Article published online: 2022-05-27

Vitals are Vital: Simpler Clinical Data Model Predicts Decompensation in COVID-19 Patients

Joanna Schneider Cavalier[1] Cara L. O'Brien[1] Benjamin A. Goldstein[2] Congwen Zhao[2] Armando Bedoya[1]

[1] Department of Medicine, Duke University, Durham, North Carolina, United States
[2] Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, United States

Address for correspondence Armando Bedoya MD, MMCi, Department of Medicine, Duke University, 2080 Duke University Road, Durham, NC 27708, United States (e-mail: armando.bedoya@duke.edu).

ACI Open 2022;6:e34–e38.

Abstract

Objective Several risk scores have been developed and tested on coronavirus disease 2019 (COVID-19) patients to predict clinical decompensation. We aimed to compare an institutional, automated, custom-built early warning score (EWS) to the National Early Warning Score (NEWS) in COVID-19 patients.

Methods A retrospective cohort analysis was performed on patients with COVID-19 infection who were admitted to an intermediate ward from March to December 2020. A machine learning–based customized EWS algorithm, which incorporates demographics, laboratory values, vital signs, and comorbidities, and the NEWS, which uses vital signs only, were calculated at 12-hour intervals. These patients were retrospectively assessed for decompensation in the subsequent 12 or 24 hours, defined as death or transfer to an intensive care unit.

Results Of 709 patients, 112 (15.8%) had a decompensation event. Using the custom EWS, decompensation within 12 and 24 hours was predicted with areas under the receiver operating curve (AUC) of 0.81 and 0.79, respectively. The NEWS score applied to the same population yielded AUCs of 0.83 and 0.81, respectively. The 24-hour negative predictive values (NPV) of the NEWS and EWS in patients identified as low risk were 99.6 and 99.2%, respectively.

Conclusion The NEWS score performs as well as a customized EWS in COVID-19 patients, demonstrating the significance of vital signs in predicting outcomes. The relatively high positive predictive value and NPV of both scores are indispensable for optimally allocating clinical resources. In this relatively young, healthy population, a more complex score incorporating electronic health record data beyond vital signs does not add clinical benefit.

Keywords
- COVID-19
- 2019 novel coronavirus
- early warning score
- predictive modeling
- decompensation

Introduction

Hospitals have faced significant challenges in caring for patients hospitalized with the novel coronavirus (coronavirus disease 2019 [COVID-19]). Patients are often admitted with moderate symptoms to an intermediate floor, but then decompensate and require an intensive care unit (ICU).

Several risk scores have been developed to predict clinical decompensation, many of which have been tested on COVID-19 patients with mixed results.[1] One of the most common scores is an aggregate-weighted track and trigger system

received May 31, 2021
accepted February 3, 2022

DOI https://doi.org/10.1055/s-0042-1749193.
ISSN 2566-9346.

entitled the National Early Warning Score (NEWS).[2] A systematic review of the use of prognostic models in COVID-19 patients found that many of the scores underestimated risk.[3] One reason for this may be that most "off-the-shelf" early warning scores (EWS) are by nature simplified scores, based on relatively few input variables.[2]

We previously tested the NEWS in our own health care system. While we found it to have moderate performance, it ultimately performed worse than a custom, machine learning–based score that incorporates significantly more inputs.[4,5] While the NEWS has been shown to be useful in COVID-19 patients,[6] we hypothesized that a customized EWS incorporating more electronic health record data (laboratory values, demographics, and comorbidities) would perform better than the NEWS in predicting deterioration in COVID-19 patients. As Villar et al discussed in their recent editorial on limitations of NEWS, "electronic scores including specialty specific calibration and patient trajectory are the next step in the journey to optimise hospital physiological surveillance."[7]

Objectives

In this study, we sought to validate our own score in COVID-19 patients and compare its performance throughout the patients' trajectories to the more commonly used NEWS. We aimed to analyze the benefit of a simple versus more complex machine learning–based prediction model in COVID-19 patients.

Methods

Study Design

We performed a retrospective analysis of patients admitted to intermediate care wards at a large academic medical center with a positive laboratory-confirmed COVID-19 test between March 25 and December 20, 2020. Patients admitted directly to the ICU were excluded.

Our primary outcome of interest—termed decompensation—was mortality while on an intermediate ward or transfer to an ICU. We compared two EWSs: the NEWS and our own custom-built, machine learning–based score. The NEWS is a widely used points-based EWS score consisting of seven variables, including vital signs and level of consciousness.[2] We have previously described the development of our own EWS and the methodology behind a three-tiered risk model.[4] In brief, the score consists of 50 predictor variables including demographics, vitals, comorbidities, and laboratory tests. The score was calibrated on non-COVID-19 patients. The model was estimated via least absolute shrinkage and selection operator (LASSO) regression and implemented directly into our Epic-based electronic health record as an acuity score, although it is not implemented in our COVID-19 wards.

Data Analysis

We evaluated the performance of the NEWS and our EWS based on their risk assessment as of 8 am and 8 pm of each day a patient was on an intermediate care floor. We performed a time-varying analysis to assess the performance of each score to accurately predict risk over the next 12 and 24 hours. Scores were calculated throughout each patient's trajectory until the patient was discharged or had the primary outcome of interest. Thus, each patient would have multiple overlapping scores predicting their chance of decompensation in the following 12 or 24 hours after each score is calculated.

We assessed each score's discrimination and calibration via the area under the receiver operating characteristic (AUROC) and calibration slope, respectively.[8] We used the bootstrap to calculate 95% confidence intervals (CIs) for the AUROC, accounting for the repeated risk assessments of each patient. We also calculated the sensitivity, positive predictive value (PPV), and negative predictive value (NPV) of each score's risk categories. All analyses were conducted in R 3.6. This work was exempt per our institutional review board (Pro00065513).

Results

Between March and December 2020, there were 709 COVID-19–positive patients admitted to our hospital who met inclusion criteria. These patients contributed to 770 unique encounters. Of these patients, 112 (15.8%) had a deterioration event, of which 47 (42%) died, 38 (33.9%) transferred to the ICU, and 27 (24.1%) both transferred to an ICU and died. Those patients who had a deterioration event were, on average, older and likely to have a history of malignancy, renal disease, congestive heart failure, history of myocardial infarction, and complicated diabetes. Those who decompensated had lower body mass indices ($p < 0.02$, standardized mean difference $= 0.30$) (►Table 1).

Overall, both the NEWS and our EWS had similar performance. The custom EWS yielded AUROCs of 0.81 (95% CI: 0.795–0.855) and 0.79 (95% CI: 0.776–0.842) when evaluated at 12- and 24-hour intervals, respectively. The NEWS yielded corresponding AUROCs of 0.83 (95% CI: 0.814–0.880) and 0.81 (95% CI: 0.791–0.854). The calibration slopes for the EWS and NEWS were 0.88 (95% CI: 0.746–1.320) and 0.66 (95% CI: 0.459–1.114) at 12 hours and 0.77 (95% CI: 0.669–1.140) and 0.73 (95% CI: 0.492–1.175) at 24 hours (►Table 2).

When risk-stratified into low-, medium-, and high-risk, the custom EWS yielded a PPV of 12.4% for the highest risk group for decompensation within 24 hours and an NPV of 99.2% for the lowest risk group. The NEWS score was used to risk stratify patients into low (<7) and high (≥ 7) risk, and gave corresponding PPVs and NPVs of 9.4 and 99% for decompensation within 24 hours (►Table 2).

Discussion

In this study, we show that a custom-built, machine learning–based EWS, which incorporates demographics, laboratory values, vital signs, and comorbidities, was equally as effective as the simpler, vitals-based NEWS in predicting deterioration in COVID-19 patients. Both scores perform well

Table 1 Baseline characteristics of COVID-19 patients with and without a decompensation event

Baseline characteristic	Patients without decompensation event ($N = 597$)	Patients with decompensation event ($N = 112$)	Total ($N = 709$)	SMD[a]	*p*-Value
Age					
Mean (SD)	57.6 (17.5)	67.5 (15.3)	59.2 (17.5)	0.60	<0.01[b]
Missing	1 (0.17%)	–	1 (0.14%)		
Sex					
Male	321 (53.8%)	71 (63.4%)	392 (55.3%)	0.20	0.16[c]
Female	275 (46.1%)	41 (36.6%)	316 (44.6%)		
Missing	1 (0.2%)	0 (0.0%)	1 (0.1%)		
Race/ethnicity					
Hispanic	112 (18.8%)	15 (13.4%)	127 (17.9%)	0.20	0.37[d]
Non-Hispanic black	217 (36.4%)	44 (39.3%)	261 (36.9%)		
Non-Hispanic white	231 (38.8%)	49 (43.8%)	280 (39.5%)		
Other/unknown	36 (6.0%)	4 (3.6%)	40 (5.6%)		
Smoking status					
Yes	34 (5.7%)	3 (2.7%)	37 (5.2%)	0.16	0.42[d]
No	396 (66.3%)	75 (67.0%)	471 (66.4%)		
Unknown	167 (28.0%)	34 (30.4%)	201 (28.3%)		
BMI					
Mean (SD)	31.6 (8.8)	29.1 (7.6)	31.2 (8.7)	0.30	0.02[b]
Missing	174 (29.2%)	34 (30.4%)	208 (29.3%)		
Comorbidities					
Any malignancy	91 (15.2%)	28 (25.0%)	119 (16.8%)	0.24	0.02[c]
Renal disease	166 (27.8%)	48 (42.9%)	214 (30.2%)	0.32	<0.01[c]
Chronic pulmonary disease	159 (26.6%)	32 (28.6%)	191 (26.9%)	0.04	0.76[c]
Congestive heart failure	107 (17.9%)	36 (32.1%)	143 (20.2%)	0.33	<0.01[c]
Myocardial infarction	49 (8.2%)	17 (15.2%)	66 (9.3%)	0.22	0.03[c]
Diabetes, uncomplicated	190 (31.8%)	43 (38.4%)	233 (32.9%)	0.14	0.21[c]
Diabetes, complicated	196 (32.8%)	52 (46.4%)	248 (35.0%)	0.28	0.01[c]
AIDS/HIV	6 (1.0%)	0 (0.0%)	6 (0.8%)	0.14	0.60[d]

Abbreviations: AIDS, acquired immunodeficiency syndrome; BMI, body mass index; HIV, human immunodeficiency virus; SD, standard deviation; SMD, standardized mean difference.
[a]Typically, an SMD greater than 0.10 indicates imbalance between the two groups.
[b]Kruskal–Wallis test.
[c]Chi-squared test.
[d]Fisher's exact test.

and give clinicians accurate, automated information upon which to act in real time throughout the hospitalization. However, both scores were undercalibrated, as they were developed in non-COVID-19 patients who do not deteriorate as frequently as those with COVID-19.

We hypothesized that our custom, more complex EWS would outperform the NEWS by incorporating more clinical data (demographics, comorbidities, and laboratory values). However, the scores performed equally well, demonstrating the importance of vital signs in predicting the clinical course of COVID-19 patients. We postulate that the scores' equal performance is due to the COVID-19 patient population being relatively homogenous and healthy as compared with our institution's typical patient population.

Both scores were undercalibrated, meaning they under-predicted risk. This highlights the challenge of predicting decompensation in a population that decompensates frequently, as compared with non-COVID-19 patients. Our earlier study using the custom EWS in our non-COVID-19 population had a much lower event rate of 2.5% ICU transfers and 0.9% inpatient death, as compared with 6.6 and 9.2% in our COVID-19 population.[4] A second challenge unique to COVID-19 patients is the varying clinical courses despite similar presentations. For example, one study showed that,

Table 2 Performance metrics of the custom early warning score versus the national early warning score

Score	EWS			NEWS	
Decompensation in 12 h					
AUROC (95% CI)	0.81 (0.80–0.86)			0.83 (0.81–0.88)	
Calibration slope (95% CI)	0.88 (0.75–1.32)			0.66 (0.46–1.11)	
Decompensation in 24 h					
AUROC (95% CI)	0.79 (0.78–0.84)			0.81 (0.79–0.85)	
Calibration slope (95% CI)	0.77 (0.67–1.14)			0.73 (0.49–1.18)	
Performance by risk stratification	**Low risk**	**Medium risk**	**High risk**	**Low risk**	**High risk**
Decompensation in 12 h					
% of patients *with* decompensation event	38.0	39.2	22.8	53.2	46.8
% of patients *without* decompensation event	84.5	13.4	2.1	93.5	6.5
PPV in 12 h (%)	–	2.6	9.2	–	6.2
NPV in 12 h (%)	99.6	97.4	–	99.5	–
Decompensation in 24 h					
% of patients *with* decompensation event	43.8	29.9	26.3	59.1	40.9
% of patients *without* decompensation event	84.7	12.3	3.0	93.6	6.4
PPV in 24 h (%)	–	3.8	12.4	–	9.4
NPV in 24 h (%)	99.2	96.2	–	99.0	–

Abbreviations: AUROC, area under the receiver operating characteristic; CI, confidence interval; EWS, early warning score; NEWS, National Early Warning Score; NPV, negative predictive value; PPV, positive predictive value.

of patients who require high-flow nasal oxygen, approximately half will require escalation to intubation or noninvasive positive pressure ventilation, whereas half will improve on only high flow.[9] Thus, it is exceedingly difficult to predict how patients will progress based on initial presentation. Both the custom EWS and NEWS do not incorporate the degree of supplemental oxygen needed, and instead include supplemental oxygen as a binary variable. Prior studies in COVID-19 patients have shown the degree of hypoxia to be an independent predictor of mortality.[10,11] In agreement with Villar et al, we theorize that one of the barriers to achieving better predictive performance in COVID-19 patients is the severe hypoxemia seen with this illness.[7]

One limitation in our study is the possible bias introduced by clinical use of the custom EWS. While it was not being formally used as a rounding tool on our COVID-19 units during the study, individual providers may have accessed and thus acted upon the EWS in caring for those patients. There was limited knowledge of or adoption of the score among the COVID-19 units, so the likelihood of a true clinically significant impact on outcomes is low.

Conclusion

We show the NEWS and a custom-built EWS are equally effective in predicting patient decompensation in COVID-19 patients, defined as mortality or transfer to an ICU. The equal performance of both scores highlights the importance of vital signs in predicting decompensation in COVID-19 patients. Both scores yield a relatively high PPV in high-risk patients and an extremely high NPV in low-risk patients, giving clinicians actionable and real-time information throughout the hospitalization. While still clinically useful, the under-calibration of both scores in this patient population points to the relative frequency of decompensation in COVID-19 patients. Implementation of the widely available NEWS can help identify patients before they decompensate and allocate limited-supply resources, without need for a new, more complex, customized EWS at each institution.

Clinical Relevance Statement

EWSs are tools that clinicians may use to predict clinical decompensation of patients. During the COVID-19 pandemic, resources such as ICU beds, stepdown beds, and ventilators have been in short supply. Therefore, it is imperative that clinicians be able to understand the most likely trajectories of their patients with COVID-19 to best allocate clinical resources. While there is a tendency to try to use informatics to make scores more and more complex, clinicians should be aware that vital signs are the greatest driver of clinical outcomes. Vital signs must be monitored very closely in this patient population.

Protection of Human and Animal Subjects

Given the retrospective nature of this work using deidentified clinical data, this work was exempt from human subjects' protections.

Conflict of Interest

None declared.

References

1 Hu H, Yao N, Qiu Y. Predictive value of 5 early warning scores for critical COVID-19 patients. Disaster Med Public Health Prep 2020. Doi: 10.1017/dmp.2020.324

2 Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS–Towards a national early warning score for detecting adult inpatient deterioration. Resuscitation 2010;81(08):932–937

3 Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020;369:m1328

4 Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA. Minimal impact of implemented early warning score and best practice alert for patient deterioration. Crit Care Med 2019;47(01):49–55

5 O'Brien C, Goldstein BA, Shen Y, et al. Development, implementation, and evaluation of an in-hospital optimized early warning score for patient deterioration. MDM Policy Pract 2020;5(01):2381468319899663

6 Subbe C, Thorpe C. Finally good NEWS: something simple is working in COVID-19!. Resuscitation 2021;159:168–169

7 Villar SS, Machiwenyika JMT, Zhu Y, Mackay JH. The performance of the National Early Warning Score in hospitalised patients infected by Covid-19. Resuscitation 2021;162:441–442

8 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21(01):128–138

9 Xia J, Zhang Y, Ni L, et al. High-flow nasal oxygen in coronavirus disease 2019 patients with acute hypoxemic respiratory failure: a multicenter, retrospective cohort study. Crit Care Med 2020;48(11):e1079–e1086

10 Choi KJ, Hong HL, Kim EJ. The association between mortality and the oxygen saturation and fraction of inhaled oxygen in patients requiring oxygen therapy due to COVID-19-associated pneumonia. Tuberc Respir Dis (Seoul) 2021;84(02):125–133

11 Dillon K, Hook C, Coupland Z, Avery P, Taylor H, Lockyer A. Pre-hospital lowest recorded oxygen saturation independently predicts death in patients with COVID-19. Br Paramed J 2020;5(03):59–65