Enhancing Quality of Care and Survival Outcomes with Deep Learning-Based Sepsis Prediction

[1]Mukesh Kumar Saini, [2]Sachin Gupta, [3]Brett Fritze

[1]Researcher and Solution Architect, [2]Senior Specialist Engineer, [3]Product and Technology Consultant
Computer Science and Software Engineering,
[1]Wipro Limited, [2]Synechron, [3]HCA Healthcare

Abstract*:* Sepsis continues to be a leading cause of death and illness globally. Early detection algorithms may enhance patient outcomes, yet their real-world effectiveness is not widely studied. This research aimed to evaluate the impact of a deep-learning model, COMPOSER, on early sepsis prediction and patient outcomes. We conducted a quasi-experimental study across two Emergency Departments (EDs) within the Health Care System, analyzing data from 6,217 adult septic patients between January 1, 2016, and April 30, 2018. The intervention tested was a nurse-facing Best Practice Advisory (BPA) activated by COMPOSER. We compared metrics including in-hospital mortality, sepsis bundle adherence, changes in the Sequential Organ Failure Assessment (SOFA) score within 72 hours of sepsis onset, ICU-free days, and ICU encounters before and after the intervention. Using a Bayesian structural time-series analysis with confounder adjustments, we found that the implementation of COMPOSER was linked to a 1.9% absolute decrease (17% relative reduction) in in-hospital sepsis mortality (95% CI: 0.3%–3.5%), a 5.0% absolute increase (10% relative increase) in sepsis bundle compliance (95% CI: 2.4%–8.0%), and a 4% reduction (95% CI: 1.1%–7.1%) in SOFA score change within 72 hours post-sepsis onset. These results indicate that COMPOSER's early prediction capabilities were associated with significant improvements in mortality and sepsis bundle compliance.

Keywords **-** Sepsis, Artificial Intelligence (AI), SERA Algorithm, Structured Data, Unstructured Data, Clinical Measures, Mortality Reduction, Healthcare Analytics, Machine Learning, Risk Assessment, Patient Outcomes.

I. INTRODUCTION

Sepsis, a dysregulated host response to infection, is estimated to afflict over 48.9 million people annually worldwide, of whom approximately 11 million die[1,2]. Early recognition of sepsis is critical, as interventions such as fluid resuscitation, antibiotic administration, and source control have greater benefits when implemented earlier in the disease course[3–9]. Detecting patients with sepsis can be challenging due to the condition's heterogeneity; thus, we and others have used predictive analytics to improve early detection[10–14]. We recently reported the performance of COMPOSER, a deep-learning model that uses real-time data from electronic health records to predict sepsis before obvious clinical manifestations[15]. Few sepsis algorithms have been rigorously tested at the bedside or evaluated regarding patient outcomes[16–19]. Existing algorithms within electronic health records (EHRs) have demonstrated relatively poor positive predictive value (PPV) and may contribute to provider mistrust of predictive models[20,21]. Notably, false positive alerts from such models often lead to alarm fatigue and provider burnout/mistrust. COMPOSER was specifically designed to reduce false alarms by flagging outliers and out-of-distribution samples as indeterminate[15].

Based on this conceptual framework, we integrated the COMPOSER algorithm into two emergency departments (ED) at UC San Diego (UCSD) Health via our EHR (Epic Systems, Verona, WI). We seek to test the hypothesis that our algorithm-based intervention was feasible in real time and that the additional information would help guide clinicians to earlier sepsis recognition and result in improved patient outcomes. To accomplish this goal, we are conducting a quasi-experimental study in which we track outcomes before and after deployment, using historical control data to account for baseline acuity, comorbidities, seasonal effects, and secular trends over time.

II. RESULTS

During the study period, January 1st, 2021 through April 30th, 2023, 6,340 ED encounters met the Sepsis-3 consensus sepsis definition, of which 123 were excluded because they were transitioned to comfort measures before sepsis onset. The final study included 6217 patients, 5065 in the pre-intervention phase, and 1152 in the post-intervention phase. Table 1 shows baseline characteristics and summary statistics for the study cohort. Baseline characteristics from each emergency department are compared in Supplementary Table 1. Most septic patients exhibited some level of chronic comorbidity (median Elixhauser of 5) and the median SOFA score at the time of sepsis was 2. We did not observe significant differences in the baseline character- istics between the pre-intervention cases and post-intervention cases.

COMPOSER alerts

During the post-intervention period, an average of 235 alerts were generated per month, equating to approximately 1.65 alerts per nurse per

Table 1. Demographics and baseline characteristics of septic patients before and after COMPOSER.

	Total	Pre-intervention	Post-intervention	*P*-value[a]
Characteristic				
Number of patients, *N* (%)	6217 (100%)	5065 (81.5%)	1152 (18.5%)	-
Age, mean (SD)	63 (17.1)	63 (17.0)	64 (17.3)	0.08
Sex, *N* (%)				
Male	3592 (57.8%)	2966 (58.6%)	626 (54.3%)	-
Female	2625 (42.2%)	2099 (41.4%)	526 (45.7%)	-
Race				
Asian	530 (8.5%)	404 (8%)	126 (10.9%)	-
Black or African American	639 (10.3%)	519 (10.2%)	120 (10.4%)	-
White	2983 (48%)	2440 (48.2%)	543 (47.1%)	-
Other[b]	2065 (33.2%)	1702 (33.6%)	363 (31.5%)	-
Ethnic group				-
Hispanic/Latino	1756 (28.2%)	1449 (28.6%)	307 (26.6%)	-
Not Hispanic/Latino	4461 (71.8%)	3616 (71.4%)	845 (73.4%)	-
Organ dysfunction				
Elixhauser Comorbidity Index, Median (IQR)	5 (0–13)	5 (0–13)	5 (0–14)	0.64
SOFA Score at Time of Sepsis, Median (IQR)	2 (1–3)	2 (1–3)	2 (1–3)	0.99
Lab values				
Lactate at the time of sepsis	2.4 (1.6–4.3)	2.4 (1.6–4.3)	2.4 (1.6–4.3)	0.76
Interventions				
Mechanical Ventilation, *N* (%)[c]	1035 (16.6%)	849 (16.8%)	186 (16.1%)	0.64
Administration of Vasoactive Medications, *N* (%)[c]	424 (6.8%)	345 (6.8%)	79 (6.9%)	1.0

[a]*P*-values for continuous variables are based on Kruskal–Wallis rank sum tests. *P*-values for categorical variables are based on Pearson's chi-squared tests.
[b]Other race corresponds to Native Hawaiian or Other Pacific Islander, American Indian or Alaska Native, Other Race or Mixed Race, or Unknown.
[c]Within 72-h of ED arrival.

month. Figure 1 visualizes the alerts by acknowledgment reason. The most common acknowledgment reason was "Will Notify MD Immediately," which accounted for over half of all acknowledgments. Only about 5.9% of BPAs (Best Practice Alerts) were exited without acknowledgment. Responses to the BPAs remained consistent throughout the 5-month intervention period.

Interventions and patient outcomes

The results from the causal impact analysis on our primary and secondary outcomes are summarized in **Table 2**. The observed in-hospital mortality rate and

Table 2. Observed outcomes in the pre-intervention period, the expected counterfactual values from causal impact analysis, and the actual post- intervention values.

Outcome	Pre-intervention value	Expected post-intervention value (95% CI)	Actual post-intervention value
In-hospital mortality %	10.3%	11.4% (9.8%−13.0%)	9.5%
Average 72-h Change in SOFA	3.71	3.71 (3.6–3.8)	3.56
Sepsis bundle compliance rate	48.3%	48.4% (45.5%−51.0%)	53.4%
Blood cultures prior to antibiotics compliance rate	71.1%	72.0% (69.9%−73.9%)	73.9%
Rate of antibiotics administered within 24 h prior and 3 h after severe sepsis onset.	82.8%	82.8% (81.3%−84.4%)	84.6%
Rate of lactate measured within 6 h prior and 3 h after severe sepsis onset	83.5%	83.4% (81.3%−85.8%)	85.6%
Rate of repeat lactate measured within 6 h after severe sepsis onset if initial lactate is elevated	97.8%	97.3% (96.2%−98.4%)	98.6%
Rate of administration of vasoactive medications within 6 h of septic shock	58.0%	57.5% (46.7%−68.2%)	55.5%
Rate of administration of 30cc/kg of fluids within 3 h of presentation of septic shock or hypotension	54.2%	53.9% (48.9%−58.8%)	59.3%
ICU transfer rate	32.6%	32.5% (30.7%−34.2%)	31.8%
Average ICU-free days	25.4	25.1 (24.6–25.6)	25.6

Significant post-intervention values against the 95% confidence interval are bolded.

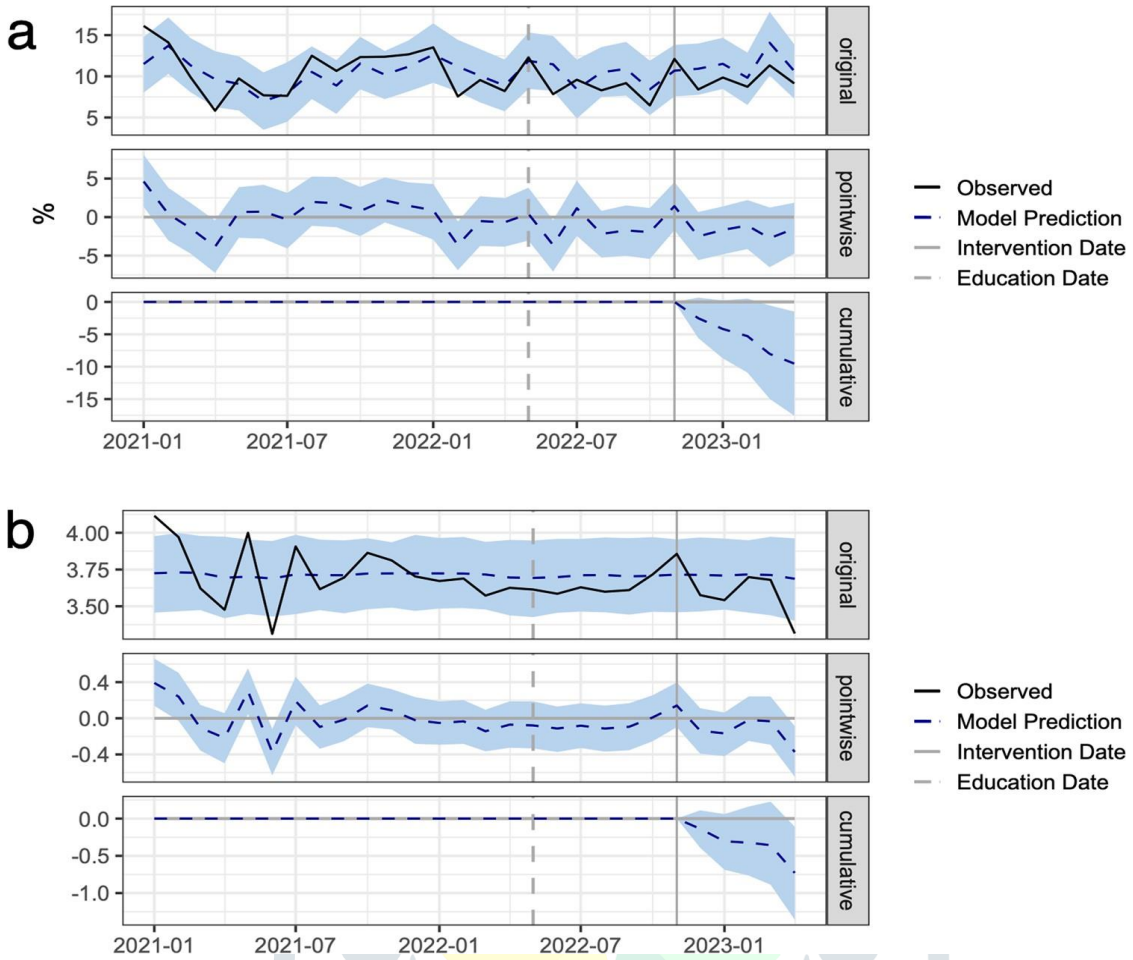

Fig. 2. Causal Impact Analysis of the COMPOSER Best Practice Advisory on Patient Outcomes. Plots illustrate the causal impact analysis using a Bayesian structural time-series model. The top subpanel ("original") shows the actual outcome (black), the average model predictions (dashed blue), and the 95% confidence limits (shaded blue) during the pre-intervention and post-intervention periods, with the intervention start indicated by the solid gray vertical line. The middle subpanel ("pointwise") depicts the difference between model predictions and observed outcomes. The bottom subpanel ("cumulative") displays the sum of pointwise differences during the post-intervention period. Preparation for the implementation of COMPOSER began in May 2022, approximately 6 months before the model's go-live date. (a) The cumulative post-intervention in-hospital sepsis mortality rate is below the 95% confidence limit. (b) The cumulative post-intervention 72-hour change in SOFA score is below the 95% confidence limit.

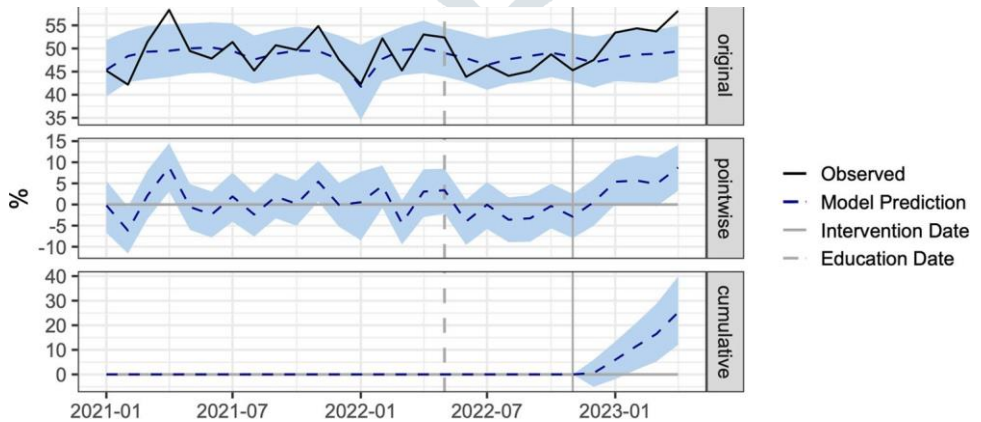

Fig. 3. Causal Impact Analysis of the COMPOSER Best Practice Advisory (BPA) on Sepsis Bundle Compliance Rate. The implementation of the COMPOSER BPA was significantly associated with an increase in sepsis bundle compliance

the corresponding predictions from the Bayesian structural time-series model are shown in **Fig. 2a**. The residual quantile-quantile and autocorrelation plots are detailed in **Supplementary Figs. 4 and 5**. The average sepsis mortality rate during the post-intervention period was 9.49%. If the COMPOSER algorithm had not been deployed, the expected counterfactual mortality rate would have been 11.39%, with a 95% confidence interval of [9.79%, 13.00%]. This corresponds to a 1.9% absolute decrease in sepsis-related in-hospital mortality, which represents a 17% relative decrease in mortality among sepsis patients and 22 additional survivors during the 5-month intervention period. The probability of this outcome occurring by chance is determined by the Bayesian one-sided tail-area probability, p=0.014p = 0.014p=0.014. Additional data regarding the difference in mortality at our two hospitals are provided in **Supplementary Figs. 6–9**. At one site (the "safety net" hospital), we observed a significant decrease in mortality in the post-intervention period, whereas no significant change was noted at the other clinical site (the quaternary care facility).

The average compliance rate during the post-intervention period was 53.42%. In the absence of the COMPOSER intervention, the expected compliance rate would have been 48.38% (95% CI, 45.46%–51.01%; **Fig. 3**), representing a 5.0% (95% CI, 2.4%–8.0%) increase in compliance. This equates to a 10% (95% CI, 5%–16%) relative increase in sepsis bundle compliance following the implementation of COMPOSER. As shown in **Supplementary Figs. 10 and 11**, compliance with our sepsis bundle increased at both emergency departments (EDs) during the intervention period. Compliance with specific bundle elements is detailed in **Table 2** and **Supplementary Figs. 12–17**, showing significant improvements in antibiotic compliance, repeat lactate compliance, and fluid administration compliance.

We also observed a reduction in the 72-hour change in SOFA score following sepsis onset (**Fig. 2b**). The average change in SOFA score during the post-intervention period was 3.56. In the absence of the COMPOSER intervention, the expected counterfactual average change in SOFA score would have been approximately 3.71, with a 95% confidence interval of [3.58, 3.83]. This represents a 4% decrease in the average 72-hour change in SOFA score following sepsis onset. The probability of this reduction occurring by chance is p=0.013p = 0.013p=0.013. Additional data on the change in SOFA score at each ED are provided in **Supplementary Figs. 18 and 19**. We also observed a downward trend in ICU admissions (**Supplementary Fig. 20**) and an upward trend in ICU-free days (**Supplementary Fig. 21**), although neither trend reached statistical significance. The temporal trends of all covariates used in the Bayesian structural time-series models are provided in **Supplementary Fig. 22**.

Associations between time-to-antibiotics in septic patients and acknowledgment reasons are shown in **Table 3**. We found that cases where nurses indicated they would notify physicians immediately had a significant reduction in time to antibiotic administration (p=0.002p = 0.002p=0.002; two-sided t-test with adjustments for ED volume, sex, baseline SOFA score, Elixhauser comorbidity score, and age).

III. DISCUSSION

In this before-and-after quasi-experimental study, we demon-Our study demonstrated that the implementation of a real-time deep-learning model to predict sepsis in two emergency departments (EDs) was associated with a 5.0% absolute increase in sepsis bundle compliance and a 1.9% absolute decrease in in-hospital sepsis-related mortality. To our knowledge, this is the first prospective use of a deep-learning model showing an association with improved patient-centered outcomes in sepsis. Our findings also suggest that utilizing such models in clinical care is associated with improvements in intermediate outcomes, such as reduced organ injury at 72 hours from sepsis onset and enhancements in sepsis bundle elements. These improvements may explain the observed mortality benefit. Importantly, in scenarios where nursing staff reported notifying providers about sepsis concerns (approximately 55% of cases), antibiotics were administered sooner, providing a plausible mechanism for the lower-than-expected in-hospital mortality we report.

Despite substantial interest in strategies to mitigate sepsis morbidity and mortality, novel therapeutics have often failed to translate into meaningful patient-centered outcomes. The potential to enhance care through artificial intelligence is promising, particularly given advances in machine learning over the past decade[22][23]. Unfortunately, most sepsis prediction algorithms do not reach clinical implementation[24]. Older models, based on clinical criteria such as SIRS criteria or hypotension, occasionally improved quality metrics (e.g., increased rates of lactate orders or time to antibiotics) but did not enhance patient-centered outcomes and had poor positive predictive value (PPV)[25][26][27].

Recent studies have employed sophisticated models across various hospitals, showing benefits for patients. Shimaburuko et al. conducted a small randomized trial of 142 ICU patients using a machine-learning algorithm to predict severe sepsis and found reductions in in-hospital mortality and length of stay in the intervention group, although this study was limited to ICU and hospital ward patients[18]. Adams et al. recently provided a prospective analysis of the TREWS model across five hospital systems, demonstrating significant decreases in mortality, organ failure, and length of stay in hospitalized patients when sepsis alerts were confirmed by a provider[16]. While not randomized, this study provides compelling evidence that proper implementation may improve patient-centered outcomes in sepsis.

In contrast, the Epic Sepsis Score (ESS), a commonly used predictive model, has not consistently demonstrated improved patient-centered outcomes. Although a small randomized quality improvement initiative at a single center found better outcomes in the ESS care group, these results have not been generalized. Researchers at the University of Michigan noted a significant drop in the ESS's test characteristics (sensitivity, specificity, PPV) compared to reported values from Epic, along with an unacceptably high rate of false positives[20]. To our knowledge, the only deep-learning model previously tested in an ED setting is the Sepsis Watch by investigators at Duke, but no patient-centered outcomes have been reported thus far[28]. Therefore, our study is the first to

report improvements in patient-centered outcomes attributable to the deployment of a deep-learning-based sepsis prediction model.

The use of deep learning for early sepsis prediction is significant because these models can capture temporal, nonlinear, and complex correlations among risk factors, enabling them to address more challenging problems. Additionally, deep-learning models can handle large volumes of multimodal data from sources such as radiology imaging, clinical notes, and wearable sensors[29][30][31][32]. These models also offer a flexible framework for transfer learning and continual learning, facilitating adaptation to local healthcare settings[33][34][35].

Importantly, the COMPOSER deep-learning model was designed to minimize false alarms through the conformal prediction framework. This approach sets a boundary around the algorithm, allowing it to determine whether it has sufficient prior knowledge of similar cases to reliably assess a patient's risk for sepsis. If the algorithm finds the data non-conformant, it adjusts its predictions acco

Table 3. Associations Between BPA Response and Antibiotics Timing.

Variable	Coefficient [95% CI]	*P* Value[a]
(Intercept)	22.42 [−45.27–90.11]	0.513
Age	0.01 [−0.15–0.18]	0.875
Male	−0.53 [−5.99–4.93]	0.848
Elixhauser Comorbidity Index	0.21 [−0.10–0.52]	0.180
Baseline SOFA	1.12 [−2.89–5.13]	0.581
Monthly ED Volume (in thousands)	0.62 [−13.33–14.58]	0.930
BPA Acknowledgement: No Infection Suspected	−13.24 [−26.60–0.127]	0.052
BPA Acknowledgement: Sepsis Treatment/Workup in Progress	−19.16 [−33.73–−4.60]	0.010
BPA Acknowledgement: Will Notify MD Immediately	−19.95 [−32.39–−7.51]	0.002

[a]*P* values are based on two-sided t-tests with adjustments for ED volume, sex, baseline SOFA, Elixhauser comorbidity score, and age.

Coefficients and *P* values are tabulated for a linear regression model with adjustments. If the nurse selected "No Infection Suspected" on an 80-year- old male with an Elixhauser of 4 and a baseline SOFA of 1, the expected "time from ED triage to antibiotics administration" would be 14.5 hours. However, if the nurse had selected "Will Notify MD Immediately" the expected "time from ED triage to antibiotics administration" would have been 7.7 hours.

To the training samples, the model will flag the case as 'indeterminate' if it does not have sufficient prior knowledge. This approach has led to a reported 75% reduction in false alarms, thereby significantly decreasing the resources and time spent on false diagnoses.

To the training samples, the model will flag the case as 'indeterminate' if it does not have sufficient prior knowledge. This approach has led to a reported 75% reduction in false alarms, thereby significantly decreasing the resources and time spent on false diagnoses.

Several factors may explain the observed reduction in mortality. First, approximately 55% of alerts were transmitted by nursing staff to physicians. In these cases, we found that patients were more likely to receive timely antibiotics, which could contribute to decreased mortality and reduced organ dysfunction at 72 hours. Artificial intelligence facilitating a shared mental model of risk between nursing staff and providers has been well-received and has improved the use of these models in other clinical areas[36]. In our system, nurses received the alerts and determined if escalation to the provider was necessary. Given that nurses frequently open patient charts, they were deemed ideal for managing these alerts. Physicians in the ED often manage 15–20 patients simultaneously and may not see a Best Practice Advisory (BPA) notification if it requires chart access. The high rate of provider notifications in our system likely benefited patient care while minimizing unnecessary alerts. Additionally, although speculative, the alert system might have enhanced situational awareness of sepsis care among our ED staff, a benefit reported with other sepsis clinical decision tools[37].

Despite these strengths, our study has several limitations. Firstly, the lack of randomization means our findings cannot establish definitive causal relationships or mechanistic insights. We performed a causal impact analysis considering common confounders, which showed significant positive associations. Nevertheless, our findings warrant further research. Secondly, our study was conducted at two EDs within a large academic center with a strong focus on sepsis and clinical informatics. While we had a large and diverse patient sample, external validation in different healthcare settings (e.g., community hospitals, varying demographics, hospitals with different IT infrastructures) is necessary. Thirdly, although abrupt interventions can raise awareness and prioritize care, their sustainability might be questioned, highlighting the need for long-term follow-up. While human interventions may suffer from fatigue and complacency, we anticipate that our automated algorithms will improve over time with more data, enhancing end-user satisfaction. Continuous education remains crucial for care optimization. Finally, we did not assess the impact on patients who did not have sepsis, such as potential adverse effects from inappropriate antibiotic use and associated healthcare costs. Additionally, there was no comparison data from the same period, as all our EDs used this model. There were no other concurrent quality improvement initiatives. Despite these limitations, we believe our findings are actionable and significant, real-time deep-learning model for sepsis prediction was associated with a significant increase in bundle compliance, a

reduction in in-hospital mortality, less organ dysfunction at 72 hours, and improved timeliness of antibiotics when nurses notified physicians of the BPA. To our knowledge, this is the first report of improved patient outcomes due to a deep-learning model for sepsis prediction. Future multicenter randomized trials are needed to validate these findings across diverse hospital and patient populations.

IV. METHODS

Study design and cohort

We conducted a prospective before-and-after quasi-experimental study to evaluate the impact of a sepsis Best Practice Advisory (BPA; **Fig. 4**) powered by the COMPOSER deep-learning model on patient outcomes and process measures. Institutional Review Board (IRB) approval was obtained from the University of California San Diego with a waiver of informed consent (IRB #805726), and additional approval was secured from the Aligning and Coordinating QUality Improvement, Research, and Evaluation (ACQUIRE) Committee (project #609). The study was conducted in accordance with STROBE guidelines[38], with a completed checklist available in **Supplementary Note 1**. The participating emergency departments (EDs) handle a combined annual volume of approximately 100,000 patients, with one located at a quaternary academic center and the other at an urban "safety net" hospital.

Patients were identified as septic according to the latest international consensus definitions for sepsis ("Sepsis-3")[1][3]. The onset time of sepsis was established using previously published methodologies, which involve evidence of organ dysfunction and clinical suspicion of infection[1][12][15]. Clinical suspicion of infection was defined by the presence of a blood culture draw and at least 4 days of non-prophylactic intravenous antibiotic therapy meeting either of the following criteria: (1) if a blood culture was ordered first, antibiotics had to be administered within the subsequent 72 hours, or (2) if antibiotics were ordered first, a blood culture had to be drawn within the next 24 hours. Evidence of organ dysfunction was defined as an increase in the Sequential Organ Failure Assessment (SOFA) score by two or more points. Specifically, organ dysfunction occurring 48 hours before to 24 hours after the suspected infection was considered, as suggested by Seymour et al.[1]. The onset time of sepsis was taken as the time of clinical suspicion of infection. The inclusion of 4 days of non-prophylactic antibiotics to enhance sepsis specificity aligns with the surveillance approach proposed by Rhee et al., which outperforms reliance on administrative coding of sepsis[39].

We included all adult patients (age ≥18 years) who met the Sepsis-3 criteria within the first 12 hours of their ED stay. Patients transitioned to comfort measures before the time of sepsis and those who developed sepsis after 12 hours of hospital admission were excluded. All data used to determine the onset time of sepsis and patient outcomes were extracted via SQL queries on Epic Clarity

Sepsis algorithm and platform

The COMPOSER algorithm for early sepsis prediction is detailed in Shashikumar et al.[15]. It is a feed-forward neural network model that integrates routinely collected laboratory data, vital signs, patient demographics (age and sex), comorbidities, and concomitant medications to generate a risk score for sepsis onset within the next 4 hours. The model employs conformal prediction methods to handle out-of-distribution samples, which might result from data entry errors or unfamiliar cases. It achieves an area under the receiver operating characteristic curve (AUROC) of 0.938–0.945 within emergency department (ED) settings[15]. We set the score threshold to ensure 80% sensitivity, which was previously shown to correspond to a positive predictive value (PPV) of 20.1%.

The COMPOSER algorithm is hosted on a cloud-based healthcare analytics platform that uses FHIR and HL7v2 standards to access data elements in real time (**Supplementary Fig. 1**)[40]. Specifically, the platform, hosted on Amazon Web Services (AWS), receives a continuous stream of Admit, Discharge, Transfer (ADT) messages from the hospital's integration engine to track active patients and their journey through care units. Data is extracted hourly for these patients using FHIR APIs with OAuth2.0 authentication and is then passed to COMPOSER. The hourly extraction frequency ensures sufficient data availability for predictions. The sepsis risk score and the top features influencing the prediction are recorded in a flowsheet within the electronic health record (EHR) via an HL7v2 outbound message. This flowsheet triggers a nurse-facing Best Practice Advisory (BPA) (**Fig. 4**), which alerts the caregiver to the risk of severe sepsis and provides the model's top reasons. The nurse can acknowledge the alert by selecting one of the following options: (i) no infection suspected, (ii) sepsis treatment/workup in progress, or (iii) will notify MD immediately. If no option is selected before exiting the patient's chart, it is recorded as "no acknowledgment." If the "will notify MD immediately" option is chosen, the nurse can use a 'secure chat' feature within the BPA to contact the provider to discuss the patient's care.

Prospectively deployed algorithms are vulnerable to model drift, where performance may degrade over time due to shifts in the patient population or changes in treatment practices[41]. To detect such drift, we implemented a data quality dashboard that monitors the median values of all input features to ensure they remain within predefined upper and lower control limits (based on quantiles from the training cohort). Additionally, we evaluate model performance, including sensitivity and PPV, on a biweekly basis to detect any performance degradation of COMPOSER (**Supplementary Note 3**). A Predetermined Change Control Plan (PCCP) is in place to trigger model retraining if performance falls below specified thresholds; however, this has not yet been necessary as of this reporting.

Implementation of COMPOSER into our electronic health record

The implementation of the COMPOSER model was conducted in stages according to the EPIS (Exploration, Preparation,

Implementation, Sustainment) framework, with frequent feedback provided to nursing staff throughout the implementation and sustainment phases[42]. The Exploration phase began 2–3 years before the model's deployment, with significant institutional support at both departmental and health system levels. Preparation commenced approximately 6 months before the go-live date and included forming a multidisciplinary team to guide the implementation, conducting surveys to identify nursing staff needs, providing educational sessions, and making iterative changes to the Best Practice Advisory (BPA) based on end-user feedback.

During the Preparation phase, a "silent mode trial" was conducted where COMPOSER outputs were reviewed in real-time by a team of physicians to evaluate the accuracy and usefulness of the alerts. Adjustments were made iteratively based on these reviews to enhance the timeliness and appropriateness of the alerts.

In the Implementation phase, ongoing feedback and education were provided to nursing staff regarding the COMPOSER model. The BPA was set to trigger for all adult patients (aged 18 years or older) in the ED who had a risk score above the threshold, with the following exclusions: patients discharged or deceased, those with comfort care measures initiated, patients no longer under the care of ED nurses, or those who had already had a sepsis bundle instituted during their stay. The BPA would only trigger if the patient's chart was open. Lockout periods for each acknowledgment reason were as follows: "No infection suspected" - 8 hours; "Will notify MD immediately" - 12 hours; "Sepsis treatment/work-up in progress" - 12 hours.

Our pre-intervention period was defined from January 1, 2021, to December 6, 2022. COMPOSER went live on December 7, 2022, and the post-implementation phase extended until April 30, 2023.

Primary and Secondary Outcomes

Our primary outcome was in-hospital mortality. Secondary outcomes included:

- Compliance with the sepsis bundle, which consists of: initial and repeat serum lactate measurements if the initial lactate was >2 mmol/L, initiation and completion of a 30 mL/kg crystalloid fluid bolus, blood cultures checked prior to antibiotics, and initiation of intravenous antibiotics within 3 hours of sepsis onset.
- The 72-hour change in Sequential Organ Failure Assessment (SOFA) score following sepsis onset.
- ICU admission and ICU-free days.

ICU-free days were calculated as 30 minus the number of ICU days, with in-hospital death and stays longer than 30 days fixed at 0. For patients who died in the ICU, this value was 0. For example, a patient who spends 4 days in the ICU and survives would have an ICU-free day value of 26. Patients who either die in the ICU or stay in the ICU for more than 29 days would have a value of 0.

Statistical methods

Descriptive statistics were provided as indicated. Differences between the pre-intervention and post-intervention cohorts were assessed using Kruskal-Wallis rank sum tests for continuous variables and Pearson's chi-squared tests for categorical variables, with significance set at a P-value of 0.05. All statistical analyses were performed using R statistical software version 4.0.4 and the CausalImpact package version 1.2.7[43][44].

To estimate the causal effect of the COMPOSER Best Practice Advisory (BPA) intervention, we employed a Bayesian structural time-series model[44]. This approach, developed by Brodersen et al. from Google Inc., is widely used to assess the impact of advertisement campaigns on product sales and the effect of economic changes on markets[45][47]. Here, we apply it to patient outcomes data to evaluate the impact of the COMPOSER algorithm while adjusting for confounders.

Briefly, a state-space model is trained on the control time-series prior to the intervention. The observed outcome is modeled as a function of the latent state and Gaussian noise, with the latent state represented by a local linear trend and a linear regression on the model covariates. We assume the effect of the regression coefficients on the outcome is independent of time, with static regression parameters sampled from a spike-and-slab prior distribution. Posterior inference is performed on the post-intervention time-series to estimate the counterfactual outcome if the intervention had not been introduced. Under the assumption that the response variable in the control time-series is independent of the intervention, the difference between the model prediction and the observed value reflects the probability density of the causal impact of the intervention over time (see Supplementary Note 3 for further details). Additionally, model residuals were assessed using quantile-quantile and autocorrelation plots to ensure adequate modeling of the time-series data.

Covariates included in the Bayesian structural time-series model were emergency department (ED) volume, sex, baseline SOFA score, comorbidity burden via the Elixhauser comorbidity score, age, COVID-19 infection status, ED location (La Jolla or Hillcrest), local trends, and season. We used 1000 samples of Markov Chain Monte Carlo for posterior inference. No imputation was performed as all covariates were fully observed. There were no missing values in the covariates. Seasons and local trends (e.g., ED volume) were included based on prior data suggesting that sepsis outcomes may be worse in winter and affected by high patient volumes[48][49]. Outcomes were predicted at a monthly resolution to mitigate the impact of random fluctuations on the outcome variable. We plotted model predictions, true outcomes data, the pointwise difference between the two, and the cumulative difference across the post-intervention period, and assessed significance against 95% confidence intervals.

We further evaluated the association between alert acknowledgement and sepsis intervention by measuring the time

from ED triage to the administration of antibiotics. Adjustments were made for the aforementioned confounders, and a two-sided t-test was performed on time-to-antibiotics as a function of the acknowledgement reason.

REFERENCES

1. Singer, M. et al. The Third International Consensus Definitions for sepsis and septic shock (sepsis-3). *JAMA* 315, 801–810 (2016).
2. Rudd, K. E. et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *Lancet* 395, 200–211 (2020).
3. Rhodes, A. et al. Surviving sepsis campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Med* 43, 304–377 (2017).
4. Kumar, A. et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit. Care Med* 34, 1589–1596 (2006).
5. Ferrer, R. 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* 42, 1749–1755 (2014).
6. Liu, V. X. et al. The timing of early antibiotics and hospital mortality in sepsis. *Am. J. Respir. Crit. Care Med* 196, 856–863 (2017).
7. Peltan, I. D. et al. ED door-to-antibiotic time and long-term mortality in sepsis. *Chest* 155, 938–946 (2019).
8. Chamberlain, D. J., Willis, E. M. & Bersten, A. B. The severe sepsis bundles as processes of care: a meta-analysis. *Aust. Crit. Care J. Confed. Aust. Crit. Care Nurses* 24, 229–243 (2011).
9. Centers for Medicare & Medicaid Services. QualityNet—inpatient hospitals specifications manual Version 5.13 (2023). https://www.qualitynet.org/inpatient/ specifications-manuals.
10. Reyna, M. A. et al. Early prediction of sepsis from clinical data: The PhysioNet/Computing in Cardiology Challenge 2019. *Crit. Care Med.* 48, 210–217 (2020).
11. Shashikumar, S. P., Li, Q., Clifford, G. D. & Nemati, S. Multiscale network representation of physiological time series for early prediction of sepsis. *Physiol. Meas.* 38, 2235–2248 (2017).
12. Lauritsen, S. M. et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records. *Nat. Commun.* 11, 3852 (2020).
13. Nemati, S. et al. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit. Care Med.* 46, 547–553 (2018).
14. Henry, K. E., Hager, D. N., Pronovost, P. J. & Saria, S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci. Transl. Med.* 7, 299ra122 (2015).
15. Shashikumar, S. P., Wardi, G., Malhotra, A. & Nemati, S. Artificial intelligence sepsis prediction algorithm learns to say "I don't know. *NPJ Digit Med.* 4, 134 (2021). Adams, R. et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. *Nat. Med.* 28, 1455–1460 (2022).
16. Giannini, H. M. et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit. Care Med.* 47, 1485–1492 (2019).
17. Shimabukuro, D. W., Barton, C. W., Feldman, M. D., Mataraso, S. J. & Das, R. 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.* 4, e000234 (2017).
18. McCoy, A. & Das, R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. *BMJ Open Qual.* 6, e000158 (2017).
19. Wong, A. et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med*. Published online June 21. https://doi.org/10.1001/jamainternmed.2021.2626 (2021).
20. Lyons, P. G. et al. Factors associated with variability in the performance of a proprietary sepsis prediction model across 9 networked hospitals in the US. *JAMA Intern. Med.* 183, 611–612 (2023).
21. Wardi, G. et al. Bringing the promise of artificial intelligence to critical care: what the experience with sepsis analytics can teach us. *Crit. Care Med.* 51, 985–991 (2023).
22. Classen, D.C., Longhurst, C. & Thomas, E.J. Bending the patient safety curve: how much can AI help?. npj Digit. Med. 6, 2 (2023). https://doi.org/10.1038/s41746- 022-00731-5.
23. Bektaş, M., Tuynman, J. B., Costa Pereira, J., Burchell, G. L. & van der Peet, D. L. Machine learning algorithms for predicting surgical outcomes after colorectal surgery: a systematic review. *World J. Surg.* 46, 3100–3110 (2022).
24. Narayanan, N., Gross, A. K., Pintens, M., Fee, C. & MacDougall, C. Effect of an electronic medical record alert for severe sepsis among ED patients. *Am. J. Emerg. Med*. 34, 185–188 (2016).
25. Berger, T., Birnbaum, A., Bijur, P., Kuperman, G. & Gennis, P. A computerized alert screening for severe sepsis in emergency department patients increases lactate testing but does not improve inpatient mortality. *Appl. Clin. Inf.* 1, 394–407 (2010).
26. Makam, A. N., Nguyen, O. K. & Auerbach, A. D. Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: a systematic review. *J. Hosp. Med*. 10, 396–402 (2015).
27. Sendak, M. P. et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. *JMIR Med. Inf.* 8, e15182 (2020).
28. Carlile, M. et al. Deployment of artificial intelligence for radiographic diagnosis of COVID-19 pneumonia in the emergency department. *J. Am. Coll. Emerg. Physicians Open* 1, 1459–1464 (2020).
29. Amrollahi, F., Shashikumar, S. P., Razmi, F. & Nemati, S. Contextual embeddings from clinical notes improves prediction of sepsis. *AMIA Annu Symp. Proc. AMIA Symp.* 2020, 197–202 (2020).
30. Goh, K. H. et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat. Commun.* 12, 711 (2021).
31. Amrollahi, F. et al. Predicting Hospital Readmission among Patients with Sepsis Using Clinical and Wearable Data. *Annu Int Conf IEEE Eng Med Biol Soc.* 2023, 1–4 (2023).
32. Wardi, G. et al. Predicting progression to septic shock in the emergency department using an externally generalizable machine-learning algorithm. *Ann. Emerg. Med* 77, 395–406 (2021).
33. Holder, A. L., Shashikumar, S. P., Wardi, G., Buchman, T. G. & Nemati, S. A locally optimized data-driven tool to predict sepsis-associated vasopressor use in the ICU. *Crit. Care Med.* 49, e1196–e1205 (2021).
34. Amrollahi, F., Shashikumar, S. P., Holder, A. L. & Nemati, S. Leveraging clinical data across healthcare institutions for continual learning of predictive risk models. *Sci. Rep.* 12, 8380 (2022).
35. Li R. C., et al. Using AI to empower collaborative team workflows: Two implementations for advance care planning and care escalation. *NEJM Catal*. 2022;3. https://doi.org/10.1056/CAT.21.0457.
36. Gibbs, K. D. et al. Evaluation of a sepsis alert in the pediatric acute care setting. *Appl. Clin. Inf.* 12, 469–478 (2021).
37. von Elm, E. et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 370, 1453–1457 (2007).
38. Rhee, C. et al. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. *JAMA* 318, 1241–1249 (2017).
39. Boussina, A. et al. "Development & Deployment of a Real-time Healthcare Predictive Analytics Platform," 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023, 1-4, https://doi.org/10.1109/EMBC40787.2023.10340351.
40. Davis, S. E., Greevy, R. A., Lasko, T. A., Walsh, C. G. & Matheny, M. E. Detection of calibration drift in clinical prediction models to inform model updating. *J. Biomed. Inf.* 112, 103611 (2020).
41. Moullin, J. C., Dickson, K. S., Stadnick, N. A., Rabin, B. & Aarons, G. A. Systematic review of the exploration, preparation, implementation, sustainment (EPIS) framework. *Implement Sci. IS* 14, 1 (2019).
42. R Core Team. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing https://www.R-project.org (2021)
43. Brodersen, K. H., Gallusser F., Koehler J., Remy N., Scott S. L. Inferring causal impact using Bayesian structural time-series models. *Ann. Appl. Stat*. 9, 247–274 (2015).
44. Takyi, P. O. & Bentum-Ennin, I. The impact of COVID-19 on stock market performance in Africa: A Bayesian structural time series approach. *J. Econ. Bus.* 115, 105968 (2021).
45. Jalan, A., Matkovskyy, R. & Urquhart, A. What effect did the introduction of Bitcoin futures have on the Bitcoin spot market? *Eur. J. Financ.* 27, 1251–1281 (2021).
46. Martin W., Sarro F. & Harman M. Causal impact analysis for app releases in Google Play. In: *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*. 435–446 (ACM, 2016).
47. Saini, Mukesh Kumar, and Yash Pal. "A STUDY OF IMPACT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN HEALTH AND HEALTH CARE." (2017).
48. Danai, P. A., Sinha, S., Moss, M., Haber, M. J. & Martin, G. S. Seasonal variation in the epidemiology of sepsis. *Crit. Care Med.* 35, 410–415 (2007).
49. Woodworth, L. Swamped: emergency department crowding and patient mortality. *J. Health Econ*. 70, 102279 (2020).