BMJ Health & Care Informatics

Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission: a prospective multicentre clinical outcomes evaluation of real-world patient data from US hospitals

Hoyt Burdick,[1,2] Eduardo Pino,[1,2] Denise Gabel-Comeau,[1] Andrea McCoy,[3] Carol Gu,[4] Jonathan Roberts,[4] Sidney Le,[4] Joseph Slote,[4] Emily Pellegrini,[4] Abigail Green-Saxena [ID] ,[4] Jana Hoffman,[4] Ritankar Das[4]

¹Cabell Huntington Hospital, Huntington, West Virginia, USA
²Marshall University School of Medicine, Huntington, West Virginia, USA
³Cape May Regional Medical Center, Cape May Court House, New Jersey, USA
⁴Dascena Inc, Oakland, California, USA

Correspondence to
Dr Jana Hoffman;
jana@dascena.com

ABSTRACT

Background Severe sepsis and septic shock are among the leading causes of death in the USA. While early prediction of severe sepsis can reduce adverse patient outcomes, sepsis remains one of the most expensive conditions to diagnose and treat.

Objective The purpose of this study was to evaluate the effect of a machine learning algorithm for severe sepsis prediction on in-hospital mortality, hospital length of stay and 30-day readmission.

Design Prospective clinical outcomes evaluation.

Setting Evaluation was performed on a multiyear, multicentre clinical data set of real-world data containing 75 147 patient encounters from nine hospitals across the continental USA, ranging from community hospitals to large academic medical centres.

Participants Analyses were performed for 17 758 adult patients who met two or more systemic inflammatory response syndrome criteria at any point during their stay ('sepsis-related' patients).

Interventions Machine learning algorithm for severe sepsis prediction.

Outcome measures In-hospital mortality, length of stay and 30-day readmission rates.

Results Hospitals saw an average 39.5% reduction of in-hospital mortality, a 32.3% reduction in hospital length of stay and a 22.7% reduction in 30-day readmission rate for sepsis-related patient stays when using the machine learning algorithm in clinical outcomes analysis.

Conclusions Reductions of in-hospital mortality, hospital length of stay and 30-day readmissions were observed in real-world clinical use of the machine learning-based algorithm. The predictive algorithm may be successfully used to improve sepsis-related outcomes in live clinical settings.

Trial registration number NCT03960203

Summary

What is already known?

► Severe sepsis and septic shock are among the leading causes of death in the USA, and sepsis remains one of the most expensive conditions to diagnose and treat.

► Accurate early diagnosis and treatment can reduce the risk of adverse patient outcomes, but the accuracy of traditional rule-based screening methods is limited.

► Machine learning-based algorithms (MLAs) have been developed for sepsis detection and prediction. However, many of these MLAs require extensive training data, laboratory test results or specialist annotation and have not been evaluated with real-world data.

What does this paper add?

► This study is a novel multisite prospective real-world data evaluation of the effect of a machine learning algorithm for severe sepsis detection and prediction on clinical outcomes.

► In an analysis across nine diverse hospitals from the Northeast, South, Midwest and Western USA, including academic centres and community hospitals, use of the MLA was associated with a statistically significant reduction of in-hospital mortality, hospital length of stay and 30-day readmissions for sepsis-related patient stays.

► Given that clinician perception of MLAs remains a barrier to their broad acceptance and use, this study advances the field of MLAs for prediction and detection of sepsis by providing clinically relevant evidence that an MLA requiring only minimal data inputs, routinely collected by the electronic health record, can improve patient outcomes without adding to clinician workload.

INTRODUCTION

Despite a high associated mortality[1 2] and high costs of treatment,[2 3] severe sepsis remains notoriously difficult to diagnose and treat. The healthcare costs of sepsis in the USA in 2013 reached nearly US$24 billion, roughly 6% of the nation's total hospital bill, while sepsis

patients represented only 3.6% of all hospital stays.[4] Prior research has emphasised the importance of timely sepsis recognition to both improving patient outcomes and reducing costs associated with treatment.[5–7] New definitions intended to improve the clinical recognition of sepsis have recently been proposed,[8 9] as the previous use of screening based on systemic inflammatory response syndrome (SIRS) criteria has been found to be nonspecific.[10] Evidence from the medical literature has shown that accurate early diagnosis and treatment can reduce the risk of adverse patient outcome from severe sepsis and septic shock.[11–13] Therefore, earlier detection of sepsis and more accurate recognition of patients at high risk of developing severe sepsis or septic shock is essential for effective sepsis treatment.

Screening tools used in clinical settings for the identification of decompensating patients include the Sequential Organ Failure Assessment (SOFA),[14] the SIRS criteria[15] and the Modified Early Warning Score (MEWS).[16] These systems have been used to recognise severe sepsis due to their ability to both identify systemic inflammation as a sign of infection, and to detect possible organ dysfunction. The utility of such systems for the identification of septic patients has been studied at length in recent literature.[17–22] However, systems, such as MEWS, SOFA and SIRS, were originally designed as generalised screening tools as opposed to explicitly identifying sepsis, and their efficacy in sepsis diagnosis is limited. For example, SOFA has been reported to be not widely applicable outside of the intensive care unit (ICU), and it often requires use of laboratory values that are not rapidly available.[17] SIRS has been reported to be nonspecific[17–23] and also may yield up to one in eight false negatives in detecting patients with organ failure and infection.[17–24] Despite their limitations, these scoring systems have established performance metrics, and serve as important comparators for newly developed severe sepsis detection and prediction systems and their effect on clinical outcomes.[25–28]

Improvement in sepsis care and adoption of electronic health record (EHR) systems have been incentivised by the Centers for Medicare & Medicaid Services in recent years.[29 30] Currently, 96% of hospitals in the USA have an EHR federally tested and certified for the government's incentive programme.[31–33] A number of methods have been developed to monitor patient EHR data for severe sepsis, but few provide predictive capabilities to enable early intervention and improve patient outcomes. Although they represent fairly new additions to the field of sepsis care, machine learning algorithms (MLAs) have the potential to significantly improve patient outcomes through advance warning of impending sepsis onset. Sepsis prediction MLAs may also serve to empower clinicians to have confidence in their sepsis diagnosis in a variety of ambiguous cases, including instances when positive culture results are not available,[34] and in cases of atypical clinical presentation among older patients who comprise a majority of sepsis cases.[35 36] Machine learning-based decision support systems, therefore, represent an important area of investigation for sepsis research.[37 38]

Table 2 Hospital characteristics; geographical region, teaching status and size of hospitals included in this study

Hospital characteristic	Clinical outcomes analysis
Geographical region	
Northeast	1
South	3
Midwest	1
West	4
Teaching status	
Teaching	7
Non-teaching	2
Hospital Size	
Small (<100 beds)	3
Medium (100–250 beds)	2
Large (>250 beds)	4

The MLA used in this study has been described in previous peer-reviewed publications both retrospectively and prospectively,[39–45] but has not been evaluated for its effect on clinical outcomes on multicentre diverse hospital settings. In this study, performance of our MLA for severe sepsis prediction and detection was evaluated using real-world data from patient EHRs at nine diverse hospitals from the northeast, southern, midwestern and western regions of USA, spanning academic centres to community hospitals. A clinical outcomes analysis was performed to evaluate the effect of the algorithm on in-hospital patient mortality, hospital length of stay (LOS) and 30-day readmissions.

METHODS
Dataset
Prospectively collected real-world patient data were abstracted from the EHR systems of Epic (Epic Systems, Verona, Wisconsin, USA), Allscripts (Allscripts Healthcare Solutions, Chicago, Illinois, USA), Cerner (Cerner Systems, North Kansas City, Missouri, USA), Meditech (Meditech, Westwood, Massachusetts, USA), Paragon (McKesson, San Francisco, California, USA) and Soarian (Cerner Systems, North Kansas City, Missouri, USA), across the nine hospitals for a clinical outcomes evaluation. These data spanned 75 147 patient encounters from early 2017 to mid-2018. Details about these nine hospitals are provided in table 2.

All patient information was deidentified prior to analysis in compliance with the Health Insurance Portability and Accountability Act. Data collection for all datasets was passive and did not impact patient safety.

In this clinical outcomes analysis, only adult (at or above age 18) EHR record data from inpatient wards and emergency departments were analysed. All genders and ethnicities were included. Patient stays that met two or more SIRS criteria at any point during their stay

were considered 'sepsis related' and included for clinical outcomes analysis. We defined the onset time of severe sepsis as the first time at which two SIRS criteria and at least one organ dysfunction criteria (online supplementary table 1) were met within the same hour. This resulted in the inclusion of 17 758 patient encounters for analysis. The design of and recruitment to this study did not involve patients and the public.

Demographic, admission and discharge times, vital sign, laboratory and drug administration data were abstracted, for each visit of a given patient, from the EHR. Online supplementary file 1 provides details on data field abstraction. Not all data fields were available at all facilities.

Machine learning algorithm

The machine learning classifier was constructed using gradient boosted trees, implemented in Python (Python Software Foundation, https://www.python.org/), with the XGBoost package.[46] The algorithm analysed the patient vital signs of systolic blood pressure, diastolic blood pressure, heart rate, temperature, respiratory rate and SpO_2 (oxygen saturation), and age. Missing values were filled using last-one carry forward imputation, wherein the most recent observation of a measurement is used to replace the missing value. This method of imputation is appropriate for clinical measurements, because observations of a given vital sign are expected to be highly dependent on previous observations.[47 48] The vector of vital sign measurements was analysed, and measurements were concatenated for up to 2 hours before the measurement time as additional features. Differences in measurement values between time steps were also concatenated where appropriate. Thus, each clinical feature represents between 3 and 5 columns in the data matrices. Our previous work has used this procedure of transforming time series problems into supervised learning problems.[49] Values were concatenated into a feature vector with 15 elements. An ensemble of decision trees was constructed using the gradient boosted trees approach, and the ensemble prediction is based on an aggregate of these scores. Vital sign measurements were discretised into two categories to determine tree branching, and patient risk scores were determined by their final categorisation in each tree. Tree branching was limited to six levels. We set the XGBoost learning rate to 0.1 and included no more than 1000 trees in the final ensemble. These hyperparameters were justified in the context of the present data with a coarse grid search and align with previous work.[39] For additional details about MLA development, see Mao *et al*.[39]

Study design

For clinical outcomes analysis, we collected data from nine hospitals that implemented the MLA for sepsis prediction and detection. Data was then evaluated to determine the effect of the algorithm on patient outcomes of in-hospital mortality, hospital LOS and 30-day readmission. Providers at the hospitals using the MLA received automated

telephonic alerts if the MLA score was above a threshold set by the hospital.

Adult 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 units where the MLA was used. 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 those patients who met the 2001 consensus severe sepsis or septic shock definition criteria or the more recent sepsis-3 criteria may have excluded patients who would have developed sepsis had they not been identified and treated early. It has been reported that sepsis-3 diagnostic criteria narrows the sepsis population at the expense of sensitivity, and that disease diagnosis may be delayed due to resulting false negatives.[50] The SIRS criteria, while non-specific, are associated with early sepsis diagnostic criteria, and their use in this study ensured that those patients most at risk for sepsis were included in our final analysis.

At study sites, patient EHR data were constantly monitored by software stored in the computational servers used for our data integration. Any changes in patient state represented in the EHR would prompt the software to apply the MLA in order to generate an MLA score. If the MLA score was above a threshold set by the hospital, an indicator of patient risk would be generated, and a parallel monitoring service would detect the indicator and send a telephonic alert for the corresponding patient. Telephonic notification volumes differed from month to month during the trial period. Months with uncharacteristically low volumes (fewer than 5) were excluded from analysis. Three sites in the study were affected by the exclusion of low volume months from analysis. Alert volumes varied as site-specific customisation was performed through PDSA (plan-do-study-act) cycles for thresholding and rules-based suppression to optimise the algorithm for the best fit into a given care setting.[51] In particular, for any patient for whom an alert had already been produced, additional alerts were uniformly suppressed. At four of the nine hospitals, we collected data prior to the implementation of the MLA for measurement of baseline outcomes and for training of the MLA once deployed. When data from this baseline period preceding implementation of the MLA were not available, the baseline period used was the month immediately following implementation. This was the case for the remaining five of the nine hospitals. The analysis was repeated including only three of the nine hospitals which had at least 1 month of baseline data preceding MLA implementation, and the outcomes were similar. Once trained on data from the baseline period, MLAs remained static and were not trained further.

Not all three patient outcomes were measured at all sites. LOS was measured at all sites, in-hospital mortality was measured at six out of nine sites, and readmission was measured at five out of nine sites. If admission and discharge time stamps were unavailable, LOS and

Table 1 Demographics—aggregated clinical and demographic characteristics of patients from nine hospitals used for clinical outcomes analysis

| | Clinical outcomes analysis | |
	Baseline	MLA
Total no	12 793	62 354
Mean age (SD)	45 (24.4)	45 (24.0)
Male	5429 (42.4)	26 126 (41.9)
Female	7364 (57.6)	36 228 (58.1)
Unknown	—	—
White	11 832 (87.7)	54 635 (81.8)
Black	594 (4.4)	2469 (3.7)
Hispanic	1063 (7.87)	9609 (14.4)
Asian American	10 (0.1)	44 (0.1)
Unknown	—	—
Temperature	36.8 (0.3)	36.8 (0.3)
Respiratory rate	18.2 (4.7)	18.2 (4.1)
Systolic blood pressure	127.0 (18.0)	129.5 (19.1)
Diastolic blood pressure	72.9 (11.0)	75.1 (11.5)
Heart rate	84.7 (17.3)	86.1 (18.5)
Lactate	1.9 (1.70)	1.9 (1.86)
Creatinine	1.4 (2.54)	1.2 (1.70)
International normalised ratio	1.2 (0.58)	1.3 (0.90)
Platelets	239.5 (77.1)	241.9 (85.0)
SpO_2	97.4 (1.6)	97.4 (1.7)
White blood count	8.4 (2.51)	8.2 (2.03)
PaO_2	101.7 (47.9)	103.8 (51.8)
Bilirubin	0.7 (1.3)	0.7 (1.1)
FiO_2	44.3 (20.4)	46.6 (22.2)
pH	7.4 (0.08)	7.4 (0.09)

Values are shown with percentages of total population or SD.
FiO_2, fractional inspired oxygen; MLA, machine learning-based algorithm; PaO_2, arterial oxygen tension (or pressure).

readmission were determined by defining new visits when all vital sign measurements for a given patient were observed to be greater than 120 hours apart.

Statistical analysis

We used the 2-proportion risk difference z-test to determine if there was a statistically significant decrease in the in-hospital mortality, LOS, or the 30-day readmission rate with the use of the MLA. All tests were two tailed with an alpha level of 0.05, and were performed using Python.

RESULTS

Aggregated patient demographic data from the nine participating hospitals in this study are presented in table 1. Seventeen per cent of patients were included for

the baseline analysis period and 83% of patients were included for the MLA analysis period. Vital sign averages and SD were not significantly different between the baseline and MLA analysis periods. Among those patients analysed by the MLA, the mean age was 45 years (41.9% male vs 58.1% female). Patients included for clinical outcomes analysis were generally representative of those at risk of developing severe sepsis in terms of gender and racial/ethnic distribution.[1 52]

Table 2 shows the variation in hospital size, location and type for the hospitals included in this analysis. The wide range of geographical and population distribution demonstrates a diverse range of hospital types included for clinical outcomes determination.

Clinical outcomes were measured for all patients over 18 years who met two or more SIRS criteria at any point during their stay, in order to ensure that those patients most at risk for sepsis were included in our final analysis. The subsequent outcomes analysis was performed in order to determine if use of the MLA had significant effects on in-hospital patient mortality, hospital LOS and/or 30-day readmissions. We emphasise that while the SIRS criteria were used to determine which patients should be included in the outcomes analysis, the MLA in this study uses only patient vital signs to predict severe sepsis.

The sepsis-related outcomes after MLA implementation were a 39.50% reduction of in-hospital mortality (p<0.001), a 32.27% reduction of LOS (p<0.001) and a 22.74% reduction in 30-day readmission (p<0.001; table 3, figure 1). These results include sites where data from the period preceding implementation of the MLA were not available, in which case the baseline period used was the month immediately following implementation.

The analysis was repeated for a subset of 3 hospitals with at least 1 month of baseline (pre-MLA implementation) data, with a total of 52 487 patients. This resulted in 3951 patients in the baseline period (971 included as sepsis related, as defined in the Study Design section), and 48 536 patients (10 646 included as sepsis related) in the MLA analysis period. The outcomes for this patient subset were a 42.50% reduction of in-hospital mortality (p<0.05) and a 23.82% reduction in LOS (p<0.05).

Table 3 Sepsis-related patient outcomes table—analysis of in-hospital mortality, hospital length of stay and 30-day readmissions, in the baseline and MLA periods for sepsis-related patient

	Baseline period	MLA period	Reduction
In-hospital mortality	3.86%	2.34%	39.50%
Length of stay	4.83 days	3.27 days	32.27%
30-day readmission	36.4%	28.12%	22.74%

There were 12 793 patients in the baseline period, of whom, 3592 were included for analysis and 62 354 patients in the MLA period, of whom, 14 166 patients were included for analysis
MLA, machine learning-based algorithm.
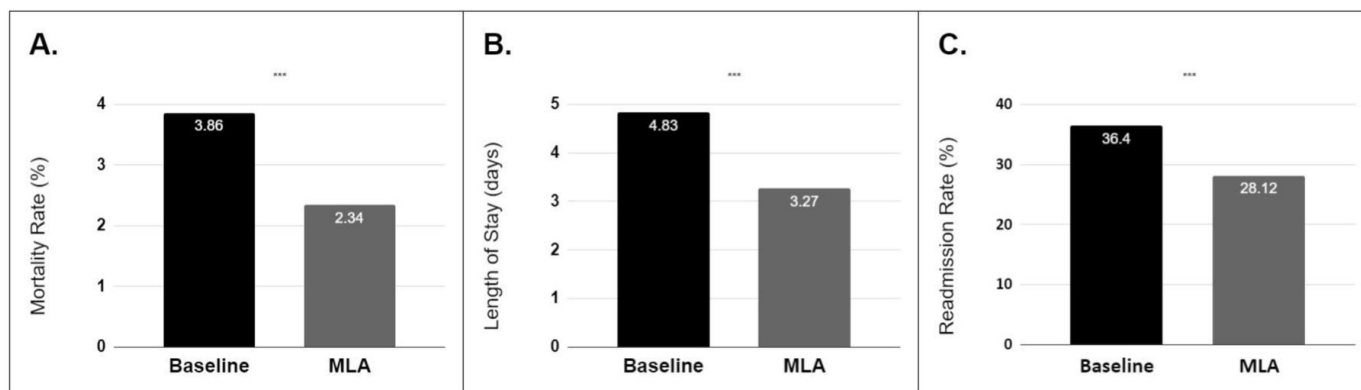
Figure 1 Patient outcomes——differences in (A) in-hospital mortality, (B) hospital length of stay and (C) 30-day readmissions in the baseline period and the MLA period for sepsis-related patients. Use of the MLA was associated with a 39.5% reduction of in-hospital mortality (p<0.001), a 32.3% reduction in length of stay (p<0.001) and a 22.7% reduction in 30-day readmissions (p<0.001). MLA, machine learning-based algorithm.

Results indicate that our machine learning algorithm for severe sepsis prediction can be successfully used to improve clinical outcomes of in-hospital mortality, LOS and 30-day readmission rates.

DISCUSSION

In this clinical outcomes study, we tested the hypothesis that use of an MLA for severe sepsis detection and prediction would result in reductions of adverse sepsis-related clinical outcomes. The design of this study involved minimal to no risk of patient harm, but offered potential benefits to both patients and providers. In particular, the algorithm's ability to identify patients with severe sepsis prior to onset provided a significant opportunity for early intervention. Prior studies have shown that early detection or prediction of sepsis and severe sepsis, respectively, can lead to a decrease in the time to administration of antibiotics,[40 53] and early intervention has been shown to reduce rates of patient mortality.[54–56] Use of the MLA in this study was associated with a 39.5% reduction of in-hospital mortality (p<0.001), a 32.3% reduction in LOS (p<0.001) and a 22.7% reduction in 30-day readmissions (p<0.001).

Improvements in clinical outcomes were calculated by comparing outcomes before algorithm implementation with outcomes after implementation. Not all data fields were available for abstraction at all nine participating hospitals. In cases where pre-implementation measurements were not available, the first month of clinical implementation was used as an approximate baseline. During this initial period, the MLA alert sensitivity, specificity and clinical response were undergoing evaluation and development, and therefore, did not represent the final state of the MLA alert and response. However, including the use of this period as a baseline may result in an underestimation of the effect of the MLA, compared with the pre-implementation period.

Results from the clinical outcomes analysis indicate that the algorithm has a more significant effect on improving clinical outcomes than other screening tools such as MEWS, SOFA and SIRS.[25–28] For example, in a prospective comparative analysis of qSOFA and SIRS for predicting adverse outcomes of patients with suspicion of sepsis, discrimination of in-hospital mortality using the SIRS score was reported to be significantly less than that of the qSOFA score, with an overall in-hospital mortality rate of 19%.[25] A pre-implementation and post-implementation study evaluating the effect of an SIRS-based sepsis early warning system that monitored SIRS criteria along with signs of organ dysfunction (based on systolic blood pressure and serum lactate thresholds), found that while the tool prompted more timely sepsis care, there was no significant reduction in mortality.[53] In a comprehensive review of peer-reviewed literature to evaluate the effect of MEWS on improving clinical outcomes, limited data and no clinical trials which linked use of MEWS scoring systems to 'robust' outcomes were found.[26] An analysis of a variety of disease severity scoring systems for the prognostic assessment of septic patients revealed that SOFA and MEWS showed only moderate discrimination in predicting 28-day mortality rates.[28] Beyond the simple heuristics of rules-based scoring systems such as MEWS, SOFA, qSOFA and SIRS, several machine learning approaches have been retrospectively evaluated for the detection and prediction of incipient sepsis.[37 38 57–66] They include dynamic Bayesian networks,[60] support vector machines,[57] survival-analytical models (TREWScore, Artificial Intelligence Sepsis Expert),[61 62] smoothed disease severity score learning,[63] hierarchical switching linear dynamical systems,[64] autoregressive hidden Markov models,[65] free-text models[38] and random-forest models.[57] These tools contribute notably to the field of sepsis detection because they offer generalisability, are scalable, and can be updated as new information is acquired.[58] However, many do not use information about measurement trends or correlations,[67] or do so ineffectively. Most machine learning approaches have been evaluated only on retrospective data as proof-of-concept.[37 57 58 60 62 64–66]

There remains an ongoing need for research which evaluates the clinical utility of sepsis prediction models in prospective and real-world settings.

Towards this end, Nelson *et al* conducted a prospective trial of a real-time electronic surveillance system to expedite early care of severe sepsis.[67] Outcome measures were rate and timeliness of sampling of blood lactate and blood cultures, performance of chest radiography and provision of antibiotics; however, only time to blood culture was significantly improved. The primary limitation of the trial was cited as the inability to detect severely septic cases before caregivers. Umscheid *et al* conducted a real-world pre-implementation and post-implementation study of an early warning and response system (EWRS) for sepsis outside of the ICU.[68] The EWRS identified at risk patients with a sensitivity of 16% and a specificity of 97%. Compared with a control period, the EWRS activated in the post-implementation period resulted in an increase in ICU transfer <6 hours after alert (p=0.06). However, additional outcome measures of hospital LOS (p=0.92), ICU transfer <24 hours after alert (p=0.20), renal replacement therapy ≤6 hours after alert (p=0.51) and all-patient mortality reductions (p=0.45) failed to reach statistical significance.[68] Austrian *et al* performed a time-series study which evaluated an electronic surveillance system on mortality and LOS on emergency department patients with severe sepsis or septic shock,[69] finding a modest decrease in LOS (16%) that did not reach statistical significance, with no difference in-hospital mortality or other intermediate outcome measures. Alert fatigue due to low positive predictive value (PPV) (0.146) was proposed as the primary contributor to these results, and researchers noted that more sophisticated approaches to early sepsis identification are needed to consistently improve patient outcomes. Importantly, the study supports the principle that high PPV is critical for effective clinical decision support interventions.[69] The early and accurate alerting system introduced in our study is associated with an LOS reduction of 32.3%, a mortality reduction of 39.50% and obtains a PPV of approximately 40% for sepsis prediction as demonstrated in prior work.[44]

In addition to clinically improving patient outcomes, the sepsis prediction tool analysed in this study also provides economic advantages. The cost of severe sepsis has been reported to extend 'well beyond' patient impact, as a large part of the sepsis economic burden is incurred after discharge and during rehospitalisation.[70] Administration of timely treatment is therefore crucial to reducing costs, reducing rates of readmission and improving treatment outcomes. This clinical outcomes study provides a prospective analysis of machine learning algorithm performance in the sepsis care domain. To the extent possible, we also calculate a first order approximation of cost reductions incurred through reductions in LOS from the use of the algorithm. The average LOS reduction was found to be 1.56 days. At an average per diem cost of care of US$2271 with 343 patients included per month at the nine locations, the reduction of LOS

translates to approximately US$14.5 million of annual cost savings across all nine hospitals included in this analysis. These findings on post-marketing real-world data confirm pre-marketing randomised clinical trial results.[40] Previous research has shown that early detection and treatment of sepsis can improve patient outcomes and reduce hospital costs.[12 13 71 72]

Our real-world data analysis has several limitations. We cannot guarantee the usefulness of resulting alerts to clinicians, which indicates a need for future studies which include qualitative analysis of algorithm utility (ie, clinician surveys or interviews). Future work including clinician surveys would also help to determine how clinicians responded to the alerts, including any diagnostics tests or treatment interventions that were ordered. This level of detail would be helpful in assessing the potential means through which the observed positive impacts on patient outcomes were achieved. Ideally, this future work would also include the clinical adjudication of sepsis onset times, instead of defining onset times in terms of gold standard criteria, so that the extent to which alerts were accurate and early could be determined. Further, although machine learning systems have made significant advances in the healthcare domain over the past decade, it is important to consider the unintended ways in which they impact clinical practice. Unintended consequences of machine learning in medicine include an over-reliance on the capabilities of automation; a lack of contextual information which may lead to diagnostic misinterpretation; and observer variability affecting the accuracy and reliability of machine learning performance.[73] However, it should be noted that risks of machine learning are minimised by screening or 'sniffer' algorithms such as this MLA, which are designed to increase clinician oversight for high-risk cases, and not to replace expert clinical judgement and standards of care. Other limitations of our study include variation in clinician and team responses to patients at possible risk for sepsis. Only adults in US hospitals were included in the study. While nine diverse hospitals were included in the analysis, these hospitals may not be representative of all US hospitals or international hospital settings. Data were not available from all hospitals for all months and outcome measurements. Baseline data were not available for all hospitals and the first month of MLA data was used as an approximation in these cases. This may lead to an underestimation of the effect of the MLA at these sites. However, the analysis was repeated on a subset of three hospitals with at least 1 month of baseline pre-MLA implementation data and outcomes were similar. This study did not follow patient mortality after hospital discharge. We cannot eliminate the possibility that implementation of a sepsis algorithm raised general awareness of sepsis within a hospital, which may lead to higher recognition of septic patients, independent of algorithm performance.

CONCLUSION

This study evaluates the effect of a machine learning algorithm for severe sepsis detection and prediction on

clinical outcomes. In an analysis of the algorithm across nine hospitals, use of the MLA was associated with a 39.5% reduction of in-hospital mortality, a 32.3% reduction in hospital LOS and a 22.7% reduction in 30-day readmissions. These results support that the implementation of an accurate machine learning algorithm for early sepsis recognition may lead to improved patient outcomes, and by extension may serve to reduce the financial burden to the US healthcare system. In future studies, we will continue to analyse the algorithm's impact on patient outcomes in other care settings.

ORCID iD
Abigail Green-Saxena http://orcid.org/0000-0002-8502-6589

REFERENCES

1. Angus DC, Linde-Zwirble WT, Lidicker J, et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med* 2001;29:1303–10.
2. Stevenson EK, Rubenstein AR, Radin GT, et al. Two decades of mortality trends among patients with severe sepsis: a comparative meta-analysis*. *Crit Care Med* 2014;42:625–31.
3. Torio CM, Celeste M, Andrews RM. "*National inpatient hospital costs: the most expensive conditions by payer 2011*", 2013.
4. Torio C, Moore B. National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2013. HCUP Statistical Brief #204. Agency for Healthcare Research and Quality, Rockville, MD, 2016. Available: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb204-MostExpensive-Hospital-Conditions.pdf
5. Dellinger RP, Levy MM, Rhodes A, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med* 2013;39:165–228.
6. Lagu T, Rothberg MB, Shieh M-S, et al. Hospitalizations, costs, and outcomes of severe sepsis in the United States 2003 to 2007. *Crit Care Med* 2012;40:754–61.
7. Gaieski DF, Edwards JM, Kallan MJ, et al. Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med* 2013;41:1167–74.
8. Levy MM, Fink MP, Marshall JC, et al. 2001 sccm/esicm/accp/ats/sis international sepsis definitions conference. *Crit Care Med* 2003;31:1250–6.
9. 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.
10. Shankar-Hari M, Phillips GS, Levy ML, et al. Developing a new definition and assessing new clinical criteria for septic shock: for the third International consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016;315:775–87.
11. Damiani E, Donati A, Serafini G, et al. Effect of performance improvement programs on compliance with sepsis bundles and mortality: a systematic review and meta-analysis of observational studies. *PLoS One* 2015;10:e0125827–24.
12. Moore LJ, Moore FA. Early diagnosis and evidence-based care of surgical sepsis. *J Intensive Care Med* 2013;28:107–17.
13. Kenzaka T, Okayama M, Kuroki S, et al. Importance of vital signs to the early diagnosis and severity of sepsis: association between vital signs and sequential organ failure assessment score in patients with sepsis. *Intern Med* 2012;51:871–6.
14. Vincent JL, Moreno R, Takala J, et al. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. on behalf of the Working group on sepsis-related problems of the European Society of intensive care medicine. *Intensive Care Med* 1996;22:707–10.
15. 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;29:1368–71.
16. Subbe CP, Slater A, Menon D, et al. Validation of physiological scoring systems in the accident and emergency department. *Emerg Med J* 2006;23:841–5.
17. McLymont N, Glover GW. Scoring systems for the characterization of sepsis and associated outcomes. *Ann Transl Med* 2016;4:527.
18. Churpek MM, Snyder A, Han X, et al. Quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit. *Am J Respir Crit Care Med* 2017;195:906–11.
19. Usman OA, Usman AA, Ward MA. Comparison of SIRS, qSOFA, and news for the early identification of sepsis in the emergency department. *Am J Emerg Med* 2019;37:1490–7.
20. Johnson AEW, Aboab J, Raffa JD, et al. A comparative analysis of sepsis identification methods in an electronic Database*. *Crit Care Med* 2018;46:494–9.
21. Bhattacharjee P, Edelson DP, Churpek MM. Identifying patients with sepsis on the hospital wards. *Chest* 2017;151:898–907.
22. van der Woude SW, van Doormaal FF, Hutten BA, et al. Classifying sepsis patients in the emergency department using SIRS, qSOFA or MEWS. *Neth J Med* 2018;76:158–66.
23. Churpek MM, Zadravecz FJ, Winslow C, et al. Incidence and prognostic value of the systemic inflammatory response syndrome and organ dysfunctions in ward patients. *Am J Respir Crit Care Med* 2015;192:958–64.
24. Kaukonen K-M, Bailey M, Pilcher D, et al. Systemic inflammatory response syndrome criteria in defining severe sepsis. *N Engl J Med* 2015;372:1629–38.
25. Finkelsztein EJ, Jones DS, Ma KC, et al. Comparison of qSOFA and SIRS for predicting adverse outcomes of patients with suspicion of sepsis outside the intensive care unit. *Crit Care* 2017;21:73.
26. Roney JK, Whitley BE, Maples JC, et al. Modified early warning scoring (MEWS): evaluating the evidence for tool inclusion of sepsis screening criteria and impact on mortality and failure to rescue. *J Clin Nurs* 2015;24:3343–54.
27. Lie KC, Lau C-Y, Van Vinh Chau N, et al. Utility of SOFA score, management and outcomes of sepsis in Southeast Asia: a multinational multicenter prospective observational study. *J Intensive Care* 2018;6:9.
28. Innocenti F, Tozzi C, Donnini C, et al. SOFA score in septic patients: incremental prognostic value over age, comorbidities, and parameters of sepsis severity. *Intern Emerg Med* 2018;13:405–12.
29. Office of the National Coordinator for Health Information Technology. 'Office-based Physician Electronic Health Record Adoption,' Health IT Quick-Stat #50, 2016. Available: http://www.webcitation.org/6rmdNMHPW
30. HealthIT.gov. EMR Incentives & Certification, 2013. Available: https://www.healthit.gov/providers-professionals/ehr-incentive-programs
31. Available: https://www.healthcare-informatics.com/news-item/ehr/survey-nearly-all-us-hospitals-use-ehrs-cpoe-systems
32. Available: https://dashboard.healthit.gov/quickstats/quickstats.php
33. Available: https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php
34. Phua J, Ngerng W, See K, et al. Characteristics and outcomes of culture-negative versus culture-positive severe sepsis. *Crit Care* 2013;17:R202.

35 Lamantia MA, Stewart PW, Platts-Mills TF, *et al*. Predictive value of initial triage vital signs for critically ill older adults. *West J Emerg Med* 2013;14:453–60.

36 Girard TD, Opal SM, Ely EW. Insights into severe sepsis in older patients: from epidemiology to evidence-based management. *Clin Infect Dis* 2005;40:719–27.

37 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;73:334–44.

38 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.

39 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.

40 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.

41 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.

42 Burdick H, Pino E, Gabel-Comeau D, *et al*. Evaluating a sepsis prediction machine learning algorithm using minimal electronic health record data in the emergency department and intensive care unit. *bioRxiv*:224014.

43 Calvert JS, Price DA, Chettipally UK, *et al*. A computational approach to early sepsis detection. *Comput Biol Med* 2016;74:69–73.

44 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 Med Inform* 2016;4:e28.

45 Calvert J, Desautels T, Chettipally U, *et al*. High-Performance detection and early prediction of septic shock for alcohol-use disorder patients. *Ann Med Surg* 2016;8:50–5.

46 Chen T, Guestrin C. *XGBoost: a scalable tree boosting system*. Paper presented at the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

47 Shao J, Zhong B. Last observation carry-forward and last observation analysis. *Stat Med* 2003;22:2429–41.

48 Ali MW, Talukder E. Analysis of longitudinal binary data with missing data due to dropouts. *J Biopharm Stat* 2005;15:993–1007.

49 Mohamadlou H, Lynn-Palevsky A, Barton C, *et al*. Prediction of acute kidney injury with a machine learning algorithm using electronic health record data. *Can J Kidney Health Dis* 2018;5:205435811877632.

50 Fang X, Wang Z, Yang J, *et al*. Clinical evaluation of Sepsis-1 and Sepsis-3 in the ICU. *Chest* 2018;153:1169–76.

51 Burdick H, Pino E, Gabel-Comeau D, *et al*. Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission. *bioRxiv* 2018;1:457465.

52 Mayr FB, Yende S, Angus DC. Epidemiology of severe sepsis. *Virulence* 2014;5:4–11.

53 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.

54 Kumar A, Roberts D, Wood KE, *et al*. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006;34:1589–96.

55 Rivers E, Nguyen B, Havstad S, *et al*. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001;345:1368–77.

56 Yealy DM, Kellum JA, ProCESS Investigators, *et al*. A randomized trial of protocol-based care for early septic shock. *N Engl J Med* 2014;370:1683–93.

57 Giannini HM, Ginestra JC, Chivers C, *et al*. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit Care Med* 2019;47:1485–92.

58 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* 2016;23:269–78.

59 Wulff A, Montag S, Marschollek M, *et al*. Clinical decision-support systems for detection of systemic inflammatory response syndrome, sepsis, and septic shock in critically ill patients: a systematic review. *Methods Inf Med* 2019:1–15.

60 Nachimuthu SK, Haug PJ. Early detection of sepsis in the emergency department using dynamic Bayesian networks. *AMIA Annu Symp Proc* 2012;2012:653–62.

61 Henry KE, Hager DN, Pronovost PJ, *et al*. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015;7:299ra122–299.

62 Nemati S, Holder A, Razmi F, *et al*. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med* 2018;46:547–53.

63 Dyagilev K, Saria S. Learning (predictive) risk scores in the presence of censoring due to interventions. *Mach Learn* 2016;102:323–48.

64 Stanculescu I, Williams CK, Freer Y, editors. *A hierarchical switching linear dynamical system applied to the detection of sepsis in neonatal condition monitoring*. UAI, 2014.

65 Stanculescu I, Williams CKI, Freer Y. Autoregressive hidden Markov models for the early detection of neonatal sepsis. *IEEE J Biomed Health Inform* 2014;18:1560–70.

66 Harrison AM, Thongprayoon C, Kashyap R, *et al*. Developing the surveillance algorithm for detection of failure to recognize and treat severe sepsis. *Mayo Clin Proc* 2015;90:166–75.

67 Nelson JL, Smith BL, Jared JD, *et al*. Prospective trial of real-time electronic surveillance to expedite early care of severe sepsis. *Ann Emerg Med* 2011;57:500–4.

68 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.

69 Austrian JS, Jamin CT, Doty GR, *et al*. Impact of an emergency department electronic sepsis surveillance system on patient mortality and length of stay. *J Am Med Inform Assoc* 2018;25:523–9.

70 Tiru B, DiNino EK, Orenstein A, *et al*. The economic and humanistic burden of severe sepsis. *Pharmacoeconomics* 2015;33:925–37.

71 Ferrer R, Martin-Loeches I, Phillips G, *et al*. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. *Crit Care Med* 2014;42:1749–55.

72 Calvert J, Hoffman J, Barton C, *et al*. Cost and mortality impact of an algorithm-driven sepsis prediction system. *J Med Econ* 2017;20:646–51.

73 Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA* 2017;318:517–8.