category 6 or 7 at baseline were excluded. Exact respiratory rates were not recorded for those already on mechanical ventilation but were the maximal score for respiratory rate when calculating NEWS and MEWS.

RESULTS

Descriptive plots of baseline NEWS against time to different end points are shown in **Figure 1**. Analogous plots for other risk scores are in the including NEWS+age (**Supplemental Fig. 1A**), MEWS (**Supplemental Fig. 1B**), MEWS+age (**Supplemental Fig. 1C**), average NEWS first 48 hours (**Supplemental Fig. 1D**), slope of NEWS first 48 hours (**Supplemental Fig. 1E**), average MEWS first 48 hours (**Supplemental Fig. 1F**), slope of MEWS first 48 hours (**Supplemental Fig. 1G**) (Supplement, <http://links.lww.com/CCX/A699>). Demographics for the ACTT-1 cohort for the placebo, remdesivir and total groups are reported in **Supplemental Table 2** (Supplement, <http://links.lww.com/CCX/A699>). For both the mortality and recovery end points, the placebo arm had 512 patients (77 deaths and 352 recoveries by day 29), and the remdesivir arm had 531 patients (59 deaths and 399 recoveries by day 29). Distribution of the cohort across the NEWS scores, and descriptive statistics of the components of the NEWS scores are reported in **Supplemental Table 3** (Supplement, <http://links.lww.com/CCX/A699>). Kaplan-Meier plots of the placebo and remdesivir arms stratified by baseline NEWS is shown in **Supplemental Figure 3** (Supplement, <http://links.lww.com/CCX/A699>). For deterioration, after excluding those with an ordinal score of 6 or 7 at baseline, the placebo arm had 265 patients (90 deteriorations by day 29) and the remdesivir arm had 306 patients (65 deteriorations by day 29). For average NEWS and change in NEWS between days 1 and 2, the placebo and remdesivir arms had 507 and 525 patients, respectively. The median time to recovery was 15 days in the placebo arm and 10 days in the remdesivir arm. **Table 1** contains c-index data for all scores (including the average and slope measurements) for all end points. The c-indexes will be described in the following sections for each end point, and then statistical evaluations for the addition of age and alternate methods of calculation will be discussed.

Validating Existing Risk Scores

NEWS, MEWS, and NEWS+age were prespecified as risk scores to evaluate. MEWS+age was not a prespecified existing risk score but is included in this section to facilitate comparisons with MEWS and NEWS+age.

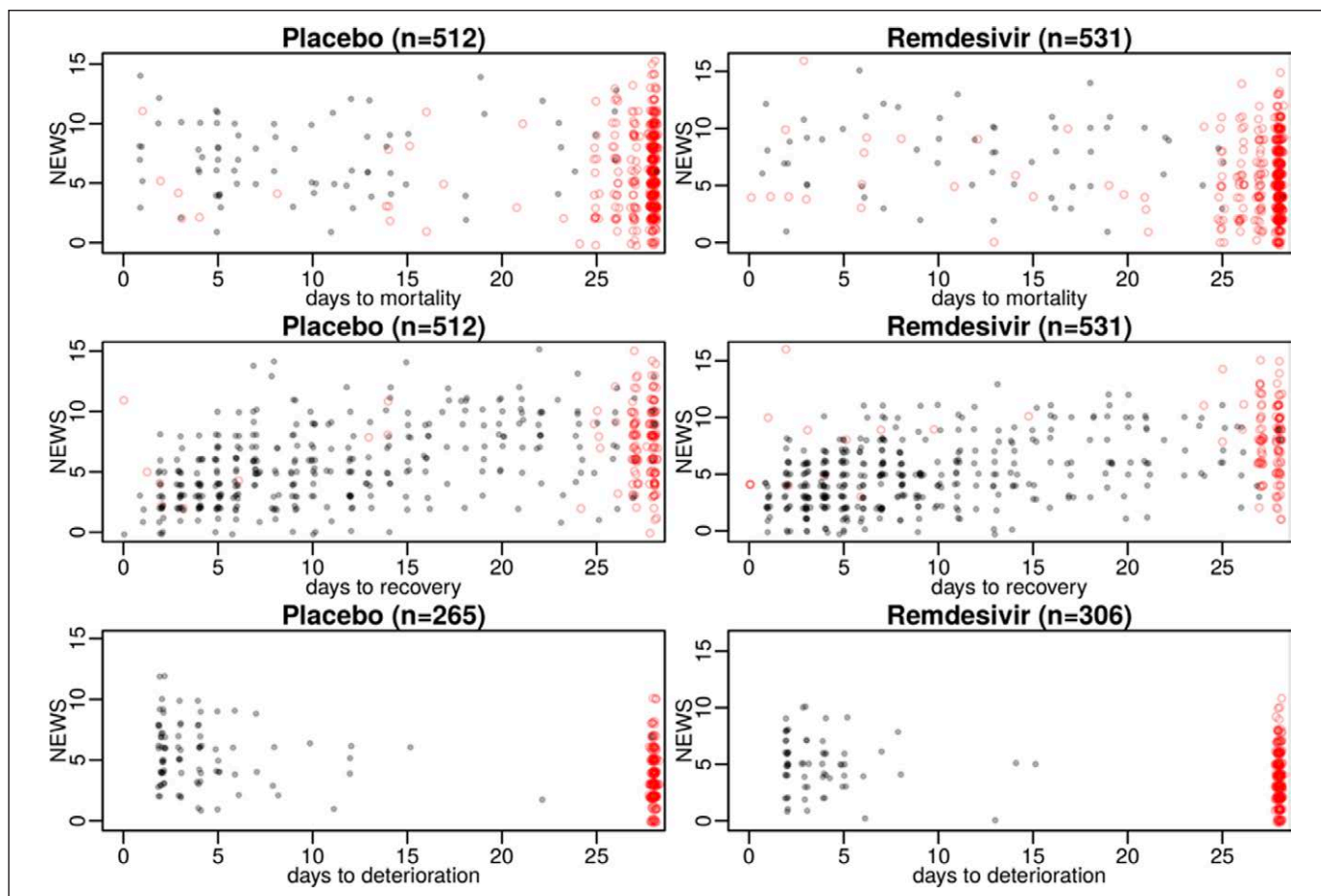


Figure 1. Illustration of baseline National Early Warning Score (NEWS) plotted against different end points, by placebo arm and remdesivir arm. *Gray dots* represent events and *red dots* represent censoring. *Points* are jittered to address overplotting.

For the mortality end point, baseline NEWS and MEWS were weakly to moderately prognostic (c-index, 0.60–0.68), whereas NEWS+age and MEWS+age were moderately prognostic (c-index, 0.66–0.74).

For the recovery end point, the risk scores demonstrated better prognostic ability compared with the mortality end point. Baseline NEWS and MEWS were moderately prognostic (c-index, 0.65–0.69). Unlike the mortality end point, NEWS+age and MEWS+age only modestly improved prognostic performance (c-index, 0.68–0.71).

For the deterioration end point, baseline NEWS and MEWS were weakly to moderately prognostic (c-index, 0.59–0.69), whereas NEWS+age and MEWS+age modestly improved prognostic performance (c-index, 0.63–0.70).

Overall, prognostic performance between NEWS and MEWS was similar, as were NEWS+age and MEWS+age. Adding age to NEWS modestly improved prognostic performance for mortality (change in c-index in placebo and remdesivir arms: $p = 0.009$ and

0.016, respectively) and recovery (change in c-index in placebo and remdesivir arms: $p = 0.030$ and 0.013, respectively). Similarly, for MEWS, adding age modestly improved prognostic performance for mortality (change in c-index in placebo and remdesivir arms: $p = 0.017$ and 0.027, respectively) and recovery (change in c-index in placebo and remdesivir arms: $p = 0.047$ and 0.076, respectively). Complete results for NEWS, MEWS, NEWS+age, and MEWS+age are shown in Table 1. Graphics of the ROC curves for 14-day events are shown in **Figure 2**, whereas ROC curves for 28-day events are in the **Supplemental Figure 2, A and B** (Supplement, <http://links.lww.com/CCX/A699>).

Although the original NEWS proposed a score of 0–4 as low risk for clinical worsening (22), a median presenting score of 3 has been correlated with discharge from an emergency department and higher scores with hospital admission (36). We thus refer to a NEWS of 3 or lower as “low NEWS” and 7 or higher as “high NEWS.” Study population 14-day event rates and event rates in the high and low NEWS groups, along with

TABLE 1.
C-Index (95% CI) for 14-d Cumulative/Dynamic Area Under Receiver Operating Curve and 28-d Cumulative/Dynamic Area Under Receiver Operating Curve for Each Risk Score, End Point, and Treatment Arm

	Placebo C-Index	Remdesivir C-Index	Placebo 14-d AUC	Remdesivir 14-d AUC	Placebo 28-d AUC	Remdesivir 28-d AUC
Risk score for mortality end point						
NEWS	0.60 (0.54–0.66)	0.68 (0.61–0.74)	0.59 (0.51–0.66)	0.70 (0.60–0.79)	0.61 (0.55–0.68)	0.68 (0.61–0.76)
NEWS+age	0.66 (0.60–0.72)	0.73 (0.67–0.79)	0.65 (0.58–0.71)	0.77 (0.69–0.84)	0.67 (0.61–0.74)	0.74 (0.68–0.81)
MEWS	0.59 (0.53–0.65)	0.66 (0.60–0.73)	0.56 (0.48–0.64)	0.65 (0.56–0.73)	0.60 (0.53–0.67)	0.67 (0.60–0.74)
MEWS+age	0.67 (0.61–0.73)	0.74 (0.68–0.80)	0.64 (0.57–0.71)	0.76 (0.68–0.83)	0.69 (0.63–0.75)	0.75 (0.68–0.81)
NEWS avg	0.66 (0.60–0.71)	0.71 (0.64–0.78)	0.66 (0.58–0.72)	0.73 (0.63–0.82)	0.67 (0.61–0.73)	0.72 (0.64–0.79)
NEWS slope	0.63 (0.56–0.69)	0.55 (0.47–0.63)	0.67 (0.59–0.76)	0.54 (0.42–0.65)	0.62 (0.56–0.70)	0.56 (0.47–0.64)
MEWS avg	0.65 (0.59–0.70)	0.71 (0.64–0.77)	0.64 (0.56–0.70)	0.71 (0.61–0.79)	0.66 (0.60–0.72)	0.72 (0.65–0.78)
MEWS slope	0.57 (0.49–0.64)	0.53 (0.45–0.61)	0.62 (0.54–0.70)	0.60 (0.51–0.71)	0.56 (0.48–0.64)	0.53 (0.44–0.61)
Risk score for recovery end point						
NEWS	0.68 (0.65–0.71)	0.69 (0.67–0.72)	0.76 (0.72–0.81)	0.79 (0.74–0.83)	0.67 (0.62–0.71)	0.76 (0.71–0.81)
NEWS+age	0.70 (0.67–0.72)	0.71 (0.69–0.74)	0.78 (0.74–0.82)	0.80 (0.76–0.84)	0.70 (0.65–0.75)	0.78 (0.73–0.82)
MEWS	0.65 (0.62–0.68)	0.67 (0.64–0.70)	0.72 (0.68–0.77)	0.76 (0.72–0.80)	0.66 (0.60–0.71)	0.74 (0.68–0.79)
MEWS+age	0.68 (0.65–0.71)	0.69 (0.67–0.72)	0.75 (0.71–0.79)	0.78 (0.73–0.81)	0.70 (0.65–0.75)	0.76 (0.70–0.81)
NEWS avg	0.72 (0.69–0.74)	0.73 (0.71–0.76)	0.82 (0.78–0.85)	0.82 (0.78–0.85)	0.72 (0.67–0.77)	0.8 (0.75–0.84)
NEWS slope	0.58 (0.55–0.61)	0.56 (0.53–0.59)	0.60 (0.55–0.65)	0.56 (0.51–0.61)	0.62 (0.56–0.68)	0.56 (0.49–0.61)
MEWS avg	0.70 (0.67–0.72)	0.72 (0.69–0.74)	0.79 (0.75–0.82)	0.81 (0.77–0.85)	0.72 (0.67–0.77)	0.79 (0.74–0.83)
MEWS slope	0.56 (0.53–0.59)	0.53 (0.50–0.56)	0.57 (0.52–0.62)	0.52 (0.47–0.58)	0.58 (0.52–0.63)	0.52 (0.46–0.58)
Risk score for deterioration end point						
NEWS	0.69 (0.64–0.75)	0.65 (0.58–0.71)	0.71 (0.65–0.78)	0.65 (0.58–0.73)	0.72 (0.65–0.78)	0.65 (0.58–0.72)
NEWS+age	0.70 (0.65–0.75)	0.66 (0.59–0.73)	0.73 (0.66–0.79)	0.67 (0.59–0.75)	0.74 (0.67–0.80)	0.68 (0.61–0.74)
MEWS	0.62 (0.56–0.68)	0.59 (0.53–0.66)	0.62 (0.56–0.69)	0.61 (0.54–0.68)	0.64 (0.57–0.71)	0.60 (0.53–0.67)
MEWS+age	0.64 (0.58–0.70)	0.63 (0.57–0.70)	0.66 (0.59–0.73)	0.65 (0.56–0.72)	0.68 (0.61–0.74)	0.65 (0.57–0.72)
NEWS avg	0.78 (0.73–0.83)	0.71 (0.65–0.77)	0.82 (0.76–0.87)	0.73 (0.66–0.79)	0.82 (0.76–0.87)	0.73 (0.66–0.80)
NEWS slope	0.61 (0.55–0.67)	0.59 (0.51–0.66)	0.63 (0.56–0.70)	0.59 (0.51–0.68)	0.63 (0.56–0.70)	0.59 (0.51–0.67)
MEWS avg	0.71 (0.65–0.77)	0.67 (0.60–0.74)	0.73 (0.66–0.79)	0.68 (0.60–0.75)	0.73 (0.66–0.80)	0.68 (0.60–0.76)
MEWS slope	0.57 (0.51–0.64)	0.56 (0.49–0.64)	0.60 (0.52–0.68)	0.56 (0.47–0.63)	0.58 (0.51–0.66)	0.56 (0.48–0.64)

AUC = area under receiver operating curve, MEWS = Modified Early Warning Score, MEWS avg = average Modified Early Warning Score, MEWS slope = change in Modified Early Warning Score, NEWS = National Early Warning Score, NEWS avg = average National Early Warning Score, NEWS slope = change in National Early Warning Score.

PPV and NPV interpretations, are shown in **Table 2**. Among participants receiving placebo, NEWS of less than or equal to 3 had a high NPV for mortality (0.95), whereas the PPV of a high NEWS (7 or greater) was 0.14 for mortality. Similar results for mortality were noted in the remdesivir arm, with a PPV for mortality of 0.12 for a NEWS of 7 or greater and a NPV of 0.97 for NEWS less than or equal to 3.

Among participants receiving placebo, NEWS of less than or equal to 3 had a moderately high PPV for recovery (0.77) and the NPV of a high NEWS (7 or greater) was 0.75 for recovery. Similar results for recovery were noted in the remdesivir arm, with a PPV for recovery of 0.84 for a NEWS of less than or equal to 3, whereas the NPV for recovery of a NEWS of 7 or greater was 0.68.

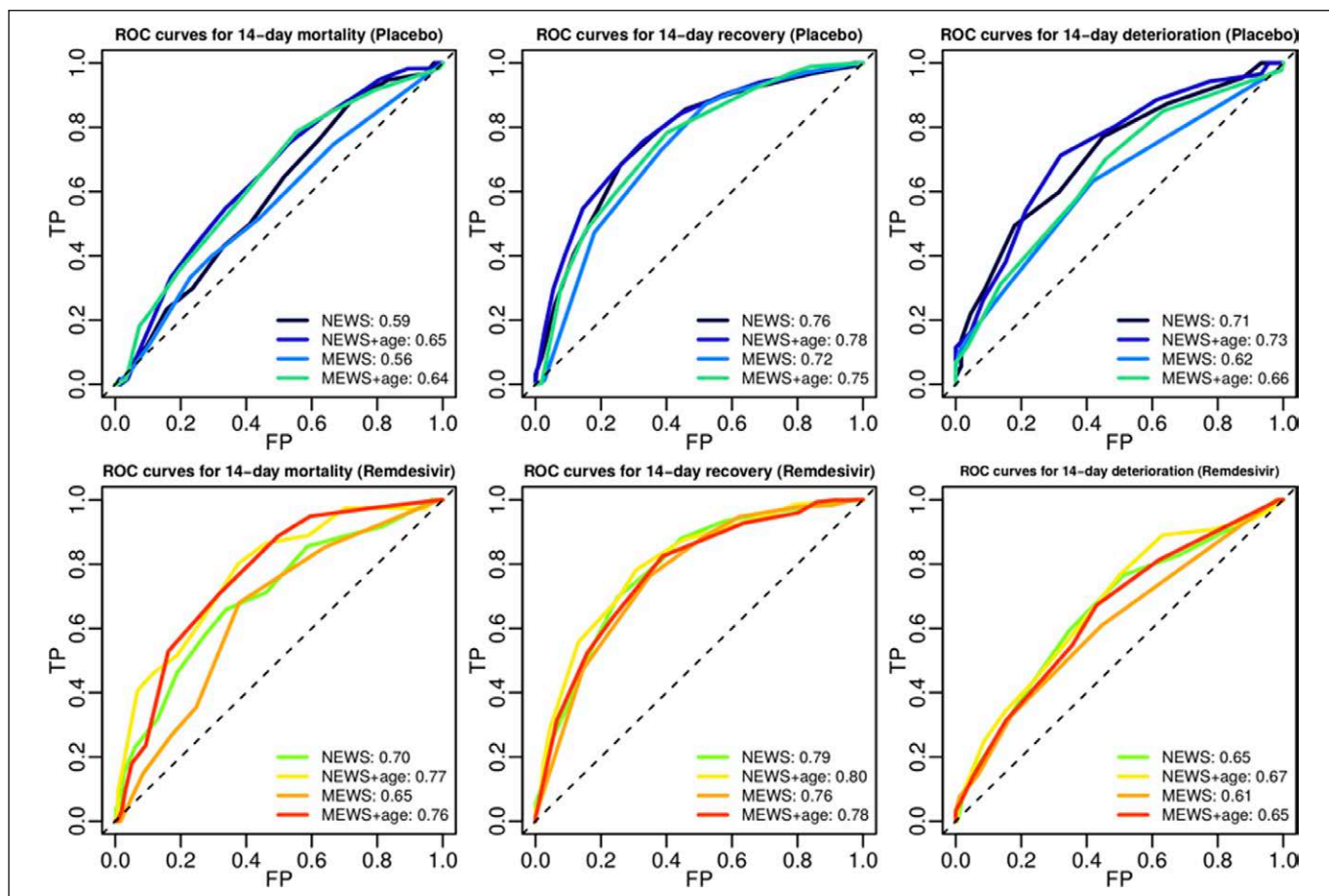


Figure 2. Receiver operating characteristic (ROC) curves plotting false positives (FP) on the x-axis and true positives (TP) on the y-axis summarizing sensitivity and specificity of National Early Warning Score (NEWS), Modified Early Warning Score (MEWS), NEWS+age, and MEWS+age for the end points of 14-d mortality, recovery, and deterioration. The area under receiver operating curve for each curve is shown in the legend.

Among participants receiving placebo, NEWS of 7 or greater had a PPV of 0.62 for deterioration and a NEWS of less than or equal to 3 had a NPV of 0.83. For participants in the remdesivir group, NEWS of 7 or greater had a PPV of 0.37 and NEWS of less than or equal to 3

showed an NPV of 0.89 for deterioration. Values shown for PPV and NPV are at cutoffs we selected as optimal. **Supplemental Tables 4–6** (Supplement, <http://links.lww.com/CCX/A699>) provide PPV and NPV for 14 day mortality, recovery and deterioration respectively.

TABLE 2.
14-d Event Rates According to Treatment and National Early Warning Score Groups

Group	Mortality Rate (%)	Recovery Rate (%)	Deterioration Rate (%)
Placebo	12	49	33
High NEWS	14 (0.14 PPV)	25 (0.75 NPV)	62 (0.62 PPV)
Low NEWS	5 (0.95 NPV)	77 (0.77 PPV)	17 (0.83 NPV)
Remdesivir	7	61	21
High NEWS	12 (0.12 PPV)	32 (0.68 NPV)	37 (0.37 PPV)
Low NEWS	3 (0.97 NPV)	84 (0.84 PPV)	11 (0.89 NPV)

High NEWS = National Early Warning Score greater than or equal to 7, Low NEWS = National Early Warning Score less than or equal to 3, NPV = negative predictive value, PPV = positive predictive value.

Evaluating Modifications to Existing Risk Scores

For the mortality end point, longitudinal averages of NEWS and MEWS were moderately prognostic (c-index, 0.65–0.71), comparable with adding age to baseline NEWS and MEWS. The slope of NEWS and MEWS was weakly to moderately prognostic (c-index, 0.53–0.63).

For the recovery end point, longitudinal averages of NEWS and MEWS were moderately prognostic (c-index, 0.70–0.73) and slightly better than adding age to baseline NEWS and MEWS. The slope of NEWS and MEWS was weakly prognostic (c-index, 0.53–0.58).

For the deterioration end point, longitudinal averages of NEWS and MEWS were moderately prognostic (c-index, 0.67–0.78) and slightly better than adding age to baseline NEWS and MEWS. The slope of NEWS and MEWS was weakly prognostic (c-index, 0.59–0.61).

Overall, prognostic performance between longitudinal summaries of NEWS and MEWS was similar. Compared with baseline NEWS alone, averaging the first two NEWS scores modestly improved prognostic performance for mortality (change in c-index in placebo and remdesivir arms: $p < 0.001$ and 0.017 , respectively), recovery (change in c-index in placebo and remdesivir arms: $p < 0.001$ and < 0.001 , respectively), and deterioration (change in c-index in placebo and remdesivir arms: $p < 0.001$ and 0.002 , respectively).

Similarly, for MEWS, averaging the first two MEWS scores modestly improved prognostic performance for mortality (change in c-index in placebo and remdesivir arms: $p = 0.004$ and 0.018 , respectively), recovery (change in c-index in placebo and remdesivir arms: $p < 0.001$ and < 0.001 , respectively), and deterioration (change in c-index in placebo and remdesivir arms: $p < 0.001$ and 0.003 , respectively). The slope of the first two observations did not improve performance.

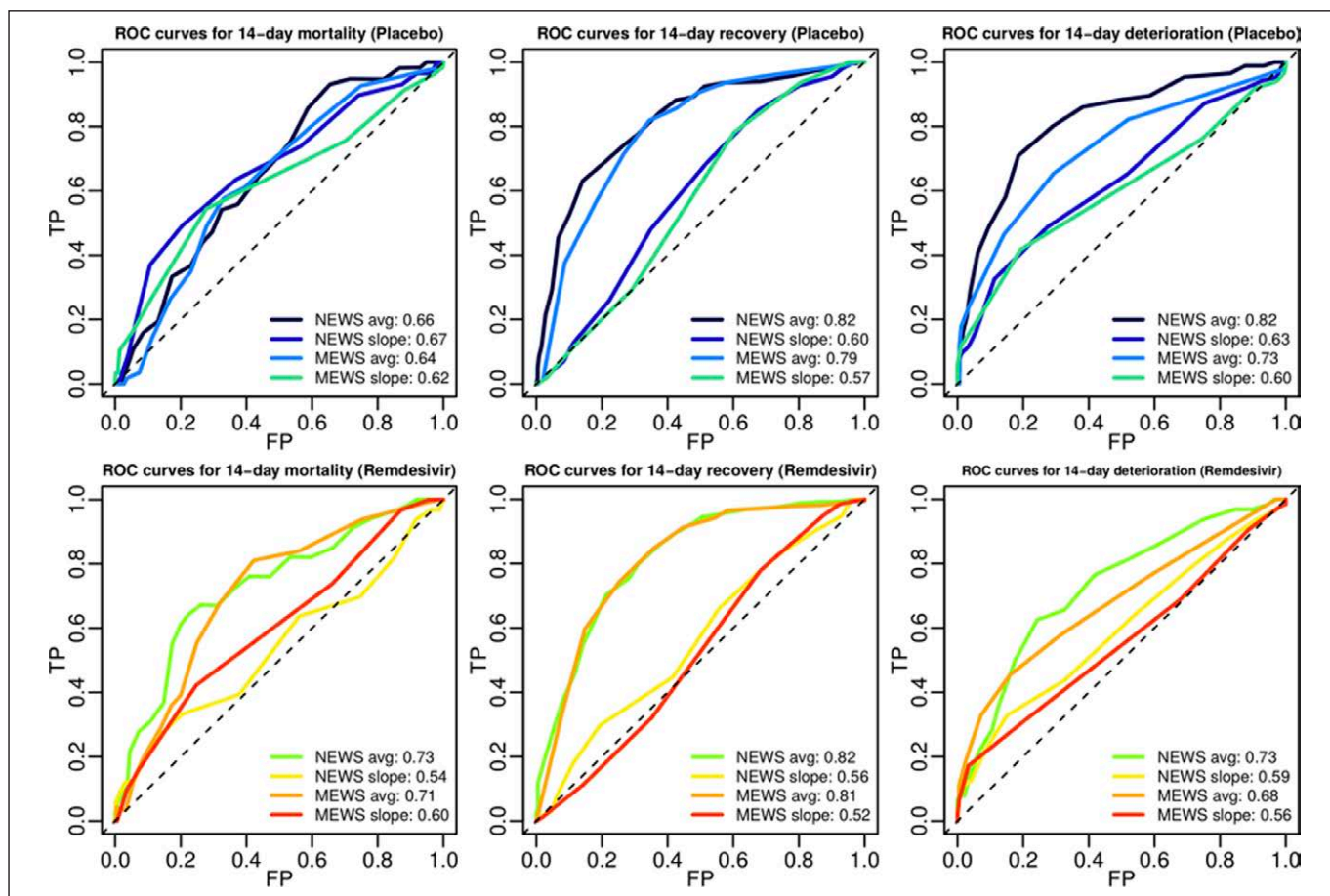


Figure 3. Receiver operating characteristic (ROC) curves plotting false positives (FP) on the x-axis and true positives (TP) on the y-axis summarizing sensitivity and specificity of average NEWS (NEWS avg), average MEWS (MEWS avg), change in NEWS (NEWS slope), and change in MEWS (MEWS slope) for the end points of 14-d mortality, recovery, and deterioration. The area under receiver operating curve for each curve is shown in the legend.

Graphics of the ROC curves for 14-day events are shown in **Figure 3**. ROC curves for 28-day events are in the Supplement (<http://links.lww.com/CCX/A699>).

DISCUSSION

Prognostic score performance depends greatly on the intended function and clinical context. They may attempt to precisely define a syndrome, provide prognostic information in seriously ill patients, or identify patients at risk for decompensation. A growing number of prognostic scores have been described in patients with COVID-19 (21).

An effective clinical prediction score for COVID-19 should identify increased risk for progression to respiratory failure and/or death and should occur early enough to permit preventive interventions. The score must not only be correlated with outcome but should guide and hone a clinician's experience in judging so-called edge cases through regular use. Systematic assessment of clinical prediction scores has shown many have poor accuracy and methodological issues (36). Additionally, some studies have demonstrated that clinician judgment can predict clinical outcomes as well as predictive scores and that increased experience improved accuracy (37, 38). In cases where experienced clinician judgment and prediction score methods correlate, the score should provide a ready means to bridge this "experience gap" for early-career clinicians. Readily available physiologic parameters, (e.g., vital signs and level of consciousness) are ideally suited for inclusion in early warning scores. Laboratory studies are time- and resource-intensive but may be useful if readily accessible. Assays with long turnaround times, such as interleukin-6, are unlikely to be useful at the point of care. Finally, a prognostic score that identifies patients based on obvious features (e.g., severe hypoxemia as a risk factor for endotracheal intubation) has little clinical value.

Early warning scores, including NEWS and MEWS, have the advantage of being quickly assessed at the bedside, not requiring laboratory assays, and having reasonable prognostic performance in several settings (39–42). Multiple studies evaluating the performance of early warning scores in COVID-19 (2, 3, 12–14, 24–26, 43, 44) show mixed results. NEWS appeared to outperform qSOFA in COVID-19 in a retrospective series of 110 patients in South Korea (14) and a similar series of 673 patients in Wuhan (44). A NEWS2 score of 7 or greater was associated with increased risk of ICU

admission in a series of 68 patients in Italy (2). NEWS and NEWS2 had generally consistent AUC scores of less than 0.75 in these series, suggesting only moderate prognostic performance in COVID-19. An analysis of in the U.K.'s National Health Service showed NEWS and NEWS2 tended to underestimate in-hospital mortality in COVID-19 versus non-COVID-19 admissions, with similar c-statistics (0.64 for NEWS and NEWS2 in COVID-19 patients) (3) as our study.

Many COVID-19-specific novel predictive models have been described (15–20), although several are based on small retrospective series with a high risk of bias (21). The COVID-GRAM risk score was derived using a 1,590-patient development cohort and a 710-patient validation cohort; although predictive of severe disease, the most highly predictive features (including unconsciousness and hemoptysis) would demand clinical attention in the absence of a risk score (15). The 4C Mortality Score used a 34,463-patient derivation cohort and 22,361-patient validation cohort, with age, gender, comorbid disease, respiratory rate, oxygen saturation, Glasgow Coma Scale, blood urea nitrogen, and C-reactive protein as its constituent components (17). Despite these additional variables, the 4C Mortality Score AUC was 0.78 (17), little better than NEWS+age alone in our cohort.

A potential limitation of these scores is the frequency of measurement. With scores based on vital signs, measurement can happen much more frequently than scores that rely on laboratory testing. It is unclear whether this potential to measure short-term variability results in improved accuracy or if the trend over time is more accurate. In our study, we lacked multiple daily measurements due to the limitations of data collection in the original ACTT-1 study.

PPV and NPV are key characteristics of a test's performance from a clinical decision-making standpoint. Although prognostic scores are usually valued for their ability to predict future occurrences, equally important is their ability to predict nonoccurrences, such as a patient not requiring ICU admission or mechanical ventilation. With overwhelmed health systems and limited resources, it may be particularly useful to identify patients at low risk of deterioration who may be sent home safely or managed on a general medical ward. Thus, the NPV of a score to rule out a high risk of deterioration could be the most useful application of a score like NEWS. In our cohort, a NEWS of 3 or less had a NPV of 95–97% for 14-day mortality, and 83–89% for 14-day deterioration,

suggesting that a low NEWS is reassuring in COVID-19. Supporting decisions like this could aid in resource utilization and planning at the local level by helping to conserve capacity, and at the system or regional level by determining the need for greater resources based on features beyond strict case numbers.

We note an improved AUC for both NEWS and MEWS in patients receiving remdesivir versus placebo early in the pandemic, in terms of recovery and mortality. The reasons for this are unclear, but it is noteworthy that patients requiring only low-flow supplemental oxygen in ACTT-1 appeared to have the greatest benefit from antiviral therapy (33). Given the prolonged illness time seen in COVID-19, one may hypothesize that the use of remdesivir early in the course of serious illness may modify the disease trajectory back to a more “normal” acute illness model, wherein predictive scoring is more accurate.

Our study has limitations. Baseline data were collected at enrollment, potentially days after hospital admission. Mortality prediction is limited by the relatively small number of deaths during our observation period. Patients with severely compromised kidney function, a known negative prognostic factor, were excluded. Additionally, due to the once daily recording of NEWS in ACTT, we were only able to use NEWS at one consistent time daily, neither accounting for variations throughout the day nor determining the potential impact of the best or worst score in a 24-hour period. Our study also does not address other, perhaps more common, methods of utilizing NEWS, such as more frequent measurements (up to and including hourly measurements). Although averaging frequent scores may result in a similar outcome as averaging two scores obtained once daily, further comparative study would be required to evaluate this concept. Additionally, the scores were originally designed to be used to predict risk prior to admission, and our entire cohort was enrolled in the ACTT-1 trial after admission.

CONCLUSIONS

Compared with study population event rates, those with high NEWS are more likely to die or deteriorate and less likely to recover, and those with low NEWS are less likely to die or deteriorate and more likely to recover. However, the extent of the risk stratification is insufficient for clinical decision-making based solely

on NEWS. Our prospective data confirmed that the addition of age to NEWS or MEWS improves the performance of these prognostic scores in patients with severe COVID-19. Averaging the NEWS or MEWS scores obtained on the first 2 days may also enhance prognostic ability. However, none of these adjustments to scores are sufficiently predictive to independently guide clinical decisions. More complex models that incorporate other clinical characteristics, including comorbidities and laboratory markers of inflammation or organ dysfunction, may have stronger predictive performance but at the expense of complexity or a longer turnaround time. Further research is needed to determine the optimal method to accurately identify patients at risk for critical illness and death from COVID-19.

- 1 Madigan Army Medical Center, Tacoma, WA.
- 2 Infectious Disease Clinical Research Program, Uniformed Services University of the Health Sciences, Bethesda, MD.
- 3 The Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD.
- 4 Naval Medical Center, San Diego, CA.
- 5 University of Rochester Medical Center, Rochester, NY.
- 6 University of California Davis, Sacramento, CA.
- 7 University of Florida Gainesville, Gainesville, FL.
- 8 Saint Louis University and St. Louis VA Medical Center, Saint Louis, MO.
- 9 Department of Community Health Systems, School of Nursing and Center for Nursing Excellence and Innovation, University of California San Francisco, San Francisco, CA.
- 10 University of Nebraska Medical Center, Omaha, NE.
- 11 Brooke Army Medical Center, San Antonio, TX.
- 12 The Center for Health Research, Kaiser Permanente Northwest, Portland, OR.
- 13 University of Maryland School of Medicine, Baltimore, MD.
- 14 Cedars-Sinai Medical Center, Los Angeles, CA.
- 15 Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, MD.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (<http://journals.lww.com/ccejjournal>).

Drs. C. J. Colombo and R. E. Colombo contributed equally to the conception, execution, writing, editing, and leading of this article.

Drs. C. J. Colombo and R. E. Colombo are coequal first authors.

The Adaptive COVID-19 Treatment Trial was sponsored and primarily funded by the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, MD. This trial has been funded, in part, by federal funds from the NIAID and the National Cancer Institute, NIH, under contract

HHSN261200800001E 75N910D00024 and task order number 75N91019F00130/75N91020F00010, and by the Department of Defense, Defense Health Program. This trial has been supported, in part, by the NIAID of the NIH under award numbers UM1AI148684, UM1AI148576, UM1AI148573, UM1AI148575, UM1AI148452, UM1AI148685, UM1AI148450, and UM1AI148689. The trial has also been funded, in part, by the governments of Denmark, Japan, Mexico, and Singapore. The trial site in South Korea received funding from the Seoul National University Hospital. Support for the London International Coordinating Centre was also provided by the United Kingdom Medical Research Council (MRC_UU_12023/23).

Mention of trade name, commercial products, or organizations does not imply endorsement by the U.S. Government. Some of the authors (C.J.C., R.E.C., R.C.M., D.A.L., C.J.L.) are military service members or employees of the U.S. Government. This work was prepared as part of their official duties. Title 17 U.S.C. §105 provides that copyright protection under this title is not available for any work of the U.S. Government. Title 17 U.S.C. §101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person's official duties. The investigators have adhered to the policies for protection of human subjects as prescribed in 45 CFR 46. This analysis used data from the Adaptive COVID-19 Treatment Trial (doi:10.1056/NEJMoa2007764). Collaborating investigators involved with collection of data during the ACTT-1 trial are noted in the Supplementary Appendix (<http://links.lww.com/CCX/A699>).

The opinions expressed in this article are solely those of the authors and do not represent those of the organizations with which the authors are affiliated, including but not limited to Madigan Army Medical Center, Brooke Army Medical Center, the National Institutes of Health, Uniformed Services University of the Health Sciences, the Henry M. Jackson Foundation for the Advancement of Military Medicine, the Department of the Army, the Department of the Air Force, the Department of the Navy, Department of Defense, Department of Health and Human Services, or the United States Government.

For information regarding this article, E-mail: colombocj@aol.com

REFERENCES

- Zhou F, Yu T, Du R, et al: Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet* 2020; 395:1054–1062
- Gidari A, De Socio GV, Sabbatini S, et al: Predictive value of National Early Warning Score 2 (NEWS2) for intensive care unit admission in patients with SARS-CoV-2 infection. *Infect Dis (Lond)* 2020; 52:698–704
- Richardson D, Faisal M, Fiori M, et al: Use of the first National Early Warning Score recorded within 24 hours of admission to estimate the risk of in-hospital mortality in unplanned COVID-19 patients: A retrospective cohort study. *BMJ Open* 2021; 11:e043721
- Reese H, Iuliano AD, Patel NN, et al: Estimated incidence of COVID-19 illness and hospitalization – United States, February–September, 2020. *Clin Infect Dis* 2021; 72:e1010–e1017
- Galloway JB, Norton S, Barker RD, et al: A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: An observational cohort study. *J Infect* 2020; 81:282–288
- Rosenthal N, Cao Z, Gundrum J, et al: Risk factors associated with in-hospital mortality in a US national sample of patients with COVID-19. *JAMA Netw Open* 2020; 3:e2029058
- Anderson MR, Geleris J, Anderson DR, et al: Body mass index and risk for intubation or death in SARS-CoV-2 infection: A retrospective cohort study. *Ann Intern Med* 2020; 173:782–790
- Izcovich A, Ragusa MA, Tortosa F, et al: Prognostic factors for severity and mortality in patients infected with COVID-19: A systematic review. *PLoS One* 2020; 15:e0241955
- Asai N, Shiota A, Ohashi W, et al: The SOFA score could predict the severity and prognosis of infective endocarditis. *J Infect Chemother* 2019; 25:965–971
- Nguyen DT, Jenkins HE, Graviss EA: Prognostic score to predict mortality during TB treatment in TB/HIV co-infected patients. *PLoS One* 2018; 13:e0196022
- Raith EP, Udy AA, Bailey M, et al; Australian and New Zealand Intensive Care Society (ANZICS) Centre for Outcomes and Resource Evaluation (CORE): Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit. *JAMA* 2017; 317:290–300
- Hu H, Yao N, Qiu Y: Predictive value of 5 early warning scores for critical COVID-19 patients. *Disaster Med Public Health Prep* 2020; 9:1–8
- Volff M, Tonon D, Bourenne J, et al: No added value of the modified NEWS score to predict clinical deterioration in COVID-19 patients. *Anaesth Crit Care Pain Med* 2020; 39:577–578
- Jang JG, Hur J, Hong KS, et al: Prognostic accuracy of the SIRS, qSOFA, and NEWS for early detection of clinical deterioration in SARS-CoV-2 infected patients. *J Korean Med Sci* 2020; 35:e234
- Liang W, Liang H, Ou L, et al: Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Int Med* 2020; 180:1–9
- Guo Y, Liu Y, Lu J, et al: Development and validation of an early warning score (EWAS) for predicting clinical deterioration in patients with coronavirus disease 2019. *medRxiv Preprint* posted online April 21, 2020. doi:10.1101/2020.04.17.20064691
- Knight SR, Ho A, Pius R, et al; ISARIC4C investigators: Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Development and validation of the 4C Mortality Score. *BMJ* 2020; 370:m3339
- Sharp AL, Huang BZ, Broder B, et al: Identifying patients with symptoms suspicious for COVID-19 at elevated risk of adverse events: The COVAS Score. *Am J Emerg Med* 2020 Nov 5. [online ahead of print]
- Haimovich AD, Ravindra NG, Stoytchev S, et al: Development and validation of the quick COVID-19 severity index: A prognostic tool for early clinical decompensation. *Ann Emerg Med* 2020; 76:442–453
- Covino M, De Matteis G, Burzo ML, et al; GEMELLI AGAINST COVID-19 Group: Predicting in-hospital mortality in COVID-19

- older patients with specifically developed scores. *J Am Geriatr Soc* 2021; 69:37–43
21. Wynants L, Van Calster B, Collins GS, et al: Prediction models for diagnosis and prognosis of covid-19 infection: Systematic review and critical appraisal. *BMJ* 2020; 369:m1328
 22. Royal College of Physicians. National Early Warning Score (NEWS): Standardising the Assessment of Acute Illness Severity in the NHS. Report of a Working Party. 2012. RCP. Available at: <https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2>. Accessed December 1, 2020
 23. 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. 2017. RCP. Available at: <https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2>. Accessed December 1, 2020
 24. Myrstad M, Ihle-Hansen H, Tveita AA, et al: National Early Warning Score 2 (NEWS2) on admission predicts severe disease and in-hospital mortality from Covid-19 - a prospective cohort study. *Scand J Trauma Resusc Emerg Med* 2020; 28:66
 25. Gupta RK, Marks M, Samuels THA, et al: Systematic evaluation and external validation of 22 prognostic models among hospitalized adults with COVID-19: An observational cohort study. *Eur Respir J* 2020; 56:2003498
 26. Kostakis I, Smith GB, Prytherch D, et al. The performance of the National Early Warning Score and National Early Warning Score 2 in hospitalised patients infected by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Resuscitation* 2021; 159:150–157
 27. Ihle-Hansen H, Berge T, Tveita A, et al: COVID-19: Symptoms, course of illness and use of clinical scoring systems for the first 42 patients admitted to a Norwegian local hospital. *Tidsskr Nor Laegeforen* 2020; 140:7
 28. Wu Z, McGoogan JM: Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 2020; 323:1239–1242
 29. Liao X, Wang B, Kang Y: Novel coronavirus infection during the 2019-2020 epidemic: Preparing intensive care units-the experience in Sichuan Province, China. *Intensive Care Med* 2020; 46:357–360
 30. Carr E, Bendayan R, Bean D, et al: Evaluation and improvement of the National Early Warning Score (NEWS2) for COVID-19: A multi-hospital study. *medRxiv* Preprint posted online September 30, 2020. doi:10.1101/2020.04.24.20078006
 31. Subbe CP, Kruger M, Rutherford P, et al: Validation of a modified Early Warning Score in medical admissions. *QJM* 2001; 94:521–526
 32. Burdick H, Lam C, Mataraso S, et al: Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial. *Comput Biol Med* 2020; 124:103949
 33. Beigel JH, Tomashek KM, Dodd LE, et al; ACTT-1 Study Group Members: Remdesivir for the treatment of Covid-19 - final report. *N Engl J Med* 2020; 383:1813–1826
 34. Harrell FE Jr, Lee KL, Mark DB: Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15:361–387
 35. Heagerty PJ, Lumley T, Pepe MS: Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; 56:337–344
 36. Gerry S, Bonnici T, Birks J, et al: Early warning scores for detecting deterioration in adult hospital patients: Systematic review and critical appraisal of methodology. *BMJ* 2020; 369:m1501
 37. Rojas JC, Lyons PG, Jiang T, et al: Accuracy of clinicians' ability to predict the need for intensive care unit readmission. *Ann Am Thorac Soc* 2020; 17:847–853
 38. Arnold J, Davis A, Fischhoff B, et al: Comparing the predictive ability of a commercial artificial intelligence early warning system with physician judgement for clinical deterioration in hospitalised general internal medicine patients: A prospective observational study. *BMJ Open* 2019; 9:e032187
 39. Bilben B, Grandal L, Søvik S: National Early Warning Score (NEWS) as an emergency department predictor of disease severity and 90-day survival in the acutely dyspneic patient – a prospective observational study. *Scand J Trauma Resusc Emerg Med* 2016; 24:80
 40. Malycha J, Farajidavar N, Pimentel MAF, et al: The effect of fractional inspired oxygen concentration on early warning score performance: A database analysis. *Resuscitation* 2019; 139:192–199
 41. 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
 42. Pimentel MAF, Redfern OC, Gerry S, et al: A comparison of the ability of the National Early Warning Score and the National Early Warning Score 2 to identify patients at risk of in-hospital mortality: A multi-centre database study. *Resuscitation* 2019; 134:147–156
 43. Meylan S, Akrou R, Regina J, et al: An Early Warning Score to predict ICU admission in COVID-19 positive patients. *J Infect* 2020; 81:816–846
 44. Liu FY, Sun XL, Zhang Y, et al: Evaluation of the risk prediction tools for patients with coronavirus disease 2019 in Wuhan, China: A single-centered, retrospective, observational study. *Crit Care Med* 2020; 48:e1004–e1011