nature medicine

ARTICLES

https://doi.org/10.1038/s41591-022-01895-z

Check for updates

Factors driving provider adoption of the TREWS machine learning-based early warning system and its effects on sepsis treatment timing

Katharine E. Henry[1,2,16], Roy Adams[2,3,16], Cassandra Parent[4,16], Hossein Soleimani[5], Anirudh Sridharan[6], Lauren Johnson[7], David N. Hager[8], Sara E. Cosgrove[8], Andrew Markowski[9], Eili Y. Klein[10], Edward S. Chen[8], Mustapha O. Saheed[10], Maureen Henley[7], Sheila Miranda[11], Katrina Houston[7], Robert C. Linton II[6], Anushree R. Ahluwalia[7], Albert W. Wu[8,12,13,14] ✉ and Suchi Saria[1,2,8,12,15] ✉

Machine learning-based clinical decision support tools for sepsis create opportunities to identify at-risk patients and initiate treatments at early time points, which is critical for improving sepsis outcomes. In view of the increasing use of such systems, better understanding of how they are adopted and used by healthcare providers is needed. Here, we analyzed provider interactions with a sepsis early detection tool (Targeted Real-time Early Warning System), which was deployed at five hospitals over a 2-year period. Among 9,805 retrospectively identified sepsis cases, the early detection tool achieved high sensitivity (82% of sepsis cases were identified) and a high rate of adoption: 89% of all alerts by the system were evaluated by a physician or advanced practice provider and 38% of evaluated alerts were confirmed by a provider. Adjusting for patient presentation and severity, patients with sepsis whose alert was confirmed by a provider within 3 h had a 1.85-h (95% CI 1.66–2.00) reduction in median time to first antibiotic order compared to patients with sepsis whose alert was either dismissed, confirmed more than 3 h after the alert or never addressed in the system. Finally, we found that emergency department providers and providers who had previous interactions with an alert were more likely to interact with alerts, as well as to confirm alerts on retrospectively identified patients with sepsis. Beyond efforts to improve the performance of early warning systems, efforts to improve adoption are essential to their clinical impact and should focus on understanding providers' knowledge of, experience with and attitudes toward such systems.

Clinical decision support (CDS) tools that leverage machine learning techniques are becoming more common. They have been used to facilitate early recognition of disease states, reduce diagnostic errors and improve patient outcomes[1–4]. Of particular interest are tools that can identify at-risk patients early in the progression of a disease, allowing providers to intervene earlier and potentially improve outcomes. While traditional CDS tools use a small number of criteria to assess patient risk, tools informed by machine learning techniques use large amounts of high-dimensional historical data to learn patterns indicative of the disease of interest[1,5]. They can also incorporate individual-specific features (such as comorbid conditions and patient history) in the algorithm. In retrospective evaluations, these systems are generally more precise and identify patients earlier in their disease trajectory[6–10]. Improved identification of disease, however, contributes little if the tool is not adopted and used by providers[11–15], making user adoption key to improving patient outcomes. Studies to date have shown limited success gaining widespread adoption[16–22], with users typically responding to only 6–45% of alerts or requiring dedicated staff to review alerts and having low to moderate impact on provider practice[23–26]; however, as these evaluations typically combine the questions of adoption and clinical impact, it is challenging to distinguish the extent to which each contributes to the limited impact on provider practice. Moreover, there is limited evidence on how best to design and integrate such tools to improve adoption and increase impact on clinical practice.

Adoption of automated systems in non-clinical settings depends on several factors, including personal characteristics and preferences of the user, characteristics of the automated system (such as a CDS tool) and the environment in which the technology is used[27]. In clinical simulations in a 'laboratory' setting, where providers are shown simulated CDS recommendations for exemplar patients, studies have found that interface design[28], provider expertise[29] and clinical time constraints[30] all play a role in adoption of the tool;

[1]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA. [2]Malone Center for Engineering in Healthcare, Johns Hopkins University, Baltimore, MD, USA. [3]Department of Psychiatry and Behavioral Science, Johns Hopkins School of Medicine, Baltimore, MD, USA. [4]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA. [5]Health Informatics, University of California, San Francisco, San Francisco, CA, USA. [6]Howard County General Hospital, Columbia, MD, USA. [7]Department of Quality Improvement, The Johns Hopkins Hospital, Baltimore, MD, USA. [8]Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA. [9]Suburban Hospital, Bethesda, MD, USA. [10]Department of Emergency Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA. [11]Department of Medicine, The Johns Hopkins Hospital, Baltimore, MD, USA. [12]Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [13]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [14]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [15]Bayesian Health, New York, NY, USA. [16]These authors contributed equally: Katharine E. Henry, Roy Adams, Cassandra Parent. ✉e-mail: awu@jhu.edu; ssaria@cs.jhu.edu

however, in the real-world clinical setting, there are additional barriers to system adoption, including unpredictable variations in workflow, changes in personnel and high-stakes consequences of incorrect decisions that are difficult to replicate in simulations[31]. In this study, we sought to identify which patient, provider and environmental factors were associated with adoption of a CDS tool in the real-world setting and could be modified to increase adoption of these systems.

In this study, we examined the predictive performance and clinical adoption of a deployed CDS tool for early recognition of sepsis called the Targeted Real-time Early Warning System (TREWS). Early recognition of sepsis is critical for successful treatment and, in particular, early administration of antibiotics is associated with decreased mortality[32–34]. Using electronic health record (EHR) data collected following the initial deployment of TREWS, we measured adoption of TREWS based on the degree to which providers reviewed and confirmed alerts within the EHR system. Using these data we analyzed the extent to which those actions were associated with changes in the timing of antibiotic ordering, as well as various patient, provider and environmental factors. As analyzing the impact of the tool on patient outcomes requires a separate study design, we have prepared a companion manuscript to address this question[35]. While the current manuscript analyzes the association between provider response to the alert and the timing of antibiotic orders, the companion manuscript examines the association between provider response to the alert and patient outcomes[35].

Results

Retrospective performance characterization of TREWS. To provide context, we first characterized how well TREWS identified EHR-confirmed sepsis in a retrospective, pre-deployment cohort composed of patients admitted to one academic and two community hospitals in the Maryland/Washington DC area between January 2016 and March 2018. Of 173,931 patient encounters included in the retrospective cohort, 3,858 sepsis cases were retrospectively identified using EHR-based sepsis phenotyping, which, consistent with the third sepsis consensus definition of sepsis, identifies sepsis cases based on the co-occurrence of suspicion of infection and related organ dysfunction, while also accounting for common sources of confounding (Online Methods provides the details)[36–38]. In the retrospective cohort, the model identified sepsis with an area under the curve (AUC) of 0.97. At a sensitivity of 0.8, the model had a positive predictive value (PPV) of 0.27 and 7% of patient encounters met alert criteria. The complete receiver operating characteristic and sensitivity–PPV curves are shown in Extended Data Fig. 1. Patients with sepsis not immediately recognized upon admission were identified a median 3.6 h before first antibiotic order and 5.7 h before first antibiotic order among patients with sepsis who died in hospital (Extended Data Table 1). The timing of the alert relative to the patient's first antibiotic order is further summarized in Extended Data Table 1.

Description of the deployed TREWS alert system. Beginning in 2018, TREWS was deployed to several hospitals in the Johns Hopkins Health System. When a TREWS alert occurred on a patient, the alert was passively displayed in the EHR as a clickable icon on the emergency department (ED) track board or the patient list. The bedside provider (physician or advanced practice provider) then had the option to open the alert and view the tool's analysis, which included current indicators of organ dysfunction and a list of factors considered by the model (Methods provides additional details about the model, deployment and interface). The bedside provider could then choose to enter an evaluation via the TREWS interface indicating whether or not they believed the patient had sepsis at that time (Methods provides further details).

Definition of adoption. A primary goal of TREWS is to trigger providers to open the page and review the alert (referred to here as an evaluation) and to enter in the page whether the patient had sepsis (referred to here as confirmation or dismissal) to prompt earlier consideration of treatment for patients deemed to be septic. Our primary measure of adoption, therefore, was whether or not the provider (either a physician or advanced practice provider) entered a patient evaluation (either confirmed as having sepsis or dismissed as not having sepsis) within the tool following the alert. We considered an evaluation 'timely' if it was entered within 3 h after the alert appeared in the EHR. The 3-h window was chosen to match the treatment window recommended by the Centers for Medicare and Medicaid Services (CMS) sepsis core measure (SEP-1) and the Surviving Sepsis Campaign guidelines[38–40]. As providers were not required to respond to alerts within the TREWS interface, this definition may not capture all patient evaluations resulting from the TREWS alerts. For example, a provider may see the alert in the EHR and choose to document and initiate sepsis treatment without documenting it in the tool interface; however, as the majority of alerts had an evaluation entered during our study (see 'study question 1' below), we consider this to be a strong proxy measure. We used the percent of alerts evaluated and the percent confirmed as our primary measures of system adoption throughout.

Study question 1: study population and overall adoption. During the post-deployment study period, the TREWS system screened 469,419 patient encounters (Fig. 1). Screened encounters included all patients who presented to the ED as well as those who were admitted to an observation or inpatient unit. Overall, the system flagged 31,591 (6.7%) patient encounters for sepsis screening; average daily alert counts for each hospital are shown in Extended Data Table 2. Among screened patients, 9,805 (2.1%) were retrospectively identified as having sepsis using EHR-based sepsis phenotyping (Fig. 1)[36,37]. Of the patient encounters with sepsis, 8,033 (82%) were flagged by the tool. The sample characteristics for these encounters are reported in Extended Data Table 3.

To assess the extent to which clinicians adopted the alert, we measured the percentage of patients with an alert who had an evaluation entered within the TREWS page. Among all patient encounters with an alert, 28,243 (89%) had a provider evaluation entered in the TREWS page, with 16,768 (53%) and 22,982 (73%) receiving evaluation within 1 h and 3 h, respectively (Table 1). A total of 1,965 unique providers entered at least one evaluation in TREWS during the study period. Alerts on patients with EHR-confirmed sepsis were evaluated at similar rates to the general population. Of the patients who had their alert evaluated, 10,644 (38%) had their alert confirmed (the provider recorded that the patient had sepsis at the time of evaluation) (Table 1). Among patients who had their alert evaluated and were retrospectively identified as having sepsis, 5,388 (71%) had their alert confirmed (Table 1). The rate of confirmation among alerts on sepsis patients was similar across all time windows considered.

Study question 2: timing of antibiotics relative to alerts. To assess the association between tool adoption and patient care, we examined the extent to which recording an evaluation for sepsis within 3 h after the alert was associated with the timing of a patient's first antibiotic order, a key element of sepsis treatment[33,34,41]. All results were adjusted for patient demographics, medical history, laboratory measurements, vital signs, comorbidities and admitting hospital (Methods). A total of 3,775 patients had an alert, had sepsis and had antibiotics first ordered and administered in the 24 h following their alert and were thus included in our primary analysis (Fig. 1). Compared to patients for whom a timely evaluation was not entered, patients with a timely evaluation entered had a 1.12-h (95% CI 0.87−1.30) lower adjusted median time from alert to first

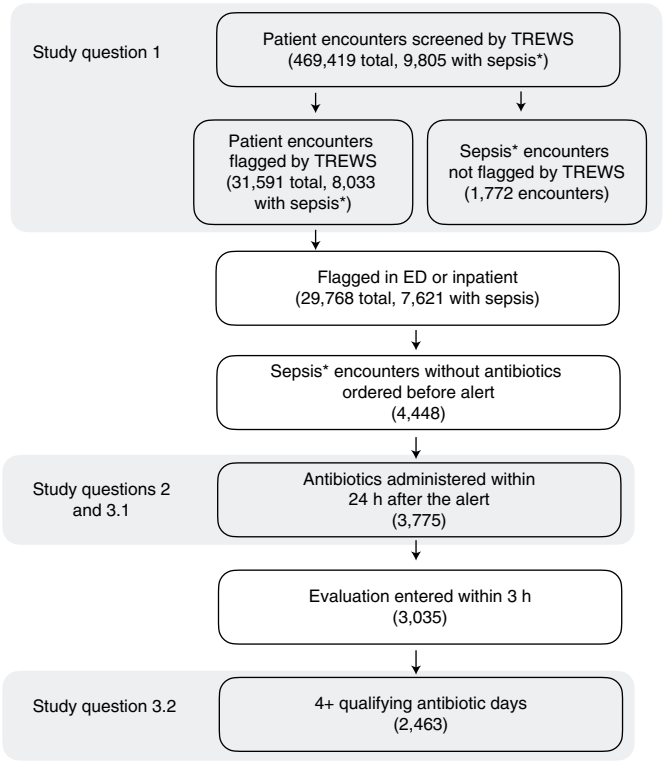


Fig. 1 | Included study population by study question. The waterfall diagram shows the included population for each study question. Study question 1 included 469,419 screened patients. This included all patients presenting to the ED or admitted to an observation or inpatient unit. Study questions 2 and 3.1 included 3,775 patients with sepsis who received an alert and who had no antibiotic orders before the alert, but who received antibiotics within 24 h after the alert. Study question 3.2 included the 2,463 of these patients who had an evaluation of their alert entered by a provider within 3 h of the alert and who also received antibiotic treatment over the course of 4 d or more.

antibiotic order (Table 2). Further, patients whose alert was confirmed (not just evaluated) within 3 h had a 1.85-h (95% CI 1.66–2.00) lower adjusted median time from alert to antibiotic order compared to patients who either did not have a timely evaluation or had their alert dismissed (Table 2).

Study question 3.1: factors associated with alert adoption. To further understand alert adoption, we examined which patient, provider and environmental factors were associated with timely alert evaluation among patients with sepsis. A complete list of factors considered is provided in Table 3. For each of these factors, we estimated the association between that factor and whether the alert was evaluated within 3 h, adjusting for all other considered factors. Among the 3,775 patients who met the criteria for our primary analysis, 3,035 had an evaluation entered within 3 h and 740 did not. Among patient factors (Table 3), only advanced age (>70 years) was significantly associated with an increased likelihood of entering a timely evaluation (Table 4). Among environmental factors (Table 3), alerts occurring between 7:00 and 15:00 were associated with increased likelihood of timely evaluation. Alerts between 15:00 and 23:00 and 23:00 and 3:00 and high admission volumes were not significantly associated with timely evaluation. A high number of alerts in the previous 24 h ('alert volume') significantly decreased the likelihood of timely evaluation. Provider factors (Table 3) had the strongest associations with timely evaluation, with ED providers and providers with a recent interaction with the alert having the highest likelihood of entering a timely evaluation with adjusted risk ratios of 1.22 (95% CI 1.14–1.32) and 1.22 (95% CI 1.19–1.26), respectively.

Study question 3.2: Factors associated with alert dismissal. Even when a provider responds to an alert, sepsis may not be immediately recognized or providers may prefer to manage sepsis without assistance from the system. We defined alert dismissal as occurring when a provider entered the TREWS page and entered an evaluation indicating that the patient did not currently have sepsis. Restricting our population to retrospectively identified patients with sepsis who received a substantial antibiotic course (4+ continuous days of an antibiotic or antibiotics until death or transfer to an acute care facility), indicating suspicion of infection, we examined which patient, provider and environmental factors were associated with alert dismissal on patients with sepsis (Table 3). Among the alerts on included patients with EHR-confirmed sepsis ($n=7,621$), 2,463 received a timely evaluation and met the additional 4+ antibiotic day restriction (1,751 confirmed alerts and 712 dismissed alerts). Among patient factors, the absence of key sepsis symptoms and younger age were significantly associated with an increase in the likelihood of dismissing evaluated alerts (Table 5). High acute general severity (Table 3) was also associated with an increase in the likelihood of dismissing the alert. Other patient factors were not significantly associated with alert dismissal. Among provider factors, working in the ED and recent interactions with alerts were both associated with decreased likelihood of dismissal and, among environmental factors, alerts occurring during the evening or overnight shifts (15:00–23:00 or 23:00–7:00) were more likely to be dismissed.

Discussion

In this study, we characterized the adoption and clinical impact of TREWS, a machine learning-based CDS system for recognizing and treating sepsis early in its progression, and evaluated the extent to which patient presentation, environmental and provider-related factors were associated with provider response to the alert. Based on a retrospective evaluation, TREWS achieved an AUC of 0.97 and alerts triggered an average of 3.6 h before antibiotic ordering for sepsis patients not immediately recognized and treated upon admission. Throughout the prospective deployment, TREWS was adopted at a high rate, with providers entering evaluations for 89% of alerts (73% of alerts within 3 h) and with 37–38% of those patients confirmed by the provider as having sepsis. Note that the deployed alert had a higher confirmation rate than might be expected based on the retrospective PPV (27%). This may reflect tuning of the alert threshold that occurred during the deployment process, as well as providers using the alert interface to document suspected rather than confirmed sepsis. In a separate analysis, we found that the sensitivity and PPV of the deployed alert was similar between sex and racial groups, but the rate of alert confirmation by providers differed across racial groups[42].

Timely confirmation of alerts was associated with a shorter time from alert to first antibiotic order among patients with sepsis (−1.85 h; 95% CI −2.00 to −1.66). The observed reduction in time to antibiotics suggests that use of TREWS as intended can lead to faster treatment among patients with sepsis[33,34,43]. In a companion paper, we examined the association between provider response to TREWS and patient outcomes[35]. Analysis of the associations between patient presentation, alert environment and provider characteristics and real-time provider response to alerts, showed that provider characteristics had the strongest association with the decision to evaluate the alert; however, among alerts with a timely evaluation, certain patient, provider and environmental factors were significantly associated with a provider's confirmation of the alert.

Table 1 | Provider interaction with the TREWS interface following an alert

	Alerts with provider evaluation entered ($n = 31,591$)		Alerts on sepsis cases* ($n = 8,033$)	
Time from alert to response	All alerts (% of alerts)	Confirmed alerts (% of evaluated alerts)	All alerts (% of alerts on sepsis cases)	Confirmed alerts (% of evaluated alerts on sepsis cases)
Within 1 h	16,768 (53%)	6,184 (37%)	4,343 (54%)	3,162 (73%)
Within 3 h	22,982 (73%)	8,587 (37%)	5,943 (74%)	4,311 (73%)
Within 6 h	25,020 (79%)	9,337 (37%)	6,485 (81%)	4,680 (72%)
Ever	28,243 (89%)	10,644 (38%)	7,603 (95%)	5,388 (71%)

*Identified retrospectively

Table 2 | Association between response to alert and time from alert to first antibiotic order

Difference in median hours from alert to antibiotics between:	Unadjusted difference, hours (95% CI)	Adjusted difference, hours (95% CI)
Evaluation entered within 3 h ($n = 22,982$) **— no evaluation entered within 3 h** ($n = 8,609$)	−1.28 (−1.50 to −1.02)	−1.12 (−1.30 to −0.87)
Alert confirmed within 3 h ($n = 8,587$) **– not confirmed within 3 h** ($n = 14,395$)	−1.90 (−2.02 to −1.74)	−1.85 (−2.00 to −1.66)

In sepsis, based on promising retrospective validation, a growing number of tools have been deployed to clinical settings[21,44–49]. A subset of these have shown impact on treatment processes[44,46,47] but all relied on dedicated staff to manage the high alert volumes and false alarm rates. General deterioration alerts such as the system described by Escobar et al. have also shown promise for influencing care; however, the system described by Escobar et al. is not specific to sepsis and requires a centralized deployment strategy using dedicated staff[26]. Employing dedicated staff can ensure adoption, but poses challenges for scaling CDS to monitoring multiple conditions and may not be possible at sites with fewer resources. Instead, deploying reliable CDS with low alert volumes that are designed to integrate into the clinical workflow and encourage adoption can enable bedside implementation that improves responsiveness and alert value, while both reducing alert burden and the cost of additional staff[30,50,51]; however, this introduces the question of whether clinicians would in fact use and adopt such a system and what factors may impact that use.

The high overall rate of provider response to the TREWS alert observed in this study (a provider entered an evaluation in response to 89% of alerts) is promising given the documented challenges to gaining adoption of such systems[21,22,52–55]. It demonstrates that a bedside alert system can be used to disseminate clinical alerts and still attain high rates of adoption. As alert burden and the perceived accuracy of a CDS tool both play major roles in tool adoption and trust[16,17,31]; one reason for the high observed adoption of TREWS may be the high predictive performance and low alert burden of TREWS relative to comparable deployed systems. Even with a sensitivity of 82%, precision was high with one in three evaluated alerts confirmed by a provider to be sepsis. Past deployed systems have reported significantly lower predictive performance on similar hospital populations[21,48,55]. For example, one of the most widely deployed sepsis early warning systems had a sensitivity of only 33% and a precision of 2.4% (1 in 46 alerts within 24 h of sepsis onset)[48]. Additionally, ease of use and integration into the workflow have been noted as important factors influencing adoption[28,30,56,57]. Availability of TREWS within a provider's EHR workflow and the inclusion of alert context to avoid 'black box' presentation may also have improved overall adoption of the tool.

Provider characteristics had the strongest association with the likelihood of evaluating a TREWS alert. While to a lesser extent, environmental factors such as time of day were also associated with the likelihood of evaluation, we did not find associations between patient presentation variables and alert evaluation. Providers who work in the ED or who had previously interacted with the tool and entered an evaluation, were most likely to evaluate a new alert. There are several possible reasons for these results. Some providers may be more willing to adopt new CDS tools than others; this tendency is sometimes referred to as 'dispositional trust'[58,59]. Additionally, increased familiarity with the system may add to its perceived ease of use or accuracy. As most first alerts occurred in the ED, those providers may naturally get more exposure to TREWS and be more familiar with the system. This is an example of learned trust[59]. Alternatively, the higher patient load and greater degree of uncertainty around patients in the ED may increase provider willingness to utilize the alert, which is an example of situational trust[59]. Creating opportunities to interact with and practice using TREWS in a simulated setting or adapting the alert policy and interface design for different types of providers could help increase familiarity and increase adoption.

The lack of an association between most patient factors and the likelihood of a provider entering a timely evaluation could also be viewed as promising (Table 4). It suggests that providers are willing to engage with the system even in cases that do not display an obvious presentation of sepsis; however, patient presentation was associated with alert dismissal in patients with sepsis. We found that alerts occurring on patients with sepsis who did not have specific key sepsis symptoms or with higher acute complexity at the time of the alert, were more likely to have their alert dismissed. It makes sense that alerts are more likely to be confirmed when there is clear support for the diagnosis and a lack of alternate explanation; however, this may pose a problem in cases where patients have less typical presentations of sepsis or where the alert occurs in advance of those symptoms developing. Further, if TREWS is perceived as less accurate in cases with high general acute severity, adoption may be lower in these cases as well. Education to increase awareness about alternative presentations of sepsis, or situations where patient complexity may mask developing sepsis symptoms, may help improve provider trust that the alerts are delivering valuable information.

Among environmental factors, alert dismissal (evaluating the alert and indicating that the patient did not have sepsis) on patients retrospectively identified as having sepsis was most strongly associated with time of day, with alerts occurring during the time ranges of 15:00–23:00 or 23:00–7:00 more likely to be evaluated and dismissed, even after accounting for patient presentation (Table 5). This may reflect an association between time of day and unit volume. Greater workload during the later shifts could contribute to

Table 3 | Potential factors influencing provider response to alerts

Factor	Definition	Rationale for inclusion
Patient presentation factors		
Absence of key sepsis symptoms	True if no more than one of the following were met before the alert: lactate > 2.0 mmol/L, WBC > 12 10^9/L or WBC < 4 10^9/L and temperature >38.0 °C or temperature <36.0 °C	These three criteria are commonly associated with infection and sepsis. Providers may be more willing to dismiss alerts that present without multiple of these symptoms
Alternative diagnosis	True if any of the following diagnoses were made during the patient's stay based on the presence of ICD-10 codes: myocardial infarction, stroke, acute respiratory failure	Presence of an alternative diagnosis may increase the complexity of the diagnostic process by masking sepsis symptoms
Condition at risk for fluid overload	True if any of the following chronic conditions were present based on the presence of ICD-10 codes: COPD, CKD and CHF	Confirming the alert is related to initiation of the sepsis bundle. Providers may dismiss the alert on patients who are at risk for fluid overload because they do not want to initiate the sepsis bundle fluid requirement
Acute general severity	The adjustment used the raw SAPS II score. For the risk ratio estimation, this feature was true if SAPS II was above the observed median	Patients with higher SAPS II may be more complex and have other conditions that mask sepsis symptoms
Chronic complexity	The adjustment used the raw CCI score computed without age as a factor, as age is included as a separate factor. For the risk ratio estimation, this feature was true if CCI excluding age, was above the observed median CCI in the population	Providers may have a higher threshold for dismissing an alert on a patient with more comorbidities because they are at a higher risk of deterioration
Advanced age	Age > 70 years	Providers may have a higher threshold for dismissing an alert on an older patient because they are at a higher risk of deterioration
Environmental factors		
High alert level	True if the total number of TREWS alerts in the past 24 h in that unit exceeded the median for that unit and was greater than two alerts in the past 24 h	Providers may have alert fatigue if there have been a lot of alerts in the past day and be less likely to respond to new alerts
High admit volume	True if the total number of admissions in the past 3 h in that unit exceeded the median for that unit and the number of new admissions was greater than two	Providers are busier when there are many new admissions to the unit and may be less likely to respond to alerts in a timely way
Alert occurred 7:00–15:00	True if alert occured between 7:00 and 15:00	This corresponds to the morning/early afternoon hospital shift, which tends to have fewer new admissions in most units
Alert occurred 15:00–23:00	True if alert occured between 15:00 and 23:00	This corresponds to the late afternoon/evening shift, which tends to have increased rates of new admissions and buildup of volume in the ED
Alert occurred 23:00–7:00	True if alert occured between 23:00 and 7:00	This corresponds to the overnight shift, which tends to have higher total patient volume in the ED from buildup through the day, sparser provider coverage and fewer new admissions
Provider factors		
ED provider	True if provider caring for the patient at the time of the alert was an ED provider	ED providers interact with patients earlier in their stay when there is more uncertainty and have a higher patient load per hour
Provider experience with alert	True if provider evaluated a previous alert within the past 30 d	Providers who are more familiar with the alert, may be more aware of the alert and be more likely to respond again

CCI, Charlson comorbidity index; CHF, congestive heart failure; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disorder; SAPS, Simplified Acute Physiology Score; WBC, white blood cell.

a perception that dismissing the alert is faster than evaluating and completing related documentation on the TREWS page. Increasing awareness about the benefits of timely evaluation and implementing alternative workflows during peak hours such as creating supplemental support teams, could improve uniformity of adoption of machine learning-based CDS systems.

Adoption of CDS has been studied across a wide range of clinical applications[25,60]. These studies generally report low to moderate adoption, with clinicians responding to anywhere from 6–45% of alerts depending on the clinical task, interface design and workflow integration[23–25,61]. During this study, 1,965 providers entered an evaluation of a TREWS alert and 89% of alerts received an evaluation

Table 4 | Associations between patient, environmental and provider factors and provider evaluation of TREWS alerts

Factor (number of patients with that factor present out of 3,775 patients in the study population)	Unadjusted risk ratio (95% CI)	Adjusted risk ratio (95% CI)
Patient presentation factors		
Absence of key sepsis symptoms (n=968)	1.01 (0.98–1.04)	0.99 (0.96–1.03)
Alternative diagnosis (n=2,114)	0.99 (0.96–1.02)	1.00 (0.97–1.03)
Condition at risk for fluid overload (n=1,926)	1.02 (1.00–1.04)	1.01 (0.98–1.04)
Acute general severity (n=1,887)	0.98 (0.96–1.01)	0.97 (0.94–1.01)
Chronic complexity (n=2,733)	1.04 (1.00–1.08)	1.02 (0.97–1.08)
Advanced age (n=1,810)	**1.05 (1.02–1.10)**	**1.06 (1.03–1.10)**
Environmental factors		
High alert level (n=1,749)	**0.96 (0.93–0.99)**	**0.94 (0.91–0.96)**
High admit volume (n=1,557)	1.01 (0.98–1.05)	0.99 (0.96–1.03)
Alert occurred 7:00–15:00 (n=1,310)	**1.06 (1.04–1.09)**	**1.03 (1.01–1.06)**
Alert occurred 15:00–23:00 (n=1,686)	**0.94 (0.92–0.97)**	0.98 (0.95–1.00)
Alert occurred 23:00–7:00 (n=779)	1.00 (0.95–1.03)	1.01 (0.97–1.04)
Provider factors		
ED provider (n=3,455)	**1.35 (1.24–1.49)**	**1.22 (1.14–1.32)**
Provider experience with alert (n=1,574)	**1.25 (1.21–1.29)**	**1.22 (1.19–1.26)**

Associations in bold indicated confidence intervals that exclude zero.

Table 5 | Associations between patient, environmental and provider factors and provider dismissal of TREWS alerts

Factor (number of patients with that factor present out of 2,463 patients in the study population)	Unadjusted risk ratio (95% CI)	Adjusted risk ratio (95% CI)
Patient presentation factors		
Absence of key sepsis symptoms (n=576)	1.01 (0.86–1.19)	**1.28 (1.06–1.45)**
Alternative diagnosis (n=1,409)	**1.27 (1.14–1.42)**	1.11 (0.97–1.32)
Condition at risk for fluid overload (n=1,286)	1.10 (0.97–1.21)	1.08 (0.97–1.22)
Acute general severity (n=1,271)	**1.39 (1.23–1.56)**	**1.46 (1.28–1.66)**
Chronic complexity (n=1,823)	**0.87 (0.76–0.98)**	0.90 (0.75–1.05)
Advanced age (n=1,232)	**0.74 (0.65–0.81)**	**0.69 (0.60–0.75)**
Environmental factors		
High alert level (n=1,113)	0.91 (0.80–1.01)	1.01 (0.90–1.13)
High admit volume (n=1,031)	**0.83 (0.73–0.94)**	0.98 (0.86–1.12)
Alert occurred 7:00–15:00 (n=885)	**0.87 (0.74–0.99)**	1.12 (0.99–1.28)
Alert occurred 15:00–23:00 (n=1,079)	1.04 (0.92–1.16)	**1.20 (1.09–1.33)**
Alert occurred 23:00–7:00 (n=499)	**1.15 (1.03–1.29)**	**1.19 (1.07–1.36)**
Provider factors		
ED provider (n=2,297)	**0.39 (0.34–0.43)**	**0.47 (0.40–0.54)**
Provider experience with alert (n=1,167)	**0.58 (0.48–0.64)**	**0.66 (0.56–0.73)**

Associations in bold indicated confidence intervals that exclude zero.

(Table 1). By using real-time interactions with a deployed clinical support system, we were able to assess the extent to which different factors influenced real-time decision making and treatment at a large scale, thereby informing future system design. In a separate paper, we qualitatively analyzed additional factors impacting provider perception of TREWS and its integration into clinical workflows using semi-structured interviews[62].

This study had several limitations. First, there is a lack of consensus on how best to identify sepsis retrospectively. To maximize the reliability of the sepsis labels, we identified sepsis cases using an EHR-based sepsis phenotype that accounts for confounding comorbidities and has shown increased sensitivity and precision compared to alternatives[36,37]. We also added requirements for a substantial antibiotic course when analyzing dismissals of alerts on identified sepsis cases to ensure that they were being treated for infection; however, we cannot completely exclude the possibility that some patients had non-infectious syndromes mimicking sepsis and are examples of overtreatment. Second, we relied on International Classification of Diseases-10 (ICD-10) codes to identify the presence of chronic conditions and alternative diagnoses. While common in large retrospective studies, this may introduce some bias from coding practices. Third, all hospitals in this study were part of the same health system, which may limit its generalizability to other settings; however, the study includes a large cohort representing a diverse patient population from both academic and community hospitals. Fourth, this study focuses on quantitative evaluation of provider interactions that were recorded within the tool itself and does not capture any sepsis-related discussions or actions that occurred outside the tool. As such, there may be alerts labeled as not evaluated that were still considered and discussed. We are unable to capture those interactions, but based on the high adoption rates observed, the majority of patients with sepsis received a timely evaluation within the tool. Fifth, this study assesses the extent to which each of the factors is associated with adoption in the context of TREWS, which has specific performance characteristics, interface presentation and policy decisions about how to integrate alerts into clinical workflow. The relative importance of different factors may vary depending on the performance characteristics of the system. Increased deployment of data-driven CDS systems may change provider attitudes in the future.

Finally, while we included a variety of features related to patient presentation, hospital environment and provider characteristics, there may remain additional sources of confounding that impact the adoption of alerts. For instance, we incorporated information about provider type and experience in the tool to the extent available, but we were unable to access additional information about provider background and attitudes toward CDS. These results should not be taken as causal claims, but rather as hypothesis-generating associations to be later confirmed by causal studies. Further work is needed to understand how these additional characteristics may affect overall adoption and the potential for alert adoption to lead to over-reliance on the alerts (such as over-prescription of antibiotics

in response to sepsis alerts). Quantifying over-prescription resulting from an alert system is important for understanding the potential harms[63]; however, we currently lack metrics to assess over-prescription and leave this to future work.

Using real-time interactions with a machine learning-based sepsis support system, we characterized the adoption and clinical impact of the tool and identified key factors related to failure to use the tool. Overall, TREWS showed high provider adoption that was, in turn, associated with improvements in a key clinical process metric for patients with sepsis: time to antibiotics. Analysis of factors driving adoption showed provider-related factors, such as experience with the system and working in the ED, where providers had increased exposure to the system and had the strongest association with willingness to evaluate alerts. While patient presentation factors such as patient severity and absence of key sepsis symptoms were not significantly associated with the likelihood of evaluation, they did impact the likelihood of dismissing the alert. Education to increase awareness of variation in patient presentation may encourage providers to accept recommendations on sepsis cases with less common presentation. In addition to improving model performance, future machine learning-based systems should focus on the provider in their design choices to encourage adoption and realize the potential benefit of these systems.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-022-01895-z.

Received: 16 September 2021; Accepted: 8 June 2022;
Published online: 21 July 2022

References

1. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A. & Escobar, G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff.* **33**, 1123–1131 (2014).
2. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
3. Liang, H. et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat. Med.* **24**, 443–448 (2019).
4. Bauer, M. et al. Automation to optimise physician treatment of individual patients: examples in psychiatry. *Lancet Psychiat.* **6**, 338–349 (2019).
5. Beam, A. L. & Kohane, I. S. Big data and machine learning in health care. *JAMA* **319**, 1317–1318 (2019).
6. Castaneda, C. et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. *J. Clin. Bioinforma.* https://doi.org/10.1186/s13336-015-0019-3 (2015).
7. 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).
8. Churpek, M. M., Adhikari, R. & Edelson, D. P. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation* **102**, 1–5 (2016).
9. Amland, R. C. & Sutariya, B. B. Quick sequential (sepsis-related) organ failure assessment (qSOFA) and St John sepsis surveillance agent to detect patients at risk of sepsis: an observational cohort study. *Am. J. Med. Qual.* **33**, 50–57 (2018).
10. Tomašev, N. et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* **572**, 116–119 (2019).
11. Sittig, D. F. et al. Grand challenges in clinical decision support. *J. Biomed. Inform.* **41**, 387–392 (2008).
12. Abramoff, M. D., Tobey, D. & Char, D. S. Lessons learned about autonomous AI: finding a safe, efficacious, and ethical path through the development process. *Am. J. Ophthalmol.* **214**, 134–142 (2020).
13. Khairat, S. et al. Reasons for physicians not adopting clinical decision support systems: critical analysis. *JMIR Med. Inform.* https://doi.org/10.2196/medinform.8912 (2018).
14. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G. & King, D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* **17**, 1–9 (2019).
15. Celi, L. A., Fine, B. & Stone, D. J. An awakening in medicine: the partnership of humanity and intelligent machines. *Lancet Digit. Health* **1**, e255–e257 (2019).
16. Mertz, L. From Annoying to appreciated: turning clinical decision support systems into a medical professional's best friend. *IEEE Pulse* **6**, 4–9 (2015).
17. Greenes, R. A. et al. Clinical decision support models and frameworks: seeking to address research issues underlying implementation successes and failures. *J. Biomed. Inform.* **78**, 134–143 (2018).
18. Wright, A. et al. Analysis of clinical decision support system malfunctions: a case series and survey. *J. Am. Med. Inform. Assoc.* **23**, 1068–1076 (2016).
19. Ruppel, H. & Liu, V. To catch a killer: electronic sepsis alert tools reaching a fever pitch? *BMJ Qual. Saf.* https://doi.org/10.1136/bmjqs-2019-009463 (2019).
20. Downing, N. L. et al. Electronic health record-based clinical decision support alert for severe sepsis: a randomised evaluation. *BMJ Qual. Saf.* **28**, 762–768 (2019).
21. Giannini, H. M. et al. A machine learning algorithm to predict severe sepsis and septic shock. *Crit. Care Med.* https://doi.org/10.1097/ccm.0000000000003891 (2019).
22. Moja, L. et al. Effectiveness of a hospital-based computerized decision support system on clinician recommendations and patient outcomes: a randomized clinical trial. *JAMA Netw. Open* **2**, 1–16 (2019).
23. Khan, S. et al. Improving provider adoption with adaptive clinical decision support surveillance: an observational study. *JMIR Hum. Factors* **6**, 1–10 (2019).
24. Mann, D. et al. Impact of clinical decision support on antibiotic prescribing for acute respiratory infections: a cluster randomized implementation trial. *J. Gen. Intern. Med.* **35**, 788–795 (2020).
25. Kwan, J. L. et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. *BMJ* https://doi.org/10.1136/bmj.m3216 (2020).
26. Escobar, G. J. et al. Automated identification of adults at risk for in-hospital clinical deterioration. *N. Engl. J. Med.* **383**, 1951–1960 (2020).
27. Schaefer, K. E., Chen, J. Y. C., Szalma, J. L. & Hancock, P. A. A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. *Hum. Factors* **58**, 377–400 (2016).
28. Cai, C. J. et al. Human-centered tools for coping with imperfect algorithms during medical decision-making. *Proc of the CHI Conference on Human Factors in Computing Systems,* pp. 1–14 (2019).
29. Gaube, S. et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ Digit. Med.* https://doi.org/10.1038/s41746-021-00385-9 (2021).
30. Jacobs, M. et al. Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. *Proc of the Conference on Human Factors in Computing Systems (CHI),* pp. 1–14 (2021).
31. Shortliffe, E. H. & Sepúlveda, M. J. Clinical decision support in the era of artificial intelligence. *JAMA* **10025**, 9–10 (2018).
32. 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).
33. Seymour, C. W. et al. Time to treatment and mortality during mandated emergency care for sepsis. *N. Engl. J. Med.* **376**, 2235–2244 (2017).
34. 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).
35. 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.* https://doi.org/10.1038/s41591-022-01894-0 (2022).
36. Henry, K. E., Hager, D. N., Osborn, T. M., Wu, A. W. & Saria, S. Comparison of automated sepsis identification methods and electronic health record–based sepsis phenotyping: improving case identification accuracy by accounting for confounding comorbid conditions. *Crit. Care Explor.* https://doi.org/10.1097/cce.0000000000000053 (2019).
37. Saria, S. & Henry, K. E. Too many definitions of sepsis: can machine learning leverage the electronic health record to increase accuracy and bring consensus? *Crit. Care Med.* https://doi.org/10.1097/CCM.0000000000004144 (2020).
38. Singer, M. et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* **315**, 801–810 (2016).
39. Levy, M. M., Evans, L. E. & Rhodes, A. The surviving sepsis campaign bundle: 2018 update. *Crit. Care Med.* **46**, 997–1000 (2018).
40. Hospital Inpatient - Specifications Manuals - Sepsis Resources. *QualityNet* https://qualitynet.cms.gov/inpatient/specifications-manuals/sepsis-resources. (2021)
41. Rhee, C. et al. Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. *JAMA Netw. Open* **2**, e187571 (2019).

42. Adams, R. et al. 1405: assessing clinical use and performance of a machine learning sepsis alert for sex and racial bias. *Crit. Care Med.* **50**, 705 (2022).
43. Evans, I. V. R. et al. Association between the New York sepsis care mandate and in-hospital mortality for pediatric sepsis. *JAMA* https://doi.org/10.1001/jama.2018.9071 (2018).
44. Guy, J. S., Jackson, E. & Perlin, J. B. Accelerating the clinical workflow using the sepsis prediction and optimization of therapy (SPOT) tool for real-time clinical monitoring. *NEJM Catal. Innov. Care Deliv*. https://doi.org/10.1056/CAT.19.1036 (2020).
45. Perlin, J. B. et al. 2019 John M. Eisenberg patient safety and quality awards: SPOTting sepsis to save lives: a nationwide computer algorithm for early detection of sepsis: innovation in patient safety and quality at the national level (Eisenberg award). *Jt. Comm. J. Qual. Patient Saf.* **46**, 381–391 (2020).
46. 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*. https://doi.org/10.1136/bmjoq-2017-000158 (2017).
47. 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*. https://doi.org/10.1136/bmjresp-2017-000234 (2017).
48. Wong, A. et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern. Med.* **48109**, 1–6 (2021).
49. Sendak, M. P. et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. *JMIR Med. Inform.* **8**, e15182 (2020).
50. Khairat, S., Marc, D., Crosby, W. & Al Sanousi, A. Reasons for physicians not adopting clinical decision support systems: critical analysis. *JMIR Med. Informatics* https://doi.org/10.2196/medinform.8912 (2018).
51. Kitzmiller, R. R. et al. Diffusing an innovation: clinician perceptions of continuous predictive analytics monitoring in intensive care. *Appl. Clin. Inform.* https://doi.org/10.1055/s-0039-1688478 (2019).
52. Abramoff, M. D., Lavin, P. T., Birch, M., Shah, N. & Folk, J. C. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit. Med*. https://doi.org/10.1038/s41746-018-0040-6 (2018).
53. Ginestra, J. C. et al. Clinician perception of a machine learning–based early warning system designed to predict severe sepsis and septic shock. *Crit. Care Med*. https://doi.org/10.1097/ccm.0000000000003803 (2019).
54. Topiwala, R., Patel, K., Twigg, J., Rhule, J. & Meisenberg, B. Retrospective observational study of the clinical performance characteristics of a machine learning approach to early sepsis identification. *Crit. Care Explor*. https://doi.org/10.1097/cce.0000000000000046 (2019).
55. Carspecken, C. W., Sharek, P. J., Longhurst, C. & Pageler, N. M. A clinical case of electronic health record drug alert fatigue: consequences for patient outcome. *Pediatrics* **131**, e1970–e1973 (2013).
56. Bansal, G. et al. Beyond accuracy: the role of mental modelsin human–AI team performance. *Proc. AAAI Conference on Human Computation and Crowdsourcing* **7**, 2–11. (2019).
57. Jacobs, M. et al. How machine learning recommendations influence clinician treatment selections: the example of the antidepressant selection. *Transl. Psychiatry* https://doi.org/10.1038/s41398-021-01224-x (2021).
58. Lee, J. D. & See, K. A. Trust in automation: designing for appropriate reliance. *Hum. Factors* **46**, 50–80 (2004).
59. Hoff, K. A. & Bashir, M. Trust in automation: integrating empirical evidence on factors that infuence trust. *Hum. Factors* **57**, 407–434 (2015).
60. Murphy, E. V. Clinical decision support: effectiveness in improving quality processes and clinical outcomes and factors that may influence success. *Yale J. Biol. Med.* **87**, 187–197 (2014).
61. Mann, D. et al. Adaptive design of a clinical decision support tool: what the impact on utilization rates means for future CDS research. *Digit. Heal.* **5**, 1–12 (2019).
62. Henry, K. E. et al. Human-machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. *NPJ Digit. Med*. https://doi.org/10.1038/s41746-022-00597-7 (2022).
63. Rhee, C. et al. Infectious Diseases Society of America position paper: recommended revisions to the national severe sepsis and septic shock early management bundle (SEP-1) sepsis quality measure. *Clin. Infect. Dis.* **72**, 541–552 (2021).

Methods

This study was approved by the Johns Hopkins University Institutional Review Board (no. 00252594) and a waiver of consent was obtained.

EHR-based phenotyping definition of sepsis. Sepsis was defined consistent with the third sepsis consensus definition (Sepsis-3) as suspicion of infection and related organ dysfunction[38]. This definition was implemented using EHR-based sepsis phenotyping (ESP)[36,37], which accounts for certain confounding comorbid conditions that can cause patients to be mistakenly identified as having sepsis by automated systems[64,65] and improves the reliability of automated or EHR-confirmed sepsis labels[36,37]. Described in full in Henry et al.[36,37], suspicion of organ dysfunction was determined based on administration of new intravenous antibiotic with at least four days of antibiotic treatment and a culture within 48 h or the presence of both documentation of sepsis and two of the systemic inflammatory response syndrome criteria within 6 h of each other[36,37]. As in the Centers for Disease Control and Prevention's Adult Sepsis Event Toolkit, sepsis-related organ dysfunction was determined based on the initiation of vasopressors or mechanical ventilation or an acute change in serum lactate, serum creatinine, total bilirubin level or platelet count from baseline[65,66]; however, ESP expands on the Adult Sepsis Event Toolkit criteria by also including persistent hypotension, altered mental status and an acute change in international normalized ratio as indicators of organ dysfunction, as in Sepsis-3 and by filtering out changes in these organ dysfunction criteria that are more likely due to a confounding comorbid condition[36,37].

Targeted Real-time Early Warning System. *Description of the TREWS model.* TREWS is a machine learning-based early warning system and decision support tool that was trained using historical EHR data to recognize sepsis early in its progression. To further improve alert performance relative to the original model proposed in Henry et al., the system uses several machine learning-based techniques for tuning to patient context[67], handling missing data[68], suppressing untrustworthy alerts[69] and improving reliability and transportability[70–73]. TREWS consists of a mixture model of Cox proportional hazard models to account for patient heterogeneity[67]. The mixture model uses patient demographics, age and chronic history to assign patients to one or more groups (Extended Data Table 4). Within each group, data from patients assigned to that group are used to learn a Cox proportional hazards model as described by Henry et al.[7,74]. The model is trained iteratively to simultaneously learn the optimal assignment of patients to groups and the optimal model parameters to predict risk of sepsis within each group. To predict risk for an individual patient, each model that a patient is assigned to outputs a predicted risk value and these values are then combined using a weighted average based on group assignment. The risk score is then used by the alert policy to generate alerts. The alert policy used during this deployment is described in the 'TREWS workflow' section.

As described by Henry et al., the current model uses routinely collected laboratory measurements, vital signs, notes, medication history (excluding antibiotics), procedure history and clinical history from the EHR as inputs (Extended Data Table 4)[7,74]. The individual Cox proportional hazards models use the same feature processing and training sample creation as described by Henry et al.[7,74]; however, the model used here extends the model reported by Henry et al. in several key ways. First, the model reported by Henry et al. was trained exclusively on intensive care unit data and used to predict septic shock among patients in the intensive care unit[7,74]. The model in this paper was trained using data across the hospital and used to predict the risk of sepsis in all patients in the ED, observation, general ward and intensive care units[7,74]. Second, instead of learning a single Cox proportional hazard model, the current model combines several Cox proportional hazards models using a mixture of experts model as described above. Finally, additional features were identified and added based on discussions with providers. These additional features include narcotics blood tests, orders for transfusions and sedatives that were identified as common confounders of sepsis.

Deployment process. Before deployment at a new hospital, the alert threshold was tuned to achieve an 80% sensitivity at that hospital based on applying the model to historical data from that hospital. The same model parameters were used at each site. The deployment at each hospital was conducted in three steps. First, a team of educators including clinicians from the site and members of the tool development team, met with clinicians to explain the tool's functionality, identify clinical champions and to verify the process for clinical workflow integration. Hospital administrators, quality and safety officers and clinical experts advocated for the tool and helped determine how best to integrate it into the clinical workflow based on hospital policies. For example, at one site, hospital staff would review any pending TREWS alerts with the ED physician before admitting the patient to a general ward unit. During this period, the alert was active in the background and the technical team monitored the alert volume across different subpopulations in the hospital. Second, deployment was piloted to verify the integration of the system at each site with a subset of the users. Finally, the alert was activated in all ED and inpatient units and the deployment entered a maintenance stage. Throughout the deployment process and maintenance period, the technical and clinical teams monitored alert performance and provider use in different units through weekly emails summarizing alert interactions, performance and volume. Emails reporting frequency of providers addressing the alerts and completing the bundle items by unit were circulated to quality and safety officers, clinical department heads, the deployment team and clinical champions. Alerts included in the analyses in this study occurred after activating the system in all units.

TREWS workflow. To minimize workflow interruptions and alert fatigue, TREWS uses a passive approach to signal new alerts. Instead of triggering a pop-up box or a pager message, the system flags patients visually within the EHR, but does not actively interrupt the provider or require an immediate response before allowing the provider to continue using the EHR. Design choices about the alert and its timing were made in collaboration with the clinical team and refined based on initial feedback. Once the alert appears (for example, as an icon within the clinician's patient list), a provider (physician or advanced practice provider) can click an icon to address the alert leading to a real-time workflow within the patient chart. After an alert is triggered, it remains active for at least 15 min or until the indicator of organ dysfunction is resolved or the predicted risk of sepsis drops below the alert threshold.

From within the TREWS interface, the provider can view summary data gathered by TREWS, including factors leading to why the alert was generated, probability measures indicating likelihood of mortality and sepsis and the status of sepsis-related treatments. Providers are asked to enter an evaluation of whether or not they believe the patient currently has sepsis; however, the response is not mandatory. An evaluation consists of entering a suspected source of infection or clicking a button to indicate that no new or worsening infection is present and then reviewing the list of organ dysfunction indicators and removing any that the provider believes should not be attributed to sepsis (Extended Data Fig. 2). If a provider leaves the TREWS page without entering an evaluation, the interaction is not counted as an evaluation in the context of this study. A nurse can also pre-screen an alert and escalate it to a provider if there are indications of new or worsening infection or altered mental status.

If the provider enters a suspected source of infection and affirms that at least one indicator of organ dysfunction is likely due to sepsis, the alert is considered to be confirmed and a treatment panel is activated to track steps toward completion of the recommended treatments included in the CMS SEP-1 core measure's sepsis bundle[40]. Providers are able, but not required, to place orders for tests and treatments in the sepsis bundle. Orders placed outside of the TREWS interface are also monitored by TREWS and included in its determination of bundle completeness. No further alerts are generated after a patient has a confirmed sepsis alert, as current treatment protocols focus on the first episode of sepsis.

After an alert is dismissed based on the provider evaluating the patient as not having sepsis, no further alerts are generated on that patient for at least 72 h. After that time period, the alert resets and may trigger again.

Assessing the retrospective model performance of TREWS. *Study population.* To evaluate the retrospective accuracy and lead time of TREWS, we used a set of EHRs from adult ED, medical and surgical patients at one academic and two community hospitals in Maryland with admission between 1 January 2016 and 31 March 2018. The included hospitals were Howard County General Hospital, Johns Hopkins Hospital and Bayview Medical Center. We treated each time a patient presented to the ED or was admitted as a unique patient encounter and included each encounter separately. Patients were excluded if they were discharged from an ED or were admitted to a labor or maternity unit.

Analysis. We measured accuracy using per-encounter area under the receiver operating characteristic curve, PPV and sensitivity. As described above, positive sepsis cases were identified using ESP[36,37]. Additionally, we measured alert lead time as the median time (in hours) from the alert to the patient's first antibiotic order and the percentage of patients with at least 3, 6 and 12 h between the alert and the patient's first antibiotic order. We measured lead time only among patients with sepsis and an alert. We used an alert threshold chosen to give 0.8 sensitivity on the original development data. As our primary interest is in lead time for patients who were not recognized immediately by clinicians, we excluded from our lead-time analyses any patient who received an antibiotic order within 3 h of arrival (defined as the earlier of ED triage or admission to an inpatient unit). We report lead time among all patients, as well as among patients who died in hospital. During the period covered by this pre-deployment data, providers used a rule-based alert based on the SEP-1 criteria, which may have influenced antibiotic timing in the data.

Prospective analysis of alert response. *Study population.* The primary study population included all adults who presented to the ED or were admitted to a medical or surgical unit at any of five hospitals (three community and two academic hospitals) in the Maryland and DC areas that either (1) had a prospective TREWS alert or (2) were retrospectively identified as having sepsis based on specified criteria. The included hospitals and date ranges were Howard County General Hospital (1 April 2018 to 31 March 2020), Suburban Hospital (1 October 2018 to 31 March 2020), Bayview Medical Center (1 February 2019 to 31 March 2020), Johns Hopkins Hospital (1 April 2019 to 31 March 2020) and Sibley

Memorial Hospital (1 May 2019 to 31 March 2020). The start date at each hospital was based on the timing of the staggered deployment across the five sites. As described above, we treated each time that a patient presented to the ED or was admitted as a unique patient encounter and included each encounter separately. Population characteristics and overall adoption rates (study question 1) were estimated using all patient encounters with an alert or who were retrospectively identified as having EHR-confirmed sepsis based on ESP during this period[36,37].

When evaluating the association between adoption and clinical care, and association between various factors and evaluation or confirmation (study questions 2, 3.1 and 3.2) we included all patients wth sepsis who received an alert in the ED or an inpatient unit and who had not received an antibiotic order at the time of their alert. This criterion was used to restrict the analysis to cases where there was opportunity for the alert to impact care decisions. To further ensure that the antibiotic order and the alert were related to the same episode of sepsis, we only included patients who received antibiotics within 24 h after the alert. While initiating sepsis treatment is recommended within 3 h of diagnosing sepsis, a 24-h window was chosen to avoid excluding cases where there was delayed recognition (for example, if a provider reviewed and responded to the alert after the patient was transferred to a new unit) thereby biasing the results toward patients who received prompt treatment within 3 h of the alert.

Study question 1: assessing overall alert adoption. To understand the adoption of TREWS, we report the number and percentage of alerts with evaluations entered within 1, 3 and 6 h after the alert or ever entered. Additionally, among alerts with an evaluation entered, we report the percentage that were confirmed by a provider in real-time as having sepsis. We report these numbers for all patients with an alert and patients who were retrospectively identified as having sepsis, as described above[36].

Study question 2: timing of antibiotics relative to alerts. To assess the association between tool adoption and patient care, we examined the extent to which using the TREWS page to record an evaluation for sepsis within 3 h after the alert was associated with the timing of a patient's first antibiotic order, a key element of sepsis treatment[33,34,41]. We estimated the unadjusted and adjusted differences in median time from alert to first antibiotic order between patients who had an evaluation entered within 3 h of the alert and those who did not. The adjustment variables included patient demographics, medical history, laboratory measurements, vital signs, comorbidities and admitting hospital. As in previous studies[33,34,75], we adjusted for patient age, documented sex and comorbidities as defined by the CCI as well as the presence of ICD-10 codes for history of diabetes (with and without complications), dementia, malignant tumors, metastatic solid tumors, end-stage renal disease, CHF, acute liver disease, gastrointestinal bleeding and COPD. We also adjusted for acute severity based on individual sepsis-related organ failure assessment score components and APACHE II score[76,77], as well as several sepsis-relevant laboratory measurements, vital signs and treatments, including systolic blood pressure, altered mental status indicated by Glasgow Coma Score below 15 (ref. [75]), temperature, WBC count, lactate above 2 mmol/L and indicators for vasopressors and mechanical ventilation. Additionally, to account for potential differences in clinical practice, we adjusted for which hospital a patient was admitted to and, for ED patients, whether the trauma team was activated upon arrival. For laboratory measurements and vital signs, the most recent measurement taken in the 24 h before the alert was used except in the case where the alert occurred within 12 h of patient arrival, in which case the first available measurement taken within 12 h of arrival was used. Sepsis-related organ failure assessment and APACHE II scores were calculated using the worst measurements taken in the 24 h before the alert except in the case where the alert occurred within 12 h of patient arrival, in which case the worst measurement taken within 12 h of arrival was used. To account for non-linearities, continuous laboratory values and vital signs were included as piecewise linear terms according to the thresholds used in APACHE II. We repeated this analysis to compare the unadjusted and adjusted differences in median time from alert to antibiotic order between confirmed alerts versus alerts that were either not evaluated within 3 h or were dismissed.

Study question 3.1: factors associated with alert adoption. To assess the association between patient, environmental and provider factors and provider response to alerts, we measured the association between these factors and whether or not a patient evaluation was entered within 3 h after the alert. Specific factors that might affect alert response were identified based on clinical feedback from ED, intensive care unit and general ward providers actively using the tool and who had experience managing patients with sepsis (Table 3). Patient factors included age, chronic complexity as measured by age and CCI[78] and SAPS II (Table 3). We also accounted for presence of sepsis-related symptoms, an alternative diagnosis that may complicate sepsis diagnosis and the presence of chronic condition(s), such as COPD, CHF or CKD, which may make a provider hesitant to follow the sepsis bundle guidelines for giving high-volume fluids. We characterized environmental factors based on the shift during which the alert occurs, the TREWS alert burden in the unit, computed as the number of alerts that occurred in that unit in the past 24 h and the admit volume computed as the number of new patients admitted to that unit in the past 3 h. Provider factors included previous experience with TREWS and location of care provision (ED versus inpatient) (Table 3). Due to the low number of inpatient alerts, we were unable to further divide inpatient providers into medical and surgical providers.

For each factor, we estimated the adjusted and unadjusted risk ratio of whether the alert would be evaluated within 3 h with versus without that factor present. Due to the smaller number of patients with each given factor and alert response type, we used a simplified set of adjustment variables compared to study question 2. We also added additional environmental and provider adjustment variables based on the factors considered in Table 3. Specifically, we adjusted for patient demographics (age and sex), chronic comorbidities as measured by CCI and acute severity as measured by SAPS II. We adjusted for chronic conditions associated with increased risk of fluid overload, indicated by the presence of an ICD-10 code for CKD, COPD or CHF. We accounted for acute comorbidities that may impact the diagnosis of sepsis by adjusting for the presence an ICD-10 code for myocardial infarction, stroke or acute respiratory failure, as well acute symptoms of sepsis (lactate > 2.0 mmol/L, WBC > $12\,10^9$/L or WBC < $4\,10^9$ and temperature > 38.0 °C or < 36.0 °C). Based on the list of environmental and provider factors in the analysis, we adjusted for alert level in the past 24 h, number of admissions in that unit in the past 3 h, an indicator of whether the alert occurred between 7:00–15:00, 15:00–23:00 or 23:00–7:00, whether the provider was an ED provider and whether the provider had previous experience with the alert.

Study question 3.2: factors associated with alert dismissal. To assess which patient, environmental and provider factors were associated with a provider's decision to dismiss an alert on a patient later identified as having sepsis, we estimated the association between these factors and the evaluation entered for patients with sepsis with an evaluation entered within 3 h. As before, we excluded all patients who received an antibiotic order before the alert and also excluded patients with no evaluation entered within 3 h. As this question examines factors related to potentially incorrect dismissal of alerts on retrospectively identified patients with sepsis, we chose to use a more conservative inclusion criteria and restricted the study population to only the patients with sepsis who received a substantial antibiotic course, namely 4 d consecutively of antibiotics or antibiotics up until the time of in-hospital death, discharge to hospice or transfer to another acute care facility[65]. All antibiotics included in the CMS SEP-1 core measure were included. We referred to this criteria as having '4+ qualifying antibiotic days'. We assessed all patient factors described above and adjusted for all patient, provider and environmental factors.

Statistical analyses. Adjusted median time to first antibiotic order (study question 2) was estimated using quantile regression as implemented in the statsmodels Python package (v.0.12.2)[79]. Adjusted risk ratios (study question 3) were estimated using logistic regression as described by Norton et al.[80]. The binary outcome was regressed onto the factor of interest and all adjustment variables, then the adjusted risk ratio was estimated as the sample average regression output with the factor of interest set to one for all patients over the sample average regression output with the factor of interest set to zero for all patients. All confidence intervals were estimated using nonparametric bootstrapping with 3,775 bootstrap samples used to estimate percentile-based 95% CIs. All models and statistics were computed using Python (v.3.7.6).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data are not publicly available because they are from electronic health records approved for limited use by Johns Hopkins University investigators. Making the data publicly available without additional consent, ethical or legal approval might compromise patients' privacy and the original ethical approval. To perform additional analyses using these data, researchers should contact A.W.W. or S.S. to apply for an institutional review board-approved research collaboration and obtain an appropriate data-use agreement.

Code availability

The TREWS early warning system described in this study is available from Bayesian Health. The underlying source code is proprietary intellectual property and is not available. Code for the primary statistical analyses can be found at https://github.com/royadams/henry_et_al_2022_code.

References

64. Rhee, C. et al. Diagnosing sepsis is subjective and highly variable: a survey of intensivists using case vignettes. *Crit. Care* **20**, 1–8 (2016).
65. Rhee, C., Dantes, R. B., Epstein, L. & Klompas, M. Using objective clinical data to track progress on preventing and treating sepsis: CDC's new 'Adult Sepsis Event' surveillance strategy. *BMJ Qual. Saf.* https://doi.org/10.1002/sca.20201 (2018).
66. Rhee, C. et al. Sepsis surveillance using adult sepsis events simplified eSOFA criteria versus sepsis-3 sequential organ failure assessment criteria∗. *Crit. Care Med.* **47**, 307–314 (2019).

67. Jordan, M. I. & Jacobs, R. A. Hierarchical mixtures of experts and the EM algorithm. *Neural Comput.* **6**, 181–214 (1994).
68. Soleimani, H., Hensman, J. & Saria, S. Scalable joint models for reliable uncertainty-aware event prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 1948–1963 (2018).
69. Schulam, P. & Saria, S. Can you trust this prediction? Auditing pointwise reliability subsequent to training. In *Proc. of the 22nd International Conference on Artificial Intelligence and Statistics, 89*:1022-1031 (2019).
70. Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S. & Obermeyer, Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nat. Med.* **27**, 136–140 (2021).
71. Subbaswamy, A. & Saria, S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics* **21**, 345–352 (2020).
72. Finlayson, S. G. et al. The clinician and dataset shift in artificial intelligence. *N. Engl. J. Med.* **385**, 283 (2021).
73. Subbaswamy, A., Schulam, P. & Saria, S. Preventing failures due to dataset shift: learning predictive models that transport. In *Proc. of the International Conference on Artificial Intelligence and Statistics*, PMLR 89:3118-3127 (2019).
74. Henry, K., Wongvibulsin, S., Zhan, A., Saria, S. & Hager, D. Can septic shock be identified early? Evaluating performance of a targeted real-time early warning score (TREW score) for septic shock in a community hospital: global and subpopulation performance. *Am. J. Respir. Crit. Care Med.* **195**, 7016 (2017).
75. Peltan, I. D. et al. ED door-to-antibiotic time and long-term mortality in sepsis. *Chest* **155**, 938–946 (2019).
76. Vincent, J. L. et al. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. *Intensive Care Med.* **22**, 707–710 (1996).
77. Knaus, W. A., Draper, E. A., Wagner, D. P. & Zimmerman, J. E. APACHE II: a severity of disease classification system. *Crit. Care Med.* **13**, 818–829 (1985).
78. Metcalfe, D. et al. Charlson and Elixhauser coding mapping. *BMC Med. Res. Methodol.* **19**, 1–9 (2019).
79. Seabold, S. & Perktold, J. statsmodels: econometric and statistical modeling with Python. *Proc. 9th Python in Science Conference* (2010).
80. Norton, E. C., Miller, M. M. & Kleinman, L. C. Computing adjusted risk ratios and risk differences in Stata. *Stata J.* **13**, 492–509 (2013).

Acknowledgements

The authors thank Y. Ahmad, A. Zhang, M. Yeo and Y. Karklin whose work significantly contributed to early iterations of the development of the deployed system. We also thank R. Demski, K. D'Souza, A. Kachalia, A. Chen and clinical and quality stakeholders who contributed to tool deployment, education and championing the work. The authors gratefully acknowledge the following sources of funding: the Gordon and Betty Moore Foundation (award 3926), the National Science Foundation Future of Work at the Human-technology Frontier (award 1840088) and the Alfred P. Sloan Foundation research fellowship (2018). This information or content and conclusions are those of the authors and should not be construed as the official position or policy of, nor should any endorsements be inferred by the NSF the US Government.

Author contributions

K.E.H., R.A., C.P., E.S.C., A.W.W. and S.S. contributed to the initial study design and preliminary analysis plan. S.S. led the development and deployment efforts for the TREWS software. K.E.H., H.S., A.S., R.C.L., L.J., M.H., S.M., D.N.H., A.W.W. and S.S. contributed to the system development and deployment. K.E.H., R.A., C.P., E.Y.K., S.E.C., A.R.C., E.S.C., D.N.H., A.W.W. and S.S. contributed to the review and analysis of the results. All authors contributed to the final preparation of the manuscript.

Competing interests

Under a license agreement between Bayesian Health and the Johns Hopkins University, K.E.H., S.S. and Johns Hopkins University are entitled to revenue distributions. Additionally, the University owns equity in Bayesian Health. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict-of-interest policies. S.S. also has grants from Gordon and Betty Moore Foundation, the National Science Foundation, the National Institutes of Health, Defense Advanced Research Projects Agency, the Food and Drug Administration and the American Heart Association; she is a founder of and holds equity in Bayesian Health; she is the scientific advisory board member for PatientPing; and she has received honoraria for talks from a number of biotechnology, research and health-tech companies. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict-of-interest policies. D.N.H. discloses salary support and funding to his institution from the Marcus Foundation for the conduct of the vitamin C, thiamine and steroids in sepsis trial. S.E.C. declares consulting fees from Basilea for work on an infection adjudication committee for an *S. aureus* bacteremia trial. The other authors declare no competing interests.

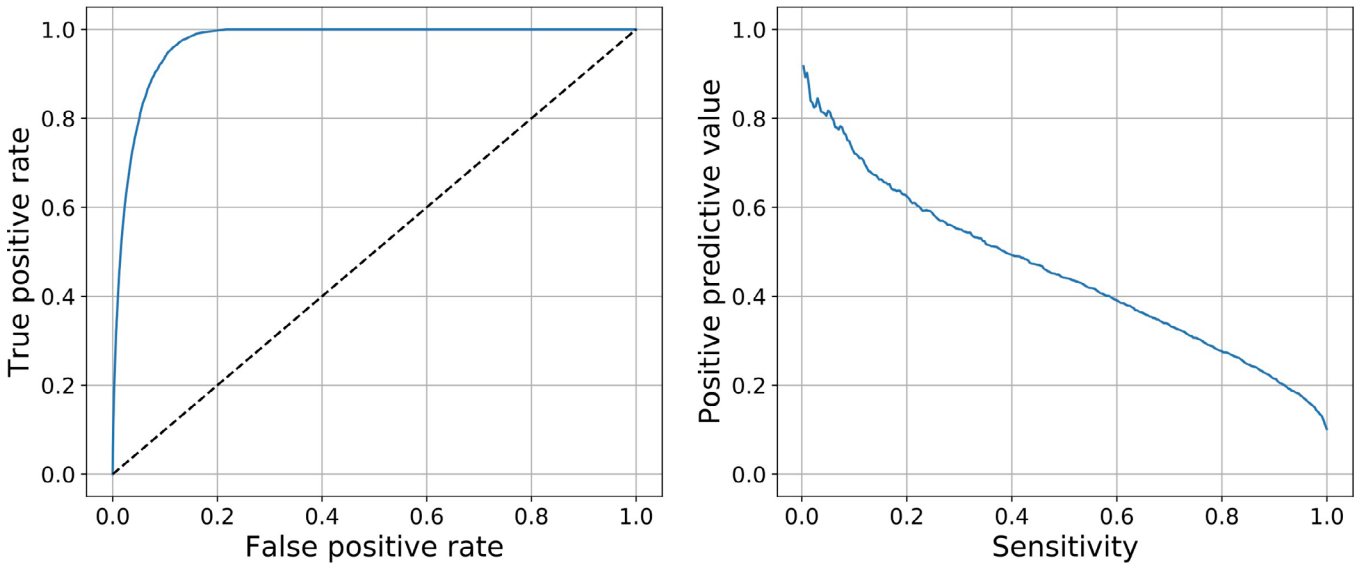
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-022-01895-z.

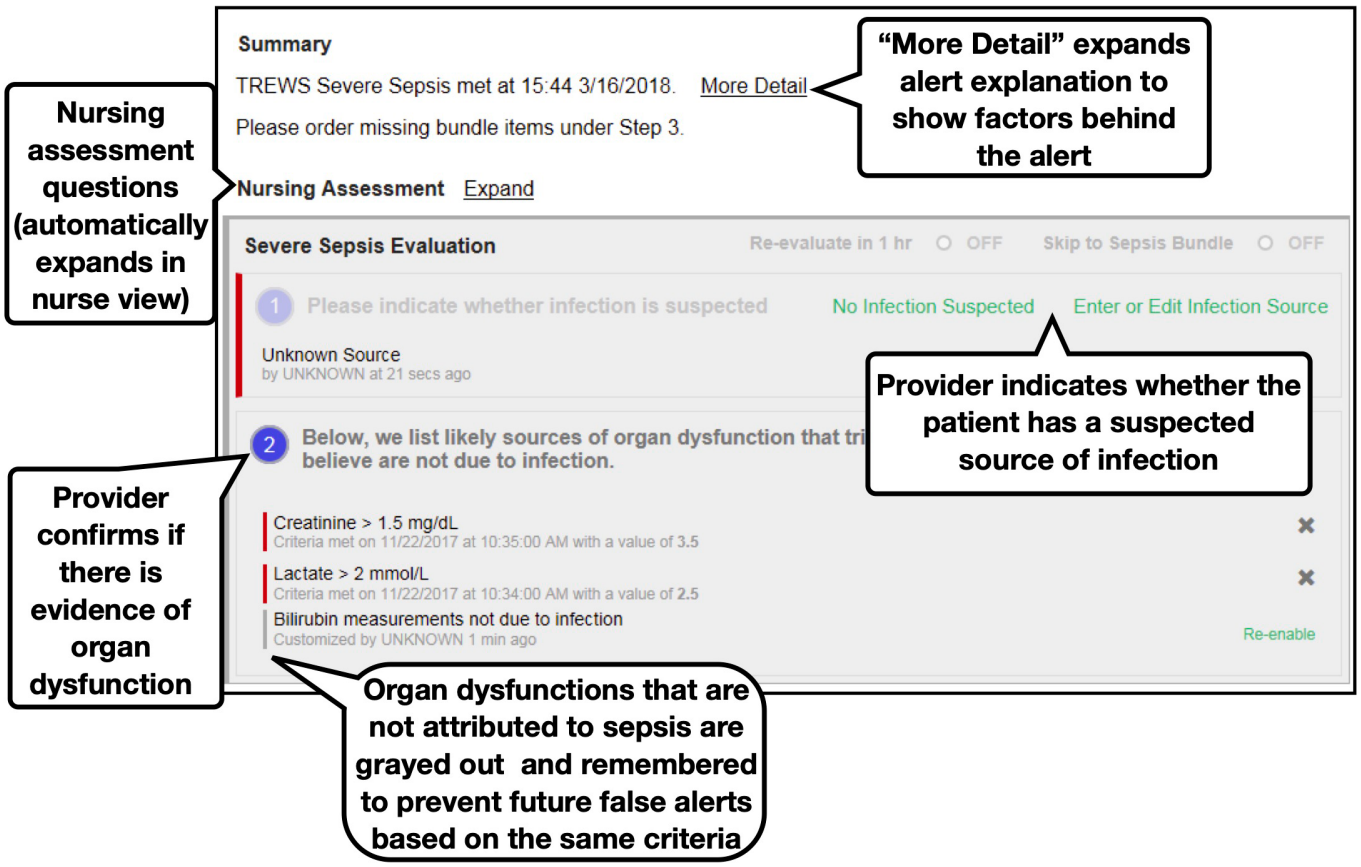
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-022-01895-z.

Correspondence and requests for materials should be addressed to Albert W. Wu or Suchi Saria.

Peer review information *Nature Medicine* thanks Derek Angus, Melanie Wright and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. **Primary Handling editor**: Michael Basson, in collaboration with the *Nature Medicine* team.



Extended Data Fig. 1 | Retrospective predictive performance of the TREWS model. Performance of the TREWS model on retrospective data. Figure (a) shows the receiver operating characteristic curve and Figure (b) shows the sensitivity–PPV curve (also referred to as the precision-recall curve).



Extended Data Fig. 2 | Annotated screenshot of the TREWS interface. Annotated screenshot of the TREWS provider evaluation page. Annotations show the main provider actions: reviewing the alert explanation, indicating whether the patient has a suspected source of infection and reviewing sources of organ dysfunction.

Extended Data Table 1 | Time from alert to first antibiotic order among retrospectively identified sepsis patients

	All sepsis patients (N = 1,182)	Sepsis patients who died in-hospital (N = 205)
% of patients detected prior to any antibiotic order	73%	86%
Median hours alert to antibiotic order	3.6	5.7
Percent with an alert		
3 hours before antibiotic order	56%	68%
6 hours before antibiotic order	38%	50%
12 hours before antibiotic order	27%	41%

Extended Data Table 2 | TREWS alert volume per day during the study period including re-alerts and alerts flagging patients who are candidates for escalation

Hospital	Number of beds	Average alerts per day	Average emergency department alerts per day	Average inpatient alerts per day
HCGH	243	18.8	9.9	9.0
SH	228	15.9	7.7	8.2
BMC	463	21.1	11.3	9.8
JHH	1162	38.2	12.1	26.2
SMH	288	8.9	4.1	4.8

Extended Data Table 3 | Population characteristics

	All Patients with an Alert (N=31,591)	Patients Included in Factor Analysis* (N=3,775)
Median age, years (IQR)	66 (53-78)	69 (56-80)
Female (%)	16,336 (52%)	1,993 (53%)
Median SAPS II (IQR)	37 (27-49)	46 (35-62)
Median CCI (IQR)	5 (2-7)	5 (3-8)
CHF (%)	5,624 (20%)	714 (20%)
CKD (%)	8,859 (32%)	1,127 (32%)
ESRD (%)	2,039 (7%)	269 (8%)
COPD (%)	6,550 (24%)	806 (23%)
Died in-hospital	2,276 (7%)	659 (17%)
Median Hospital LOS, days (IQR)	4 (2-8)	6 (4-12)
Ever admitted to the ICU (%)	8,781 (28%)	1,858 (49%)
Discharged from ED**	3,053 (10%)	52 (1%)
Trauma admission	306 (1%)	10 (0%)
Sepsis Case***	8,033 (25%)	3,775 (100%)
4+ Qualifying Antibiotic Days****	11,396 (36%)	3,060 (81%)
Number of distinct providers who entered an evaluation on an alert	1,965	627

IQR: interquartile range, SAPS: Simplified Acute Physiology Score, CCI: Charlson Comorbidity Index, CHF: congestive heart failure, CKD: chronic kidney disease, ESRD: end-stage renal disease, COPD: chronic obstructive pulmonary disorder, LOS: length of stay, ICU: intensive care unit, ED: emergency department

* Includes all sepsis patients with an alert who did not have an antibiotic order prior to the alert and who received antibiotics within 24 hours after the alert

** Includes patients who died in the ED or who were transferred to hospice or another acute care facility

*** Sepsis identified retrospectively based on clinical presentation (see Methods)

**** 4+ continuous days of antibiotics or antibiotics given up until day of discharge to hospice, another acute care facility, or in-hospital death

Extended Data Table 4 | Model features

Feature Category	Features Used by Individual Risk Models	Features Used by Mixture Model
Patient Information	Documented sex, weight, year of admission	Age
Comorbidities (defined using ICD-10 codes)	Arrhythmias, chronic airway obstruction, chronic bronchitis, chronic pancreatitis, chronic kidney disease, chronic liver disease, chronic pulmonary disease, diabetes, emphysema, end-stage renal disease, heart failure, hematologic malignancy, immunodeficiency, metastatic carcinoma, organ insufficiency, renal insufficiency	Cardiac arrest, dementia, drug overdose, fall, gastrointestinal bleed, seizure, sickle cell anemia, stroke
Chief Complaint Documented at Presentation to Emergency Department (if present)	Biliary complaint, cardiac complaint, dementia complaint, fall complaint, gastrointestinal bleed complaint, seizure complaint, sickle cell anemia complaint, stroke complain	N/A
Laboratory Measurements and Vital Signs	Alanine transaminase, amylase, arterial pH, aspartate aminotransferase, bicarbonate, blood urea nitrogen (BUN), BUN/creatinine ratio, diastolic blood pressure, fraction of inspired oxygen (FiO2), Glasgow coma scale, heart rate, hematocrit, hemoglobin, international normalized ratio, lactate, lipase, mean arterial pressure, oxygen saturation, partial pressure of carbon dioxide, partial pressure of oxygen/FiO2 ratio, platelet count, potassium, Richmond Agitation-Sedation Scale, respiratory rate, serum creatinine, shock index, sodium, systolic blood pressure, temperature, white blood cell count	Bilirubin, creatinine, Glasgow coma scale, hypotension, international normalized ratio, lactate, platelets
Procedures, medications, and consult orders	ACE inhibitor, anticoagulant, anticonvulsant, antihyperglycemic, antiplatelet, beta blocker, biliary medication, calcium channel blocker, cardiac catheterization, cardiac procedure, cardiology consult, dementia medication, diuretic, gastroenterology consult, general anesthesia, mechanical ventilation, neurology consult, stroke protocol initiation, vasodilator, vasopressor	Mechanical ventilation, vasopressor

nature portfolio

Corresponding author(s): Albert Wu and Suchi Saria

Last updated by author(s): Mar 24, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No specialized software was used for data collection.
Data analysis	The TREWS algorithm cannot be made publicly available because it is proprietary intellectual property. If other investigators are interested in performing additional studies involving TREWS, requests can be made to Suchi Saria. Code for the primary statistical analyses can be made available upon request to Suchi Saria.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data are not publicly available because they are from electronic health records approved for limited use by Johns Hopkins University investigators. Making the data publicly available without additional consent, ethical, or legal approval might compromise patients' privacy and the original ethical approval. To perform additional analyses using this data, researchers should contact AWW or SS to apply for an IRB-approved research collaboration and obtain an appropriate data use agreement.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample used was a convenience sample including all data available from the deployment of the tool under study until the COVID-19 pandemic and, as such, no power analysis was performed.
Data exclusions	No data was excluded beyond the inclusion criteria stated in the paper.
Replication	Due to the cost and time involved in deploying the clinical decision support tool under study and gathering electronic medical record data no attempts were made to replicate this prospective cohort study.
Randomization	The study was not randomized and thus covariate adjustment was used. In our analysis of the association between alert adoption and antibiotics timing, we adjusted for patient age, documented sex, and comorbidities as defined by the Charlson Comorbidity Index (CCI) as well as the presence of International Classification of Diseases, Tenth Revision (ICD-10) codes for history of diabetes (with and without complications), dementia, malignant tumors, metastatic solid tumors, end stage renal disease, congestive heart failure, acute liver disease, gastrointestinal bleeding, and chronic obstructive pulmonary disease. We also adjusted for acute severity based on individual SOFA score components and APACHE II score, as well as several sepsis-relevant laboratory measurements, vital signs, and treatments including systolic blood pressure, altered mental status indicated by Glasgow Coma Score (GCS) below 15, temperature, white blood cell count, lactate above 2 mmol/L, and indicators for vasopressors and mechanical ventilation. Additionally, to account for potential differences in clinical practice, we adjusted for which hospital a patient was admitted to, and for ED patients, whether the trauma team was activated upon arrival. For lab measurements and vital signs, the most recent measurement taken in the 24 hours prior to the alert was used except in the case where the alert occurred within 12 hours of patient arrival, in which case the first available measurement taken within 12 hours of arrival was used. SOFA and APACHE II scores were calculated using the worst measurements taken in the 24 hours prior to the alert except in the case where the alert occurred within 12 hours of patient arrival, in which case the worst measurement taken within 12 hours of arrival was used. To account for non-linearities, continuous lab values and vital signs were included as piecewise linear terms according to the thresholds used in APACHE II. In our analysis of factors associated with alert adoption, we adjusted for patient demographics (age and gender), chronic comorbidities as measured by CCI, and acute severity as measured by SAPS II. We adjusted for chronic conditions associated with increased risk of fluid overload, indicated by the presence of an ICD-10 code for CKD, COPD, or CHF. We accounted for acute comorbidities that may impact the diagnosis of sepsis by adjusting for the presence an ICD-10 code for myocardial infarction, stroke, or acute respiratory failure, as well acute symptoms of sepsis (lactate > 2.0 mmol/L, WBC > 12 mu/L or WBC < 4 mu/L, and temperature > 38.0oC or temperature < 36.0oC). Based on the list of environmental and provider factors in the analysis, we adjusted for alert level in the past 24 hours, number of admissions in that unit in the past 3 hours, an indicator of whether the alert occurred between 7am-3pm, 3pm-11pm, or 11pm-7am, whether the provider was an ED provider, and whether the provider had prior experience with the alert.
Blinding	This study was observational and thus no blinding was used.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	3,775 patient encounters were included in the primary analyses. The median age was 69.0 (IQR (56-80)) and 53.0% were document as male. The median Charlson Commorbidity Index was 5 (IQR (3-8)) and the median SAPS II score was 46 (IQR (35-62)). 17% of encounters ended in in-hospital mortality and the median hospital length-of-stay was 6 days (IQR (4-12)). For additional cohort details, see Supplementary Table 2.
Recruitment	This study used electronic health records gathered naturally over the course of patient care and thus patients were not directly recruited to the study. A waiver of consent was obtained from the Johns Hopkins University internal review board (IRB No. 00252594).
Ethics oversight	The Johns Hopkins University internal review board (IRB No. 00252594).

Note that full information on the approval of the study protocol must also be provided in the manuscript.