Published in final edited form as:

Crit Care Med. 2019 November ; 47(11): 1485–1492. doi:10.1097/CCM.0000000000003891.

A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Development, Implementation and Impact on Clinical Practice

Heather M. Giannini, MD[1], **Jennifer C. Ginestra, MD**[1], **Corey Chivers, PhD**[2], **Michael Draugelis**[2], **Asaf Hanish, MPH**[2], **William D. Schweickert, MD**[2,3], **Barry D. Fuchs, MD, MS**[2,3], **Laurie Meadows, RN, CCRN**[1], **Michael Lynch, RN, CEN**[1], **Patrick J. Donnelly, RN, MS, CCRN**[4], **Kimberly Pavan, MSN, CRNP**[5], **Neil O. Fishman, MD**[2,3], **C. William Hanson III, MD**[2,3], **Craig A. Umscheid, MD, MSCE**[2,3]

[1]Hospital of the University of Pennsylvania, Philadelphia, PA

[2]University of Pennsylvania Health System, Philadelphia, PA

[3]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

[4]Pennsylania Hospital, Philadelphia, PA

[5]Penn Presbyterian Medical Center, Philadelphia, PA

Abstract

OBJECTIVE—Develop and implement a machine learning algorithm to predict severe sepsis and septic shock and evaluate the impact on clinical practice and patient outcomes

DESIGN—Retrospective cohort for algorithm derivation and validation, pre-post impact evaluation

SETTING—Tertiary teaching hospital system in Philadelphia, PA

PATIENTS—All non-ICU admissions; algorithm derivation July 2011-June 2014 (n= 162,212); algorithm validation October-December 2015 (n=10,448); silent versus alert comparison January 2016-February 2017 (silent n= 22,280; alert n= 32,184).

INTERVENTIONS—A random-forest classifier, derived and validated using electronic health record data, was deployed both silently and later with an alert to notify clinical teams of sepsis prediction.

Correspondence and requests for reprints should be addressed to: Craig A. Umscheid, M.D., University of Chicago Medicine, American School Building, 850 E. 58th Street, Suite 123, Office 128, MC 1135, Chicago, IL 60637. craigumscheid@medicine.bsd.uchicago.edu.

Author Contributions:

Conception and design: CAU, BDF, WDS, CC, MD, LM, ML, PD, KP, HMG, JCG, NOF, CWH

Data collection: CC, MD, LM, HMG, JCG

Analysis and interpretation of data: HMG, JCG, CAU, CC, MD, NOF, CWH

Drafting of manuscript: HMG, JCG, CAU

Critical revision of article for important intellectual content: HMG, JCG, CAU, NOF, CWH

The remaining authors have disclosed that they do not have any potential conflicts of interest.

MEASUREMENT and MAIN RESULT—Patients identified for training the algorithm were required to have ICD9 codes for severe sepsis or septic shock and a positive blood culture during their hospital encounter with either a lactate > 2.2 mmol/L or a systolic blood pressure < 90 mm Hg. The algorithm demonstrated a sensitivity of 26% and specificity of 98%, with a positive predictive value of 29% and positive likelihood ratio of 13. The alert resulted in a small statistically significant increase in lactate testing and intravenous fluid administration. There was no significant difference in mortality, discharge disposition, or transfer to ICU, although there was a reduction in time-to-ICU transfer.

CONCLUSIONS—Our machine learning algorithm can predict, with low sensitivity but high specificity, the impending occurrence of severe sepsis and septic shock. Algorithm-generated predictive alerts modestly impacted clinical measures. Next steps include describing clinical perception of this tool, and optimizing algorithm design and delivery.

Keywords

severe sepsis; septic shock; electronic medical record; predictive medicine; machine learning; early warning system

INTRODUCTION

Sepsis continues to be a leading cause of death among hospitalized patients, affecting up to 6% of all admissions and conferring in-hospital mortality greater than 15% (1, 2). Early detection of sepsis has the potential to reduce mortality by facilitating timely implementation of evidence-based interventions (3).

Many studies have used multivariate models based on electronic health record (EHR) data for detection of sepsis or clinical deterioration (4–8). Our team previously developed and implemented one such detection algorithm based on the systemic inflammatory response syndrome (SIRS) (7). In that study, a non-significant improvement in mortality and an increase in discharge to home was observed. More recently, our team as well as others have begun to use machine learning (ML) approaches (9) to improve the accuracy of sepsis detection and prediction, both in the emergency department (10–12) and the inpatient setting (13–15).

When applied to retrospective data, ML algorithms designed to predict sepsis have performed well (10, 14, 16, 17, 19). However, implementation has largely been focused on intensive care unit (ICU) populations, where robust staffing and a high index of suspicion already prompt early recognition of sepsis. For example, ML algorithms have been linked to decreased mortality and length of stay in a small ICU-based randomized trial (16) and decreased sepsis-related mortality at a small private hospital (18). However, to date, the large-scale application of ML algorithms to predict sepsis in the non-ICU inpatient setting has not been reported. Here, we describe the development of a machine learning algorithm and alert for prediction of severe sepsis and septic shock in hospitalized non-ICU patients, and the subsequent clinical impact of this tool when implemented across our multi-hospital healthcare system.

METHODS

Setting and Data Sources

At the time of the study, the University of Pennsylvania Health System (UPHS) included three urban acute care hospitals with a capacity of over 1,500 beds and 70,000 annual admissions. All hospitals used the EHR Sunrise Clinical Manager version 5.5 (Allscripts, Chicago, Illinois). Data were retrieved from the Penn Data Store, which includes clinical data from our EHR and administrative data coded by clinical documentation specialists.

Algorithm Derivation and Validation

A cohort of all inpatients discharged between July 2011-June 2014 (n=162,212) from our three hospitals was used to train our algorithm. From this population, a total of 950 inpatient encounters met "Sepsis Training Criteria," which required: 1) ICD9 codes 995.92 (severe sepsis) or 785.52 (septic shock), 2) a positive blood culture, and 3) a lactate > 2.2 mmol/L or a systolic blood pressure <90, all occurring within a 1 hour window. The time of earliest measured elevated lactate or hypotension was considered "sepsis onset" for training purposes. These cases were used to train a random-forest classifier to predict severe sepsis and septic shock. The random forest approach has been described previously and used in similar studies (11). Our model considered a total of 587 features, consisting of demographics, vital signs, and lab results. For selected labs and vitals, we also derived time-series features, describing the minimum, maximum, mean, and rate of change over the preceding 24 hours. We used 100 estimators (trees) and gini criteria for splits.

The resulting algorithm was retrospectively validated on hospitalized non-ICU patients from October 1 to December 1 2015 (n=10,448 discharges). A set point from the algorithm derivation AUC was selected to produce an average of 10 alerts per day across the three hospitals in our healthcare system. During the validation period, the algorithm identified 347 patients predicted to develop severe sepsis or septic shock. The outcome of "Severe Sepsis" was defined as having: 1) >2 SIRS criteria, 2) lactate >2.2, and 3) positive blood or urine culture. The outcome of "Septic Shock" was defined as having "Severe Sepsis" plus a systolic blood pressure <90 mm Hg. These variables had to collectively occur within a one-hour time window. We will refer to patients who trigger the algorithm prediction as "screen positive".

Area under the curve (AUC) with k-fold cross validation (k=10) was estimated using the derivation population. Test characteristics, including sensitivity, specificity, predictive values and likelihood ratios were estimated from the validation population. All model construction and analyses were conducted using the open source Python programming language and Scikit-learn v0.15.2 (http://ogrisel.github.io/scikit-learn.org/sklearn-tutorial/about.html#citing-scikit-learn).

Implementation of "Early Warning System 2.0"

As a successor to our prior SIRS-based sepsis detection tool, Early Warning System (EWS) 1.0 (7), we named this new ML algorithm-based sepsis prediction tool "EWS 2.0". After derivation and validation, EWS 2.0 was deployed in the production environment over a

14-month period. Patients eligible for algorithm screening included non-ICU inpatients who were in the hospital > 24 hours (which included any time spent in the ED). Patient data was resampled hourly, with new predictions made any time a new observation (data point) was recorded.

During an initial 6-month "silent period" (January 1 – June 15, 2016), process and outcome measures were collected on screen positive patients, but no accompanying alert was sent to the care team. For the subsequent 8-month "alert period" (June 16, 2016 – February 6, 2017), the algorithm was paired with an automated alert sent to the covering care team. Alerts stated that EWS 2.0 had fired for a given patient, and included relevant recent laboratory data along with 48 hours of vital sign trends. Nurses received EHR-based alerts. Text messages were sent to providers and a rapid response coordinator (a critical-care nursing professional who monitors and responds to hospital emergencies 24-hours daily). The team was asked to perform a bedside assessment of the patient, but no specific interventions were required.

Clinical characteristics of screen positive patients were compared to those of a random population of screen negative non-ICU inpatients during the alert period. Hourly data following algorithm trigger for screen positive patients was compared to hourly data from screen negative patients following a randomly selected time. Process and outcome measures were collected for screen positive patients until discharge from the hospital during both implementation periods. Because one of our three hospitals transitioned to a new EHR during the intervention study period, it was excluded from implementation analysis; data from the two other hospitals in our system (including our flagship teaching hospital) was used for silent and alert period analyses.

To assess the alert's impact on care, we estimated proportions with confidence intervals, means with standard deviations, and medians with interquartile ranges for descriptive characteristics, process measures, and clinical outcomes in the silent and alerted periods. Unadjusted analyses using the chi-square test for dichotomous variables and the Wilcoxon rank sum test for continuous variables compared demographics and process and outcome measures in all study populations, including training, validation, silent and alert periods. P-values <0.05 were considered significant.

To better assess the impact of the alert when the care team suspected sepsis, we also performed analyses stratified by "Suspected" versus "Unsuspected Sepsis". "Suspected Sepsis" was defined by active orders for at least two of the following within 12 hours prior to the alert: broad-spectrum antibiotics, blood cultures, and/or lactate testing.

Institutional Review

This study received expedited approval, HIPAA waiver, and informed consent waiver from the University of Pennsylvania Institutional Review Board (protocol number 826028).

RESULTS

Algorithm Derivation and Validation

Demographics of the derivation, validation, and implementation period populations were clinically similar (Supplemental Table 1, Table 2). During algorithm derivation, the estimated AUC for the study outcomes of Severe Sepsis or Septic Shock was 0.88 (SD $^{+}/_{-}$ 0.03) following k-fold validation (k=10). Test characteristics estimated with the validation cohort demonstrated sensitivity of 26% and specificity of 98%. Positive and negative predictive values were 29% and 97%, respectively. Positive and negative likelihood ratios were 13 and 0.75, respectively. The clinical variables with the greatest contribution to the algorithm predictions are shown in Table 1 (see Supplemental Table 2 for a full list of included variables). SIRS criteria and markers of end-organ dysfunction contributed most to the prediction, consistent with recent sepsis consensus guidelines and definitions (20).

Algorithm and Alert Implementation

Clinical Characteristics of Screen Positive Patients—Demographics of the total study population in the silent and alert periods were clinically similar, as were the characteristics of screen positive patients from each group (Table 2). EWS 2.0 triggered for 7.4% of admissions (n=1,540) during the silent period and 7.1% of admissions (n=2,137) during the alert period. During the silent period, the tool triggered a median of 6 hours and 34 minutes (IQR: 0hrs:50min-53hrs:19min) prior to the onset of severe sepsis or septic shock. This was similar to the alert period (median of 5 hours and 25 minutes (IQR: 0hrs:45min-45hrs:0min). Almost 60% of screen positive patients met two of four SIRS criteria at the time of alert, increasing to 84% by 48-hours post-alert (Figure 1). Only 11% of patients met study criteria for the outcomes of severe sepsis or septic shock at time of alert. By 48-hours after the alert, 30% of screen positive patients met study criteria for the outcomes of severe sepsis or septic shock. Screen positive patients demonstrated marked abnormalities in vital signs and laboratory data compared to those who did not trigger EWS 2.0 (Supplemental Figure 1).

Process Measures—The alert prompted a modest but statistically significant increase in lactate testing, administration of IV fluid boluses, and CBC or BMP testing within three hours following the alert (Table 3). Increases in lactate testing and IV fluid bolus administration were sustained at six hours post-alert, but only lactate testing remained significantly increased at the 48-hour mark (Supplemental Table 3). Transfusion of packed red blood cells was also significantly increased in the first 6 hours post-alert (Supplemental Table 3). Frequency of blood cultures or initiation of antibiotics did not significantly differ between the silent and alert periods (Supplemental Table 3). Time to administration of broad-spectrum sepsis antibiotics also did not differ significantly (silent period: median 11hrs:12min, IQR 2hrs:27min-36hrs:34min; alert period: median 9hrs:47min, IQR 2hrs:37min-35hrs:39min; p=0.59).

In the alert period, 27% of screen positive patients met criteria for Suspected Sepsis. Post-alert increases in lactate testing were significant in both the Suspected and Unsuspected Sepsis groups (Supplemental Tables 4 and 5). Increases in IV fluid bolus administration,

telemetry and lab testing were primarily observed in Unsuspected Sepsis (Supplemental Table 5). Antibiotic initiation did not significantly differ for either group.

Outcome Measures—Compared to screen positive patients during the silent period, screen positive patients during the alert period had a statistically significant decrease in time to ICU transfer, but no significant change in the frequency of ICU transfer or median length of stay in the ICU (Table 4). There were also no statistically significant differences in the development of severe sepsis or septic shock, all-cause mortality, or discharge disposition.

The observed decrease in median time-to-ICU-transfer among patients in the alert group was primarily driven by the Unsuspected Sepsis cohort (24 (IQR 3–117) hours vs. 8 (IQR 2–73) hours, $p<0.01$) (Supplemental Tables 6 and 7). There was no significant change observed for the Suspected Sepsis cohort (Supplemental Table 6). Neither cohort had post-alert changes in frequency of ICU transfer, median length of stay in ICU, or mortality; however, we did observe increased frequency of discharge to inpatient hospice among patients with Suspected Sepsis at the time of the alert (3.0% vs. 5.9%, $p = 0.04$) (Supplemental Table 6).

DISCUSSION

We developed a machine learning algorithm to predict severe sepsis and septic shock and implemented the tool on non-ICU services across our multi-hospital healthcare system. Here, we confirmed the feasibility of widespread implementation of a machine learning predictive alert, but observed a limited impact on clinical practice and outcomes.

Algorithm Design

To train our algorithm, we sought to identify patients with unequivocal sepsis physiology. Our selected Sepsis Training Criteria included hypotension and lactic acidosis as markers of impaired perfusion and shock (20) and a positive blood culture as a specific marker of infection. Though recent sepsis definitions do not include bacteremia, and in fact up to 50% of sepsis cases have no confirmed source of infection, we used narrower criteria to improve the specificity and predictive value of our resulting algorithm. SIRS criteria and clinical data related to end-organ dysfunction were heavily weighted in the algorithm, thus supporting our approach to algorithm development. However, this study's results may be limited by our use of more specific sepsis definitions that have not been externally validated.

The resulting algorithm accurately identified hospitalized patients at risk for developing severe sepsis or septic shock, despite the inherent limitations of EHR data, which can be plagued by missingness, inaccuracies, and changes in practice patterns over time. Importantly, the sensitivity of the tool was limited to minimize alert fatigue given that hospital providers are estimated to receive greater than 50 EHR alerts on average per day (21), leading to providers declining, ignoring or deferring a majority of the alerts they encounter (22). Our lower sensitivity resulted in higher specificity and an excellent positive likelihood ratio.

Alert Impact

Ultimately, EWS 2.0 did not significantly improve our main outcome measures. We hypothesize that the alert's impact on clinical processes and patient outcomes was limited by multiple factors, including a lack of pre-specified interventions, limited alert format, long alert lead-times, and perhaps most importantly, minimal predictive value beyond predictions already made by the clinical teams.

Clinician Response

Despite good predictive values, in many cases, the post-alert bedside evaluation resulted in minimal changes to clinical care or outcomes. Ambiguity may have arisen about how to manage patients in the setting of a positive screen but apparent clinical stability. Prior to apparent disease, the utility of further laboratory testing, fluid administration and empiric antibiotics is unclear. This may contribute to the low level of practice change following the alert. In addition, our study is limited in that we did not evaluate whether observed practice changes were appropriate for the clinical context of each patient.

We previously reported that with our prior alert system, EWS 1.0, clinical teams already strongly suspected sepsis in >50% of cases (23). A survey administered to alerted providers and nurses during the alert period of this current study revealed a similar sentiment (24). However, in the subgroup analysis of patients who were not suspected of having sepsis at the time of alert, there was a statistically significant decrease in time to ICU transfer. On the other hand, there was an increase in referral to inpatient hospice for patients already suspected of sepsis. Thus, it appears that the algorithm may have provided information that supported either escalating care for those not suspected of having sepsis, or adjusting overall goals of care for those initially suspected of having sepsis.

Intervention Format

The format of our intervention, as a one-time alert, may have affected the alert's impact on clinical care and outcomes. There may be multiple critical opportunities for clinical teams to integrate clinical information with a sepsis risk assessment. Yet, we did not require re-evaluation of alerted patients at later time-points, even though in many cases our one-time alert triggered hours prior to the onset of sepsis physiology. For these cases, the lead-time of the alert and evaluation prior to clinically overt disease may have been too long. The question remains as to whether alerts are the most effective method of communicating real-time predictive information, or whether a continuous score may more dynamically support clinical decision-making. Moreover, while some clinical data were reported with the alerts, the variables and logic leading to alert trigger were not clearly delineated, creating what has been referred to as a machine learning "block box model" (25). This lack of transparency may have reduced overall trust in the algorithm and may have affected the clinician perception of the reliability of the prediction.

Future Algorithm Optimization

Considerations must also be made for optimizing algorithm design. Recent studies have shown that machine learning predictions in sepsis and critical care may be strengthened by incorporating free text from provider documentation using natural language processing (10,

26, 27). Importantly, we derived our algorithm using criteria that although guided by sepsis consensus guidelines, were defined by the study team, and not externally validated. While we prioritized specificity, a more sensitive algorithm may pick up subtle clinical trends for patients who are less likely to be captured by clinician's usual risk assessments (though at the risk of alert fatigue). Additionally, the most actionable moment in the course of a patient's sepsis trajectory may be the time just prior to, or during, the onset of clinical change. In this case, our alert frequently fired at a time when the patient appeared clinically well, sometimes many hours ahead of later decompensation. Finally, severe sepsis and septic shock may not be the most relevant outcomes to target when predicting unsuspected active clinical deterioration requiring a response from frontline providers. Algorithms trained for general decline, which may predict ICU transfer (17) or even mortality (28), might be more impactful with respect to changing process and outcome measures, and preventing these critical events.

CONCLUSIONS

This study demonstrates the feasibility of implementing a machine learning algorithm for real-time analysis of EHR data to accurately predict the development of severe sepsis or septic shock. We have also shown the potential implications of alerting clinicians to this prediction throughout a multi-hospital healthcare system. In this study, the alert did not significantly alter clinical practice or outcomes. Training the algorithm on more traditional definitions of clinical deterioration, enhancing ML algorithms through incorporation of natural language processing, and effectively communicating risk while avoiding alerts in patients already suspected of clinical deterioration, represent potential opportunities to improve the impact of sepsis prediction on clinical care outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors thank Mark E. Mikkelsen MD, and Joanne Resnic MBA, BSN, RN for their contributions to the design, testing, and implementation of the clinical decision support intervention examined in this manuscript.

Financial Support:

Dr. Umscheid's contribution to this project was supported in part by the National Center for Research Resources (grant no. UL1RR024134), which is now at the National Center for Advancing Translational Sciences (grant no. UL1TR000003). The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Copyright form disclosure: Dr. Giannini, Ginestra, Umscheid received support for article research from the National Institutes of Health. Dr. Schweickert received funding from Arjo, Hill Rom, the Society of Critical Care Medicine (consulting), and American College of Physicians. Dr. Umscheid's institution received funding from National Center for Research Resources (grant no. UL1RR024134), which is now at the National Center for Advancing Translational Sciences (grant no. UL1TR000003); AHRQ EPC Contracts; and FDA. He received funding from PCORI Advisory Panel and Northwell Health (honoraria for grand rounds).

REFERENCES

1. Rhee C, Dantes R, Epstein L, et al. : Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009–2014. JAMA 2017;
2. Torio CM, Andrews RM: National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011: Statistical Brief #160 [Internet]. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006. [cited 2017 Oct 3] Available from: http://www.ncbi.nlm.nih.gov/books/NBK169005/
3. Liu VX, Fielding-Singh V, Greene JD, et al. : The Timing of Early Antibiotics and Hospital Mortality in Sepsis. Am J Respir Crit Care Med 2017; 196:856–863 [PubMed: 28345952]
4. Escobar GJ, LaGuardia JC, Turk BJ, et al. : Early detection of impending physiologic deterioration among patients who are not in intensive care: development of predictive models using data from an automated electronic medical record. J Hosp Med 2012; 7:388–395 [PubMed: 22447632]
5. Bellomo R, Ackerman M, Bailey M, et al. : A controlled trial of electronic automated advisory vital signs monitoring in general hospital wards. Crit Care Med 2012; 40:2349–2361 [PubMed: 22809908]
6. Churpek MM, Yuen TC, Winslow C, et al. : Multicenter development and validation of a risk stratification tool for ward patients. Am J Respir Crit Care Med 2014; 190:649–655 [PubMed: 25089847]
7. Umscheid CA, Betesh J, VanZandbergen C, et al. : Development, implementation, and impact of an automated early warning and response system for sepsis. J Hosp Med 2015; 10:26–31 [PubMed: 25263548]
8. Khurana HS, Groves RH, Simons MP, et al. : Real-Time Automated Sampling of Electronic Medical Records Predicts Hospital Mortality. Am J Med 2016; 129:688–698.e2
9. Deo RC. Machine Learning in Medicine. Circulation 2015;132(20): 1920–1930. [PubMed: 26572668]
10. Horng S, Sontag DA, Halpern Y, et al. : Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PloS One 2017; 12:e0174708
11. Taylor RA, Pare JR, Venkatesh AK, et al. : Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. Acad Emerg Med Off J Soc Acad Emerg Med 2016; 23:269–278
12. Berger T, Birnbaum A, Bijur P, et al. : A Computerized Alert Screening For Severe Sepsis In Emergency Department Patients Increases Lactate Testing But Does Not Improve Inpatient Mortality. Appl Clin Inform 2010; 01:394–407
13. Churpek MM, Yuen TC, Winslow C, et al. : Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. Crit Care Med 2016; 44:368–374 [PubMed: 26771782]
14. Henry KE, Hager DN, Pronovost PJ, et al. : A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015; 7:299ra122
15. Hackmann G, Chen M, Chipara O, et al. : Toward a Two-Tier Clinical Warning System for Hospitalized Patients. AMIA Annu Symp Proc 2011; 2011:511–519 [PubMed: 22195105]
16. Shimabukuro DW, Barton CW, Feldman MD, et al. : Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial [Internet]. BMJ Open Respir Res 2017; 4Available from: http://bmjopenrespres.bmj.com/content/4/1/e000234.abstract
17. Wellner B, Grand J, Canzone E, et al. : Predicting Unplanned Transfers to the Intensive Care Unit: A Machine Learning Approach Leveraging Diverse Clinical Elements. JMIR Med Inform 2017; 5:e45 [PubMed: 29167089]
18. McCoy A, Das R: Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. BMJ Open Qual 2017; 6:e000158
19. Mao Q, Jay M, Hoffman JL, et al. : Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open 2018; 8:e017833

20. Singer M, Deutschman CS, Seymour CW, et al. : The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016; 315:801–810 [PubMed: 26903338]

21. Murphy DR, Reis B, Sittig DF, et al. : Notifications Received by Primary Care Practitioners in Electronic Health Records: A Taxonomy and Time Analysis. Am J Med 2012; 125:209.e1–209.e7

22. Ancker JS, Edwards A, Nosal S, et al. : Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak 2017; 17:36 [PubMed: 28395667]

23. Guidi JL, Clark K, Upton MT, et al. : Clinician Perception of the Effectiveness of an Automated Early Warning and Response System for Sepsis in an Academic Medical Center. Ann Am Thorac Soc 2015; 12:1514–1519 [PubMed: 26288388]

24. 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. (In review)

25. Cabitza F, Rasoini R, Gensini G: Unintended consequences of machine learning in medicine. JAMA 2017; 318:517–518 [PubMed: 28727867]

26. Weissman GE, Hubbard RA, Ungar LH, et al. : Inclusion of Unstructured Clinical Text Improves Early Prediction of Death or Prolonged ICU Stay. Crit Care Med 2018; 46:1125–1132 [PubMed: 29629986]

27. Marafino BJ, Park M, Davies JM, et al. : Validation of Prediction Models for Critical Care Outcomes Using Natural Language Processing of Electronic Health Record Data. JAMA Netw Open 2018; 1:e185097

28. Ryan DP, Daley BJ, Wong K, et al. : Prediction of ICU in-hospital mortality using a deep Boltzmann machine and dropout neural net [Internet]. IEEE; 2013. p. 1–4.[cited 2018 Apr 4] Available from: http://ieeexplore.ieee.org/document/6618491/

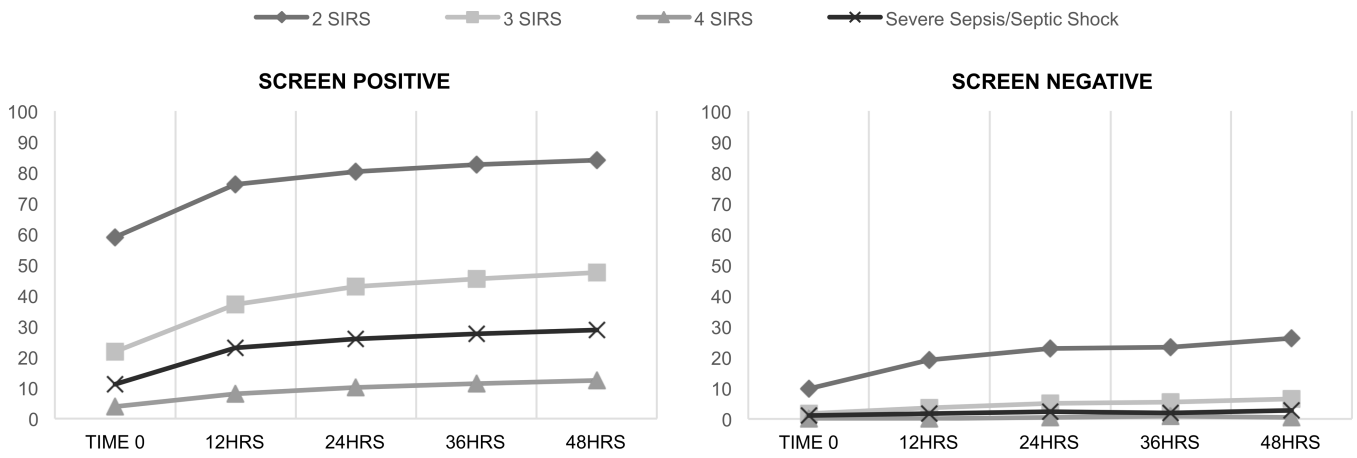

Figure 1:

Proportion of Screen Positive Patients Meeting Systemic Inflammatory Response Syndrome (SIRS) Criteria in the Hours Following Algorithm Detection, Compared with Controls

SIRS (Systemic Inflammatory Response Syndrome) Criteria include:

(1) Temp >38°C (100.4°F) or < 36°C (96.8°F)

(2) Heart rate > 90

(3) Respiratory rate > 20 or $PaCO_2$ < 32 mm Hg

(4) WBC > 12,000/mm^3, < 4,000/mm^3, or > 10% bands

Criteria for Severe Sepsis: >2 SIRS and positive blood or urine culture and lactate >2.2;

Septic Shock: Severe Sepsis AND systolic blood pressure <90 mm Hg

Table 1:

Top Twenty Variables Contributing to Algorithm Prediction and Corresponding Weight

Variable	Time Variation	Weight
BP Noninvasive Diastolic (mm Hg)	Most Recent	0.01624320
BP Noninvasive Systolic (mm Hg)	24hr Minimum	0.01606099
Pulmonary Service	N/A	0.01554681
Heart Rate (beats/min)	24hr Rate of Change	0.01455791
Blood Urea Nitrogen	Most Recent	0.01372632
BP Noninvasive Systolic (mm Hg)	24hr Variation from the Mean	0.01370169
Temperature	Most Recent	0.01358034
Temperature	24hr Maximum	0.01325225
% Monocytes	Most Recent	0.01315597
Temperature	24hr Variation	0.01266883
Blood Urea Nitrogen	24hr Mean	0.01264225
Heart Rate (beats/min)	Most Recent	0.01182879
Blood Urea Nitrogen	24hr Minimum	0.01165007
Blood Urea Nitrogen	24hr Maximum	0.01141574
Age	Most Recent	0.01108977
BP Noninvasive Diastolic (mm Hg)	24hr Minimum	0.01092703
Carbon Dioxide	Most Recent	0.01057971
Creatinine	Most Recent	0.01047186
Absolute Lymphocyte Count	Most Recent	0.01046288
Temperature (degrees F)	24hr Variation from the Mean	0.00959570

BP = blood pressure, F = Fahrenheit, mm Hg = millimeters of mercury, min = minute, hr = hour.

Table 2:

Demographics of Intervention Population

	Total Intervention Population			Screen Positive Population		
Demographic	**Silent Period (n=22,280)**	**Alert Period (n=32,184)**	**P-value**	**Silent Period (n=1,540)**	**Alert Period (n=2,137)**	**P-value**
Age, mean, yr	58.5	58.7	0.22	61.3	62.5	0.02
Female, %	48.7	49.0	0.60	47.7	47.4	0.88
BMI, mean	29.0	28.8	0.07	27.9	28.0	0.93
Race/Ethnicity, n (%)			<0.01			0.66
White	11,802 (53.0)	16,987 (52.8)		825 (53.6)	1,186 (55.5)	
Black	8,664 (38.9)	12,234 (38.0)		551 (35.8)	739 (34.6)	
Other	492 (2.2)	679 (2.1)		51 (3.3)	62 (2.9)	
Unknown	1,322 (5.9)	2,284 (7.1)		113 (7.3)	150 (7.0)	
Hospital, n (%)			0.03			0.67
HUP	13,990 (62.8)	19,918 (61.9)		1,189 (77.2)	1,636 (76.6)	
PPMC	8,290 (37.2)	12,266 (38.1)		351 (22.8)	501 (23.4)	
Admission Type, n (%)			<0.01			0.02
Elective	8,560 (38.4)	12,086 (37.6)		433 (28.1)	515 (24.1)	
Emergency	9,583 (43.0)	13,750 (42.7)		831 (54.0)	1,197 (56.0)	
Transfer	3,963 (17.8)	6,336 (19.7)		275 (17.9)	424 (19.8)	
Hosp LOS, median (IQR), d	4 (2–7)	4 (2–7)	<0.01	9 (5–18)	9 (5–18)	0.39
DRG Weight, median (IQR)	1.45 (0.97–2.23)	1.48 (0.97–2.20)	0.27	1.88 (1.35–4.23)	1.79 (1.38–3.49)	0.05

BMI = body mass index, d = days, DRG = diagnosis related group, HUP = Hospital of the University of Pennsylvania, IQR = interquartile range, Hosp LOS = hospital length of stay, PPMC = Penn Presbyterian Medical Center, yr = years.

Table 3:

Clinical Process Measures in Screen Positive Patients Within Three Hours of Alert

Process Measure, %	Silent (n=1,540)	Alert (n=2,137)	P-value
CBC or BMP	46.9	51.0	0.01
IV Fluid Bolus	21.7	25.5	<0.01
Any antibiotic	17.3	16.9	0.76
Sepsis Antibiotic(s)[a]	15.6	15.2	0.76
Blood Cultures	14.0	15.7	0.18
Telemetry or ECG	12.8	14.5	0.15
Chest Radiograph	9.4	10.0	0.62
Lactate	8.0	11.7	<0.01
CT Imaging[b]	5.3	4.6	0.38
RBC Transfusion	3.8	4.2	0.67
Diuretic	3.2	3.8	0.43
AV Nodal Blockade	2.9	3.4	0.39
Arterial Blood Gas	2.8	3.5	0.29
Vasopressors	2.2	2.8	0.30
Naloxone	0.1	0.2	0.40

AV = atrioventricular, BMP = basic metabolic panel, CBC = complete blood count, CT = computed tomography, ECG = electrocardiogram, IV = intravenous, RBC = red blood cell.

See Supplemental Table 3 for process measures at 3, 6, and 48-hour time intervals.

[a] List available in Supplemental Table 8.

[b] Includes CT chest, head, or abdomen.

Table 4:

Outcomes in Screen Positive Patients

Outcome Measure	Silent (n=1,540)	Alert (n=2,137)	P-value
Hospital Length of Stay, median (IQR), d	9 (5–18)	9 (5–18)	0.39
ICU Transfer < 6 hours After Alert, %	9.2	12.0	0.14
ICU Transfer < 24 hours After Alert, %	14.4	16.8	0.19
ICU Transfer < 48 hours After Alert, %	16.4	18.9	0.20
ICU Transfer Any Time After Alert, %	25.6	26.1	0.80
Time to ICU Transfer After Alert, median (IQR), h	16 (2–108)	8 (2–62)	<0.01
ICU Length of Stay, median (IQR), h	71 (38–163)	85 (43–179)	0.11
Mortality ≤ 30 days After Trigger, %	9.8	9.4	0.81
In-hospital Mortality, %	10.6	10.3	0.88
Discharged to Home, %	59.9	58.4	0.42
Discharged to Nursing Facility, %	15.3	15.2	0.93
Discharged to Inpatient Hospice, %	3.4	4.6	0.51
Severe Sepsis or Septic Shock[a], %	20.5	18.6	0.32

ICU = intensive care unit, IQR = interquartile range.

[a]Severe Sepsis: >2 SIRS and positive blood or urine culture and lactate >2.2; Septic Shock: Severe Sepsis AND systolic blood pressure <90 mm Hg.