Early detection of in-patient deterioration: one prediction model does not fit all

* Jacob N. Blackwell, MD¹

* Jessica Keim-Malpass, PhD, RN^{2, 3}

Matthew T. Clark, PhD⁴

Rebecca L. Kowalski¹

Salim N. Najjar¹

Jamieson M. Bourque, MD^{1,2}

Douglas E. Lake, Ph.D.^{1, 3}

J. Randall Moorman, MD^{1, 3,4}

* contributed equally as first author

¹Division of Cardiovascular Medicine, Department of Medicine, University of Virginia, Charlottesville, VA, USA.

²Department of Acute and Specialty Care, University of Virginia School of Nursing Charlottesville, VA, USA.

³Center for Advanced Medical Analytics, University of Virginia, Charlottesville, VA, USA.

⁴AMP3D, Charlottesville, VA, USA

Correspondence:

J Randall Moorman, M.D.

Division of Cardiovascular Medicine

University of Virginia

Box 800158

Charlottesville VA 22908

434-982-3367 voice

434-982-1998 fax

rm3h@virginia.edu

Conflicts of interest:

JRM is Chief Medical Officer and shareholder and MTC is Chief Scientific Officer and shareholder in Advanced Medical Predictive Devices, Diagnostics, and Displays, Charlottesville, VA.

Grant support: Frederick Thomas Advanced Medical Analytics Fund

Word count: 3500

Number of tables and figures: 6

Key Words:

critical care; deterioration; predictive models; predictive analytics; ICU transfer; clinical computing; continuous predictive analytics monitoring

Running title: No single predictive model for patient deterioration

Abstract

Objectives: Early detection of subacute potentially catastrophic illnesses using available data is a clinical imperative, and scores that report risk of imminent events in real time abound. Patients deteriorate for a variety of reasons, and it is unlikely that a single predictor such as an abnormal National Early Warning Score (NEWS) will detect all of them equally well. The objective of this study was to test the idea that the diversity of reasons for clinical deterioration leading to ICU transfer mandates multiple targeted predictive models.

Design: Individual chart review to determine the clinical reason for ICU transfer; determination of relative risks of individual vital signs, lab tests and cardiorespiratory monitoring measures for prediction of each clinical reason for ICU transfer; logistic regression modeling for the outcome of ICU transfer for a specific clinical reason.

Setting: Cardiac medical-surgical ward; tertiary care academic hospital.

Patients: 8111 adult patients, 457 of whom were transferred to an ICU for clinical deterioration.

Interventions: None.

Measurements and main results: We calculated the contributing relative risks of individual vital signs, lab tests and cardiorespiratory monitoring measures for prediction of each clinical reason for ICU transfer, and used logistic regression modeling to calculate ROC areas and relative risks for the outcome of ICU transfer for a specific clinical reason. The reasons for clinical deterioration leading to ICU transfer were varied, as were their predictors. For example, the three most common reasons – respiratory instability, infection and suspected sepsis, and heart failure requiring escalated therapy – had distinct signatures of illness. Statistical models trained to target specific reasons for ICU transfer performed better than one model targeting combined events, and both performed better than the untrained NEWS score.

Conclusions and relevance: A single predictive model for clinical deterioration does not perform as well as having multiple models trained for the individual specific clinical events leading to ICU transfer.

Introduction

Patients who deteriorate on the hospital ward and are emergently transferred to the ICU have poor outcomes.^{1–6} Early identification of subtly worsening patients might allow quicker treatment and improved outcome. To aid clinicians, early warning scores have been advanced that compare vital signs and lab tests to thresholds. The reception has been mixed.^{7,8} Vital signs are often delayed, incorrect, or never measured,^{9,10} and lab tests require blood draws and time. Nonetheless, these are the major (or only) inputs into popular track-and-trigger systems such as the National Early Warning Score (NEWS) and others.^{11–20} Thresholds for awarding risk points are based on clinical experience, but suffer from lack of validation and the loss of information that forced dichotomization inevitably brings.^{21–24} Only some are trained on clinical events. For example, eCART is trained for early detection of hospital patients with cardiac arrest¹⁴, and the Rothman Index is trained to detect patients who will die in the next 12 months¹⁵.

NEWS has been used widely as an early warning score of clinical deterioration since it was adopted in 2012 by the Royal College of Physicians and the UK National Health Service. A primary strength of general track-and-trigger systems like NEWS is in promoting inter-provider communication.^{25–27} On the other hand, a meta-analysis²⁸ of 232 studies emphasized multiple shortcomings of this kind of early warning tool, including the intermittent sampling, vulnerability to inaccuracies and user error, limited sensitivity, and lack of specialty-specific predictors.^{29–32}

Importantly, most systems ignore the wealth of information that is present in cardiorespiratory monitoring. While many patients are continuously monitored, these data are bulky and require mathematical analysis. Nonetheless, they contribute equally to vital signs and lab tests in detection of patients at risk of clinical deterioration and ICU transfer from a cardiac medical and surgical ward.³⁹ We note, moreover, that risk estimation for neonatal sepsis using advanced mathematical analysis that considers only continuous cardiorespiratory monitoring saves lives.³³

Finally, all of these early detection systems present one score as an omniscient risk marker for all conditions and patients.³⁴ Clinicians know, though, that there are many paths to clinical deterioration. For example, the most common forms of deterioration leading to ICU transfer are respiratory instability,^{35–41} hemodynamic instability^{35,38,39}, sepsis^{40,42}, bleeding^{38,42}, neurological decompensation⁴², unplanned surgery³⁸, and acute renal failure or electrolyte abnormalities.⁴² The multiplicity of candidate culprit organ systems suggests that a single model is not likely to catch all the patients who are worsening.^{38,39}

More recent evidence of the heterogeneity of human illness and the elusive nature of early detection lies in the results of the 2019 Computing in Cardiology Challenge.⁴³ Contestants trained models on data from 2 hospitals and tested them on a held-out data from a third hospital. We note two important results. First, the best scores on the training data were on the order of 40 to 50% of the highest possible score. Second, the scores fell a great deal – most to negative numbers, signifying worse performance than an algorithm that never predicted sepsis – when tested in the third hospital. We take this as more evidence of the difficulty of detection of even a single medical diagnosis, sepsis, let alone all possible ones at the same time.

Here, we tested three approaches to predictive analytics monitoring. The first is untrained models that apply thresholds and cutoffs to vital signs, and we represent the class with NEWS. The second was to train a universal predictive model on all patients who went to the ICU using measured values of vital signs, lab tests and continuous cardiorespiratory monitoring. This approach has advantages of learning from continuous cardiorespiratory

monitoring and from the range of measured vital signs and lab test values, avoiding problems of dichotomization by thresholds. The third is a set of models trained on patients who had specific reasons for ICU transfer identified by clinician review, which has the additional advantage of learning signatures of specific target illnesses.

Approaches that might improve track-and-trigger systems include the use of libraries of predictive models that are tailored for specific reasons for clinical deterioration within the widely varied venues within the modern hospital, risk estimation along the continuum of measured values rather than threshold crossings, and analysis of continuous data in addition to vital signs and lab tests. Here we test these ideas.

Methods

Study design

In early detection of hospital patients at risk of clinical deterioration and the need of escalation to ICU care, we wished to test the idea that a single predictive analytics model using thresholds of vital signs and lab tests could be improved upon by a set of target-specific models that used continuous risk estimates and cardiorespiratory monitoring.

We examined the empiric risk profiles of individual vital signs, lab tests and cardiorespiratory monitoring parameters for 7 clinical deterioration phenotypes. We made predictive statistical models based on logistic regression adjusted for repeated measures for each phenotype as well as their composite, and examined their performance in detecting other deterioration phenotypes. The University of Virginia Institutional Review Board approved the study.

Study population

We studied 8111 consecutive admissions from October 2013 to September 2015 on a 73-bed adult acute care cardiac and cardiovascular surgery ward at the University of Virginia Hospital.³⁹ We used an institutional electronic data warehouse to access electronic medical record (EMR). Six patients were added to the original cohort (n= 8105) who did not have complete continuous cardiorespiratory monitoring data.

We reviewed the charts of the 457 patients who were transferred to the ICU because of clinical deterioration. Five clinical reviewers developed and implemented clinical definitions indicating reasons for deterioration and reviewed records for the 48 hours prior to and following ICU transfer. To evaluate inter-rater reliability, we calculated a weighted kappa coefficient on a nested 50 events that were evaluated by all reviewers. The new clinical review was independent of our earlier work,³⁹ which used a combined outcome of ICU transfer, urgent surgery, or unexpected death.

Clinical and continuous cardiorespiratory monitoring data

We analyzed vital signs and laboratory tests recorded in the EMR, and the continuous 7-lead EKG signal that is standard of practice. Vital signs (Figure 3) and laboratory tests (Figure 4) were sampled and held. We calculated cardiorespiratory dynamics at 15-minute intervals over 30-minute windows.

Statistical analyses

To estimate the relative risk of an event as a function of measured variables, we constructed predictiveness curves⁴⁴. In order to reduce bias due to repeated measures and missing data, we used a bootstrapping technique. We sampled one measurement within 12 hours before each event and one measurement from all non-event patients at a random time during their stay. The distributions of time since admission were not significantly different for these two groups (data not shown). We calculated the relative risk of an event at each decile of the sampled variable, then interpolated the risk to 20 points evenly spaced in the range of the variable. We repeated this process of sampling, calculating relative risk, and interpolating 30 times. Finally, we averaged the 30 risk estimate curves to obtain a bootstrapped predictiveness curve at the 20 evenly spaced points, and we display the result as a heat map.

To relate the predictors to the outcomes, we used multivariable logistic regression analysis adjusted for repeated measures. We labelled data within 12 hours of ICU transfer as outcome=1 and the rest as outcome=0. The output of the resulting regression expressions is the probability of ICU transfer in the next 12 hours.

To test for the significance of differences between the outputs of predictive models, we calculated the area under the receiver operating characteristic (ROC) curves for NEWS, a model trained to target all the ICU transfers, and 7 additional models, one for each reason for ICU transfer. The ROC curves report on discrimination between data belonging to patients with and without ICU transfer in the next 12 hours. For each reason for transfer, therefore, we calculated 3 ROC areas. For the group of 7 specific reasons for ICU transfer, we tested the difference between NEWS and the outputs of the general model, and between the outputs of the general model and the outputs of the 7 specifically targeted models using the paired t-test.

In model development, we adhered to the Transparent Reporting of multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement recommendations.⁴⁵

Results

Study population

The main admitting services were cardiology (49%) and cardiac surgery (19%); the remainder were equally from other medical or surgical services. The median age was 59 (IQR 55 to 75). The mortality rate was 0.4% and 17% for patients who did not or did require ICU transfer.

Figure 1 is an UpSet plot⁴⁶ of the numbers of patients and their reasons for ICU transfer. The most common reason for transfer was respiratory instability alone, followed by respiratory instability and suspected sepsis. The kappa statistic ranged from 0.616 to 1.0 indicating moderate to excellent agreement among the reviewers.

Relative risks for ICU admission

Figure 2 shows the relationship of risk (color) to reason for admission to ICU (row) and measured variable (column). Each of the three panels shows a set of predictiveness curves as heat maps for kinds of measurements – a lab test (Na) in panel A, a vital sign (mean arterial pressure, MAP) in panel B, and a measure from continuous cardiorespiratory monitoring (breathing rate derived from the EKG⁴⁷ in panel C. The gray-scale bars at the top show the densities of the measurements.

The major finding of this subset of the overall results is that serum Na and the MAP, when low, signify impending ICU transfer for escalation of HF therapy, but they do not portend ICU transfer for any of the other reasons. Panel C shows that a rise in the continuously measured respiratory rate identifies patients with respiratory instability or infection to such a degree that ICU transfer takes place. These maps indicate the heterogeneity of patient characteristics, but are not quantitative tests. For that purpose, we used logistic regression models.

The top rows of the colored matrices show the risk profiles of a predictive model trained on all of the events combined. The smaller ranges of color demonstrate that the predictive utility of Na for HF, MAP for HF, and breaths from EKG for respiratory instability or infection are blunted. Thus, by combining outcomes, the ability of a multivariable statistical model to detect any individual clinical cause for ICU transfer is diminished.

Figures 3 to 5 show the complete catalog of relationship of risks to clinical causes for ICU transfer and measured variable. The finding is that the phenotypes of reasons for ICU transfer differ from each other. Consider bleeding (bottom row) as an example – it is uniquely marked by hypokalemia, hypocalcemia, thrombocytopenia and, of course, anemia. As a result, we would expect a predictive model trained to detect bleeding would not perform well in detecting other reasons for ICU transfer, nor would we expect models trained on other reasons for ICU transfer to do well at detecting bleeding.

Statistical models

We limited the number of predictor variables to one per ten events. Vital signs figured in all 8 of the trained models; lab tests and continuous cardiorespiratory monitoring measures each figured in 6 models. The continuous cardiorespiratory monitoring measure most commonly included was EKG-derived respiratory rate (4 models). The CAD model had the

most continuous cardiorespiratory monitoring measures, 2 of its 3 (detrended fluctuation analysis and coefficient of sample entropy) components.

The abilities of models to detect on- and off-target causes for ICU transfer are demonstrated in Figure 6, a matrix of ROC areas that result when the model targeting the event in the rows is used for early diagnosis of the event listed in the columns. The diagonal from bottom left to top right shows the ROC areas of the models as trained for the target events; off-diagonal values give the values for detecting off-target events. For example, the model to detect bleeding has a ROC area of 0.78 (the value in the cell at the lower left of the matrix), but that model has no ROC area of >0.6 in detecting other reasons for ICU transfer (bottom row), and the ROC areas of models trained for other events (left-hand column) exceed 0.6 in only case, the model trained on all events combined.

Additionally, Figure 6 shows variation of ROC areas from 0.66 to 0.84 along the diagonal, meaning some events are easier to detect than others. For examples, models to detect ICU transfer due to infection or neurological events have ROC area >0.8 , while that for myocardial ischemia was <0.7 . While reasonably common on this cardiac and cardiac surgery ward, we note that unstable coronary syndromes are not particularly good candidates for early detection. Arguably, there is no consistent prodrome for the instability of a coronary plaque.

We can directly compare the general model trained on all events to the models targeting specific reasons for ICU transfer by comparing the values in the second row (ROC area of the model targeting all events combined) to the values in the diagonal from lower left to upper right (ROC areas of the individual models). The combined event model performs about as well as the individually trained model infection (0.83 vs 0.83), and respiratory instability (0.80 vs 0.76). These are the most common reasons for ICU transfer (Figure 1), accounting for the finding.

The results of using the NEWS score to detect the individual or combined events is shown in the top row. The highest ROC area for NEWS was for infection, 0.73. Use of this single score for the entire population of combined events yielded ROC area 0.67. Of models trained for individual classes of ICU transfer, whose ROC areas appear in the diagonal, only that for unstable coronary symptoms was lower (0.66).

The mean NEWS ROC area was 0.63, significantly lower than the mean ROC area of 0.72 of the general model trained on combined events ($p<0.01$, paired t-test), which was lower than the mean ROC area of 0.78 of the specifically targeted models ($p<0.05$).

Discussion

We studied statistical models for early detection of subacute potentially catastrophic illnesses leading to ICU transfer from a cardiac and cardiac surgery ward. This is a high-profile clinical event – the mortality rose by more than 40-fold in those who deteriorated³⁹. We targeted specific reasons for clinical deterioration, used predictors that were continuous rather than thresholded, and included mathematical analyses of continuous cardiorespiratory monitoring data. Our major findings are that phenotypes vary among the myriad of clinical conditions that lead to ICU transfer, and that statistical models trained specifically performed better than a model trained on all events combined, which in turn performed better than the untrained NEWS score.

We have used multivariable logistic regression models to test hypotheses about the difference in characteristics of endotypes of patients that deteriorated and required escalation to ICU care. The major hypothesis is that models trained on one endotype of patient do not necessarily perform well in detecting other kinds of patients. This directly tests the prevailing practice of using a single predictive model throughout the hospital. It is important to note that we here use the statistical models only for hypothesis testing and not for their clinical impact. We have not tested these models in a clinical setting, and a reasonable question is how to implement so many. We favor the idea of calculating all the model results for all the patients all the time, and presenting the clinician with a summary statistic. We tested ways to combine model results in the setting of early detection of sepsis in the neonatal ICU, where we found that the most pessimistic model was the most accurate. Other ways such as the ensemble mean or median might be equally or more useful.

It is a natural exercise of clinicians to synthesize disparate data elements into a clinical picture of the patient. After 1981 when Knaus, Wagner and colleagues introduced the APACHE (Acute Physiology and Chronic Health Evaluation) score, more methods have been introduced that combine elements of clinical data to yield quantitative estimates of patients' health status and risk.⁴⁸ While varied in inputs, targets, patient populations, and mathematical tools, most are single scores meant for universal use.³⁴ This is not altogether unreasonable. For example, we found that large abrupt spikes in risk estimation using a single model to identify patients at risk for ICU transfer had a positive predictive value 25% for imminent acute adverse event.⁴⁰ But use of only a single model is a limited approach, as there are many paths of deterioration – accurate capture of all of these paths at the same time is challenging.

The need to look for multiple modes of critical illness was recently supported by the finding of four sepsis phenotypes of responses and outcomes.⁴⁹ We found different sepsis signatures in the medical as opposed to the surgical ICU.³⁸ Thus, no single predictive model for sepsis seems adequate. Likewise, there are distinct respiratory deterioration phenotypes within COPD and acute respiratory distress syndrome.^{50,51} This heterogeneity of manifestations of illness, a point that resonates with bedside clinicians, argues against models that use the same thresholds for all situations. (An exception is sepsis in premature infants, where heart rate characteristics index monitoring identifies an abnormal phenotype that is common to many acute neonatal illnesses,⁵² and in a large RCT, was found to save lives.⁵³)

We emphasize the importance of continuous cardiorespiratory monitoring in early detection of subacute potentially catastrophic illnesses. We found distinct physiological signatures for sepsis, hemorrhage leading to large unplanned transfusion and respiratory failure leading to urgent unplanned intubation in the adult Medical and Surgical ICUs. Display of risk estimates using two of the resulting statistical models based solely on cardiorespiratory monitoring³⁸ was associated with a 50% reduction in the rate of septic shock.⁵⁴ Moreover, cardiorespiratory monitoring data add to vital signs and lab tests in the population reported here.³⁹ We propose that no scheme that omits cardiorespiratory monitoring data will perform as well as those that include it.

A strength of this study is the individual review of charts to identify reasons for ICU transfer. Clinicians recognize that illness presentations are complex, nuanced, and not well-documented.⁵⁵ Statistical models trained for one kind of deterioration had poor performance in detecting other kinds. As examples, bleeding – a common form of deterioration on our mixed medical-surgical cardiac ward – was well-detected by a model trained specifically for it, but no other model had reasonable performance; hemoglobin was, logically, a powerful prediction of ICU transfer for bleeding but not at all useful when myocardial ischemia was the problem;

coefficient of sample entropy, a detector of atrial fibrillation, was useful only for identifying patients transferred to the ICU for arrhythmia but nothing else.

There are other approaches. For example, Redfern and colleagues reported an early warning system for ICU transfer based on vital signs alone.¹⁰ Their approach differed from ours in several key ways. First, there was no clinical review of the reasons that patients deteriorated and required escalation to ICU care. We found that fully 50% of ICU transfers were not due to clinical deterioration³⁸. Addition of a large number of non-relevant cases negatively impacts the relevance of predictive models that result by inflating confidence intervals. Second, they studied only vital signs and lab tests. They also reported the large proportion of late or missed vital sign measurements – a median of 44%, presumably from the same time period and from one of the study hospitals. Third, bright cut-offs were used, thus a patient who, on successive measurements, had an increase in respiratory rate from 12 to 20 breaths per minute, a decrease in O₂ saturation from 100% to 96%, an increase in temperature from 36.2 to 38.0, a fall in BP from 218 to 112 mmHg, an increase in heart rate from 51 to 90 beats per minute – all clinically worrisome when simultaneous – has an unchanging NEWS score of 0. Finally, they made no distinction among the clinical reasons for transfer.

Here, we show large differences in signatures of illness and performance of statistical models in early identification of patients at risk of ICU transfer. The approach of Redfern and colleagues is exemplary of most efforts in clinical surveillance for early detection of subacute potentially catastrophic illnesses. We propose more precise methods: clinician review of events so that predictive models are more focused, addition of continuous cardiorespiratory monitoring when it is available to fill in the timeline between nurse visits and blood draws, and the use of predictive models that individually target the diagnoses that lead to ICU transfer in our patients.

Our results argue against a one-size-fits-all approach, no matter how Big the Data nor how Deep the Learning. In this era of increasingly precise and personalized medicine, we propose *precision predictive analytics monitoring* as a more focused and clinically informed approach to early detection of subacute potentially catastrophic illnesses in the hospitalized patient.

Bibliography

1. Escobar GJ, Gardner MN, Greene JD, Draper D, Kipnis P. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Med Care*. 2013;51(5):446-453. doi:10.1097/MLR.0b013e3182881c8e
2. Rosenberg AL, Hofer TP, Hayward RA, Strachan C, Watts CM. Who bounces back? Physiologic and other predictors of intensive care unit readmission. *Crit Care Med*. 2001;29(3):511-518.
3. Delgado MK, Liu V, Pines JM, Kipnis P, Gardner MN, Escobar GJ. Risk factors for unplanned transfer to intensive care within 24 hours of admission from the emergency department in an integrated healthcare system. *J Hosp Med*. 2013;8(1):13-19. doi:10.1002/jhm.1979
4. Escobar GJ, Greene JD, Gardner MN, Marelich GP, Quick B, Kipnis P. Intra-hospital transfers to a higher level of care: contribution to total hospital and intensive care unit (ICU) mortality and length of stay (LOS). *J Hosp Med*. 2011;6(2):74-80. doi:10.1002/jhm.817
5. O'Callaghan DJP, Jayia P, Vaughan-Huxley E, et al. An observational study to determine the effect of delayed admission to the intensive care unit on patient outcome. *Crit Care*. 2012;16(5):R173. doi:10.1186/cc11650
6. Reese J, Deakynne SJ, Blanchard A, Bajaj L. Rate of preventable early unplanned intensive care unit transfer for direct admissions and emergency department admissions. *Hosp Pediatr*. 2015;5(1):27-34. doi:10.1542/hpeds.2013-0102
7. Ginestra JC, Giannini HM, Schweickert WD, et al. Clinician Perception of a Machine Learning-Based Early Warning System Designed to Predict Severe Sepsis and Septic Shock. *Crit Care Med*. 2019;47(11):1477-1484. doi:10.1097/CCM.0000000000003803
8. van Galen LS, Struik PW, Driesen BEJM, et al. Delayed recognition of deterioration of patients in general wards is mostly caused by human related monitoring failures: A root cause analysis of unplanned ICU admissions. *PLoS One*. 2016;11(8):e0161393. doi:10.1371/journal.pone.0161393
9. Redfern OC, Griffiths P, Maruotti A, Recio Saucedo A, Smith GB, Missed Care Study Group. The association between nurse staffing levels and the timeliness of vital signs monitoring: a retrospective observational study in the UK. *BMJ Open*. 2019;9(9):e032157. doi:10.1136/bmjopen-2019-032157
10. Redfern OC, Pimentel MAF, Prytherch D, et al. Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: Development and validation of a multivariable model. *Resuscitation*. 2018;133:75-81. doi:10.1016/j.resuscitation.2018.09.021
11. The Royal College of Physicians. The National Early Warning Score (NEWS) thresholds and triggers. 2012:1.
12. Smith MEB, Chiovaro JC, O'Neil M, et al. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. *Annals of the American Thoracic Society*. 2014;11(9):1454-1465. doi:10.1513/AnnalsATS.201403-102OC
13. Churpek MM, Wendlandt B, Zdravetz FJ, Adhikari R, Winslow C, Edelson DP. Association between intensive care unit transfer delay and hospital mortality: A multicenter investigation. *J Hosp Med*. 2016;11(11):757-762. doi:10.1002/jhm.2630
14. Churpek MM, Yuen TC, Park SY, Meltzer DO, Hall JB, Edelson DP. Derivation of a cardiac arrest prediction model using ward vital signs*. *Crit Care Med*. 2012;40(7):2102-2108. doi:10.1097/CCM.0b013e318250aa5a

15. Rothman MJ, Rothman SI, Beals J. Development and validation of a continuous measure of patient condition using the Electronic Medical Record. *J Biomed Inform.* 2013;46(5):837-848. doi:10.1016/j.jbi.2013.06.011
16. Bittman J, Nijjar AP, Tam P, Khan N. Early warning scores to predict noncritical events overnight in hospitalized medical patients: A prospective case cohort study. *J Patient Saf.* September 2017. doi:10.1097/PTS.0000000000000292
17. Alam N, Hobbelink EL, van Tienhoven AJ, van de Ven PM, Jansma EP, Nanayakkara PWB. The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. *Resuscitation.* 2014;85(5):587-594. doi:10.1016/j.resuscitation.2014.01.013
18. Bartkowiak B, Snyder AM, Benjamin A, et al. Validating the Electronic Cardiac Arrest Risk Triage (eCART) Score for Risk Stratification of Surgical Inpatients in the Postoperative Setting. *Ann Surg.* 2018;XX(Xx):1.
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. doi:10.1016/j.resuscitation.2017.10.028
20. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting Intensive Care Unit Readmission with Machine Learning Using Electronic Health Record Data. *Annals of the American Thoracic Society.* 2018;15(7):846-853. doi:10.1513/AnnalsATS.201710-787OC
21. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* 2006;25(1):127-141. doi:10.1002/sim.2331
22. Wynants L, van Smeden M, McLernon DM, Steyerberg EW, Van Calster B. Three myths about risk thresholds for prediction models. *BMC Med.* 2019;17(192).
23. Cohen J, criti- F. The Cost of Dichotomization . *Applied Psychological Measurement.* 1983;7(3):249-253.
24. Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. *Pharm Stat.* 2009;8(1):50-61. doi:10.1002/pst.331
25. Brangan E, Banks J, Brant H, Pullyblank A, Le Roux H, Redwood S. Using the National Early Warning Score (NEWS) outside acute hospital settings: a qualitative study of staff experiences in the West of England. *BMJ Open.* 2018;8(10):e022528. doi:10.1136/bmjopen-2018-022528
26. Mackintosh N, Rainey H, Sandall J. Understanding how rapid response systems may improve safety for the acutely ill patient: learning from the frontline. *BMJ Qual Saf.* 2012;21(2):135-144. doi:10.1136/bmjqs-2011-000147
27. Elliott D, Allen E, Perry L, et al. Clinical user experiences of observation and response charts: focus group findings of using a new format chart incorporating a track and trigger system. *BMJ Qual Saf.* 2015;24(1):65-75. doi:10.1136/bmjqs-2013-002777
28. Downey CL, Tahir W, Randell R, Brown JM, Jayne DG. Strengths and limitations of early warning scores: A systematic review and narrative synthesis. *Int J Nurs Stud.* 2017;76:106-119. doi:10.1016/j.ijnurstu.2017.09.003
29. Mohammed MA, Rudge G, Watson D, et al. Index blood tests and national early warning scores within 24 hours of emergency admission can predict the risk of in-hospital mortality: a model development and validation study. *PLoS One.* 2013;8(5):e64340. doi:10.1371/journal.pone.0064340

30. Jarvis SW, Kovacs C, Badriyah T, et al. Development and validation of a decision tree early warning score based on routine laboratory test results for the discrimination of hospital mortality in emergency medical admissions. *Resuscitation*. 2013;84(11):1494-1499. doi:10.1016/j.resuscitation.2013.05.018
31. Nickel CH, Kellett J, Cooksley T, Bingisser R, Henriksen DP, Brabrand M. Combined use of the National Early Warning Score and D-dimer levels to predict 30-day and 365-day mortality in medical patients. *Resuscitation*. 2016;106:49-52. doi:10.1016/j.resuscitation.2016.06.012
32. Jo S, Jeong T, Lee JB, Jin Y, Yoon J, Park B. Validation of modified early warning score using serum lactate level in community-acquired pneumonia patients. The National Early Warning Score-Lactate score. *Am J Emerg Med*. 2016;34(3):536-541. doi:10.1016/j.ajem.2015.12.067
33. Moorman JR, Delos JB, Flower AA, et al. Cardiovascular oscillations at the bedside: early diagnosis of neonatal sepsis using heart rate characteristics monitoring. *Physiol Meas*. 2011;32(11):1821-1832. doi:10.1088/0967-3334/32/11/S08
34. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI. Review and performance evaluation of aggregate weighted “track and trigger” systems. *Resuscitation*. 2008;77(2):170-179. doi:10.1016/j.resuscitation.2007.12.004
35. Bapojé SR, Gaudiani JL, Narayanan V, Albert RK. Unplanned transfers to a medical intensive care unit: causes and relationship to preventable errors in care. *J Hosp Med*. 2011;6(2):68-72. doi:10.1002/jhm.812
36. Cohen RI, Eichorn A, Motschwiller C, et al. Medical intensive care unit consults occurring within 48 hours of admission: a prospective study. *J Crit Care*. 2015;30(2):363-368. doi:10.1016/j.jcrc.2014.11.001
37. Hillman KM, Bristow PJ, Chey T, et al. Duration of life-threatening antecedents prior to intensive care admission. *Intensive Care Med*. 2002;28(11):1629-1634. doi:10.1007/s00134-002-1496-y
38. Moss TJ, Lake DE, Calland JF, et al. Signatures of subacute potentially catastrophic illness in the ICU: model development and validation. *Crit Care Med*. 2016;44(9):1639-1648. doi:10.1097/CCM.0000000000001738
39. Moss TJ, Clark MT, Calland JF, et al. Cardiorespiratory dynamics measured from continuous ECG monitoring improves detection of deterioration in acute care patients: A retrospective cohort study. *PLoS One*. 2017;12(8):e0181448. doi:10.1371/journal.pone.0181448
40. Keim-Malpass J, Clark MT, Lake DE, Moorman JR. Towards development of alert thresholds for clinical deterioration using continuous predictive analytics monitoring. *J Clin Monit Comput*. July 2019. doi:10.1007/s10877-019-00361-5
41. Le Guen MP, Tobin AE, Reid D. Intensive care unit admission in patients following rapid response team activation: call factors, patient characteristics and hospital outcomes. *Anaesth Intensive Care*. 2015;43(2):211-215. doi:10.1177/0310057X1504300211
42. Dahn CM, Manasco AT, Breaud AH, et al. A critical analysis of unplanned ICU transfer within 48 hours from ED admission as a quality measure. *Am J Emerg Med*. 2016;34(8):1505-1510. doi:10.1016/j.ajem.2016.05.009
43. Reyna MA, Josef C, Jeter R, et al. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. *Crit Care Med*. 2019.
44. Pepe MS, Feng Z, Huang Y, et al. Integrating the predictiveness of a marker with its

- performance as a classifier. *Am J Epidemiol.* 2008;167(3):362-368.
doi:10.1093/aje/kwm305
45. Collins GS, Reitsma JB, Altman DG, Moons KGM, TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. *Circulation.* 2015;131(2):211-219.
doi:10.1161/CIRCULATIONAHA.114.014508
46. Lex A, Gehlenborg N, Strobel H, Vuilleumot R, Pfister H. Upset: visualization of intersecting sets. *IEEE Trans Vis Comput Graph.* 2014;20(12):1983-1992.
doi:10.1109/TVCG.2014.2346248
47. Moody GB, Mark RG, Zoccola A, Mantero S. Derivation of respiratory signals from multi-lead ECGs. *Computers in Cardiology.* 1985;12(1985):113-116.
48. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med.* 1981;9(8):591-597.
49. Seymour CW, Kennedy JN, Wang S, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA.* 2019;321(20):2003-2017.
doi:10.1001/jama.2019.5791
50. Lawler PR, Fan E. Heterogeneity and phenotypic stratification in acute respiratory distress syndrome. *Lancet Respir Med.* 2018;6(9):651-653. doi:10.1016/S2213-2600(18)30287-X
51. Augustin IML, Spruit MA, Houben-Wilke S, et al. The respiratory physiome: Clustering based on a comprehensive lung function assessment in patients with COPD. *PLoS One.* 2018;13(9):e0201593. doi:10.1371/journal.pone.0201593
52. Moss TJ, Lake DE, Moorman JR. Local dynamics of heart rate: detection and prognostic implications. *Physiol Meas.* 2014;35(10):1929-1942. doi:10.1088/0967-3334/35/10/1929
53. Moorman JR, Carlo WA, Kattwinkel J, et al. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. *J Pediatr.* 2011;159(6):900-6.e1. doi:10.1016/j.jpeds.2011.06.044
54. Ruminski CM, Clark MT, Lake DE, et al. Impact of predictive analytics based on continuous cardiorespiratory monitoring in a surgical and trauma intensive care unit. *J Clin Monit Comput.* 2019;33(4):703-711. doi:10.1007/s10877-018-0194-4
55. Despins LA. Automated deterioration detection using electronic medical record data in intensive care unit patients: A systematic review. *Comput Inform Nurs.* 2018;36(7):323-330. doi:10.1097/CIN.0000000000000430
56. Lake DE, Moorman JR. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *Am J Physiol Heart Circ Physiol.* 2011;300(1):H319-25. doi:10.1152/ajpheart.00561.2010
57. Carrara M, Carozzi L, Moss TJ, et al. Classification of cardiac rhythm using heart rate dynamical measures: validation in MIT-BIH databases. *J Electrocardiol.* 2015;48(6):943-946. doi:10.1016/j.jelectrocard.2015.08.002
58. Carrara M, Carozzi L, Moss TJ, et al. Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy. *Physiol Meas.* 2015;36(9):1873-1888. doi:10.1088/0967-3334/36/9/1873

Figure 1. Reasons for ICU transfer. The horizontal bars show the number of patients in each category of the reason for ICU transfer; the vertical bars show the numbers of patients arranged by the presence of one or more reasons. The classifications were respiratory instability (increase in oxygen requirement, oxygen desaturation, respiratory distress or increased work of breathing); infection workup (culture taken and antibiotics started or broadened, proven infection that could have reasonably contributed to the deterioration); heart failure exacerbation (pulmonary edema, persistent hypotension, poor diuresis, severe pulmonary hypertension); arrhythmia (ventricular or supraventricular tachyarrhythmia, bradyarrhythmia); unstable coronary artery disease (unstable angina, acute myocardial infarction with or without ST segment elevation); neurological event (stroke, seizure, or altered mental status); and bleeding (transfusion requirement, hematoma, post-surgical bleed, GI bleed). Event classifications were not mutually exclusive, and many patients met more than one classification.

Figure 2. Heat maps that relate the relative risk (color bar; 1 means the risk is average for the ward) to the reason for ICU transfer (rows) and measured value of a predictor (abscissa) for a representative laboratory value (A, serum Na in meq/dl), vital sign (B, mean arterial pressure, MAP in mm Hg), and cardiorespiratory measure from time series analysis of continuous cardiorespiratory monitoring (C, breaths per minute derived from the EKG). The gray-scale bars show the density of the measurements.

The Figure may be interpreted as follows. The change in color from deep red to blue from left to right in the fourth row of panel A signifies that low serum Na concentration was associated with a high risk for ICU transfer for escalation of heart failure (HF) therapy.

Figure 3. Heat maps that relate the relative risk of the reason for ICU transfer to the measured values of predictor laboratory values.

Figure 4. Heat maps that relate the relative risk of the reason for ICU transfer to the nurse-entered values of predictor vital signs. HR heart rate, RR respiratory rate, SpO2 oxygen saturation, MAP non-invasive mean arterial pressure.

Figure 5. Heat maps that relate the relative risk of the reason for ICU transfer to measured values of predictors derived from cardiorespiratory monitoring. HRxEDR cross-correlation of heart rate and breathing rate derived from EKG; S.D. standard deviation, COSEn coefficient of sample entropy 56,57, DFA detrended fluctuation analysis 58, LDd local density score 39.

Figure 6. Receiver operating characteristic curve areas for models trained for the events listed in rows, and tested for events arranged as columns. The top row is the performance of the National Early Warning Score (NEWS); the next-to-top row and the final column represent results of a model trained on all the ICU transfer events combined. The gray scale reflects the value of the ROC area.