Evaluating a sepsis prediction machine learning algorithm using minimal electronic health record data in the emergency department and intensive care unit

Hoyt Burdick[1,2], Eduardo Pino[1,2], Denise Gabel-Comeau[1], Carol Gu[3], Heidi Huang[3], Anna Lynn-Palevsky[3*], Ritankar Das[3]

[1] Cabell Huntington Hospital, Huntington, WV
[2] Marshall University School of Medicine, Huntington, WV
[3] Dascena Inc., Hayward, CA

* Corresponding author
Email: anna@dascena.com
22710 Foothill Blvd., Suite #2
Hayward, CA 94541
Tel: (872) 228-5332

Abstract

Introduction Sepsis is a major health crisis in US hospitals, and several clinical identification systems have been designed to help care providers with early diagnosis of sepsis. However, many of these systems demonstrate low specificity or sensitivity, which limits their clinical utility. We evaluate the effects of a machine learning algodiagnostic (MLA) sepsis prediction and detection system using a before-and-after clinical study performed at Cabell Huntington Hospital (CHH) in Huntington, West Virginia. Prior to this study, CHH utilized the St. John's Sepsis Agent (SJSA) as a rules-based sepsis detection system.

Methods The Predictive algoRithm for EValuation and Intervention in SEpsis (PREVISE) study was carried out between July 1, 2017 and August 30, 2017 (NCT03235193). All patients over the age of 18 who were admitted to the emergency department or intensive care units at CHH were monitored during the study. We assessed pre-implementation baseline metrics during the month of July, 2017, when the SJSA was active. During implementation in the month of August, 2017,

SJSA and the MLA concurrently monitored patients for sepsis risk. At the conclusion of the study period, the primary outcome of sepsis-related in-hospital mortality and secondary outcome of sepsis-related hospital length of stay were compared between the two groups.

Results Sepsis-related length of stay decreased from 2.99 days in the pre-implementation phase to 2.48 days in the post-implementation phase, a 17.1% relative reduction ($P < 0.001$), and sepsis-related in-hospital mortality decreased from 3.97% to 2.64%, a 33.5% relative decrease ($P = 0.038$).

Conclusion Reductions in patient mortality and length-of-stay were observed with use of a machine learning algorithm for early sepsis detection in the emergency department and intensive care units at Cabell Huntington Hospital, and may present a method for improving patient outcomes.

Introduction

Sepsis is a common, life-threatening syndrome that arises from the body's dysregulated response to infection, and has been declared a global health priority by the World Health Organization [1]. In the United States, sepsis is responsible for a cost of over $20 billion [2] and affects a population of 750,000 annually [3]. Severe sepsis, distinguished by organ failure, may progress to septic shock, presenting with refractory hypotension. An increase in mortality rate from over 10% to near 40% accompanies this escalation in condition severity [4]. In spite of the high prevalence of sepsis syndromes and the associated poor outcomes, the variations in host response and disease progression often inhibits the critical early and accurate diagnosis of sepsis. As demonstrated by the recent proposal of changes to the stages of sepsis (Sepsis-3) [5], there is some controversy in establishing unanimous definitions of clinical sepsis presentations. Yet, numerous studies have reached the consensus that early detection of sepsis and compliance with sepsis treatment bundles can positively impact patient mortality and length of stay [6].

Healthcare systems grapple with accurately identifying sepsis early in disease progression. The increasing availability of data from patients' electronic health records (EHR) may provide valuable insight into the processes of sepsis disease progression. Existing prospective studies of EHR data-derived tools in clinical settings have been primarily rules-based

[7], applying preset score thresholds to classify risk level [8]. However, these studies have often demonstrated subpar sensitivity and specificity [9]. Machine learning algorithms have the potential to improve on rules-based systems through flexibility and learning from patient data, clinical response patterns, and correlative trends. Previous work conducted on sepsis detection machine learning algorithms constructed from EHR data include the retrospective studies of Henry et al. [10], Nachimuthu et al. [11], and Sawyer et al [12].

West Virginia provider Cabell Huntington Hospital (CHH), a 303-bed facility, partnered with Dascena (Hayward, CA) to improve sepsis-related outcomes using a machine learning algodiagnostic (MLA). The Dascena MLA was validated for sepsis prediction and detection in several studies [13-15], demonstrating an area under the receiver operator characteristic (ROC) curve (AUROC) over 0.90 using only six vital signs, in a multicenter cohort study of over 650,000 encounters [16]. In a recent randomized clinical trial, mortality decreased by 12.4 percentage points with use of the MLA, a relative reduction of 58% [17]. Comparison of the MLA to rules-based scores such as Systemic Inflammatory Response Syndrome (SIRS) [18] criteria have shown superior sensitivity and specificity up to four hours in advance of sepsis onset [13]. In this study, we evaluate improvements in CHH sepsis-related in-hospital mortality rate and hospital length of stay with the use of the machine learning algorithm in the emergency and ICU patient populations, using a before-and-after study design.

Methods

Study Design

This study was designed as a prospective before-and-after study (study registration: ClinicalTrials.gov NCT03235193). Approval for the study was granted by the institutional review board (IRB) at Marshall University. We measured pre-implementation baseline metrics as well as post-implementation metrics in order to determine the effect of the algorithm.

Prior to and during this trial, CHH used Cerner's St. John's Sepsis Agent (SJSA). St. John's issues two types of alerts: 1) a SIRS alert fires when three or more vital signs or lab results fall out of range [18] and 2) a sepsis alert fires when at least two SIRS criteria are met and lab results indicate organ dysfunction. When the criteria are met, an alert appears on an

electronic health record (EHR) screen and persists until the appropriate orders are obtained. Although the criteria are designed to discern sepsis progression, SJSA produces a high false alarm rate due to its low specificity [19]. Low specificity often leads to alarm fatigue, or clinician indifference to the alerts, which results in a delay of treatment during the critical early intervention period.

Only CHH's SJSA was active during the pre-implementation period; during the post-implementation period, both the SJSA and the machine learning sepsis predictor were actively monitoring patients. Pre-implementation data were measured during the period July 1 to July 30, 2017, and post-implementation data were measured during the period spanning from August 1 to August 30, 2017. All data were collected through CHH's EHR system, CARE Connect (Cerner Corp, North Kansas City, Missouri).

During the post-implementation phase, all patients over the age of 18 who were admitted to either the emergency department or intensive care units (ICU) were monitored by the MLA for sepsis risk. The MLA assessed each patient for sepsis by extracting real-time data from each patient's EHR and analysing trends in the patient measurements. Risk prediction scores were computed hourly throughout the duration of each patient's stay. The MLA used in this study is described in detail in prior prospective [16, 17] and retrospective work [13].

The algorithm was designed to compare trends in each patient's EHR measurements to confirmed prior sepsis cases in order to accurately detect and predict sepsis. The classifier used to perform the comparison was an ensemble of decision trees. After patient data passed through the classifier, the MLA generated a sepsis risk prediction score between 0 and 100. Healthcare providers were called and informed of a possible sepsis case when a patient's score exceeded 80. At this point, patients were examined and treated under CHH's standard sepsis protocol. Patients were monitored by the algorithm for the duration of their stay in the emergency department or ICU. Additionally, patients continued to be monitored by the SJSA in the post-implementation period. This design ensured that minimal risk was incurred by patients; if the MLA failed to detect a case of sepsis, the SJSA may still have detected sepsis and alerted a clinician.

The primary and secondary outcomes assessed in this study were the sepsis-related in-hospital mortality rate and the sepsis-related hospital length of stay (LOS), respectively, at

CHH.

Data Collection and Analysis

Demographic and clinical information was collected for each patient monitored during the study period. Patients were monitored and clinical data was collected during the duration of their stay in the participating hospital units, and participants were followed until hospital discharge in order to determine overall sepsis-related in-hospital mortality and LOS. Patients were considered to be "sepsis-related" and included for analysis if they met two or more SIRS criteria at any point during their stay in participating units and were over the age of 18. We classified patients in this manner due to the predictive nature of the MLA. Because the algorithm is designed to identify patients likely to develop sepsis, including only patients who met the 2001 consensus definition criteria could have excluded patients who would have developed sepsis had they not been identified and treated early. The SIRS criteria are closely linked to sepsis diagnostic criteria, and their use in this study ensured that only patients with sepsis or closely related conditions were included in our final analysis.

The MLA determined each patient's sepsis risk through real-time abstraction of data in the patient EHR. At least one measurement each of systolic blood pressure, diastolic blood pressure, heart rate, temperature, respiratory rate, and peripheral oxygen saturation (SpO_2) were required for sepsis prediction. Any vital signs not recorded during a given hour were gap-filled using a forward-filling imputation process in which the most recently recorded past measurement was used for sepsis risk score computation. Additionally, although not necessary, the algorithm was able to incorporate lab results such as pH, white blood cell count, and glucose levels when they were available. The MLA analyzed each patient's clinical measurements as well as hourly changes in measurements in order to determine sepsis risk.

After the conclusion of the study period, the primary outcome of sepsis-related in-hospital mortality and the secondary outcome of average sepsis-related hospital LOS were calculated. No interim analyses were conducted during the study period.

Results

Patient Characteristics

Our final analysis included a total of 2,296 septic cases, which included 1,160 patients in the pre-implementation phase and 1,136 patients in the post-implementation phase. Patient demographics for each period are displayed in Table 1. Because all patients were tracked throughout the duration of their hospital stay, no patients were lost to follow-up.

Table 1. Demographic characteristics for patients monitored during the study period, based on data abstracted from the EHR.

Demographic Overview	Characteristic	Pre-implementation (%)	Pre-implementation (%)
Gender	Female	43.33	39.20
	Male	56.67	60.80
Age	18-29	20.47	27.08
	30-39	16.74	13.54
	40-49	13.96	12.55
	50-59	16.74	13.12
	60-69	15.35	21.01
	70+	16.74	12.69
Ethnicity	Asian or Pacific Islander	0	0.13
	Black	12.45	4.88
	White	86.27	94.37
	Native Hawaiian	0.43	0
	Unknown	0.86	0.63

Outcomes

We evaluated the number of sepsis-related in-hospital mortality cases and the mean length of stay for sepsis patients before and after MLA integration. The pre-implementation baseline mortality rate was 46/1160 (3.97%, standard error (SE) 0.57%). After MLA implementation, the mortality rate was 30/1136 (2.64%, SE 0.48%) representing a 33.5% reduction (P=0.038). During the baseline period, average sepsis-related length of stay was 2.99 days (SE 0.028); post MLA implementation, sepsis-related length of stay was 2.48 days (SE 0.051), a 17.1% reduction (P<0.001). These results are shown in Figure 1.

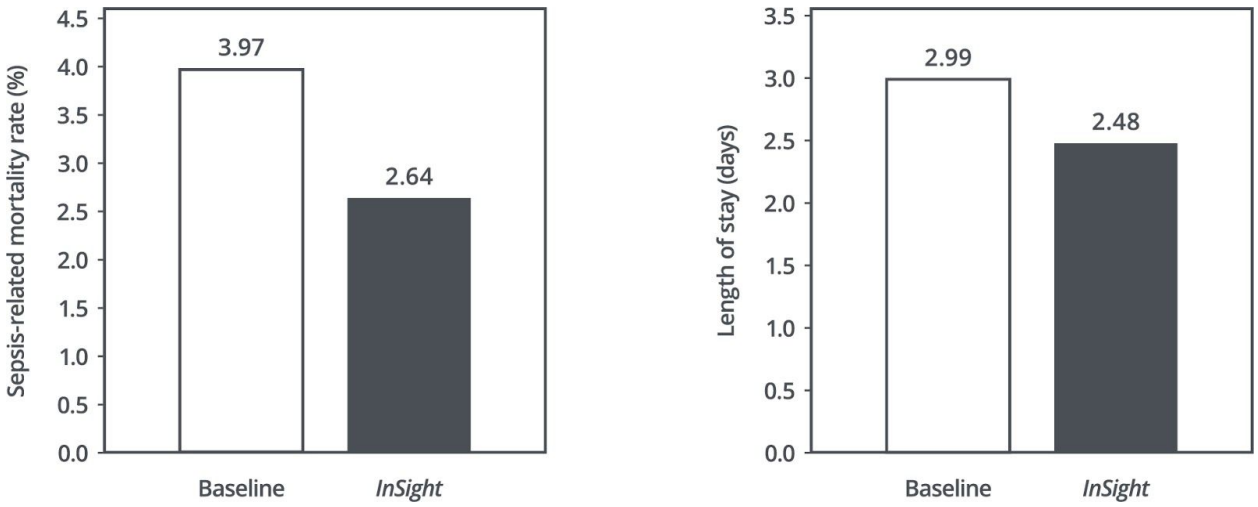

Figure 1: a) Sepsis-related in-hospital mortality during the baseline period and after machine learning algorithm (MLA) implementation. The use of the MLA was associated with a reduction of in-hospital mortality of 33.5%. b) Sepsis-related hospital length of stay during the baseline period and after MLA implementation. The use of the MLA was associated with a reduction in length of stay of 17.1%.

In addition to analyzing patient outcomes, we compared the performance of the algorithm to that of the Modified Early Warning Score (MEWS) [20], SIRS, and the Quick Sequential Organ Failure Assessment (qSOFA) [4] for sepsis detection. On a retrospective set of 1,912 patients (70 meeting severe sepsis criteria) admitted to CHH during the pre-implementation

period, the MLA demonstrated higher Area Under the Receiver Operating Characteristic (AUROC) curve, sensitivity, and specificity as compared to all three rules-based scores (Table 2).

Table 2. Comparison of area under the receiver operator characteristic curve (AUROC), sensitivity, and specificity for the machine learning algorithm (MLA), SIRS score, MEWS, and qSOFA score for sepsis detection. 95% Confidence Intervals (CI) were calculated only for the MLA.

	MLA	**SIRS**	**MEWS**	**qSOFA**
AUROC (95% CI)	0.95 (0.94, 0.97)	0.68	0.78	0.74
Sensitivity	0.92 (0.90, 0.94)	0.59	0.91	0.53
Specificity	0.92 (0.90, 0.94)	0.76	0.24	0.94

Discussion

In this prospective study, sepsis-related patient outcomes were improved through the implementation of a machine-learning based sepsis prediction algorithm. When deployed together with SJSA in the emergency department and intensive care units at CHH, the algorithm resulted in decreases in sepsis-related in-hospital mortality and sepsis-related length of stay compared to outcomes using SJSA alone (Figure 1).

The MLA also fired a sepsis alert an average of two hours earlier than the SJSA, which is likely a result of the predictive design of the algorithm. The early sepsis warning provided by the MLA, coupled with its high accuracy, potentially enabled earlier clinical intervention to identify sepsis cases, provide supportive treatment, and possibly prevent progression of the condition. Studies have shown that early treatment of sepsis can improve patient outcomes [6, 21], and that confirmation of a positive microbiology and correspondingly targeted antibiotic therapy can improve mortality rates [22]. The early, accurate warning of sepsis onset may have provided

clinicians an opportunity both to begin early treatments and to identify the causal agent in a timely manner.

The MLA's ability to maintain currently high sensitivity and specificity, as demonstrated by its performance on retrospective data, is of clinical importance. Alarm systems with low specificity can generate high numbers of false alarms, contributing to the problem of alarm fatigue [23]. Alarm fatigue presents a patient safety concern, as providers may begin to ignore alarms which they deem unreliable. The high specificity maintained by this MLA may help mitigate alarm fatigue in clinical settings.

The algorithm assessed in this study has previously been examined in several retrospective studies, where it has been validated for detection of sepsis [14], severe sepsis [13], and septic shock [15]. The algorithm has also been previously evaluated in prospective studies, including a randomized controlled trial where use of the MLA resulted in statistically significant decreases in in-hospital mortality and average length of stay [17]. The present study presents further evidence that machine-learning methods for sepsis detection and prediction can provide routes towards improving sepsis-related patient outcomes.

Limitations

The present study examines the algorithm for sepsis detection in a single medical center located in West Virginia. Other settings, with different patient demographic characteristics and EHR recording practices, may experience different outcomes with utilizing this MLA. Further, this MLA was assessed only in the emergency department and the intensive care units at CHH. The MLA may perform differently in other hospital settings, such as speciality cancer centers. The limited period of monitoring both pre- and post-implementation metrics additionally limit the generalizability of these results.

Further, confounding factors may have influenced the differences in patient outcomes noted between the pre- and post-implementation periods. Variations in the clinical staff present during the two periods may have influenced the improvements in patient outcomes noted during the two periods. Increased awareness of sepsis risk may have resulted in more rigorous bedside monitoring for sepsis during the post-implementation period. Clinicians may have more closely

monitored those patients who generated an MLA alert in the post-implementation period of the study; increased attention to at-risk patients may therefore have been at least partially responsible for the improved outcomes noted.

Conclusion

This clinical trial demonstrates improved patient outcomes through use of a machine learning-based sepsis prediction algorithm. Statistically significant reductions in the in-hospital mortality rate and hospital length of stay were obtained with this algorithm, deployed concurrently with a rules-based sepsis monitoring system, over the rules-based system alone. These results are consistent with prior clinical results demonstrating improved patient outcomes with the use of machine learning-based sepsis prediction algorithms. Limitations of this study include a focus on only the emergency department and intensive care units in a single medical center, and a limited period of analysis. The MLA's performance in a broader range of geographic regions and patient groups will be investigated in future studies.

References

1. Fleischmann C, Scherag A, Adhikari NK, et al. Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. *Am. J. Respir. Crit. Care Med.* 2016;**193**:259-72.
2. Chalupka AN, Talmor D. The economics of sepsis. *Crit. Care Clin.* 2012;28(2):57-76.
3. Angus DC, Linde-Zwirble WT, Lidicker, et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit. Care Med* 2001;**29**:1303-1310.
4. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* 2016;**315**:801-10.
5. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;**315**:762-74.
6. Nguyen HB, Corbett SW, Steele R, et al. Implementation of a bundle of quality indicators for the early management of severe sepsis and septic shock is associated with decreased mortality. *Crit. Care Med.* 2007;**35**:1105-1112.
7. Hooper MH, Weavind L, Wheeler AP, Martin JB, Gowda SS, Semler MW, Hayes RM, Albert DW, Deane NB, Nian H, Mathe JL. Randomized trial of automated, electronic

monitoring to facilitate early detection of sepsis in the intensive care unit. *Crit. Care Med*. 2012;**40**:2096-101.
8. Herasevich V, Afessa B, Chute CG, Gajic O. Designing and testing computer based screening engine for severe sepsis/septic shock. *AMIA Annu. Symp. Proc*. 2007;**2007**: 966.
9. 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
10. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci. Transl. Med*. 2015;7:299ra122.
11. Nachimuthu SK, Haug PJ. Early detection of sepsis in the emergency department using Dynamic Bayesian Networks. *AMIA Annu. Symp. Proc.* 2012;**2012**:653.
12. Sawyer AM, Deal EN, Labelle AJ, et al. Implementation of a real-time computerized sepsis alert in nonintensive care unit patients. *Crit. Care Med*. 2011;**39**:469-73.
13. Desautels T, Calvert J, Hoffman J, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Medical Informatics* 2016;4(3):28. PMID: 27694098
14. Calvert JS, Price DA, Chettipally UK, et al. A computational approach to early sepsis detection. *Comput. Biol. Med.* 2016;**74**:69-73.
15. Calvert JS, Price DA, Chettipally UK, et al. High-performance detection and early prediction of septic shock for alcohol-use disorder patients. *Ann. Med. Surg.* 2016;**8**:50-55.
16. Mao Q, Jay M, et al. Multicenter validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open 2017, In Press.
17. Shimabukuro D, 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 Randomized Clinical Trial. BMJ Open Respiratory Research 2017;4:e000234.
18. Rangel-Frausto MS, Pittet D, Costigan M, Hwang T, Davis CS, Wenzel RP. The natural history of the systemic inflammatory response syndrome (SIRS): a prospective study. *JAMA*. 1995;**273**:117-23.
19. Jaimes F, Garcés J, Cuervo J, et al. The systemic inflammatory response syndrome (SIRS) to identify infected patients in the emergency room. *Intensive Care Med*. 2003;**29**:1368-1371.
20. Subbe CP, Slater A, Menon D, Gemmell L. Validation of physiological scoring systems in the accident and emergency department. *Emergency Medicine Journal* 2006 Nov;23(11):841-845. PMID: 17057134
21. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, Peterson E, Tomlanovich M. Early goal-directed therapy in the treatment of severe sepsis and septic shock. New England Journal of Medicine. 2001 Nov 8;345(19):1368-77. PMID: 11794169
22. Singer M. Antibiotics for sepsis- does each hour really count? Or is it incestuous amplification? Am. J. Respir. Respir. Crit. Care Med. Forthcoming 2017.
23. Cvach M. Monitor alarm fatigue: an integrative review. *Biomedical Instrumentation & Technology*. 2012 Jul;46(4):268-77. PMID: 22839984