Observational/Cohort Study

Retrospective Observational Study of the Clinical Performance Characteristics of a Machine Learning Approach to Early Sepsis Identification

Raj Topiwala, MPH1,2; Kanak Patel, MD1; Joan Twigg, RN1; Jane Rhule, RN1; Barry Meisenberg, MD1

Objectives: To estimate performance characteristics and impact on care processes of a machine learning, early sepsis recognition tool embedded in the electronic medical record.

Design: Retrospective review of electronic medical records and outcomes to determine sepsis prevalence among patients about whom a warning was received in real time and timing of that warning compared with clinician recognition of potential sepsis as determined by actions documented in the electronic medical record.

Setting: Acute care, nonteaching hospital.

Patients: Patients in the emergency department, observation unit, and adult inpatient care units who had sepsis diagnosed either by clinical codes or by Center for Medicare and Medicaid Services Severe Sepsis and Septic Shock: Management Bundle (SEP-1) criteria for severe sepsis and patients who had machine learning–generated advisories about a high risk of sepsis.

Interventions: Noninterventional study.

Measurements and Main Results: Using two different definitions of sepsis as “true” sepsis, we measured the sensitivity and early warning clinical utility. Using coded sepsis to define true positives, we measured the positive predictive value of the early warnings. Sensitivity was 28.6% and 43.6% for coded sepsis and severe sepsis, respectively. The positive predictive value of an alert was 37.9% for coded sepsis. Clinical utility (true positive and earlier advisory than clinical recognition) was 2.2% and 1.6% for the two different definitions of sepsis. Use of the tool did not improve sepsis mortality rates.

Conclusions: Performance characteristics were different than previously described in this retrospective assessment of real-time warnings. Real-world testing of retrospectively validated models is essential. The early warning clinical utility may vary depending on a hospital’s state of sepsis readiness and embrace of sepsis order bundles.

Key Words: early warning system; machine learning; sepsis; sepsis recognition

The high toll of sepsis has generated worldwide awareness campaigns aimed at the public and clinicians (1), large-scale studies of optimal treatments resulting in new treatment guidelines (2), and a decades long search for early recognition tools (3).

A recent evolution of the early warning tools is electronic medical record (EMR)-embedded machine learning techniques which can detect ominous trends in vital signs, laboratory values, and oxygenation status to issue a prediction of sepsis likelihood (4). In theory, clinicians could then be notified before more overt deterioration occurs allowing for either earlier intervention with antisepsis measures or more intensive clinical and microbial monitoring. We retrospectively tested whether initiation of one such machine learning system, “Insight” (Dascena, Hayward, CA), had an impact on care delivery for hospitalized patients. Insight is a proprietary machine learning system that continuously scans the EMR for vital sign patterns consistent with sepsis. Later versions also incorporated laboratory values. The system is said to be capable of interpolating missing values and identifying clinical patterns where sepsis is more likely. The threshold for alerting can be adjusted to modulate sensitivity or specificity based on an institution’s experiences. The company’s website (5) identifies 12 published studies pertaining to derivation of the model and two
Prospective studies testing the system against different early warning systems, manual or electronic (6, 7).

The objective of this study was to evaluate the real-world performance characteristics of a sepsis early warning tool with regard to sensitivity, positive predictive value, and clinical utility.

**METHODS**

**Setting**

Anne Arundel Medical Center is a 385-bed acute care hospital in Annapolis, MD. The emergency department (ED) volume is 97,000 visits annually, with 50 nonelective medical or surgical admissions daily. During the time of this study, acute care hospitalists, without house staff, cared for 250 medical inpatients not including the 30-bed medical-surgical ICU which had its own 24/7 intensivist coverage. The Insight tool was given access to all adult patient’s EMRs to perform ongoing surveillance of vital signs and some basic laboratory values. When a pattern consistent with sepsis was detected through a proprietary algorithm, an alert was sent to a critical care-trained “resource nurse” and unit charge nurse who were tasked with re-evaluating the patient at the bedside. Nurses contacted medical staff subsequently based on their own judgment.

**Study Design**

This study followed a retrospective observational cohort design. We analyzed, by manual chart review, consecutive patients in the ED, observation unit, or hospital who received Insight alerts for high sepsis risk from February 22, 2018, the day the tool was initiated, to June 12, 2018. We measured sensitivity, positive predictive value, and clinical utility by performing manual chart review in all patients for whom an alert had been received and who had sepsis during the same time period. Two different working definitions of sepsis were used. The first included all patients with any sepsis or septic shock codes identified in the International Statistical Classification of Diseases and Related Health Problems (ICD), 10th edition (ICD-10) based on medical documentation (see Appendix for specific codes). Coding was performed by professional medical abstractors based on physician documentation. ICD-10 coding does not include laboratory analysis, only clinical documentation. Because there is controversy about sepsis definitions (8), a second definition of sepsis was also used in this study; a subgroup of the coded sepsis that met Severe Sepsis and Septic Shock: Management Bundle (SEP-1) criteria used by Center for Medicare and Medicaid Services (CMS) for national reporting in the United States (9). This definition includes evidence of two or more systemic inflammatory response syndrome (SIRS) criteria plus a known or suspected infection source and certain types of organ dysfunction. Cases for review in this category represent a random sample of severe sepsis cases supplied by CMS reviewers to the hospital based on the hospital’s submitted data. Clinical utility, defined as both the tool correctly alerting and alerting before clinical suspicion was arrived at independently, was measured in both defined sepsis patients. Clinical utility assessments compared the time of Insight alert and the time of clinical suspicion based on chart review. Clinical suspicion was operationally defined as the earliest specific documentation by clinicians of suspicion of sepsis. This most frequently was the ordering of a serum lactate test, or less often, high-volume IV fluids which are components of the sepsis bundle, but not part of “usual care.” Other components of the bundle, such as administration of antibiotics and the collection of blood cultures, were judged too nonspecific to signify clinical suspicion of sepsis; inclusion of these in the definition of clinical utility would have prejudiced the results against the early recognition tool. We recorded first use of antibiotics in relation to the early warning in true-positive cases. Sensitivity was determined by evaluating whether sepsis cases as defined above had an Insight alert generated during the admission. Positive predictive value was determined using the standard statistical definition: the number of true positives divided by the total number of positives (all alerts).

Clinical utility was defined as the tool accurately alerting in patients with sepsis and earlier than clinical suspicion provided for.

Comparison of false-positive vital signs was evaluated using a chi-square test.

Sepsis-related mortality for coded sepsis cases was compared for the 6 months before the Insight tool initiation and for the 7 months after the tool was initiated, not including the first month of intervention. A two-sample t-test was used to compare differences in the means of the pre- and postintervention time period.

The Institutional Clinical Research Committee reviewed the study before data collection commenced and determined that it met criteria for quality improvement project and not research.

**RESULTS**

Two hundred three alerts were received during the study period. The location of patients receiving alerts included ICU 51%, ED 22%, and other adult medical or surgical units 27%. During the study period, a total of 269 patients had sepsis based on ICD-10 coding from clinical documentation. Review of each of these charts showed that 77 of the 269 received Insight alerts (sensitivity, 29%). Six of these alerts came before clinical suspicion (clinical utility, 2.2%) (Table 1). Among the 71 patients in whom the alert fired after clinical suspicion, mean and median time of the alert delay were 23.5 and 11.3 hours, respectively.

Among the randomly chosen 61 patients with CMS-defined severe sepsis, 26 received Insight alerts (sensitivity, 43%), one of which came before clinical suspicion (clinical utility, 1.6%). Among the 25 patients, receiving alerts after clinical suspicion was already documented; the mean and median time of delay were 26.7 and 15.7 hours, respectively.

Among the 203 Insight alerts, there were 126 false positives and 77 true positives for coded sepsis equivalent to a positive predictive value of 38%. The more common diagnoses associated with false positives were hypertension (78), type 2 diabetes mellitus (63), end-stage renal disease (42), alcoholic cirrhosis (13), and gastrointestinal bleeding. Statistically significant differences (p < 0.05) in vital signs between true and false positives were found for tachycardia, hypoxia, and abnormal body temperature (Fig. 1).

Among the 77 true-positive septic patients, all patients had already received antibiotics a median of 18 hours before the warning was received (range, 2–124 hr).
Delays of more than 15 minutes in entering vital signs into the EMR of up to 1 hour occurred in 15% of patients for vital signs and 5% for laboratory values, although these were not concentrated in those with sepsis or in those receiving alerts.

After the implementation of the early warning tool, sepsis mortality rates dropped from 14.2% (sd, 3.8%; 95% CI, 13.2–15.1) to 12.1% (sd, 4.2%; 95% CI, 10.6–13.1). The difference in the means was not significant ($p = 0.2$).

**DISCUSSION**

In this retrospective chart review of actual real-time alert performance, we did not find the tool to be helpful with regard to detection of sepsis cases (sensitivity), positive predictive value, or clinical utility (both accurate detection and early detection). Furthermore, 51% of alerts were received for patients already in the ICU and being treated for sepsis. All patients had already been started on antibiotics by the time the alerts came. It should be recognized that antibiotic use in hospital among febrile or ill-appearing patients is common so that failure to alert before antibiotic use should not be seen as evidence of a failed tool. There was no statistical differences in sepsis mortality in a pre-post intervention comparison.

These results differ from prior published data using this tool. Most publications involve validation studies where the alerts were applied retrospectively to data already obtained. In this study, the alerts were delivered in real time allowing a direct comparison with sepsis recognition as determined by clinician action documented in the EMR. Beyond the validation studies, Insight has been shown to improve outcomes in two studies. In a single-institution, nonteaching hospital, a postimplementation trial of Insight improved sepsis outcomes compared with the preexisting method of screening for sepsis: either no screening (ED patients) or with twice-daily manual nurse screening.

### TABLE 1. Performance Characteristics of Machine Learning Tool for Early Identification of Sepsis

| Measure | Coded sepsis | CMS-defined sepsis |
|---------|--------------|--------------------|
| Sensitivity | 77/269 = 28.6% | 26/61 = 42.6% |
| Positive predictive value | 37.9% |
| Clinical utility | 6/29–2.2% | 1/61 = 1.6% |
| Duration of delay vs clinical suspicion for true positives | Mean, 23.5 hr; median, 11.3 hr | Mean, 26.7 hr; median, 15.7 hr |
| False positive | Either coded or CMS defined | 126/203 = 62% |

CMS = Center for Medicare and Medicaid Services.

**Figure 1.** Comparison of abnormal vital signs among true-positive and false-positive Insight sepsis warnings.
for SIRS criteria for hospitalized patients. The 1-month long steady-state postimplementation results showed a dramatic reduction of sepsis mortality of 75% along with improvements in length of stay and readmission rates compared with baseline (6). In a randomized trial limited to 142 evaluable ICU patients at a single institution, insight compared favorably to a preexisting sepsis recognition tool based on SIRS criteria and evidence of infection. ICU length of stay and hospital mortality were less in those monitored by Insight compared with the control EMR-based tool (7). Clinical utility was not reviewed in either of these studies nor were sustained results reported. The importance of timing in evaluating early warning systems has recently been emphasized and should be a part of every early warning tool assessment (10).

Although the exact process Insight uses to identify risk of sepsis has not been disclosed, the tool places a large amount of weight on vital sign and vital sign trends although the version in place at our hospital also considered some laboratory values. Differences in hospital or unit practices of when vital signs and laboratory values are entered into the chart may affect the ability to detect sepsis early. In our study, delays in entering results into the EMR of up to 1 hour did occur in 15% for vital signs and 5% for laboratory values. The operational delays in vital sign approval and transmission to the EMR may have reduced the early warning capability but would not account for the low sensitivity or positive predictive value. Furthermore, these small delays account for all of the delay in sepsis recognition of 11.3 and 15.7 hours for the two types of sepsis definitions, respectively. Furthermore, the tool was built to be robust to missing data (5), a common problem in acute care hospitals.

This study found significant differences in several vital signs between true-positive and false-positives sepsis cases, indicating a potential area for refinement. Hypertensive patients who experience relative drop in blood pressure during hospitalization may have generated false positives. Ultimately, vital signs may be too nonspecific to identify deterioration due to infection, the hallmark of sepsis. Other causes of hypoperfusion or hypotension, such as can occur with dialysis, may have also provoked false-positive warnings. Differentiating abnormal vital signs of septic patients from other patients undergoing hypoperfusion may not be reliable, a problem common to other early warning tools (3).

This was a single-institution study of a limited number of cases, an issue which effects much of the more recent descriptions of computer models for sepsis prediction. Generalizability is further limited by the fact that the nursing and medical staff had already undergone considerable education and training to increase awareness of sepsis, so the early alert function of the tool may have had less of a measurable impact. Indeed, adherence to the CMS sepsis bundle was 75% pre- and postimplementation. Furthermore, there are differences in how a hospital system receives and responds to alerts which can influence the value of alerts. In our hospital, alerts were received by critical care–trained “resource” nurses who then went to the bedside to help bedside nurses to reevaluate patients. Medical staff were contacted subsequently as needed. We did not measure the effectiveness or timeliness of this process nor of the medical staff response once they were contacted, but it is obvious that the afferent function of an early warning will be rendered immaterial if the efferent human response function is ineffective or delayed. Finally, the definitions we used for sepsis to measure sensitivity are known to be flawed and a subject of ongoing reevaluation (11).

Our data on the poor performance of early alerting are similar to the results of a study of a different tool, the pediatric Rothman Index, which uses vital signs, laboratory values, and nursing assessments to predict clinical deterioration (12). That study, using a similar chart review methodology to determine clinical utility, found that most clinical deterioration warnings came after clinicians had identified the problem through usual care processes. In addition, a similar finding of low clinical utility of a different early warning system was reported in a survey study of clinical staff at a large academic medical center (13). Surveys of medical and nursing practitioners were obtained within 2 hours of each alert to determine clinician perception of value in individual cases. Medical respondents found value in only 25% of alerts. Potential sepsis was the suspected trigger of the alert in only one third. More commonly, clinicians assessed the alerts as erroneous due to abnormal but stable baseline vital signs or inconsequential changes.

CONCLUSIONS
Early warning of sepsis likelihood remains an important goal even though the importance and appropriateness of early antibiotics have been challenged (14). Computer-assisted interpretation of trends in vital signs and laboratory values remains an aspirational goal particularly for patients being cared for in non-ICU environments where monitoring capability and experience with critical illness are not as developed. The current generation of machine learning algorithms might be improved by combining with single or combination biomarkers (15) for more precise and earlier diagnosis. Even with refinements, the limitation of algorithms, neural networks, and machine learning, especially those trained on limited datasets, has recently been reemphasized (16). Accordingly, clinical utility of real-time warnings should be included in future measurements of early warning tools.

REFERENCES
1. Society of Critical Care Medicine: Surviving Sepsis Campaign. Available at: http://www.survivingsepsis.org/Pages/default.aspx. Accessed March 18, 2019.
2. Howell MD, Davis AM: JAMA clinical guidelines synopsis: Management of sepsis and septic shock. JAMA 2017; 317:847–848.
3. McGaughey J, Alderice F, Fowler R, et al: Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. Cochrane Database Syst Rev 2007; (3):CD005529.
4. Delahanty RJ, Alvarez J, Flynn LM, et al: Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. Ann Emerg Med 2019; 73:334–344.
5. Dascena: Insight Dascena. Available at: https://www.dascena.com/insight. Accessed May 19, 2019.
6. 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.
7. 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. BMJ Open Respir Res 2017;4:e000234
8. Rhee C, Gohil S, Klompas M: Regulatory mandates for sepsis care-reasons for caution. N Engl J Med 2014;370:1673–1676
9. Rhee C, Filbin MR, Massaro AF, et al; Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program: Compliance with the national SEP-1 quality measure and association with sepsis outcomes: A multicenter retrospective cohort study. Crit Care Med 2018;46:1585–1591
10. Rolnick JA, Weissman GE: Early warning systems: The neglected importance of timing. J Hosp Med 2019;14:445–447.
11. Kalantari A, Mallemat H, Weingart SD: Sepsis definitions: The search for gold and what CMS got wrong. West J Emerg Med 2017;18:951–956
12. Winter MC, Kubis S, Bonaﬁde CP: Beyond reporting early warning score sensitivity: The temporal relationship and clinical relevance of "true positive" alerts that precede critical deterioration. J Hosp Med 2019;14:138–143
13. Guidi JL, Clark K, Upton M, 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
14. Klompas M, Calandra T, Singer M: Antibiotics for sepsis-finding the equilibrium. JAMA 2018;320:1433–1434
15. Larsen FF, Petersen JA: Novel biomarkers for sepsis: A narrative review. Eur J Intern Med 2017;45:46–50.
16. Coiera E: On algorithms, machines and medicine. Lancet Oncol 2019;2092:166–167.
## APPENDIX

| Sepsis Diagnoses                                      | International Statistical Classification of Diseases and Related Health Problems, 10th Edition, Codes |
|-------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Salmonella sepsis                                      | A021                                                                                             |
| Anthrax sepsis                                         | A227                                                                                             |
| Erysipelothrix sepsis                                   | A267                                                                                             |
| Listerial sepsis                                       | A327                                                                                             |
| Sepsis due to *Streptococcus*, group A                  | A400                                                                                             |
| Sepsis due to *Streptococcus*, group B                  | A401                                                                                             |
| Sepsis due to *Streptococcus pneumoniae*                | A403                                                                                             |
| Other streptococcal sepsis                             | A408                                                                                             |
| Streptococcal sepsis, unspecified                       | A409                                                                                             |
| Sepsis due to methicillin-susceptible *Staphylococcus aureus* | A4101                                                                                           |
| Sepsis due to methicillin-resistant *S. aureus*         | A4102                                                                                            |
| Sepsis due to other specified *Staphylococcus*          | A411                                                                                             |
| Sepsis due to unspecified *Staphylococcus*              | A412                                                                                             |
| Sepsis due to *Hemophilus influenzae*                   | A413                                                                                             |
| Sepsis due to anaerobes                                 | A414                                                                                             |
| Gram-negative sepsis, unspecified                      | A4150                                                                                            |
| Sepsis due to *Escherichia coli*                       | A4151                                                                                            |
| Sepsis due to *Pseudomonas*                            | A4152                                                                                            |
| Sepsis due to *Serratia*                               | A4153                                                                                            |
| Other Gram-negative sepsis                             | A4159                                                                                            |
| Sepsis due to *Enterococcus*                           | A4181                                                                                            |
| Other specified sepsis                                 | A4189                                                                                            |
| Sepsis, unspecified organism                           | A419                                                                                             |
| Actinomycotic sepsis                                   | A427                                                                                             |
| Gonococcal sepsis                                      | A5486                                                                                             |
| Candidal sepsis                                        | B377                                                                                             |
| Severe sepsis without septic shock                     | R6520                                                                                             |
| Severe sepsis with septic shock                        | R6521                                                                                             |
| Postprocedural septic shock, initial encounter          | T8112XA                                                                                            |