medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



# Validating a Widely Implemented Deterioration Index Model Among Hospitalized COVID-19 Patients

Karandeep Singh, MD, MMSc[1,8]; Thomas S. Valley, MD, MSc[2,8]; Shengpu Tang, BSE[3]; Benjamin Y. Li, BS[3]; Fahad Kamran, BA[3]; Michael W. Sjoding, MD[2,8]; Jenna Wiens, PhD[3,8]; Erkin Otles, MEng[4]; John P. Donnelly, PhD[1,8]; Melissa Y. Wei, MD, MPH, MS[2,8]; Jonathon P. McBride, BS[5]; Jie Cao, MPH[6]; Carleen Penoza, MHSA, RN[7]; John Z. Ayanian, MD, MPP[2,8]; Brahmajee K. Nallamothu, MD, MPH[2,8]

1. Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI

2. Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI

3. Division of Computer Science and Engineering, University of Michigan College of Engineering, Ann Arbor, MI

4. Department of Industrial Engineering, University of Michigan College of Engineering, Ann Arbor, MI

5. Department of Cellular and Molecular Biology, University of Michigan Medical School, Ann Arbor, MI

6. Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI

7. Nursing Informatics, Michigan Medicine, Ann Arbor, MI

8. Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



Word count (abstract): 346

Word count (text): 2857

**Corresponding author:** Karandeep Singh, MD, MMSc

Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI

Address: 1161H NIB, 300 N. Ingalls St., Ann Arbor, MI 48109

Phone: 734-936-1649, E-mail: kdpsingh@umich.edu

**Keywords:** coronavirus disease, deterioration index

**Funding and Disclosures**: TSV was supported by grant K23 HL140165 from the National Heart, Lung, and Blood Institute. JPD was supported by grant K12-HL138039 from the National Heart, Lung, and Blood Institute. MYW was supported by grant K23 AG056638 from the National Institute on Aging. MWS, JW, and BKN were supported by grant R01 LM013325-01 from the National Library of Medicine and by the Michigan Institute for Data Science. MWS was supported by grant K01 HL136687 from the National Heart, Lung, and Blood Institute.

**Acknowledgements:** We would like to acknowledge Vikas Parekh, MD, Rich Medlin, MD, Max Garifullin, MSc, Justin Pestrue, MEcon, and Judy Konye, MSN for their contributions to this work.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.




## Abstract

**Introduction:** The coronavirus disease 2019 (COVID-19) pandemic is straining the capacity of U.S. healthcare systems. Accurately identifying subgroups of hospitalized COVID-19 patients at high- and low-risk for complications would assist in directing resources.

**Objective:** To validate the Epic Deterioration Index (EDI), a predictive model implemented in over 100 U.S. hospitals that has been recently promoted for use in COVID-19 patients.

**Methods:** We studied adult patients admitted with COVID-19 to non-ICU level care at a large academic medical center from March 9 through April 7, 2020. We used the EDI, calculated at 15-minute intervals, to predict a composite adverse outcome of ICU-level care, mechanical ventilation, or death during the hospitalization. In a subset of patients hospitalized for at least 48 hours, we also evaluated the ability of the EDI (range 0-100) to identify patients at low risk of experiencing this composite outcome during their remaining hospitalization. We evaluated model discrimination and calibration using both raw EDI scores and their slopes.

**Results:** Among 174 COVID-19 patients meeting inclusion criteria, 61 (35%) experienced the composite outcome. Area under the receiver-operating-characteristic curve (AUC) of the EDI was 0.76 (95% CI 0.68-0.84). Patients who met or exceeded an EDI of 64.8 made up 17% of the study cohort and had an 80% probability of experiencing the outcome during their hospitalization with a median lead time of 28 hours from when the threshold was first exceeded to the outcome. Employing the EDI slope lowered the AUCs to 0.68 (95% CI 0.60-0.77) and 0.67 (95% CI 0.59-0.75) for slopes calculated over 4 and 8 hours, respectively. In a subset of 109 patients hospitalized for at least 48 hours and who had not experienced the composite outcome, 14 (13%) patients who never exceeded an EDI of 37.9 had a 93% probability of not experiencing the outcome throughout the rest of their hospitalization, suggesting low risk.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



**Conclusion:** In this validation study, we found the EDI identifies small subsets of high- and low-risk patients with fair discrimination. These findings highlight the need for hospitals to carefully evaluate prediction models before widespread operational use among COVID-19 patients.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



## Introduction

The coronavirus disease 2019 (COVID-19) pandemic is straining the capacity of hospitals and healthcare systems across the United States. This strain has led to an expansion of both advanced ICU capabilities within hospitals as well as the opening of lower-acuity hospital beds (e.g., field hospitals) with limited resources for more stable patients. For example, Michigan has one of the largest number of COVID-19 caseloads in the country and recently approved over 4,000 new hospital beds to provide care to COVID-19 patients.[1] Accurately identifying subgroups of COVID-19 patients at high- and low-risk for adverse outcomes using automated risk-score systems could help to alleviate this strain. In addition to assisting with directing care and resources, these systems could reduce direct contact and use of personal protective equipment for low-risk patients.[2] This could be particularly valuable for less experienced clinicians if provided through the electronic health record (EHR). Recent work, however, has demonstrated a paucity of validated high-quality models in COVID-19 patients.[3]

The Epic Deterioration Index (EDI) is a proprietary early-warning prediction model that assesses the probability of hospitalized patients requiring a rapid response, resuscitation, intensive care unit (ICU)-level care, or dying. The EDI 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 there is no published research on the validity of the EDI, ubiquitous use of the Epic EHR across the U.S. has led to growing anecdotal reports of its value in critically ill patients.[4,5] Implemented in over 100 hospitals across the U.S., advocates have further promoted its use among COVID-19 patients despite the absence of published data on its validity in this population.[6] Based on an internal evaluation, an Epic Systems spokesperson recently stated that "the model performed well [in COVID-19 patients] and…some hospitals are now using the

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



model with confidence."[7] A press release by Epic noted that validation in COVID-19 patients was ongoing but also quoted a physician who noted the EDI is "helping save lives."[8]

In this context, we sought to independently evaluate the ability of the EDI to predict adverse outcomes among patients hospitalized with COVID-19 at a large academic medical center. Our findings have potential implications for how this widespread tool may be validated and used by other healthcare systems during the COVID-19 pandemic.

## 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 April 7, 2020 from the emergency department, outpatient clinics, and outside hospital transfers. We excluded patients admitted directly to an ICU (n=124), patients discharged to hospice (n=3), and patients without available scores for the EDI (n=8). We also excluded patients who remained hospitalized but had not yet experienced the composite outcome described below (n=78) 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.

### *The Epic Deterioration Index Model*

The EDI is generated from a proprietary early-warning prediction model developed by Epic Systems Corporation (Verona, WI) using data that are routinely recorded within its EHR. Epic is one of the largest healthcare software vendors in the world, and its EHR is used by the

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



majority of 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).[9]

The EDI aims to detect patients who deteriorate and require higher levels of care. Its score ranges from 0 to 100, where higher numbers denote a greater risk of experiencing a composite adverse outcome of requiring rapid response, resuscitation, intensive care unit (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. Calculations of the EDI begin immediately following hospital admission and then continue at regular 15-minute intervals until discharge. While the algorithm was developed prior to the existence of COVID-19, it includes several predictors that may be clinically relevant in COVID-19 patients: 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). The model's structure and coefficients are proprietary and not openly available to institutions.

*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: requirement for ICU-level care, initiation of mechanical ventilation, and in-hospital death. Our primary outcome

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



was reached if these events occurred at any time during the hospitalization. We chose to include these adverse events for the composite outcome because they are highly relevant in the clinical care of COVID-19 patients where rapid respiratory decline is frequently described.

*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 prior to 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 based on the entire trajectory of predictions using 2 strategies. First, we used the raw EDI scores to identify any time a patient crossed a decision threshold and deemed them 'high-risk' at that point, regardless of subsequent scores.[10] This was done to reflect how such a model would likely be used in clinical practice to identify high-risk patients by flagging them for further clinical evaluation. For our second approach, we evaluated EDI slopes. Recent anecdotal reports have suggested that the rate of change of the EDI score may have predictive value that goes beyond the score itself.[11] For these analyses, we evaluated the model discrimination using the EDI slope calculated on a rolling basis every 15 minutes using the prior 4 and 8 hours of raw EDI scores.

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 to

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



when they experienced the outcome. In all cases, we calculated empirical 95% confidence intervals (CI) for the AUC using 1,000 bootstrap replicates of our study cohort. Although it is unknown whether the EDI can be interpreted as a probability, model calibration was assessed using a calibration curve by comparing quintiles of all predicted EDI to the observed risk.

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

Another purported 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. 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. Our goal was to evaluate how well the EDI at the end of 48 hours could identify patients at 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. 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. In consultation with intensivists on our research team (TSV and MWS), we identified two clinically actionable thresholds: one for identifying high-risk patients who will likely need ICU-level care (based on the EDI throughout

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



the hospitalization) and one for identifying low-risk patients who may be appropriate for lower-acuity care (based on the 48-hour analysis).

*Software*

We used R 3.6.0 for all analyses, employing the pROC and rms packages, and a revised version of R code made available by Darren Dahly, PhD on GitHub for assessing model discrimination and calibration.[12–14] We have made our statistical code available at https://github.com/ml4lhs/edi_validation.

**Results**

We identified 174 patients with COVID-19 who met inclusion criteria for our study cohort. Their median age was 61 years (IQR 50-71), 73 (42%) were women and 77 (44%) were African Americans (**Table 1**). Within this cohort, 61 patients (35%) experienced the composite adverse outcome during their hospitalization with a median length of follow-up of 4.5 days (IQR 2.7-7.8, max 25). The outcome occurred at a median of 1.3 days (IQR 0.55-3.5, max 10) after admission. Of these patients, 60 patients (34%) required ICU-level care, 34 (20%) required mechanical ventilation, and 7 (4%) died. Those who experienced an adverse outcome were older, more likely to be men, and more likely to have a history of cardiac arrhythmias or chronic kidney disease (all $p<0.05$) (**Table 1**).

Overall, the raw EDI score had an AUC of 0.76 (95% CI 0.68-0.84) 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 64.8 had an 80% probability of experiencing the primary outcome (i.e., positive predictive value) with a sensitivity of 39%, and they comprised 17% of the study cohort. The median lead time from when the threshold was first exceeded to

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



when the outcome occurred was 28 hours (IQR 2.5-69). The EDI slope had AUCs of 0.68 (95% CI 0.60-0.77) and 0.67 (95% CI 0.59-0.75) when assessed on a rolling basis over the prior 4 and 8 hours, respectively. **Figure 2** evaluates model calibration and demonstrates that the EDI systematically overpredicts the risk of experiencing the primary outcome if interpreted as a probability.

In the subset of 109 patients hospitalized for at least 48 hours who did not experience the composite adverse outcome, the EDI had an AUC of 0.65 (95% CI 0.52-0.77) for predicting whether an outcome would occur at any point during the rest of their hospitalization. The performance characteristics of the 48-hour maximum EDI in this subset of patients are reported in **Figure 3**. A total of 14 (13%) patients who never exceeded an EDI of 37.9 in the first 48 hours of their hospitalization had a 93% probability of not experiencing the outcome (i.e., negative predictive value) for the remainder of the hospitalization (median remaining follow-up of 3.0 days [IQR 1.5-5.8]) with a sensitivity of 96% above this threshold.

**Figure 4** demonstrates four examples of EDI patterns in COVID-19 patients overlaid with the identified high- and low-risk thresholds (≥ 64.8 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 results suggest that the EDI exhibits fair discrimination for the prediction of adverse outcomes in a diverse COVID-19 patient 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 80% but a relatively low sensitivity of 39%. We also found that the maximum EDI in the first 48 hours of the hospitalization may help identify a small subset of

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



low-risk patients who may be safely transferred to lower-acuity settings, thereby conserving resources. 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. Contrary to anecdotal reports,[11] we found the absolute EDI score to have better discrimination than the slope of the EDI score among patients with COVID-19.

Our findings have implications for both the EDI and the development and validation of other prediction models in the setting of COVID-19. The EDI is rapidly being adopted by U.S. hospitals to predict which Covid-19 patients will become critically ill. Yet there has been almost no reported validation of the tool's effectiveness on COVID-19 patients and little evaluation in other settings.[15] This unprecedented adoption is a marker of the challenges imposed by the COVID-19 pandemic on hospitals and healthcare systems. In traditional times, most institutions would carefully evaluate new prediction models on their cohorts in order to understand the performance of a tool like the EDI before implementing it in their systems. The COVID-19 pandemic is altering this paradigm with hospitals struggling to determine who is at highest risk and who may be safely discharged to lower-acuity settings. Our findings may suggest clinicians carefully consider this tradeoff with the EDI in real-world settings.

Yet this caution is not just true of the EDI but also other prediction models being discussed in the published literature. First, it is interesting to note the AUC reported in our study is much lower than other non-EDI models in the COVID-19 literature. This could be due in part to overfitting with these other models. 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.[16] By contrast, the EDI was developed on a cohort drawn from more than 130,000 hospitalizations.[7] Second, we noted the EDI fluctuates substantially when calculated at 15-minute intervals (**Figure 4**), in part because it only relies on the most recent

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



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 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 the model is run more frequently. Lastly, the proprietary nature of the EDI raises 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 predictive model. If the developers of this proprietary model openly shared their model parameters and validation, it could benefit the public by enabling the model to be refined and compared with other existing predictive models for specific clinical applications such as COVID-19.

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 174 patients was diverse in gender and race and larger than many earlier reports. As compared to a recently described large cohort of 5700 patients hospitalized with COVID-19 in New York, our study cohort had a higher proportion of African Americans (44% vs. 23%) and patients with chronic kidney disease (26% vs. 5%) and similar proportions of women (42% vs. 40%) and patients with hypertension (61% vs. 57%), congestive heart failure (11% vs. 7%), diabetes (37% vs. 34%), and obesity (36% vs. 42%).[17] 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. This further reinforces the importance of individual healthcare systems to validate scores like the EDI in their patients independently.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



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 COVID-19 patients. Our study supports, in part, a 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 growing data accumulates on COVID-19 patients. Echoing the findings of the recent systematic review conducted by Wynants and colleagues, accurate predictive models for COVID-19 are urgently needed.[3] However, we suspect no single institution will have a sufficiently representative sample size of COVID-19 patients to allow for the development of a widely generalizable model in this patient population. Thus, future prediction model development should draw data from diverse cohorts of COVID-19 patients across multiple institutions.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



## References

1. Barrett M. Michigan health officials approve 4,100 new emergency hospital beds for coronavirus patients. mlive. https://www.mlive.com/coronavirus/2020/04/michigan-health-officials-approve-4100-new-emergency-hospital-beds-for-coronavirus-patients.html. Published April 7, 2020. Accessed April 8, 2020.
2. How doctors can keep their families safe after providing COVID-19 care. American Medical Association. https://www.ama-assn.org/practice-management/physician-health/how-doctors-can-keep-their-families-safe-after-providing-covid. Accessed April 20, 2020.
3. Wynants L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020;369. doi:10.1136/bmj.m1328
4. Ochsner Health System Adopts New AI Technology to Save Lives in Real-time. Online Newsroom. https://news.ochsner.org/news-releases/ochsner-health-system-adopts-new-ai-technology-to-save-lives-in-real-time. Published February 28, 2018. Accessed April 8, 2020.
5. XGM 2019: Epic Experts Expand and Share Their Knowledge in Verona. https://www.epic.com/epic/post/xgm-2019-epic-experts-expand-share-knowledge-verona. Accessed April 8, 2020.
6. Robbins R, Brodwin E, Begley S, Ross C, Ostherr K. Stanford launches an accelerated test of AI to help with Covid-19 care. STAT. https://www.statnews.com/2020/04/01/stanford-artificial-intelligence-coronavirus/. Published April 1, 2020. Accessed April 8, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



7. AI Can Help Hospitals Triage COVID-19 Patients. IEEE Spectrum: Technology, Engineering, and Science News. https://spectrum.ieee.org/the-human-os/artificial-intelligence/medical-ai/ai-can-help-hospitals-triage-covid19-patients. Accessed April 21, 2020.
8. Artificial Intelligence from Epic Triggers Fast, Lifesaving Care for COVID-19 Patients. https://www.epic.com/epic/post/artificial-intelligence-epic-triggers-fast-lifesaving-care-covid-19-patients. Accessed April 23, 2020.
9. Caldwell P. We've Spent Billions to Fix Our Medical Records, and They're Still a Mess. Here's Why. Mother Jones. https://www.motherjones.com/politics/2015/10/epic-systems-judith-faulkner-hitech-ehr-interoperability/. Published October 21, 2015. Accessed April 24, 2020.
10. Oh J, Makar M, Fusco C, 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(4):425.
11. COVID-19 and AI: A Virtual Conference. Stanford Institute for Human-Centered Artificial Intelligence. https://hai.stanford.edu/events/covid-19-and-ai-virtual-conference/overview. Published March 6, 2020. Accessed April 21, 2020.
12. Display and Analyze ROC Curves [R package pROC version 1.16.2]. https://CRAN.R-project.org/package=pROC. Accessed April 23, 2020.
13. CRAN - Package rms. https://cran.r-project.org/web/packages/rms/index.html. Accessed April 23, 2020.
14. Dahly DL. Evaluating a logistic regression based prediction tool in R |. https://darrendahly.github.io/post/homr/. Published April 21, 2019. Accessed April 23, 2020.
15. Ross C, Herper M, Joseph A, et al. AI used to predict Covid-19 patients' decline before

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



proven to work. STAT. https://www.statnews.com/2020/04/24/coronavirus-hospitals-use-ai-to-predict-patient-decline-before-knowing-it-works/. Published April 24, 2020. Accessed April 24, 2020.

16. Bai X, Fang C, Zhou Y, et al. Predicting COVID-19 Malignant Progression with AI Techniques. March 2020. doi:10.2139/ssrn.3557984
17. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*. April 2020. doi:10.1001/jama.2020.6775

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



## Tables

**Table 1.** Patient characteristics, overall and stratified by adverse outcomes.

| Variable | Overall<br>n=174 | Adverse Outcome<br>n=61 (35%) | No Adverse Outcome<br>n=113 (65%) | p |
|---|---|---|---|---|
| Age, median (IQR) | 61 (50-71) | 67 (59-76) | 56 (49-68) | <0.001 |
| Female, n (%) | 73 (42) | 18 (30) | 55 (49) | 0.02 |
| Race, n (%) | | | | 0.89 |
| African American | 77 (44) | 27 (44) | 50 (44) | |
| White | 63 (36) | 25 (41) | 38 (34) | |
| Other | 17 (9.8) | 4 (6.6) | 13 (12) | |
| Unknown | 17 (9.8) | 5 (8.2) | 12 (11) | |
| Comorbidities, n (%)* | | | | |
| Cardiac arrhythmias | 45 (26) | 23 (38) | 22 (19) | 0.03 |
| Chronic kidney disease | 45 (26) | 24 (39) | 21 (19) | 0.009 |
| Chronic pulmonary disease | 41 (24) | 19 (31) | 22 (19) | 0.21 |
| Congestive heart failure | 20 (11) | 10 (16) | 10 (8.8) | 0.32 |
| Depression | 40 (23) | 17 (28) | 23 (20) | 0.52 |
| Diabetes | 65 (37) | 25 (41) | 40 (35) | 0.75 |
| Hypertension | 106 (61) | 41 (67) | 65 (58) | 0.37 |
| Liver disease | 13 (7.5) | 6 (9.8) | 7 (6.2) | 0.73 |
| Metastatic cancer | 20 (11) | 11 (18) | 9 (8.0) | 0.13 |
| Obesity | 63 (36) | 20 (33) | 43 (38) | 0.77 |
| Rheumatoid arthritis or collagen vascular disease | 14 (8.0) | 8 (13) | 6 (5.3) | 0.22 |

* Comorbidities were calculated using ICD-10 codes from all encounters prior to the COVID-19 encounter. Comorbidities were unavailable for 23 (13%) patients, 8 of whom experienced adverse outcomes and 15 who did not.

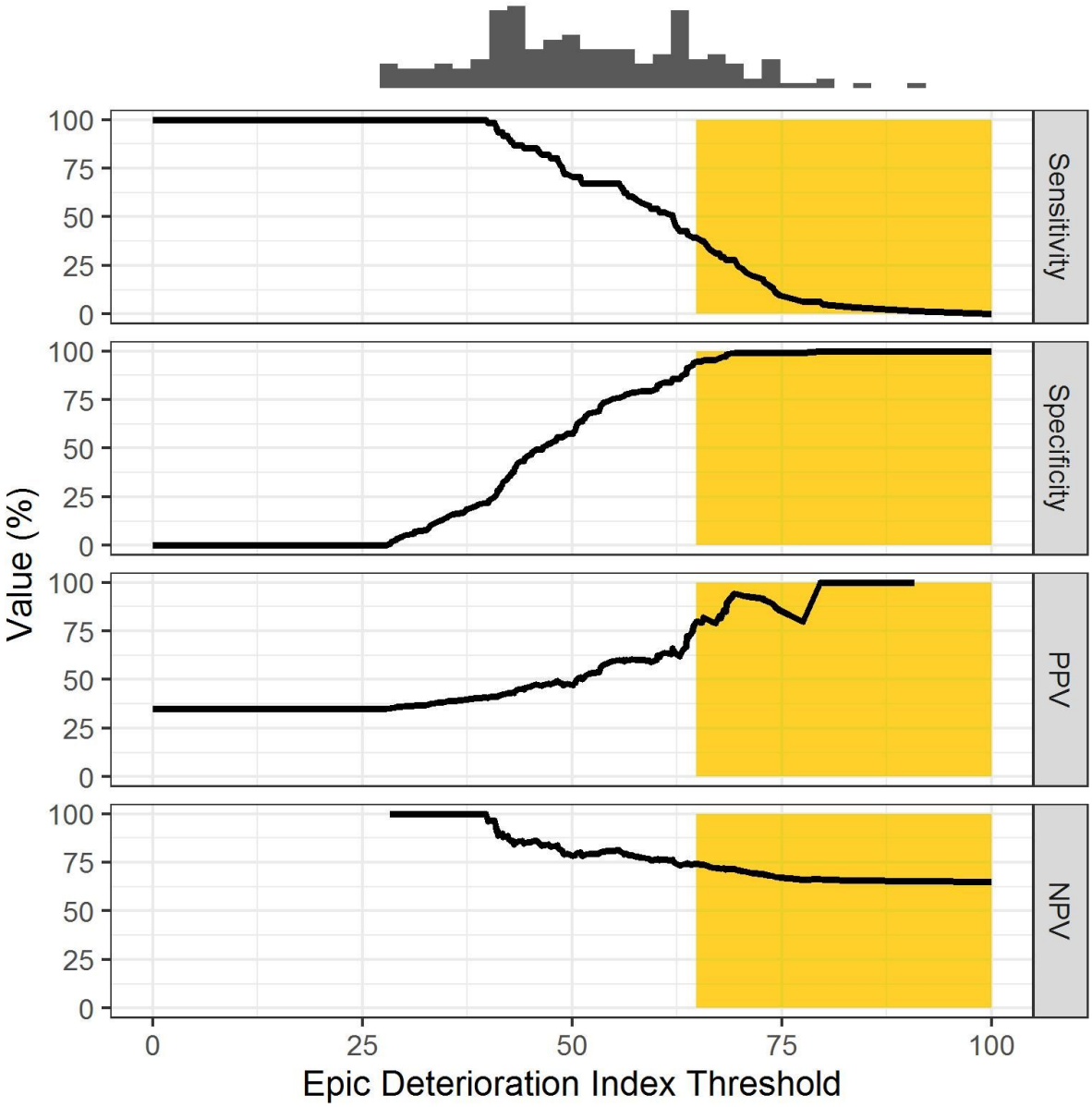
Note: Continuous variables were compared using a Wilcoxon test. Proportions were compared using a Chi-squared test.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



**Figures**

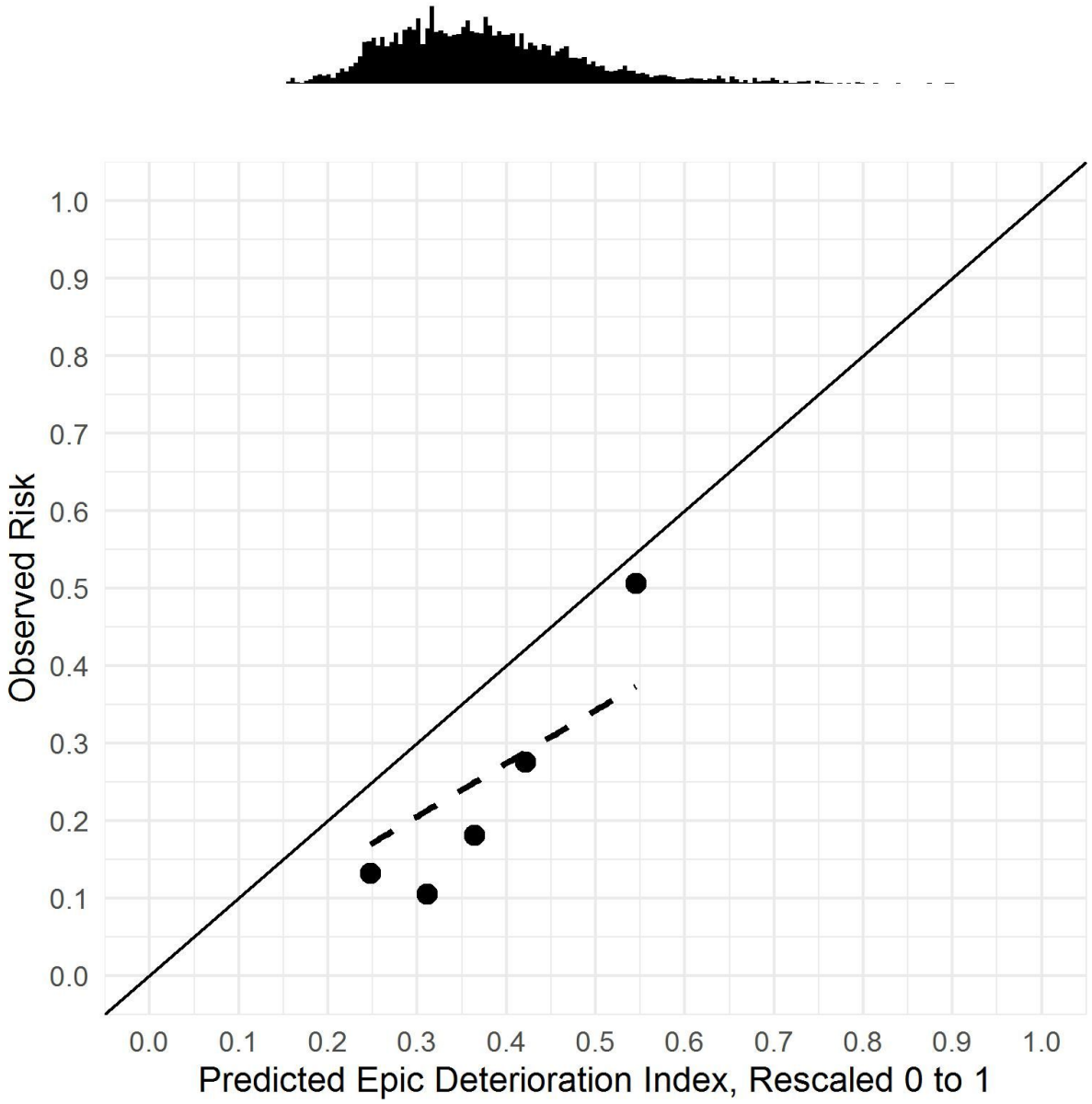
**Figure 1.** Plot showing the relationship between the EDI threshold and the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), with a histogram demonstrating the distribution of maximum EDI per patient. High-risk area shaded in orange.



medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



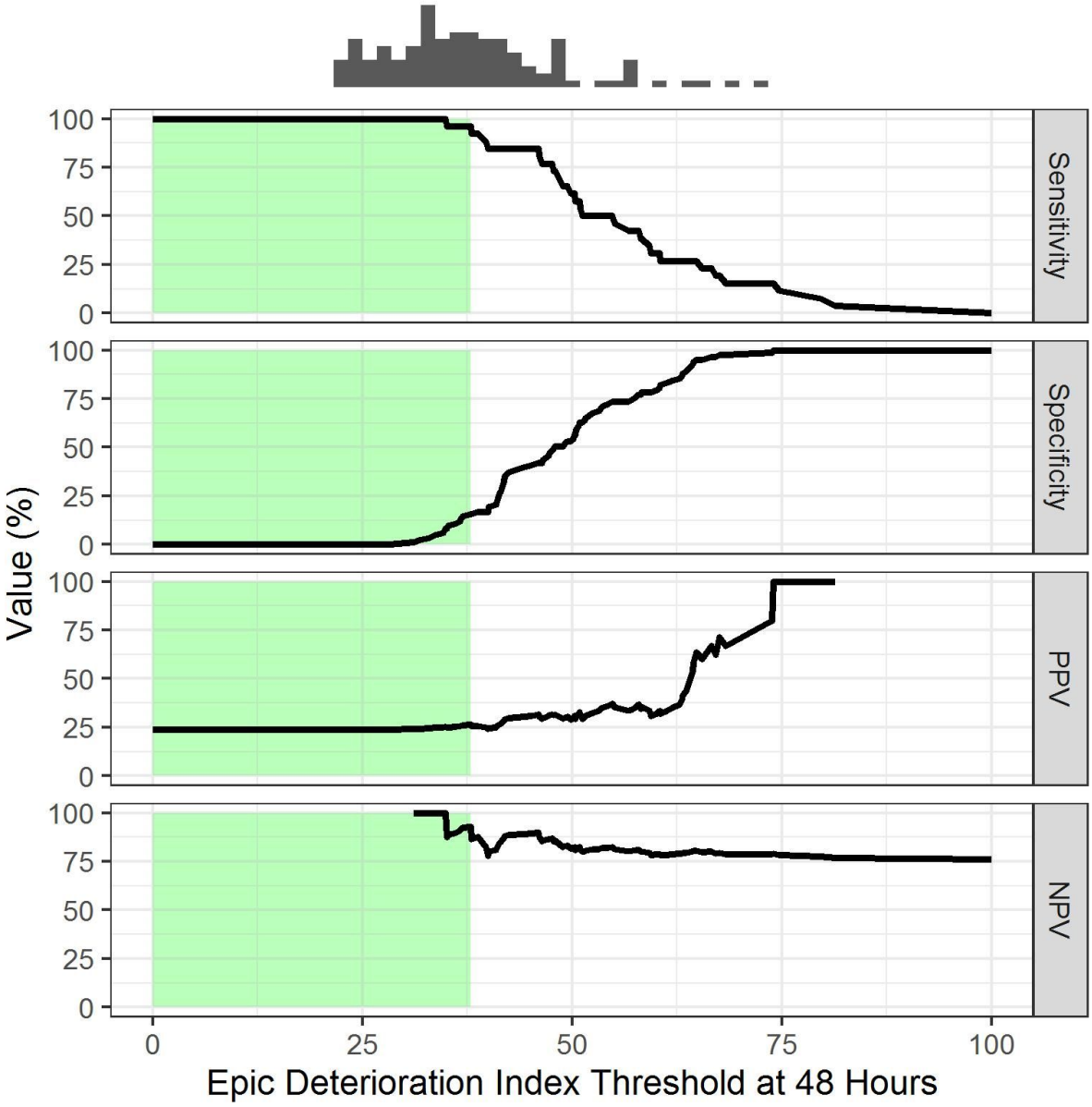
**Figure 2.** Calibration curve comparing quintiles of all predicted EDI, rescaled to 0 to 1, to the observed risk, with an ideal-fit line (solid), best-fit line (dashed), and a histogram of predicted EDI.



medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



**Figure 3.** 48-Hour Analysis: Plot showing the relationship between the EDI score threshold in the first 48 hours and the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), with a histogram demonstrating the distribution of maximum EDI per patient. Low-risk area shaded in green.



medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20079012; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

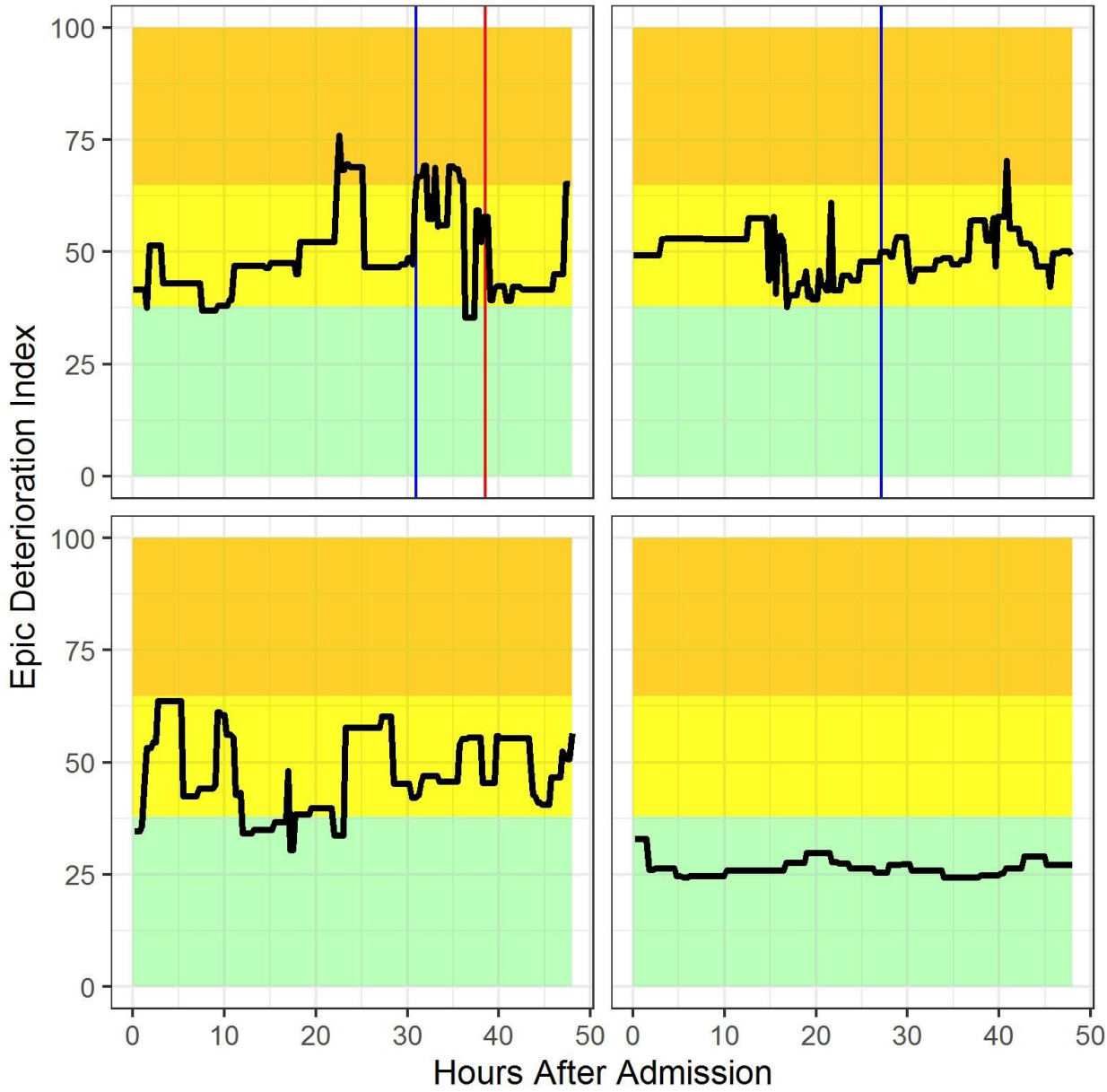



**Figure 4.** Four example patients, of whom two experienced an adverse outcome, with shaded risk thresholds. Green (EDI < 37.9): low risk, yellow: intermediate risk (≥ 37.9 to < 64.8), orange (EDI ≥ 64.8): high risk. EDI recorded after the primary outcome are shown in the top panels but were not used in the model validation. Blue line: transfer to intensive care unit, red line: onset of mechanical ventilation.