



Evaluating a Widely Implemented Proprietary Deterioration Index Model among Hospitalized Patients with COVID-19

Karandeep Singh^{1,2}, Thomas S. Valley^{2,3}, Shengpu Tang⁴, Benjamin Y. Li⁴, Fahad Kamran⁴, Michael W. Sjoding^{2,3}, Jenna Wiens^{2,4}, Erkin Otlles⁵, John P. Donnelly^{1,2}, Melissa Y. Wei^{2,3}, Jonathon P. McBride⁶, Jie Cao⁷, Carleen Penozza⁸, John Z. Ayanian^{2,3}, and Brahmajee K. Nallamothu^{2,3}

¹Department of Learning Health Sciences, ³Department of Internal Medicine, ⁶Department of Cellular and Molecular Biology, and ⁷Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan; ²Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, Michigan; ⁴Division of Computer Science and Engineering, and ⁵Department of Industrial and Operations Engineering, University of Michigan College of Engineering, Ann Arbor, Michigan; and ⁸Nursing Informatics, Michigan Medicine, Ann Arbor, Michigan.

ORCID IDs: 0000-0001-8980-2330 (K.S.); 0000-0002-5766-4970 (T.S.V.).

Abstract

Rationale: The Epic Deterioration Index (EDI) is a proprietary prediction model implemented in over 100 U.S. hospitals that was widely used to support medical decision-making during the coronavirus disease (COVID-19) pandemic. The EDI has not been independently evaluated, and other proprietary models have been shown to be biased against vulnerable populations.

Objectives: To independently evaluate the EDI in hospitalized patients with COVID-19 overall and in disproportionately affected subgroups.

Methods: We studied adult patients admitted with COVID-19 to units other than the intensive care unit at a large academic medical center from March 9 through May 20, 2020. We used the EDI, calculated at 15-minute intervals, to predict a composite outcome of intensive care unit–level care, mechanical ventilation, or in-hospital death. In a subset of patients hospitalized for at least 48 hours, we also evaluated the ability of the EDI to identify patients at low risk of experiencing this composite outcome during their remaining hospitalization.

Results: Among 392 COVID-19 hospitalizations meeting inclusion criteria, 103 (26%) met the composite outcome. The

median age of the cohort was 64 (interquartile range, 53–75) with 168 (43%) Black patients and 169 (43%) women. The area under the receiver-operating characteristic curve of the EDI was 0.79 (95% confidence interval, 0.74–0.84). EDI predictions did not differ by race or sex. When exploring clinically relevant thresholds of the EDI, we found patients who met or exceeded an EDI of 68.8 made up 14% of the study cohort and had a 74% probability of experiencing the composite outcome during their hospitalization with a sensitivity of 39% and a median lead time of 24 hours from when this threshold was first exceeded. Among the 286 patients hospitalized for at least 48 hours who had not experienced the composite outcome, 14 (13%) never exceeded an EDI of 37.9, with a negative predictive value of 90% and a sensitivity above this threshold of 91%.

Conclusions: We found the EDI identifies small subsets of high-risk and low-risk patients with COVID-19 with good discrimination, although its clinical use as an early warning system is limited by low sensitivity. These findings highlight the importance of independent evaluation of proprietary models before widespread operational use among patients with COVID-19.

Keywords: coronavirus disease; deterioration index; prediction model; validation study

(Received in original form June 19, 2020; accepted in final form December 23, 2020)

This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). For commercial usage and reprints, please contact Diane Gern (dgern@thoracic.org).

Supported by National Heart, Lung, and Blood Institute grants K23 140165 (T.S.V.), K01 HL136687 (M.W.S.), and K12-HL138039 (J.P.D.); National Institute of Aging grant K23 AG056638 (M.Y.W.); National Library of Medicine and the Michigan Institute for Data Science grant R01 LM013325-01 (M.W.S., J.W., and B.K.N.); National Institute of General Medical Sciences grant T32GM007863 (B.Y.L.); and U.S. National Institutes of Health grant T32GM007863 (E.O.). J.W. served on the board of the nonprofit organization Machine Learning for Healthcare.

Author Contributions: K.S., T.S.V., M.W.S., J.W., J.P.D., M.Y.W., J.P.M., J.C., C.P., J.Z.A., and B.K.N. made substantial contributions to the conception or design of the work. K.S., S.T., B.Y.L., F.K., M.W.S., J.W., and E.O. participated in the data acquisition and analysis. All authors participated in the interpretation of data. K.S. drafted the manuscript and all authors participated in critical revisions for intellectual content. All authors gave approval for the version being considered for publication and agreed to be accountable for the analysis, findings, and interpretation.

Correspondence and requests for reprints should be addressed to Karandeep Singh, M.D., M. M.Sc., Department of Learning Health Sciences, University of Michigan Medical School, 1161H NIB, 300 N. Ingalls Street, Ann Arbor, MI 48109. E-mail: kdpsingh@umich.edu.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Ann Am Thorac Soc Vol 18, No 7, pp 1129–1137, Jul 2021

Copyright © 2021 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.202006-698OC

Internet address: www.atsjournals.org

The coronavirus disease (COVID-19) pandemic is straining the capacity of hospitals and healthcare systems across the United States (1, 2). Accurately identifying subgroups of patients with COVID-19 at high risk and low risk for adverse outcomes could help to alleviate this strain by better directing scarce resources to those patients at greatest need. This need has led to the development and use of clinical prediction models in patients with COVID-19. Many of these models suffer from a high risk of bias due to sample sizes too small to allow for both model development *and* validation (3). Although the majority of studies have focused on newly developed models, many models already exist to detect clinical deterioration among hospitalized patients.

One of the most widely used models is the Epic Deterioration Index (EDI), which is implemented in hundreds of U.S. hospitals (4). The EDI was developed using data from three healthcare organizations between 2012 and 2016, and it uses clinical data to calculate risk scores at regular 15-minute intervals throughout a patient's stay starting from the time of hospital admission.

Although not specific to patients with COVID-19, the EDI has been widely used during the pandemic to support decision-making in patients with COVID-19 (5–7).

The widespread use of the EDI raises implementation concerns because there are no peer-reviewed publications describing its validity in any patient population. These concerns are particularly salient given that health systems are using the EDI in conflicting ways and with substantially different thresholds (7). Even before the onset of COVID-19, publicly available information about the EDI was limited to anecdotal reports of its value in critically ill patients (8, 9). The proprietary nature of models such as the EDI makes independent validation difficult because of a lack of complete information on the model's functional form and parameters (10). However, independent evaluation is needed because hospital-based models often do not perform well in external validation studies and because the performance of models erodes over time as use patterns change (11). In addition, some widely adopted proprietary models have previously been shown to be biased against Black patients even when race was not included as a predictor (12, 13). Given that COVID-19 disproportionately impacts Black individuals with respect to its incidence and

complications (14), the validity of the EDI needs to be established generally and for vulnerable subpopulations.

These concerns have not prevented its use from being advocated (5). An Epic Systems spokesperson recently stated that “some hospitals are now using the model with confidence,” (6) whereas others suggest it is “helping save lives.” (4) In this context, we sought to independently validate the ability of the EDI to predict adverse outcomes among diverse patients hospitalized with COVID-19 at a large academic medical center. We also stratified our evaluation by race, sex, and age to evaluate the model performance among key subgroups of patients. Our findings have potential implications for how the EDI—currently deployed in hundreds of U.S. hospitals (4)—may be validated and used by healthcare systems during the COVID-19 pandemic, and more broadly in how proprietary models should be evaluated.

Methods

Study Cohort

Our study cohort included adults 18 years and older diagnosed with COVID-19 who were admitted to Michigan Medicine (i.e., the academic health system of the University of Michigan in Ann Arbor) between March 9, 2020, and May 20, 2020, from the emergency department, outpatient clinics, and outside hospital transfers. We excluded encounters where patients were admitted directly to an intensive care unit (ICU) ($n = 215$) or were discharged to home or a separate facility on hospice ($n = 34$) or where EDI scores were not available ($n = 10$). Patients who transitioned to comfort care or end-of-life care in the hospital were not excluded. We also excluded patients who remained hospitalized but had not yet experienced the composite outcome described below ($n = 27$), because it was not possible to determine with certainty whether they would reach the primary outcome during their hospitalization. The study was approved by the Institutional Review Board of the University of Michigan Medical School, and the need for consent was waived.

The Epic Deterioration Index Model

The EDI is generated from a proprietary early-warning prediction model developed

by Epic Systems Corporation using data that are routinely recorded within its electronic health record. Epic is one of the largest healthcare software vendors in the world, and its electronic health record is used by most U.S. News and World Report's top-ranked healthcare systems and reportedly includes medical records for nearly 180 million Americans (or 56% of the U.S. population) (15).

The EDI aims to detect patients who deteriorate and require higher levels of care. Its score ranges from 0 to 100, in which the higher numbers denote a greater risk of experiencing a composite adverse outcome of requiring rapid response, resuscitation, ICU-level care, or dying in the next 12–38 hours. Details related to the specific cohorts, within which the model was developed, the model parameters, and its detailed performance characteristics have not been shared publicly or described in the published literature.

All hospitalized patients at Michigan Medicine have had calculations of the EDI as part of an ongoing evaluation of its clinical utility since late 2018; however, the EDI was not used in any clinical protocols during this time period and thus clinicians were blinded to the score. Calculations of the EDI begin immediately after hospital admission and then continue at regular 15-minute intervals until discharge. Although the algorithm was developed before the COVID-19 pandemic, it includes several predictors that may be clinically relevant in patients with COVID-19, including age, vital sign measurements (systolic blood pressure, temperature, pulse, respiratory rate, oxygen saturation), nursing assessments (Glasgow Coma Scale, neurological assessment, cardiac rhythm, oxygen requirement), and laboratory values (hematocrit, white blood cell count, potassium, sodium, blood pH, platelet count, blood urea nitrogen).

Definition of the Primary Outcome

We defined our primary outcome as a composite of adverse outcomes that included the first of any of the following events that occurred during the hospitalization: ICU-level care, mechanical ventilation, or in-hospital death. We chose to include these adverse events for the composite outcome because they are highly relevant in the clinical care of patients with COVID-19, in which rapid respiratory decline is frequently described.

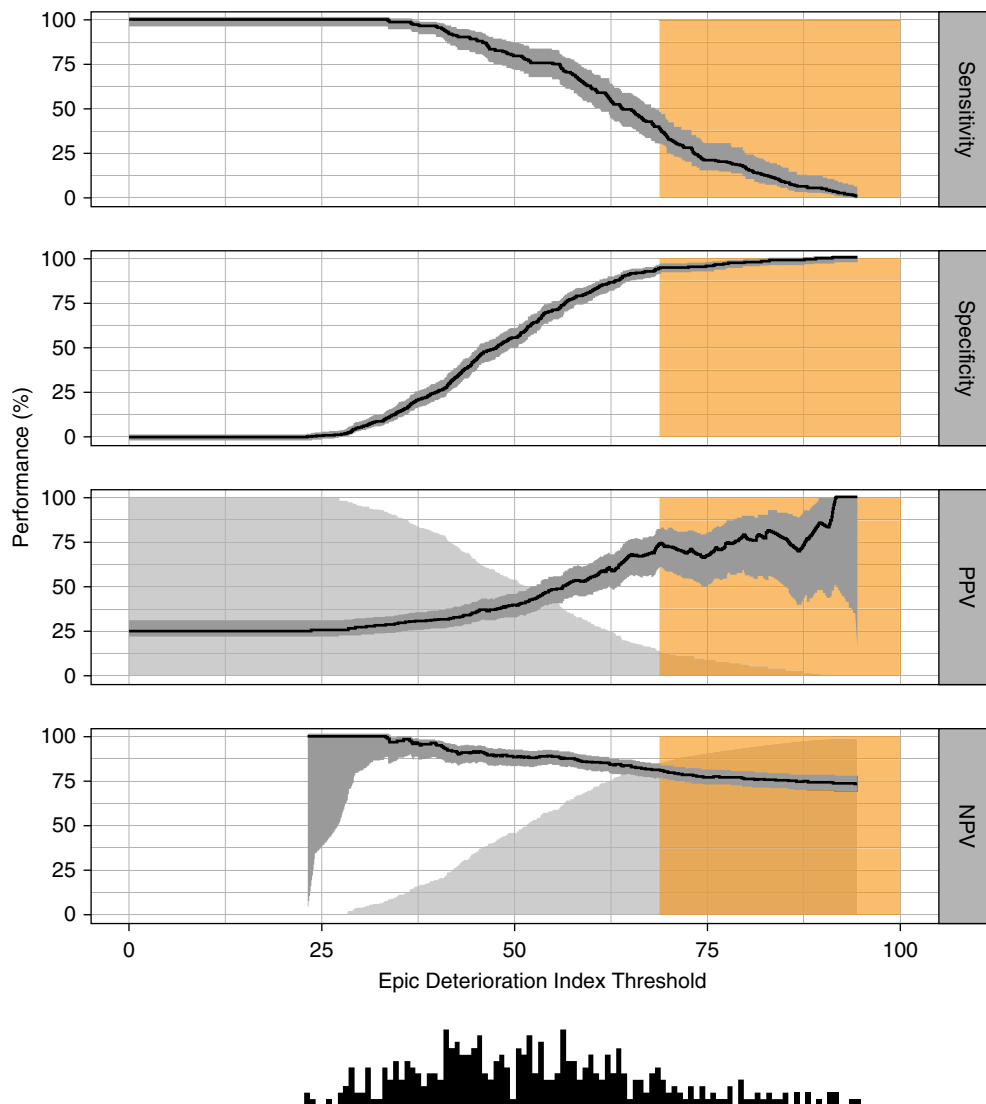


Figure 1. High-risk analysis plot showing the relationship between the Epic Deterioration Index (EDI) threshold and the sensitivity, specificity, PPV, and NPV, with a histogram demonstrating the distribution of maximum EDI per patient. The high-risk area (≥ 68.8) is shaded in orange. NPV = negative predictive value; PPV = positive predictive value.

Evaluation of the EDI to Identify High-Risk Patients

We used scores from the EDI calculated every 15 minutes throughout the hospitalization to predict the composite adverse outcome during the hospitalization. For patients who experienced the outcome, we only used EDI scores calculated before the outcome. We evaluated the discriminative performance of the EDI using the area under the receiver-operating characteristic curve (AUC). The AUC represents the probability of correctly ranking two randomly chosen individuals (one who experienced the event and one who did not). Because the model runs every

15 minutes on all hospitalized patients, we calculated the AUC on the basis of the entire trajectory of predictions.

The AUC was calculated at the hospitalization level using the strategy defined by Henry and colleagues and Oh and colleagues (16, 17). The rationale for evaluating the AUC at the hospitalization level is that if a hypothetical alert were to be linked to a score threshold, whether the alert ever fired for any given patient would depend on whether this threshold was ever exceeded during the hospitalization. The EDI is recalculated every 15 minutes, and if a patient crossed a given alerting threshold even once, this would bring the patient

to the clinician's attention if linked to an alert.

Predictions of deterioration are most beneficial when an appropriate lead time is available for action by clinicians. We therefore calculated a median lead time for the primary outcome by comparing when patients were first deemed high risk during their hospitalization (based on the "high-risk" threshold selected by intensivists) to when they experienced the outcome. In all cases, we calculated empirical 95% confidence intervals (95% CIs) for the AUC using 1,000 bootstrap replicates of our study cohort. Although it is unknown whether the EDI can be interpreted as a

Table 1. Patient characteristics overall and stratified by adverse outcomes

Variable	Overall (n = 392)	Adverse Outcome [n = 103 (26%)]	No Adverse Outcome [n = 289 (74%)]	P Value
Age, median (IQR), yr	64 (53–75)	69 (59–80)	62 (50–72)	<0.001
Female, n (%)	169 (43)	37 (36)	132 (46)	0.11
Race, n (%)				0.041
Black	168 (43)	45 (44)	123 (43)	
White	162 (36)	50 (49)	112 (39)	
Other	42 (11)	4 (3.9)	38 (13)	
Unknown	20 (5.1)	4 (3.9)	16 (5.5)	
Comorbidities, n (%)*				
Cardiac arrhythmias	181 (48)	63 (62)	118 (42)	<0.001
Chronic kidney disease	143 (38)	60 (59)	83 (30)	<0.001
Chronic pulmonary disease	125 (33)	40 (40)	85 (31)	0.13
Congestive heart failure	78 (21)	35 (35)	43 (15)	<0.001
Depression	127 (34)	43 (43)	84 (30)	0.033
Diabetes	158 (42)	57 (56)	101 (36)	<0.001
Hypertension	286 (75)	87 (86)	199 (72)	0.005
Liver disease	49 (13)	14 (15)	35 (13)	0.88
Metastatic cancer	53 (14)	22 (22)	31 (11)	0.013
Obesity	159 (42)	43 (43)	116 (42)	0.98
Rheumatoid arthritis or collagen vascular disease	30 (7.9)	14 (14)	16 (5.8)	0.018
Solid tumor without metastases	68 (18)	23 (23)	45 (16)	0.18

Definition of abbreviations: COVID-19 = coronavirus disease 2019; IQR = interquartile range; ICD-10 = *International Classification of Diseases, Tenth Revision*.

Continuous variables were compared using a Wilcoxon test. Proportions were compared using a Chi-squared test.

*Comorbidities were calculated using ICD-10 codes from all prior encounters, including the COVID-19 encounter when available. Comorbidities were unavailable for 13 (3.3%) patients, 2 of whom experienced adverse outcomes and 11 of whom did not, and were excluded from the denominator.

probability, model calibration was assessed using a calibration curve by comparing deciles of all predicted EDI to the observed risk.

Sensitivity Analysis

Recalculation of the AUC at the prediction level with varying prediction horizons. To enhance the comparability of our evaluation to related work in this domain, we recalculated the AUC using a 4-hour, 8-hour, 12-hour, and 24-hour prediction time horizon for the outcome. In this analysis, we considered patients to only have met the outcome from the time of a prediction if the outcome occurred within the prediction horizon, and we calculated the AUC at the prediction level.

Disparate Impact Analyses

To evaluate how the EDI performs in vulnerable populations, we conducted two analyses. First, we compared the mean EDI in demographic subgroups and in those with and without the following comorbidities: cardiac arrhythmias, chronic kidney disease, chronic pulmonary disease, congestive heart failure, depression, diabetes mellitus,

hypertension, liver disease, metastatic cancer, obesity, rheumatoid arthritis or other collagen vascular diseases, and solid tumors without metastases. This analysis was conducted to identify which comorbidities result in a higher EDI score, recognizing that comorbidities are not directly included in the EDI model. Then, we compared the AUC for defined by age (≥ 65 yr vs. < 65 yr), sex, and race to determine if the EDI performs equally well in these subgroups.

Evaluation of the EDI to Identify Low-Risk Patients 48 Hours after Admission

Another potential use of the EDI is to identify patients at low risk who could be sent home or to a lower-acuity setting, thereby offloading hospitals. Our goal was to evaluate how well the EDI at the end of 48 hours could identify patients at a low risk of experiencing the outcome during the remainder of their hospitalization. We selected 48 hours following admission because decisions to triage patients to lower-acuity care within this timeframe may be valuable to hospital systems struggling in response to a surge

of inpatient cases. For this analysis, we excluded patients who were discharged or experienced the composite outcome within the first 48 hours as triage decisions were not relevant for this group ($n = 65$). We did this to remove very low-risk patients who were discharged as well as very high-risk patients who experienced the primary outcome early. The AUC was calculated based on the maximum EDI in the first 48 hours, with 95% CI based on 1,000 bootstrap replicates.

Selection of Clinically Actionable Thresholds

We calculated sensitivities, specificities, positive predictive values, and negative predictive values across the entire spectrum of EDI thresholds. We identified two clinically actionable thresholds based on the threshold-performance plots in consultation with intensivists on our research team (T.S.V. and M.W.S.): one for identifying high-risk patients who will likely need ICU-level care (based on the EDI throughout the hospitalization) and one for identifying low-risk patients who may be appropriate for lower-acuity care (based on the 48-h analysis).

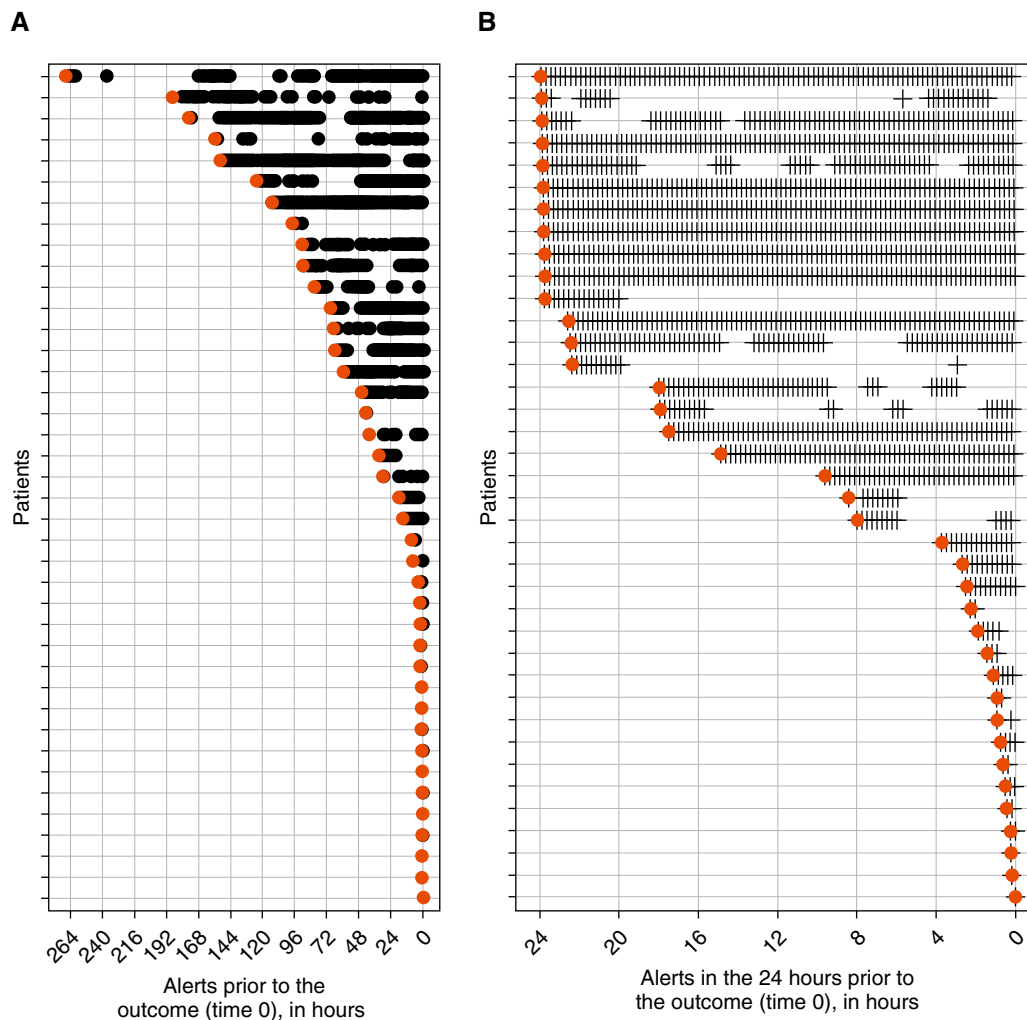


Figure 2. Distribution of alert times based on exceeding the high-risk threshold (Epic Deterioration Index ≥ 68.8), demonstrating (A) all alerts and (B) alerts in the 24 hours before the outcome, with the first alert highlighted in red. Each point represents a hypothetical alert; no actual alerts were generated.

Software

We used R 3.6.0 for all analyses, as well as the pROC and runway packages (18, 19). We have made our statistical code available at https://github.com/ml4lhs/edi_validation.

Results

We identified 392 hospitalizations for 369 patients with COVID-19 who met inclusion criteria for our study cohort. Two patients had three hospitalizations and 19 patients had two hospitalizations. At the hospitalization level ($n = 392$), the median age of patients was 64 years (interquartile range [IQR], 53–75), 169 (43%) were women, and 168 (43%) were Black (Table 1).

The composite adverse outcome occurred in 103 (26%) out of the 392 hospitalizations with a median length of follow-up of 5.3 days (IQR, 2.9–10.7; max 49). The outcome occurred at a median of 2.5 days (IQR, 0.65–5.0; max 24) after admission. Of all hospitalizations, 88 (22%) resulted in ICU-level care, 44 (11%) in mechanical ventilation, and 35 (8.9%) in death in the hospital. Those who experienced an adverse outcome were older, more likely to be white, and more likely to have a history of cardiac arrhythmias, chronic kidney disease, congestive heart failure, depression, diabetes, hypertension, metastatic cancer, and rheumatoid arthritis or other collagen vascular diseases (all $P < 0.05$) (see Table 1).

Overall, the EDI score had an AUC of 0.79 (95% CI, 0.74–0.84) as a continuous

predictor of risk. The performance characteristics of the EDI score are reported in Figure 1. Patients who met or exceeded an EDI of 68.8 had a 74% probability of experiencing the primary outcome during their hospitalization (i.e., positive predictive value) with a sensitivity of 39%, and they made up 14% of the study cohort. At this threshold, one deteriorating patient would be identified for every 1.4 patients in whom an alert was generated, a quantity also known as the number needed to evaluate (20). The median lead time from when the threshold was first exceeded to when the outcome occurred was 24 hours (IQR, 1.4–83). The entire distribution of lead times is further described in Figure 2. Figure 3 evaluates model calibration and demonstrates that the EDI systematically

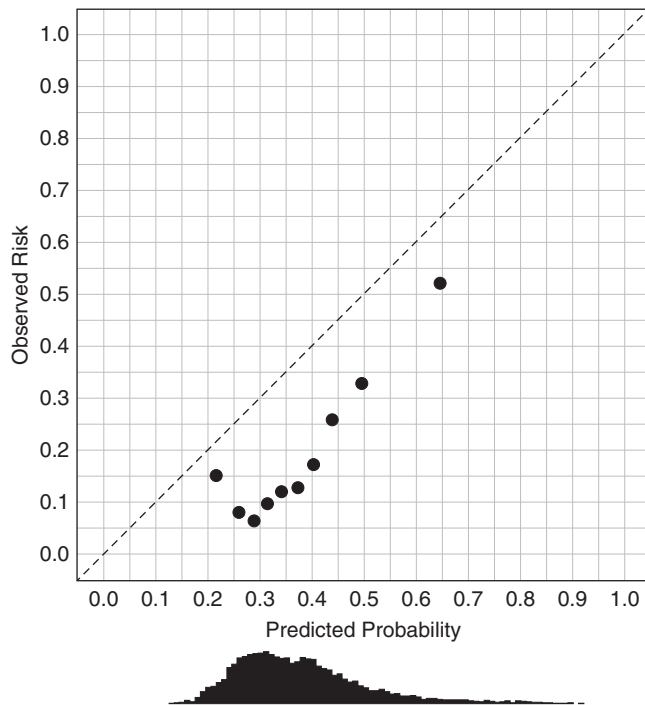


Figure 3. Calibration curve comparing deciles of all predicted Epic Deterioration Index (EDI), rescaled to 0 to 1, with the observed risk, with a line demonstrating ideal calibration (solid) and a histogram of predicted EDI.

overpredicts the risk of experiencing the primary outcome if interpreted as a probability. We found the prediction-level AUCs to be similar to those in our primary analysis although the positive predictive values were lower due to the shorter prediction horizon (Appendix Table 3 in the online supplement, Appendix Figures 1 and 2).

In the disparate impact analyses, we found that EDI predictions did not differ by sex or race (Appendix Table 1). We did find higher EDI scores for patients 65 and older and those with cardiac arrhythmias, chronic kidney disease, chronic pulmonary disease, congestive heart failure, diabetes, and hypertension. In an analysis of model performance by subgroup, the EDI performed similarly across the demographic subgroups although the model appeared to perform better in patients with liver disease as compared with those without it ($P=0.028$) (Appendix Table 2).

In the subset of 286 patients who had not been discharged or experienced the primary outcome at 48 hours, 55 (19%) experienced the composite outcome at some point during the remainder of their hospitalization. In this setting, the EDI had a hospitalization-level AUC of 0.65 (95% CI,

0.57–0.73). The performance characteristics of the 48-hour maximum EDI in this subset of patients are reported in Figure 4. A total of 14 (13%) patients who never exceeded an EDI of 37.9 in the first 48 hours of their hospitalization had a 90% probability of not experiencing the outcome (i.e., negative predictive value) for the remainder of the hospitalization (median remaining follow-up of 3.8 d [IQR, 1.7–8.9]) with a sensitivity of 91% above this threshold.

Figure 5 demonstrates four examples of EDI patterns in patients with COVID-19 overlaid with the identified high- and low-risk thresholds (≥ 68.8 for high risk and < 37.9 for low risk, respectively). As shown in the bottom-left and top-right panels of this figure, the EDI fluctuates substantially for individual patients with each assessment over the regular 15-minute intervals.

Discussion

Our study constitutes the first publicly reported independent validation of the EDI in any patient population. Our results suggest that the EDI exhibits good discrimination for the prediction of adverse outcomes in a diverse COVID-19

population. It demonstrated good performance in identifying higher-risk patients in our cohort, identifying a small proportion of patients with a positive predictive value of 74% but a relatively low sensitivity of 39%. There was no single threshold at which the EDI exhibited both a high positive predictive value and high sensitivity. Thus, from a standpoint of clinical actionability, the model's utility is somewhat limited. We found that the maximum EDI in the first 48 hours of the hospitalization may help identify a small subset of low-risk patients who may be safely transferred to lower-acuity settings (such as a field hospital), thereby conserving resources. However, 10% of the patients identified by the EDI as low risk may ultimately deteriorate, so the decision to deescalate care for patients should not be based solely on the EDI. For the vast majority of patients whose maximum EDI score falls in the intermediate-risk range, the score has limited value to guide clinical decision-making. We also noted the EDI fluctuates substantially when calculated at 15-minute intervals, in part because it only relies on the most recent value for each of the predictors. Even small changes in predictors lead to large differences in the EDI because prior normal values are ignored when a new value is obtained. Thus, we recommend that the interpretation of individual EDI scores be based on whether a patient ever exceeds specific thresholds. The substantial variation in EDI scores also underscores the notion of diminishing returns when running the model so frequently.

The proprietary nature of the EDI raises specific ethical and clinical concerns in the setting of a pandemic, in which resources may be scarce and could be allocated to higher-risk patients based on the output of this prediction model. We found no evidence that the EDI is biased against specific subgroups of vulnerable patients although the EDI was not always concordant with the observed differences in adverse outcomes (in Table 1 as compared with Appendix Table 2). Although chronic pulmonary disease was not associated with an adverse outcome, the maximum EDI score was higher for patients with chronic pulmonary disease as compared with those without it. On the other hand, although adverse outcomes were more likely in white patients and patients with depression, metastatic cancer, and rheumatoid arthritis

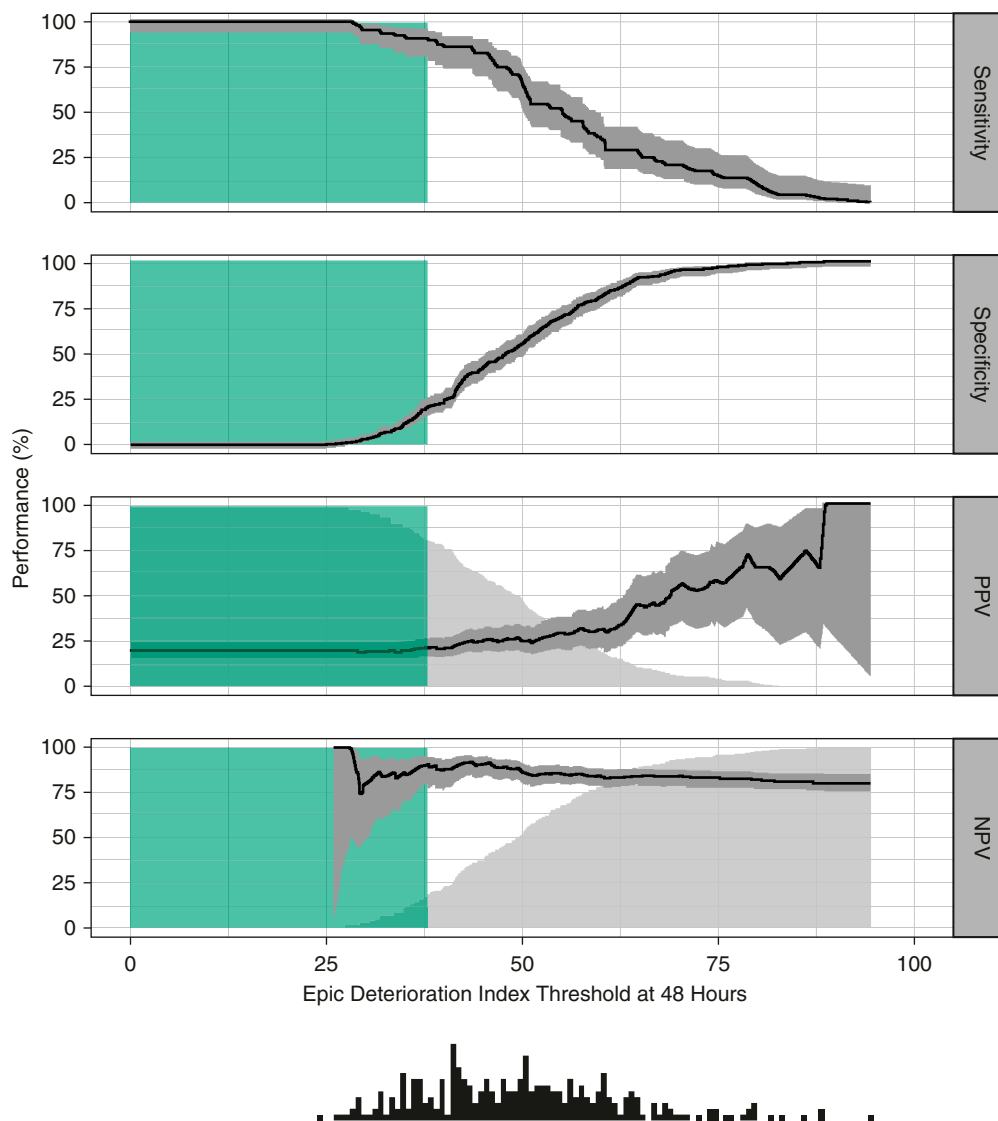


Figure 4. Low-risk analysis plot showing the relationship between the Epic Deterioration Index (EDI) score threshold in the first 48 hours and the sensitivity, specificity, PPV, and NPV, with a histogram demonstrating the distribution of maximum EDI per patient. The low-risk area (<37.9) is shaded in green. NPV = negative predictive value; PPV = positive predictive value.

or other collagen vascular diseases, the maximum EDI score was not different for these subgroups. Although we found no evidence for bias in the EDI score against Black patients, who are disproportionately impacted by COVID-19, the wide confidence intervals (Appendix Table 2) mean that such bias cannot be definitively excluded. The EDI score was higher in older individuals, which is not surprising because age is a component of the EDI. In the face of EDI being a proprietary model, many additional steps should be taken to ensure that its use is valid and improves outcomes. First, making the model parameters

available would allow for a more complete validation and could benefit the public by enabling the model to be refined and compared with other existing prediction models for specific clinical applications (21). Second, prospective validation of our proposed thresholds in other centers and clinical conditions would help validate the generalizability of the EDI. Third, linking targeted interventions to specific EDI thresholds or other clinical assessments would assess whether the EDI can improve patient outcomes.

When comparing our observed performance of the EDI in patients with

COVID-19 against other models reported in the literature, it is interesting to note the AUC reported in our study is much lower than other models in the COVID-19 literature. This could be due in part to the overfitting of other models in the setting of relatively small sample sizes. For instance, Bai and colleagues report an AUC of 0.95 for inpatient clinical deterioration with a model developed and validated in a cohort of 133 patients with 75 predictors (22). By contrast, the EDI was developed on a cohort drawn from more than 130,000 hospitalizations (6). Our findings closely match the observed performance of other deterioration indices

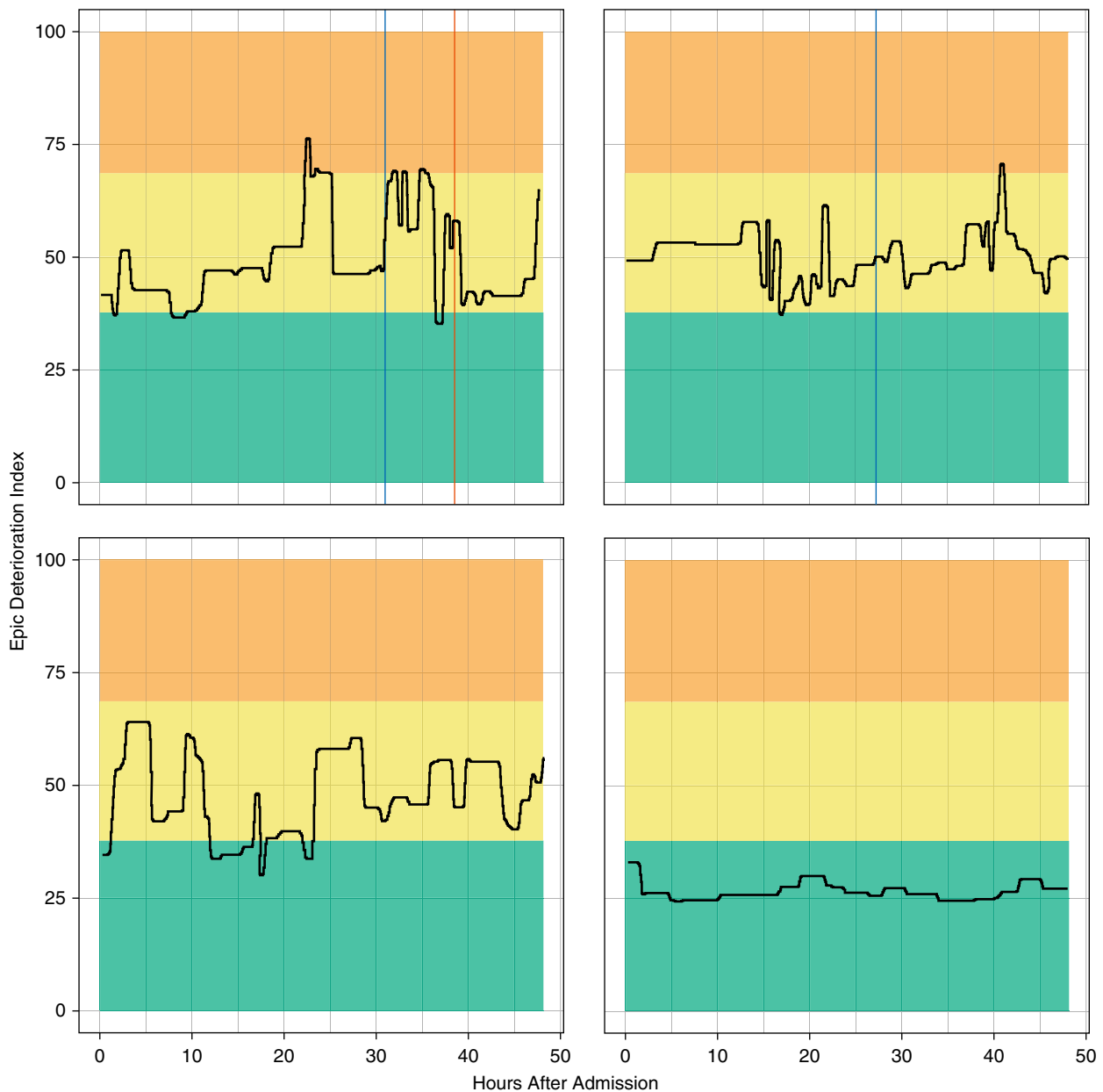


Figure 5. Four example patients, of whom two experienced an adverse outcome, with shaded risk thresholds. Green (Epic Deterioration Index [EDI] < 37.9) represents low risk, yellow represents intermediate risk ($EDI \geq 37.9$ to < 68.8), and orange ($EDI \geq 68.8$) represents high risk. EDI scores recorded after the primary outcome are shown in the top panels but were not used in the model validation. The blue line represents transfer to an intensive care unit, and the red line represents the onset of mechanical ventilation.

that have been validated in patients without COVID-19. In a study of 649,418 hospitalizations, the Advanced Alert Monitor identified deteriorating patients with an AUC of 0.82 (23). The electronic Cardiac Arrest Risk Triage score identified deteriorating patients with an AUC of 0.80 and 0.79 in two separate evaluations (23, 24). The Rothman index identified clinical deterioration with an AUC of 0.76 (25).

Our study should be interpreted in the context of the following limitations. Our evaluation was limited by its focus on a single academic medical center and a relatively small number of patients. However, our cohort of nearly 400 patients was diverse in sex and race and larger than many earlier reports. As compared with a recently described large cohort of 5,700 patients hospitalized with

COVID-19 in New York, our study cohort had a higher proportion of Black patients (43% vs. 23%) and patients with chronic kidney disease (38% vs. 5%), congestive heart failure (21% vs. 7%), and hypertension (75% vs. 57%) and similar proportions of women (43% vs. 40%) and patients with diabetes (42% vs. 34%) and obesity (42% vs. 42%) (2). Our proposed EDI

thresholds may be influenced by local factors, including patterns of COVID-19 testing, triage, and decision-making about hospital admissions and hospital-to-hospital transfers that contributed to our study cohort. These EDI thresholds should be validated in other settings to assess their generalizability. Additionally, clinically relevant EDI thresholds may differ between patients hospitalized with COVID-19 and other conditions; thus, a broader validation of the EDI among hospitalized patients is warranted.

Conclusions

Despite these limitations, our findings have important implications for hospitals with access to the EDI that may be under substantial capacity constraints and strain from managing patients with COVID-19. Our study supports, in part, the role of the EDI to identify a small subset of high-risk patients who may benefit from additional resources and higher-level care and another limited subset of low-risk patients who may be cared for safely in lower-acuity settings. It also suggests opportunities to tailor and improve risk prediction for this condition

beyond the EDI as data accumulate on patients with COVID-19. Finally, it indicates the need for institutions to independently validate widely used proprietary models where the vendor is commonly the only source of model validation. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank Vikas Parekh, M.D., Rich Medlin, M.D., Max Garifullin, M.Sc., Justin Pestrue, M.Econ., and Judy Konye, M.S.N. for their contributions to this work.

References

- Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, *et al*. Fair allocation of scarce medical resources in the time of covid-19. *N Engl J Med* 2020;382:2049–2055.
- Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, *et al*. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York city area. *JAMA* 2020;323:2052–2059. [Published erratum appears in *JAMA* 323:2098.]
- Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, *et al*. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020;369:m1328. [Published erratum appears in *BMJ* 369:m2204.]
- Epic Systems. Artificial intelligence from Epic triggers fast, lifesaving care for COVID-19 patients. Verona, WI: Epic Systems; 2020 [updated 2020 Apr 23; accessed 2020 May 1]. Available from: <https://www.epic.com/epic/post/artificial-intelligence-epic-triggers-fast-lifesaving-care-covid-19-patients>.
- Robbins R, Brodwin E, Begley S, Ross C, Ostherr K. Stanford launches an accelerated test of AI to help with covid-19 care. Boston, MA: STAT; 2020 [updated 2020 Apr 8; accessed 2020 May 1]. Available from: <https://www.statnews.com/2020/04/01/stanford-artificial-intelligence-coronavirus/>.
- Strickland, E. AI can help hospitals triage COVID-19 patients. *IEEE Spectrum*; 2020 [updated 2020 Apr 21; accessed 2020 May 1]. Available from: <https://spectrum.ieee.org/the-human-os/artificial-intelligence/medical-ai/ai-can-help-hospitals-triage-covid19-patients>.
- Ross C. Hospitals are using AI to predict the decline of Covid-19 patients—before knowing it work. Boston, MA: STAT; 2020 [updated 2020 Apr 24; accessed 2020 May 1]. Available from: <https://www.statnews.com/2020/04/24/coronavirus-hospitals-use-ai-to-predict-patient-decline-before-knowing-it-works/>.
- Ochsner Health. Ochsner health adopts new AI technology to save lives in real-time. Acadiana, LA: Ochsner Health; 2018 [updated 2020 Apr 8; accessed 2020 May 1]. Available from: <https://news.ochsner.org/news-releases/ochsner-health-system-adopts-new-ai-technology-to-save-lives-in-real-time>.
- Epic Systems. XGM 2019: Epic experts expand and share their knowledge in Verona. Verona, WI: Epic Systems; 2019 [updated 2020 Apr 8; accessed 2020 May 1]. Available from: <https://www.epic.com/epic/post/xgm-2019-epic-experts-expand-share-knowledge-verona>.
- Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works? *J Am Med Inform Assoc* 2019;26:1651–1654.
- Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. *J Am Med Inform Assoc* 2017;24:1052–1061.
- Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–453.
- Grother P, Ngan M, Hanaoka K. Face recognition vendor test part 3: demographic effects. Gaithersburg, MD: National Institute of Standards and Technology; 2019 [accessed 2020 May 1]. Available from: <https://dx.doi.org/10.6028/nist.ir.8280>.
- Gu T, Mack JA, Salvatore M, Prabhu Sankar S, Valley TS, Singh K, *et al*. Characteristics associated with racial/ethnic disparities in COVID-19 outcomes in an academic health care system. *JAMA Netw Open* 2020;3:e2025197.
- Caldwell P. We've spent billions to fix our medical records, and they're still a mess: here's why. San Francisco, CA: Mother Jones; 2015 [updated 2020 Apr 24; accessed 2020 May 1]. Available from: <https://www.motherjones.com/politics/2015/10/epic-systems-judith-faulkner-hitech-ehr-interoperability/>.
- Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015;7:299ra122.
- Oh J, Makar M, Fusco C, McCaffrey R, Rao K, Ryan EE, *et al*. A generalizable, data-driven approach to predict daily risk of *Clostridium difficile* infection at two large academic health centers. *Infect Control Hosp Epidemiol* 2018;39:425–433.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, *et al*. pROC: display and analyze ROC curves. version 1.16.2. 2020 [updated 2020 Mar 19; accessed 2020 May 1]. Available from: <https://CRAN.R-project.org/package=pROC>.
- Singh K. The runway package for R. github; 2020 [updated 2020 Oct 21; accessed 2020 Oct 21]. Available from: <https://github.com/ML4LHS/runway>.
- Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care* 2015;19:285.
- Singh K, Beam AL, Nallamothu BK. Machine learning in clinical journals: moving from inscrutable to informative. *Circ Cardiovasc Qual Outcomes* 2020;13:e007491.
- Bai X, Fang C, Zhou Y, Bai S, Liu Z, Xia L, *et al*. Predicting COVID-19 malignant progression with AI techniques. Rochester, NY: SSRN; 2020 [updated 2020 Mar 16; accessed 2020 Apr 21]. Available from: <https://papers.ssrn.com/abstract=3557984>.
- Kipnis P, Turk BJ, Wulf DA, LaGuardia JC, Liu V, Churpek MM, *et al*. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *J Biomed Inform* 2016;64:10–19.
- Green M, Lander H, Snyder A, Hudson P, Churpek M, Edelson D. Comparison of the between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients. *Resuscitation* 2018; 123:86–91.
- Wengert BC, Pei KY, Asuzu D, Davis KA. Rothman Index variability predicts clinical deterioration and rapid response activation. *Am J Surg* 2018;215:37–41.