

Real-time surveillance system for patient deterioration: a pragmatic cluster-randomized controlled trial

Received: 10 July 2024

A list of authors and their affiliations appears at the end of the paper

Accepted: 24 February 2025

Published online: 02 April 2025

 Check for updates

The COmmunicating Narrative Concerns Entered by RNs (CONCERN) early warning system (EWS) uses real-time nursing surveillance documentation patterns in its machine learning algorithm to identify deterioration risk. We conducted a 1-year, multisite, pragmatic trial with cluster-randomization of 74 clinical units (37 intervention; 37 usual care) across 2 health systems. Eligible adult hospital encounters were included. We tested if outcomes differed between patients whose care teams were and patients whose care teams were not informed by the CONCERN EWS. Coprimary outcomes were in-hospital mortality (examined as instantaneous risk) and length of stay. Secondary outcomes were cardiopulmonary arrest, sepsis, unanticipated intensive care unit transfers and 30-day hospital readmission. Among 60,893 hospital encounters (33,024 intervention; 27,869 usual care), intervention group encounters had 35.6% decreased instantaneous risk of death (adjusted hazard ratio (HR), 0.64; 95% confidence interval (CI), 0.53–0.78; $P < 0.0001$), 11.2% decreased length of stay (adjusted incidence rate ratio, 0.91; 95% CI, 0.90–0.93; $P < 0.0001$), 7.5% decreased instantaneous risk of sepsis (adjusted HR, 0.93; 95% CI, 0.86–0.99; $P = 0.0317$) and 24.9% increased instantaneous risk of unanticipated intensive care unit transfer (adjusted HR, 1.25; 95% CI, 1.09–1.43; $P = 0.0011$) compared with usual-care group encounters. No adverse events were reported. A machine learning-based EWS, modeled on nursing surveillance patterns, decreased inpatient deterioration risk with statistical significance. ClinicalTrials.gov registration: [NCT03911687](https://clinicaltrials.gov/ct2/show/NCT03911687).

Given the rise in patient acuity¹, early identification of patients' risk of deterioration is essential in preventing avoidable yet serious adverse hospital outcomes², such as mortality and sepsis. Failure to detect deterioration and intervene accordingly is an unacceptable system failure, strongly linked to information and communication breakdowns among the care team³. While several nonrandomized studies have shown that automated algorithm-based early warning systems (EWSs) positively impact patient outcomes, few randomized controlled trials have demonstrated an impact^{4,5}, and many predictions are focused on

one particular event type rather than a broad set of outcomes, such as in-hospital mortality, length of stay (LOS) and sepsis^{4,6}. Advances in EWS computational sophistication hold great promise for predicting patient 'crashing'⁷; however, algorithms typically rely on late and noisy physiologic indicators of deterioration (for example, laboratory results and vital signs)^{8,9}. Instead, our approach for the COmmunicating Narrative Concerns Entered by RNs (CONCERN) model uses electronic health record (EHR) metadata (for example, date and time stamps, and data type) of nursing surveillance activities. As a result, it identifies

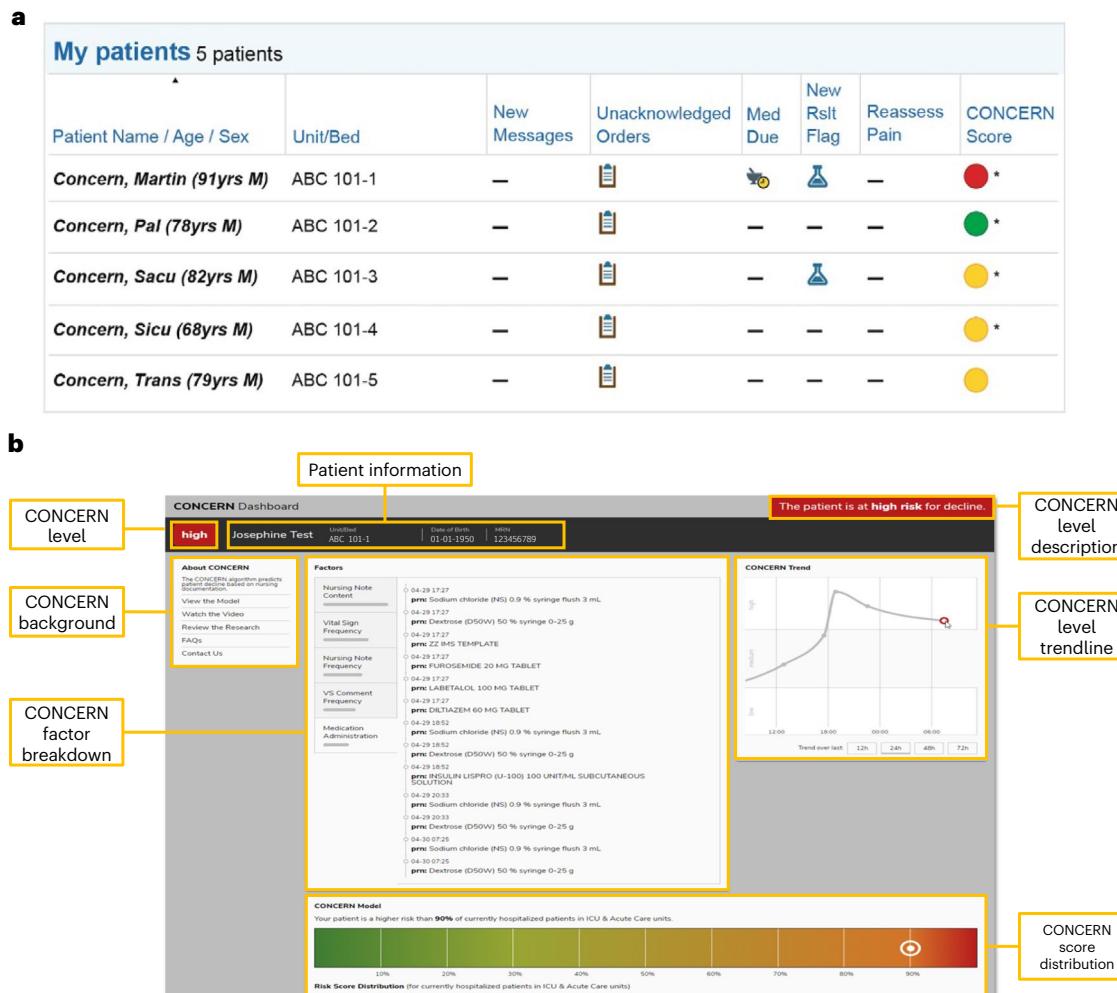


Fig. 1 | CONCERN EWS display and detailed prediction screen integrated into the EHR. a,b. An illustration of the integration of the CONCERN EWS into the EHR, displaying two key components: CONCERN EWS risk scores displayed as non-interruptive alerts in the clinicians' EHR patient list (a) and CONCERN screen with detailed information about the prediction (b). By double-clicking on the

CONCERN score icon, clinicians can access the CONCERN screen with detailed information about the prediction, which provides a detailed breakdown of the risk score, including contributing factors, trends over time, background about the CONCERN model and a score distribution on a comparative scale, illustrating the patient's risk score relative to other patients on the same unit.

all-cause deterioration up to 42 hours earlier than models reliant on physiological indicators. Therefore, CONCERN EWS can be used as clinical decision support to make the care team aware of deterioration much earlier so that more timely interventions can be performed¹⁰.

Nurse surveillance is a core component of nursing practice aimed at preventing adverse events¹¹. Increased nurse surveillance is an indicator of concern^{10,12}, and nurse concern has been shown to be a valid and frequent reason for calling a rapid response¹⁰. Nurses can recognize subtle, yet observable, clinical changes that may not be captured in physiological data or well-displayed in EHRs¹², such as pallor change with incremental increases in supplemental oxygen needs, slower recovery of arterial blood pressure after turning the patient or small changes in mental status from baseline. As part of surveillance, nurses document additional data in the EHR to highlight concerning patient changes, but the patterns of these additional data are not explicitly evident to other members of the care team^{12,13}. Unfortunately, when there is a lack of shared team situational awareness, medical interventions are delayed^{10,12–14}.

Evidence for escalation of medical interventions based on nurses' concerns has long remained classified as level 5 evidence (expert opinion)^{15–20}. The CONCERN EWS approach differs from most EWSs in two important ways. First, we are modeling the clinician (that is,

nurses' concern) to predict patient deterioration. Unlike other EWSs, the CONCERN model does not use structured clinical or physiological data as inputs to the model. Rather, to objectively measure and test nurses' concern levels in predicting patient deterioration, we created a machine learning-based predictive model that processes nurse surveillance patterns from metadata of nurse-entered documentation, with a small additional signal from natural language processing of 'mentions' of concern in nurses' narrative notes¹⁰. Our team has published on statistically significant signals detected in nursing data using this metadata modeling approach since 2012 (refs. 13,21), demonstrating its resilience, first, as EHR functionality and documentation policies change over time and across multiple sites. Second, by using an ensemble modeling²² method, we leverage crucial time-dependent signals by essentially modeling every hour of the day to continuously monitor nurses' concern levels that reflect nurses' increased surveillance, such as (1) increased frequency of assessments (for example, respiratory rate checked every 2 hours for a non-intensive care unit (ICU) acute care floor patient), (2) assessments performed at uncommon times (for example, checking vital signs in the middle of the night for a non-ICU acute care floor patient) or (3) nursing medication administration interventions, such as not administering a scheduled medication when it is due (typically because the patient is clinically unstable).

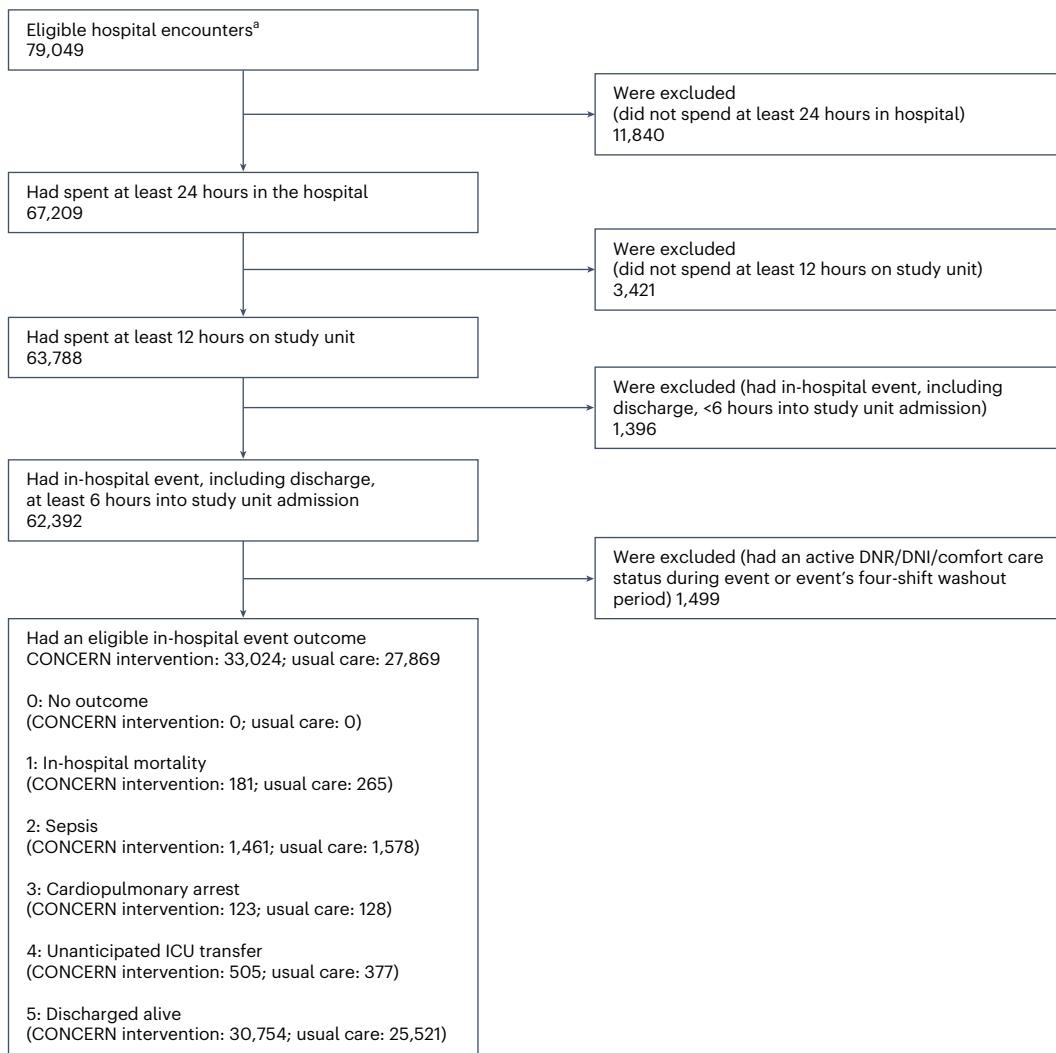


Fig. 2 | Flow diagram of hospital encounters assessed for eligibility and allocation to intervention or usual care for in-hospital event outcomes.
A total of 79,049 hospital encounters (58,994 unique patients) were screened for eligibility. After applying exclusion criteria, 60,893 hospital encounters met

eligibility requirements and were included in the analysis. From this, 33,024 encounters were allocated to the CONCERN intervention group and 27,869 encounters were assigned in the usual-care group.^a Total eligible patients: 58,994.

After calculating the nurses' concern level, the model assigns a categorical deterioration risk score of green (low), yellow (increased) or red (high), updates the score hourly and presents the score to care team members on the CONCERN clinical-decision-support EWS display in the EHR. CONCERN EWS research findings have demonstrated that it can predict patient deterioration 42 hours earlier than other leading EWSs¹⁰, and, through its use, nurses and prescribing providers perceive that their situational awareness is enhanced²³.

Earlier predictions provide clinicians with greater lead time for action²⁴, ultimately enabling the care team to identify patients who may be entering a risky state with enough time to intervene¹⁰. Therefore, the objective of this 1-year, multisite, pragmatic, cluster-randomized controlled clinical trial at two large health systems (four hospitals) was to determine at the individual-patient level per hospital encounter whether CONCERN EWS led to a decrease in coprimary outcomes (in-hospital mortality and LOS) and influenced secondary outcomes (cardiopulmonary arrest, sepsis, unanticipated ICU transfers and 30-day hospital readmission). In-hospital mortality was examined as instantaneous risk²⁵ using survival analysis (Methods). By using instantaneous risk we were able to evaluate the reduction in the risk of mortality during a nursing shift. Throughout the remainder of the article we refer to this coprimary outcome as in-hospital mortality. LOS, our other

coprimary outcome, can be used as a proxy for other outcomes, such as faster healing and quicker recovery. However, interpreting shorter LOS as positive is not universal because patients who die also can have a shorter LOS. To address this measurement challenge and account for LOS shortened due to mortality, we analyzed LOS in two different ways. The Cox proportional hazards (PH) model, a type of survival analysis, can account for deaths by examining time-to-discharged alive but is less intuitive due to its use of hazard ratios (HR). To provide clearer insights for study impact on LOS, we used both the PH model, which includes deceased patients, and a generalized linear model (GLM) analysis, which estimates adjusted mean LOS by study assignment group status and is easier to interpret. We hypothesized a priori that patients whose care team received CONCERN EWS risk scores would have a lower risk of mortality and LOS during that hospitalization than a control group of patients with non-EWS-informed teams (Fig. 1).

Results

Trial participants and hospital encounters

During the study time frames at health system sites A and B, 79,049 hospital-patient encounters (hospital admissions) occurred on our study units for 58,994 unique patients. After excluding 18,156 hospital-patient encounters that did not meet eligibility criteria, a total of

Table 1 | Characteristics of hospital encounters during the trial (N=60,893)

	CONCERN intervention (n=33,024 ^a)	Usual care (n=27,869 ^a)	Standardized mean difference ^b
Age	62.62±17.56	63.67±17.10	0.06
Male—no. (%)	16,056 (48.62)	13,966 (50.11)	0.03
Race ^c —no. (%)			
White	18,923 (57.30)	16,278 (58.41)	0.02
Black	4,878 (14.77)	4,224 (15.16)	0.01
Asian	703 (2.13)	634 (2.27)	0.01
Other or missing	8,520 (25.80)	6,733 (24.16)	0.04
Ethnic group ^c —no. (%)			
Not Hispanic or Latino	23,618 (71.52)	20,304 (72.86)	0.03
Hispanic or Latino	7,442 (22.54)	5,903 (21.18)	0.03
Unknown or not reported	1,964 (5.95)	1,662 (5.96)	0.0004
Primary language English—no. (%)	26,713 (80.89)	22,584 (81.04)	0.004
Charlson Comorbidity Index	3.11±3.16	3.51±3.34	0.12
Discharge disposition ^d —no. (%)			
Home	26,773 (81.07)	22,206 (79.68)	0.04
Other	6,251 (18.93)	5,663 (20.32)	0.04
Unit type ^e			
Acute care unit	31,050 (94.02%)	25,704 (92.23%)	0.07
ICU	1,974 (5.98%)	2,165 (7.77%)	0.07

^aReported n for intervention and usual care refers to survival analysis. ^bStandardized mean difference was defined as the difference in means or proportions (between the two arms) divided by the pooled standard deviation. All standardized mean differences are <0.2; therefore, no between-group differences. ^cRace and ethnic group were reported in the EHR. ^dDischarge disposition: home includes home with services; other includes any disposition not to home (including hospital mortality). ^eBased on allocation to intervention or usual-care unit for in-hospital event outcomes. ± Plus–minus values are mean±s.d.

60,893 were included in trial analyses, with 33,024 in the CONCERN EWS intervention group and 27,869 in the usual-care control group for our survival analyses (Fig. 2).

Overall, hospital–patient encounters across intervention and usual-care groups represented similar age, race, ethnicity and illness severity distributions (Table 1). Illness severity is demonstrated in Table 1 using two proxy variables: the Charlson Comorbidity Index and discharge disposition. Additional site comparisons are provided in Extended Data Tables 1 and 2.

Coprimary outcomes: in-hospital mortality and LOS

The intervention group had a lower in-hospital mortality and decreased LOS. A total of 181 patient encounters in the intervention group and 265 patient encounters in the usual-care group experienced in-hospital mortality in our survival analysis, with a 35.6% decrease in instantaneous risk of mortality in the hospital with the CONCERN intervention (adjusted HR, 0.64; 95% confidence interval (CI), 0.53–0.78; $P<0.0001$) (Table 2). To act as reference only and not a study outcome, we also calculated crude mortality rates, which were 10.02 per 10,000 inpatient days on intervention units and 15.84 per 10,000 inpatient days on usual-care units (Table 2). Further, a sensitivity analysis showed that any imbalance in patient encounters with a diagnosis of COVID-19 on intervention and usual-care units did not have a statistically significant impact on our findings (Extended Data Table 3). It is important to note that all mortality calculations were conducted after

applying our prespecified exclusions to our hospital–patient encounter cohort that intentionally excluded hospital–patient encounters with known high rates of mortality, such as those associated with hospice patients, those with do not resuscitate (DNR) or do not intubate (DNI) orders and admissions to specialty units (for example, oncology). Please see Table 3 for definitions of all study outcomes, treatment effects, and statistical model.

There was a 11.2% decrease in mean LOS with the CONCERN intervention (adjusted incidence rate ratio (IRR), 0.91; 95% CI, 0.90–0.93; $P<0.0001$). The estimated unadjusted mean LOS (a known underestimate given that we cannot account for patients who died) was 157.9 hours (6.6 days) for hospital–patient encounters in intervention units and 172.7 hours (7.2 days) for hospital–patient encounters in usual-care units (Table 2).

Secondary outcomes

Analyses performed between groups for each outcome, identified through our survival analysis as the first in-hospital event during that hospital–patient encounter, demonstrated that hospital–patient encounters in the intervention group had a lower instantaneous risk of sepsis and a higher instantaneous risk of unanticipated ICU transfer. Among hospital–patient encounters experiencing sepsis as the first hospital event, there were 1,461 hospital–patient encounters in the intervention group and 1,578 hospital–patient encounters in the usual-care group (adjusted HR, 0.96; 95% CI, 0.86–0.99; $P=0.0317$). For unanticipated ICU transfers, there were 505 hospital–patient encounters in the intervention group and 377 hospital–patient encounters in the usual-care group (adjusted HR, 1.25; 95% CI, 1.09–1.43; $P=0.0011$). No statistically significant differences between groups were found for cardiopulmonary arrest ($P=0.177$) or 30-day hospital readmission ($P=0.655$) (Table 2).

During the 42-hour period leading up to hospital encounter outcome events, the intervention group’s average time spent on CONCERN-assigned units was 40.27 (s.d. = 5.11) hours, and the control group’s average time spent on usual-care-assigned units was 40.19 (s.d. = 5.47) hours.

Discussion

Patients with hospital encounters during which the interprofessional care team was informed by CONCERN EWS were a third less likely to die and a quarter more likely to be transferred to intensive care. Unanticipated transfer to the ICU increased in the intervention group. Nurses can observe subtle patient changes that suggest clinical deterioration is more likely^{10,21}, and early recognition and escalation of treatment for these patients by the interprofessional care team, including transfer to the ICU, can improve outcomes. Given common challenges in interpreting LOS, we analyzed LOS in two ways and both indicated a statistically significant decrease, demonstrating the reliability of our results.

Other EWSs have influenced in-hospital mortality, sepsis or LOS, but not across the breadth of outcomes we observed^{4,5,26,27}. CONCERN EWS is a single hospital-wide intervention (that is, implemented across acute care units and ICUs) for all-cause deterioration. Most published studies target a particular condition (for example, sepsis) or hospital setting (for example, ICU)⁴ and only a minority of EWSs have been evaluated in randomized controlled trials^{4,6}. While there are other existing multifactorial machine learning clinical decision support implementations that share similarities to ours^{27–29}, the CONCERN intervention is distinct in its modeling approach by using metadata of nursing documentation as early predictors of patient deterioration in acute and intensive care settings, being fully integrated as a real-time EWS into the EHR and undergoing evaluation in a prospective multi-site, cluster-randomized, pragmatic trial that has demonstrated its positive influence on multiple patient outcomes. Additionally, a 2022 systematic review identified only 41 randomized controlled trials for machine learning interventions in health care, and none adhered to

Table 2 | CONCERN impact on coprimary and secondary outcomes

Variable	CONCERN intervention (n=33,024 ^g)	Usual care (n=27,869 ^g)	Unadjusted HR ^a (95% CI)	Adjusted HR ^b (95% CI)	Unadjusted IRR ⁱ (95% CI)	Adjusted IRR ⁱ (95% CI)	Unadjusted OR ^j (95% CI)	Adjusted OR ^j (95% CI)	P value ^k
Coprimary outcomes									
In-hospital mortality ^a	181 (0.55%)	265 (0.95%)	0.64 (0.53–0.77)	0.64 (0.53–0.78)	-	-	-	-	<0.0001
LOS	Discharged alive ^b	30,754 (93.13%)	25,521 (91.57%)	1.12 (1.10–1.14)	1.11 (1.09–1.13)	-	-	-	<0.0001
	LOS ^c	6.6 days ^l	7.2 days ^l	-	-	0.91 (0.89–0.92)	0.91 (0.90–0.93)	-	<0.0001
Secondary outcomes									
Cardiopulmonary arrest ^d	123 (0.37%)	128 (0.46%)	0.80 (0.63–1.03)	0.84 (0.66–1.08)	-	-	-	-	0.177
Sepsis ^d	1,461 (4.42%)	1,578 (5.66%)	0.82 (0.77–0.88)	0.93 (0.86–0.99)	-	-	-	-	0.0317
Unanticipated ICU transfers ^d	505 (1.53%)	377 (1.35%)	1.16 (1.02–1.33)	1.25 (1.09–1.43)	-	-	-	-	0.0011
30-day hospital readmission ^e	2,544 (7.70%)	2,234 (8.02%)	-	-	-	-	0.96 (0.90–1.02)	0.99 (0.93–1.05)	0.66
Additional analysis									
Crude mortality rate ^f	10.02 per 10,000 inpatient days	15.84 per 10,000 inpatient days	-	-	-	-	-	-	-

^aRelative risk of mortality for the CONCERN group to the usual-care group. The study assignment group was a time-dependent variable. We examined risk of mortality using a Cox PH model with time-dependent explanatory variables. We used survival analysis to model time from in-hospital admission to death. If a patient was discharged alive from the hospital, time-to-death outcome was censored at discharge. We used a frailty PH model to account for clustering at unit level. ^bRelative risk of discharged alive for the CONCERN group to usual-care group. The study assignment group was a time-dependent variable. We examined risk of discharged alive using a PH model with time-dependent explanatory variables. We used survival analysis to model time from in-hospital admission to discharged alive from hospital. If a patient died in the hospital, time-to-discharged alive outcome was censored at death. We used a frailty survival model to account for clustering at unit level. ^cRatio of time from in-hospital admission to discharged alive or death for the CONCERN group to the usual-care group. We used a GLM with log-link function (negative binomial regression). A patient's assigned study group was determined based on whether the patient was on intervention unit(s) or control unit(s) for the majority of their time during the hospital encounter. We used generalized estimating equations to account for clustering at unit level. ^dRelative risk of an in-hospital secondary event (cardiopulmonary arrest, sepsis or unanticipated ICU transfer) for the CONCERN group to the usual-care group. The study assignment group was a time-dependent explanatory variable. We examined the risk of each of these three events using a PH model with time-dependent explanatory variables. We used survival analysis to model time from in-hospital admission to the first occurrence of the in-hospital event of interest. If a patient died, had any of the other two in-hospital events or were discharged alive before the first occurrence of the in-hospital event of interest, time to the event outcome was censored at death, the first occurrence of the other in-hospital events or discharged alive. We used the frailty survival model to account for clustering at unit level. ^eOdds ratio (OR) of readmission for the CONCERN group to the usual-care group. Among patients discharged alive, we examined whether a patient was readmitted to a hospital in the same health system within 30 days of discharged alive using a GLM with logit-link function (logistic regression). A patient's assigned study group was determined based on whether the patient was on intervention unit(s) or control unit(s) for the majority of their time during the hospital encounter. ^fCrude mortality rate is reported for reference only. It is not a primary or secondary outcome; therefore, no P value or tests of statistical significance are calculated. Consistent with the outcomes reported, crude mortality rate was calculated after study patient population exclusions were applied. ^gReported n for intervention and usual care refers to survival analysis. ^hHR (CONCERN intervention as compared with usual care) and two-sided 95% CI were calculated using the Cox PH model. All HR are cause-specific. For HR, values of less than 1 indicate a lower relative risk of occurring for the CONCERN intervention group. ⁱIRR was calculated using GLM (negative binomial model). ^jOR was calculated using GLM (logistic model). ^kP values apply only to adjusted analyses. P values were computed using Wald statistics and two-sided. We determined statistical significance at 5% level. ^lEstimated unadjusted mean LOS (a known underestimate given we cannot account for patients who died in mean LOS). All analyses accounted for clustering at unit level.

all Consolidated Standards of Reporting Trials-Artificial Intelligence (CONSORT-AI) standards³⁰. We report our trial findings according to the CONSORT guidelines³¹ and extensions for AI³², cluster-randomized trials³³ and pragmatic trials³⁴ (Supplementary Table 1).

Early ICU transfer has been shown to be a 'window of critical opportunity' for timely clinical interventions that alter the trajectory of a patient's clinical progression, prevent adverse outcomes and improve survival^{35–39}. Several methodological choices allowed CONCERN EWS to have accurate and robust predictive power and sufficient lead time to alter a patient's clinical trajectory as we observed in this trial, including: (1) robust modeling of temporal data patterns^{40–42} (for example, time of day, day of week and patient hospital day) for health care processes (for example, nurses' surveillance and medication administration decisions) that reflect a nurse's level of concern about a patient's deterioration risk, and (2) the use of nursing assessment and observational data that may reflect patient condition changes earlier than other physiological values^{10,43,44} and that are information-rich^{13,21} but often not well understood outside of nursing. CONCERN EWS was developed by an interdisciplinary team that employed rigorous methods for translating a predictive model to a clinical setting, including nurse and prescribing-provider input on health care process effects

and real-time data availability, integration with existing clinical workflows, transparency and explainability to gain trust^{10,45}, as well as robust evaluation for external validity⁴⁶ and model fairness. This comprehensive implementation science approach resulted in 42-hour greater lead time than other EWSs⁸, a time frame that enables identification of patient deterioration several clinical shifts before an adverse event and provides the additional time and opportunity needed to intervene in clinically meaningful ways. By leveraging nursing surveillance patterns rather than physiological measures^{10,44}, CONCERN EWS overcomes the limitations of physiological measurements⁸ and leverages an existing expert knowledge base: nurses' autonomous decisions to take greater action and document more observations or notes than mandated by policy in direct response to concerning changes in a patient's clinical state^{10,13}. Mortality increases when care escalation is delayed⁴⁷; therefore, transparency and explainability, achieved by displaying the CONCERN score and factors driving the prediction, may overcome the dissonance between nurses' identification of patient risk based on subtle clinical indicators and prescribing providers' expectations of changes in physiological data, thereby prompting earlier decisions to escalate medical intervention. We anticipate several future clinical applications to our CONCERN EWS approach, including

Table 3 | Definitions of CONCERN study outcomes, treatment effects and statistical model

Study outcome (end point)	Definition	Time period	Treatment effect	Statistical model	
Primary	Mortality	Instantaneous risk of in-hospital all-cause mortality.	Measured during a specified time window ^a , given that death has not yet happened, during the patient's hospital encounter after being on a study unit for at least 12 h.	HR	Multivariate adjusted Cox PH model with time-dependent explanatory variables.
	LOS	In our analysis we examined LOS in two ways. (1) Discharged alive: the probability of being discharged alive. (2) Time elapsed between patient admission to an inpatient care unit and discharged alive from the hospital or died during hospitalization. This measure is also known as time to discharge.	(1) Measured during a specified time window ^a , given that discharge has not yet happened, during the patient's hospital encounter, after being on a study unit for at least 12 h. (2) Measured for patients that were on a study unit for at least 12 h. All LOS calculations excluded pre-inpatient hospital time (for example, emergency department stay).	(1) HR (2) IRR	(1) Multivariate adjusted Cox PH model with time-dependent explanatory variables. (2) GLM with log-link function (negative binomial regression). The study assignment group was determined based on whether the patient was on intervention unit(s) or control unit(s) for the majority of their time during hospitalization.
	Cardiopulmonary arrest	The instantaneous risk of a Code Blue event.	Measured during a specified time window ^a , given the event has not yet happened, during a hospital-patient encounter, after being on a study unit for at least 12 h.	HR	Multivariate adjusted Cox PH model with time-dependent explanatory variables. We used survival analysis to model time from in-hospital admission to the first occurrence of the in-hospital event of interest. If patients died, had any of the other two in-hospital events or were discharged alive before the first occurrence of the in-hospital event of interest, time to the event outcome was censored at death, the first occurrence of the other in-hospital events or discharged alive.
Secondary	Sepsis	Instantaneous risk of sepsis, defined by Sepsis-3 (ref. 51) of concurrent suspected infection-positive (body fluid cultures ordered, and oral or parenteral antibiotics administered within a specified time range) and SOFA or quick (q)SOFA-positive criteria.	Measured during a specified time window ^a , given the event has not yet happened, during the hospital-patient encounter, after being on a study unit for at least 12 h.		GLM with logit-link function (logistic regression). The study assignment group was determined based on whether the patient was on intervention unit(s) or usual-care unit(s) for the majority of their time during hospitalization.
	Unanticipated ICU transfer	Instantaneous risk of a direct transfer from an acute care unit to the ICU, not resulting from a planned procedure.	Measured during a specified time window ^a , given the event has not yet happened, during the hospital-patient encounter, after being on a study unit for at least 12 h.		
	30-day hospital readmission	Unplanned hospital readmission of at least 24 h (for multiple events, the earliest was counted) for any cause to an acute care hospital (trackable in the same health system) within 30 days of discharged alive from a hospital-patient encounter in our study, excluding planned, same-day and other specific types of readmissions ⁵² .	Among patients discharged alive, we examined whether a patient was readmitted to a hospital in the same health system within 30 days of discharge.	OR	

^aThe specified time window was measured using 12-hour increments to align with the duration of typical clinical shifts.

expansion to other clinical practice patterns and patient populations within and beyond the hospital setting.

A major purpose of hospitalization is to provide continuous, 'around-the-clock' nursing care that would not be feasible in an outpatient setting; however, health care systems and clinicians are undergoing many challenges today. Despite the well-documented poor staffing, increased patient acuity and limited hospital resources rampant across the globe during the time we conducted our study (2020 to 2022)^{1,48}, CONCERN EWS still improved outcomes over usual care across both study sites. The predictors in CONCERN EWS are based on the documented decisions that a nurse makes about how and when to provide nursing care. As such, CONCERN EWS' nurse-centered, AI-based approach to measuring nursing value and capitalizing on nursing expertise is critically needed amidst this time of clinician burnout and turnover⁴⁸.

The reported findings have important limitations. First, the trial was complicated by disruptions at the sites due to the COVID-19 pandemic, which forced a 1-year postponement at one of our sites due to implementation delays. We acknowledge that the results may not be generalizable because the trial was conducted during the COVID-19 pandemic; however, as reported, sensitivity analysis showed that any

imbalance in patients diagnosed with COVID-19 on intervention and usual-care units did not have a statistically significant impact on our findings (Extended Data Table 3). Second, the trial was conducted at two health systems located in different urban areas within the Northeastern United States, and findings may not be generalizable to hospitals in other settings, especially in other countries that may have different nursing practices. Third, the usual-care arm included 12 ICUs compared to nine ICUs in the intervention arm, although evaluation of baseline characteristics between groups indicated no differences and analyses adjusted for confounders. Future analyses will investigate additional workflow and process measures on the observed impacts. Fourth, unmeasured confounding might impact differences in trial outcomes. However, our analyses adjusted for factors commonly established as explanatory for variation in our outcomes. Fifth, there are well-known issues when reusing EHR data for secondary analyses⁴⁹. To address this, we applied a quality assessment framework to all variables⁵⁰. Sixth, data may be stored in different ways in other EHRs. Due to this variability in EHR structures, we purposely employed two approaches: (1) we used observational variables that are a proxy for nursing surveillance and are relatively simple and ubiquitous (for example, vital signs and note-writing patterns), which our prior research identified

had strong predictive relationships for deterioration events; and (2) we purposely used the Fast Healthcare Interoperability Resources methods, also known as FHIR^(R), to retrieve real-time data due to their capability for excellent standards-based mapping of observations across EHR systems. Seventh, we acknowledge the known limitations to the standard Centers for Medicare & Medicaid Services definition of 30-day readmission, including loss of follow-up, for patients readmitted within a different hospital system and patients who died and are not known to have died. Finally, eighth, while our discussion like other EWS interventional studies in hospital settings attributes mortality to a single encounter, we acknowledge that mortality may result from a series of events extending beyond a single hospital encounter. However, the use of a cluster-randomized controlled trial helps mitigate unknown confounders.

In conclusion, our study demonstrates that nursing surveillance patterns are a valuable signal to predict deterioration of hospitalized patients. A hospital-wide EWS based on nursing surveillance patterns resulted in a 35.6% decrease in in-hospital instantaneous risk of mortality, 11.2% decrease in LOS, 7.5% decrease in in-hospital instantaneous risk of sepsis and 24.9% increase in in-hospital instantaneous risk of unanticipated ICU transfers, when integrated into the care team's EHR workflow compared with usual care. CONCERN EWS research continues and plans include studies at additional sites and countries, implementation science evaluations and expansion of the predictive model to other hospital units (for example, emergency departments) and inpatient populations (for example, pediatrics).

Online content

Any methods, additional references, Nature Portfolio 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/s41591-025-03609-7>.

References

1. American Hospital Association. Report: Rising patient acuity driving up hospital costs as payments fall. *AHA News* <https://www.aha.org/news/headline/2022-08-15-report-rising-patient-acuity-driving-hospital-costs-payments-fall> (2022).
2. Hogan, H. et al. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *BMJ Qual. Saf.* **21**, 737–745 (2012).
3. Patient safety. *The Joint Commission* www.jointcommission.org/facts_about_patient_safety/ (2015).
4. Wan, Y.-K. J. et al. Information displays for automated surveillance algorithms of in-hospital patient deterioration: a scoping review. *J. Am. Med. Inform. Assoc.* **31**, 256–273 (2023).
5. Schmidt, P. E. et al. Impact of introducing an electronic physiological surveillance system on hospital mortality. *BMJ Qual. Saf.* **24**, 176–177 (2015).
6. Lee, T. C., Shah, N. U., Haack, A. & Baxter, S. L. Clinical implementation of predictive models embedded within electronic health record systems: a systematic review. *Informatics (MDPI)* **7**, 25 (2020).
7. 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. (Millwood)* **33**, 1123–1131 (2014).
8. Fu, L. et al. Development and validation of early warning score system: a systematic literature review. *J. Biomed. Inform.* **105**, 103410 (2020).
9. Drew, B. J. et al. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. *PLoS ONE* **9**, e110274 (2014).
10. Rossetti, S. et al. Healthcare process modeling to phenotype clinician behaviors for exploiting the signal gain of clinical expertise (HPM-ExpertSignals): development and evaluation of a conceptual framework. *J. Am. Med. Inform. Assoc.* **28**, 1242–1251 (2021).
11. Halverson, C. C. & Scott Tilley, D. Nursing surveillance: a concept analysis. *Nurs. Forum* **57**, 454–460 (2022).
12. Collins, S. A., Fred, M. R., Wilcox, L. & Vawdrey, D. K. Workarounds used by nurses to overcome design constraints of electronic health records. *NI 2012* **2012**, 93–97 (2012).
13. Collins, S. A. et al. Relationship between nursing documentation and mortality. *Am. J. Crit. Care.* **22**, 306–313 (2013).
14. Endsley, M. R. Toward a theory of situation awareness in Dynamic Systems. *Hum. Factors J. Hum. Factors Ergon. Soc.* **37**, 32–64 (1995).
15. Tokareva, I. & Romano, P. Failure to rescue. *PSNet* <https://psnet.ahrq.gov/primer/failure-rescue> (2025).
16. Odell, M., Victor, C. & Oliver, D. Nurses' role in detecting deterioration in ward patients: systematic literature review. *J. Adv. Nurs.* **65**, 1992–2006 (2009).
17. Burns, P. B., Rohrich, R. J. & Chung, K. C. The levels of evidence and their role in evidence-based medicine. *Plast. Reconstr. Surg.* **128**, 305–310 (2011).
18. Agency for Healthcare Research and Quality. Rapid Response Systems <https://psnet.ahrq.gov/primer/rapid-response-systems> (2019).
19. Cioffi, J. Recognition of patients who require emergency assistance: a descriptive study. *Heart Lung* **29**, 262–268 (2000).
20. Jones, L., King, L. & Wilson, C. A literature review: factors that impact on nurses' effective use of the Medical Emergency Team (MET). *J. Clin. Nurs.* **18**, 3379–3390 (2009).
21. Collins, S. A. & Vawdrey, D. K. "Reading between the lines" of flowsheet data: nurses' optional documentation associated with cardiac arrest outcomes. *Appl. Nurs. Res.* **25**, 251–257 (2012).
22. Opitz, D. & Maclin, R. Popular ensemble methods: an empirical study. *J. Artif. Intell. Res.* **11**, 169–198 (1999).
23. Hobensack, M. et al. Identifying barriers to the implementation of Communicating Narrative Concerns Entered by Registered Nurses (CONCERN), an early warning system smartapp. *Appl. Clin. Inform.* **15**, 295–305 (2024).
24. Singh, K. et al. Evaluating a widely implemented proprietary deterioration index model among hospitalized patients with COVID-19. *Ann. Am. Thorac. Soc.* **18**, 1129–1137 (2021).
25. Sashegyi, A. & Ferry, D. On the interpretation of the hazard ratio and communication of survival benefit. *Oncologist* **22**, 484–486 (2017).
26. Adams, R. et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. *Nat. Med.* **28**, 1455–1460 (2022).
27. Escobar, G. J. et al. Automated identification of adults at risk for in-hospital clinical deterioration. *N. Engl. J. Med.* **383**, 1951–1960 (2020).
28. Levin, M. A. et al. Real-time machine learning alerts to prevent escalation of care: a nonrandomized clustered pragmatic clinical trial. *Crit. Care Med.* **52**, 1007–1020 (2024).
29. Churpek, M. M. et al. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit. Care Med.* **44**, 368–374 (2016).
30. Plana, D. et al. Randomized clinical trials of machine learning interventions in health care: a systematic review. *JAMA Netw. Open* **5**, e2233946 (2022).
31. Moher, D. et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Brit. Med. J.* **340**, c869 (2010).

32. Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J. & Denniston, A. K. on behalf of SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit. Health* **2**, e537–e548 (2020).
33. Campbell, M. K., Piaaggio, G., Elbourne, D. R. & Altman, D. G. CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *Brit. Med. J.* **345**, e5661 (2012).
34. Zwarenstein, M. et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *Brit. Med. J.* **337**, a2390 (2008).
35. Kiekas, P. et al. Delayed admission to the intensive care unit and mortality of critically ill adults: systematic review and meta-analysis. *Biomed. Res. Int.* **2022**, 4083494 (2022).
36. Churpek, M. M. et al. Association between intensive care unit transfer delay and hospital mortality: A multicenter investigation. *J. Hosp. Med.* **11**, 757–762 (2016).
37. Solomon, R. S., Corwin, G. S., Barclay, D. C., Quddusi, S. F. & Dannenberg, M. D. Effectiveness of rapid response teams on rates of in-hospital cardiopulmonary arrest and mortality: A systematic review and meta-analysis. *J. Hosp. Med.* **11**, 438–445 (2016).
38. Hu, W., Chan, C. W., Zubizarreta, J. R. & Escobar, G. J. An examination of early transfers to the ICU based on a physiologic risk score. *Manuf. Serv. Oper. Manag.* **20**, 531–549 (2018).
39. Grieve, R. et al. Analysis of benefit of intensive care unit transfer for deteriorating ward patients: a patient-centered approach to clinical evaluation. *JAMA Netw. Open* **2**, e187704 (2019).
40. Hripcsak, G. & Albers, D. J. Next-generation phenotyping of electronic health records. *J. Am. Med. Inform. Assoc.* **20**, 117–121 (2013).
41. Albers, D. J. & Hripcsak, G. Using time-delayed mutual information to discover and interpret temporal correlation structure in complex populations. *Chaos* **22**, 13111 (2012).
42. Albers, D. J. & Hripcsak, G. A statistical dynamics approach to the study of human health data: resolving population scale diurnal variation in laboratory data. *Phys. Lett. A* **374**, 1159–1164 (2010).
43. Pivovarov, R., Albers, D. J., Hripcsak, G., Sepulveda, J. L. & Elhadad, N. Temporal trends of hemoglobin A1c testing. *J. Am. Med. Inform. Assoc.* **21**, 1038–1044 (2014).
44. Pivovarov, R., Albers, D. J., Sepulveda, J. L. & Elhadad, N. Identifying and mitigating biases in EHR laboratory tests. *J. Biomed. Inform.* **51**, 24–34 (2014).
45. Schwartz, J. M. et al. Factors influencing clinician trust in predictive clinical decision support systems for in-hospital deterioration: qualitative descriptive study. *JMIR Hum Factors* **9**, e33960 (2022).
46. Wong, A. et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern. Med.* **181**, 1065–1070 (2021).
47. Sankey, C. B., McAvay, G., Siner, J. M., Barsky, C. L. & Chaudhry, S. I. “Deterioration to door time”: an exploratory analysis of delays in escalation of care for hospitalized patients. *J. Gen. Intern. Med.* **31**, 895–900 (2016).
48. Office of Surgeon General. *Addressing Health Worker Burnout: The U.S. Surgeon General’s Advisory on Building a Thriving Health Workforce* (US Department of Health and Human Services, 2022).
49. Lewis, A. E. et al. Electronic health record data quality assessment and tools: a systematic review. *J. Am. Med. Inform. Assoc.* **30**, 1730–1740 (2023).
50. Weiskopf, N. G., Bakken, S., Hripcsak, G. & Weng, C. A data quality assessment guideline for electronic health record data reuse. *EGEMS (Wash DC)* **5**, 14 (2017).
51. Singer, M. et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* **315**, 801–810 (2016).
52. Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (YNNHSC/CORE). *2023 Hospital-Wide Readmission Measure Updates and Specifications Report – Version 12.0.* (2023); <https://qpp.cms.gov/resources/document/fab5e228-1667-4e8f-9b49-ea04fb3cf81>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2025

Sarah C. Rossetti  ^{1,2}✉, **Patricia C. Dykes** ^{3,4}, **Chris Knaplund** ¹, **Sandy Cho**  ⁵, **Jennifer Withall** ¹, **Graham Lowenthal** ³, **David Albers** ^{1,6}, **Rachel Y. Lee**  ¹, **Haomiao Jia** ^{2,7}, **Suzanne Bakken** ^{1,2}, **Min-Jeoung Kang** ^{3,4}, **Frank Y. Chang** ³, **Li Zhou** ^{3,4}, **David W. Bates**  ^{3,4}, **Temiloluwa Daramola** ¹, **Fang Liu** ⁸, **Jessica Schwartz-Dillard** ^{2,9}, **Mai Tran**  ¹, **Syed Mohtashim Abbas Bokhari**  ¹, **Jennifer Thatte** ¹⁰ & **Kenrick D. Cato** ^{8,11}

¹Columbia University Irving Medical Center, Department of Biomedical Informatics, New York, NY, USA. ²Columbia University Irving Medical Center, School of Nursing, New York, NY, USA. ³Brigham and Women’s Hospital, Boston, MA, USA. ⁴Harvard Medical School, Boston, MA, USA. ⁵Newton-Wellesley Hospital, Newton, MA, USA. ⁶University of Colorado, Anschutz Medical Campus, Department of Biomedical Informatics, Aurora, CO, USA. ⁷Columbia University Irving Medical Center, Mailman School of Public Health, New York, NY, USA. ⁸University of Pennsylvania, Philadelphia, PA, USA. ⁹Hospital for Special Surgery, New York, NY, USA. ¹⁰Siena College, Loudonville, NY, USA. ¹¹Children’s Hospital of Philadelphia, Philadelphia, PA, USA.

✉ e-mail: sac2125@cumc.columbia.edu

Methods

Study design, trial sites and randomization

We evaluated CONCERN EWS in a 1-year multisite, pragmatic, cluster-randomized controlled clinical trial at two large health systems in the Northeastern United States. Each system was a study site (A and B) and each comprised one academic medical center and one community hospital. Both sites used the Epic EHR. Clusters were the individual study units across the four hospitals that included all nonspecialty acute care units (ACUs) and ICUs. The ACUs were general/medical surgical standard floor units, with a mix of teaching and hospitalist services that also included step-down beds. Specialty ACUs, such as oncology, psychiatric and rehabilitation, were excluded. Typical nurse to patient ratios on our study units were one nurse to four ACU patients, two to three step-down patients or one to two ICU patients. Randomization was performed at the unit level and stratified by site before trial initiation⁵³. No clinical units dropped out during the study period. Each study unit (ACU or ICU) was randomly allocated to one of two groups (CONCERN EWS intervention or usual-care control) using a computer-generated randomization scheme⁵³. The trial included 74 clinical units (37 intervention, 37 usual care) with the following distribution: site A intervention—19 ACUs and 5 ICUs; site A usual care—17 ACUs and 7 ICUs; site B intervention—9 ACUs and 4 ICUs; site B usual care—8 ACUs and 5 ICUs (Extended Data Table 4). The intervention (CONCERN EWS prediction risk score) was displayed to the care team consisting of nurses and prescribing providers (that is, physicians, nurse practitioners and physician assistants) for each hospital–patient encounter only if that patient was admitted to an intervention-assigned unit. The CONCERN EWS prediction risk score was not generated for hospital–patient encounters on usual-care units; therefore, no EWS intervention was displayed to care teams of patients admitted to usual-care-assigned units.

Due to unavoidable delays related to the COVID-19 pandemic, the study time frame differed for each site. For the cohort at site A, the first inpatient hospital admission was 1 October 2020, the last inpatient hospital admission was 31 October 2021 and the last inpatient hospital discharge was 30 March 2022. At site B, the first inpatient hospital admission was 1 October 2021, the last inpatient hospital admission was 30 October 2022 and the last inpatient hospital discharge was 26 December 2022. The first participant was enrolled on 1 October 2020 and the last participant of the study was enrolled on 30 October 2022. A previous publication described interrupted time series analysis for our cluster-randomized pragmatic trial⁵⁴, but as described in detail in ‘Statistical analysis’, due to patient movement (for example, transfers) across units, we removed the unit-level analysis from the final analysis and analyzed all outcomes at the individual hospital–patient encounter level (see Supplementary Methods 1 for details of these modifications and the rationale from original to final analysis plan). Study design remained the same as originally planned and specified in [ClinicalTrial.gov](#). Institutional review boards at Columbia University, United States, and Brigham and Women’s Hospital, United States, approved the protocol before trial initiation with a waiver of consent. No adverse events were reported during the trial.

Trial participants and outcomes

While randomization occurred at the unit level, outcomes were assessed at the individual hospital–patient encounter level due to patient movement across units during their hospitalization. Hospital–patient encounters on study units were included in analyses if the patients were 18 years of age and older, hospitalized for more than 24 hours (EWS score begins displaying after 24 hours), admitted to a study unit for a minimum of 12 hours and free from any in-hospital event (including discharge) until at least 6 hours after study unit admission to avoid events that were influenced by previous care processes. Patient encounters that included hospice and palliative care, or DNR and/or DNI and comfort care orders activated before any trial outcome event, were excluded.

Coprimary outcomes were in-hospital mortality and LOS. Secondary outcomes were cardiopulmonary arrest, sepsis⁵¹, unanticipated ICU transfer and 30-day hospital readmission⁵². See Table 3 for definitions, including time frames used for measures. All data were collected from the EHR and evaluated and reported across study sites.

CONCERN EWS intervention

CONCERN EWS comprises a predictive model that uses an ensemble²² machine learning approach to process nursing surveillance patterns in the EHR, predicts clinical deterioration within the subsequent 24-hour period and displays the CONCERN EWS risk score as a green, yellow or red icon (non-interruptive alert) on the EHR patient list (main landing screen upon login) (Fig. 1). For all hospital–patient encounters on intervention units, the icon was visible to every nurse and prescribing provider on the patient’s clinical team, and the team could also double-click the icon to access a screen with prediction details. By targeting the entire clinical team rather than a solitary provider, CONCERN EWS promotes shared situational awareness. All existing hospital policies and procedures related to patient deterioration and escalation of care remained the same across intervention and usual-care groups.

Details of CONCERN EWS modeling approach (Extended Data Fig. 1), performance (Extended Data Table 5), factors and features (Extended Data Table 6) and EHR integration (Supplementary Methods 2) have been published elsewhere^{10,23,55}. Briefly, model development used 217,166 distinct inpatient–hospital encounters between 2015 and 2019 from two study sites, with a task of predicting a deterioration event in the next 24 hours based on the patient’s preceding 24-hour data. During model development, the first occurrence of cardiopulmonary arrest, rapid response, sepsis, transfer to ICU and death were identified as proxy measures for patient deterioration. The gradient-boosted decision-tree models were trained at site B on 70% of a retrospective dataset, with 30% used for tenfold cross-validation, with more than 97% average accuracy and 95% average area under the curve (average performance reported for our ensemble models²²) (Extended Data Table 5 and Supplementary Methods 2) and were externally validated at site A.

Our risk-prediction modeling is nonlinear, with time varying variables, which requires accuracy evaluation that must take many aspects of the model, data and clinical problem into account^{56,57}. Through experimentation on retrospective datasets of hospitalized patients’ encounters in the EHR, we refined our machine learning modeling approach, which is focused on temporal patterns of nursing observations, using ensemble learning to identify the best performing model for a patient’s specific clinical situation for each hour of the day⁵⁸. For example, ensemble learning allowed us to identify the best model for a patient who has been in an ICU for 5 hours on a Friday night and to make our EWS transparent and explainable. Every hour, the CONCERN EWS processes EHR data from the past 24 hours through the ensemble models that best fit the patient characteristics and calculates a risk score. We are not aware of any other EWS decision-support implementations that use a similar multifactorial, advanced computational and explainable modeling approach¹⁰.

Statistical analysis

The primary objective was to test the hypothesis that CONCERN EWS decreases in-hospital mortality and LOS (independent coprimary outcomes). Secondary outcomes were CONCERN EWS’s influence on in-hospital events (that is, cardiopulmonary arrest, sepsis and unanticipated ICU transfers) and 30-day hospital readmissions. A power analysis comparing mortality rates between groups estimated that a 1-year trial at our two sites’ targeted units would result in a sufficient sample size for at least 80% statistical power to detect a difference of less than 1% relative difference in mortality rates (two-sided; $\alpha = 0.05$). This is based on an expected sample of 2,000 total admissions per month with 50,000 total inpatient days based on average occupancy and a mean mortality rate of 37.5 deaths per 10,000 inpatient days (ranging from

11.9 to 48.8 across our study sites) (Extended Data Table 7). This power analysis was based on a cluster-randomized trial using intraclass correlation coefficient equal to 0.2 and generalized linear mixed models for the comparison of mortality rates between those randomized to CONCERN intervention and usual-care units.

Our study design as a cluster-randomized trial never changed. However, we modified our original analysis plan, which was to analyze outcomes both at the unit level and individual-patient level per hospital encounter (that is, each hospital admission), to instead analyze outcomes only at the individual-patient level per hospital encounter. We did so because patient movement (transfers) between clinical units occurs frequently.

It was also possible for patients to have multiple hospital–patient encounters during the trial time frame. We included all eligible hospital encounters in our analyses after applying our study inclusion and exclusion criteria consistently and independently for each encounter ('Trial participants and outcomes'). We analyzed each hospital encounter independently. For example, in the case of a patient with five study-period hospital encounters who dies during the last encounter, we included each one in the analysis (that is, four encounters without a mortality event, one encounter with a mortality event). To this point, all measures, including patient characteristics in Table 1, were analyzed for each hospital–patient encounter (and not each patient) because the reported characteristics could change between hospital encounters (for example, Charlson Comorbidity Index is assessed at each admission; discharge disposition is per encounter).

All HRs are cause-specific. Covariates were selected based on our literature review of EWS studies; we adjusted for age^{27,59–67}, sex^{27,59,62–65,67}, ethnicity⁶⁰, comorbidity^{27,59,61,63,64,66,67} and ICU status⁵⁹; as well as race and study site, all of which were identified as confounding variables that may influence outcomes. Specifically, all survival analyses performed for coprimary (in-hospital mortality and LOS) and secondary outcomes (cardiopulmonary arrest, sepsis and unanticipated transfer to ICU) were adjusted for age, Charlson Comorbidity Index, sex, race, ethnicity, study site and ICU status. GLM analyses for LOS and 30-day hospital readmission were adjusted for age, Charlson Comorbidity Index, sex, race, ethnicity and study site. Charlson Comorbidity Index, age, sex, race, ethnicity and study site were time-invariant covariates for each hospital encounter given that they are static variables during the hospital stay. ICU status was included as a time-dependent covariate in our survival analysis given that it may be influenced by other outcomes and patients may be transferred into and out of an ICU at any point during their hospital encounter. We used the updated Quan et al. version of the Charlson Comorbidity Index⁶⁸ because it is recognized as the most widely used comorbidity scoring system⁶⁹ and, therefore, could be used reliably across both study sites. All analyses accounted for clustering at the unit level. Patients' age, sex, race and ethnicity data were sourced from the EHR.

All statistical analyses were performed using SAS v.9.4 (SAS Institute Inc.). Please see detailed descriptions below for definitions of study outcomes, treatment effects and statistical models (see also Table 2 notes and Table 3).

Coprimary outcome: in-hospital mortality. We examined in-hospital mortality as the instantaneous risk of mortality. We defined this as the probability of mortality during a specified time window, given that death has not yet happened, during the patient's hospital encounter after being on a study unit for at least 12 hours. The time window was measured in 12-hour increments to align with the duration of typical clinical shifts. We aimed to detect relative risk of mortality from the CONCERN intervention group to the usual-care group. The study assignment group was a time-dependent variable. We examined risk of mortality using the Cox PH model with time-dependent explanatory variables. We used survival analysis to model time from in-hospital admission to death⁷⁰. If a patient was discharged alive from

the hospital, time-to-death outcome was censored at discharge. We used a frailty PH model to account for clustering at the unit level. Given that our study took place during the pandemic, we performed a sensitivity analysis for imbalance in COVID-19 diagnoses per patient–hospital encounters across study arms by excluding patient encounters with a COVID-19 diagnosis and repeating the survival analysis.

Coprimary outcome: LOS analyses. As discussed earlier, LOS can be used as a proxy for other outcomes (for example, shorter LOS associated with a quicker recovery). Yet a positive interpretation is not universal because patients who die can also have a shorter LOS. To address this measurement challenge, we compared LOS between intervention and usual care using two analyses. In the first LOS analysis, we examined discharged alive status, defined as the probability of being discharged alive during a specified time window, given that discharge has not yet happened, during the patient's hospital encounter after being on a study unit for at least 12 hours. The time window was measured in 12-hour increments to align with the duration of typical clinical shifts. In the second LOS analysis, we examined time elapsed between patient admission to an inpatient care unit and discharged alive from the hospital or died during hospitalization for patients that were on a study unit for at least 12 hours. LOS calculations excluded pre-inpatient hospital time (for example, emergency department stay).

The first LOS analysis determined the relative risk of being discharged alive from the CONCERN group to usual-care group by examining the risk of discharged alive using the Cox PH model with time-dependent explanatory variables. As the study assignment group was a time-dependent variable, we used a PH model with time-dependent explanatory variables. We used survival analysis to model time from in-hospital admission to discharged alive status. If a patient died in the hospital, time-to-discharged alive outcome was censored at death. We used the frailty survival model to account for clustering at unit level.

Rationales for PH model use with time-dependent explanatory variables were to censor for death and account for patient transfers between units. We aimed to estimate adjusted mean LOS by study assignment group status. However, as there were several time-varying explanatory variables and adjusting for all of them in the PH model introduces complexity to the reliability and validity of estimates, we instead selected to perform the second LOS analysis evaluating time from in-hospital admission to discharged alive or death (if died) using GLM with log-link function (negative binomial regression).

This second LOS analysis determined the ratio of time from in-hospital admission to discharged alive or death (if died) from the CONCERN group to the usual-care group (Table 2)⁷¹. We used GLM (negative binomial regression) to test the IRR for the event of staying in hospital for our LOS outcome between the two treatment groups⁷¹. As our first approach for analyzing LOS used survival analysis and censored for death, we did not censor for death in our GLM analysis. As an alternative, we used the number of hospital days a patient stayed in the hospital as a second way to measure our LOS outcome. The event we modeled was a patient staying in the hospital, as opposed to the patient being discharged alive, because the outcome variable LOS is the number of days in the hospital. We estimated mean LOS from GLM sandwich standard error estimates and used generalized estimating equations to account for clustering at the unit level. The study assignment group (intervention or usual care) for each hospital–patient encounter was determined based on whether the patient was on intervention unit(s) or usual-care unit(s) for the majority of time during that encounter (see 'Transfers between intervention and usual-care units' for a detailed description of this method).

Secondary outcomes: in-hospital events. Secondary outcomes that included in-hospital events were cardiopulmonary arrest, sepsis and unanticipated transfer to the ICU. Cardiopulmonary arrest was examined as the instantaneous risk of a Code Blue event. The

probability of a Code Blue event was measured during a specified time window, given the event has not yet happened, during a hospital–patient encounter, after being on a study unit for at least 12 hours. Sepsis, which was examined as the instantaneous risk of sepsis, was defined by Sepsis-3 (ref. 51) as concurrent suspected infection-positive (body fluid cultures ordered, and oral or parenteral antibiotics administered within a specified time range) and Sequential Organ Failure Assessment (SOFA) or quick (q)SOFA-positive criteria based on Sepsis-3. The probability of sepsis was measured during a specified time window, given that sepsis has not yet happened, during the hospital–patient encounter, after being on a study unit for at least 12 hours. Unanticipated ICU transfer was examined as the instantaneous risk of a transfer from an ACU directly to the ICU, not resulting from a planned procedure. The probability of an unanticipated ICU transfer was measured during a specified time window, given the event has not yet happened, after being on a study unit for at least 12 hours. The time window for these secondary outcome events was measured in 12-hour increments to align with the duration of typical clinical shifts, to match the coprimary outcome events procedure.

We aimed to detect a relative risk of a hospital event (cardiopulmonary arrest, sepsis and unanticipated ICU transfer) from the CONCERN group to usual-care group. The study assignment group was a time-dependent explanatory variable. We examined risk of each of the three in-hospital secondary events (cardiopulmonary arrest, sepsis and unanticipated ICU transfer) using the PH model with time-dependent explanatory variables. We used survival analysis to model time from in-hospital admission to the first occurrence of the in-hospital event of interest (cardiopulmonary arrest, sepsis or unanticipated ICU transfer). If patients died, had any of the other two in-hospital events or were discharged alive before the first occurrence of the in-hospital event of interest, time to the event outcome was censored at death, the first occurrence of the other in-hospital events or discharged alive. We used the frailty survival model to account for clustering at the unit level.

Secondary outcome: 30-day hospital readmission. The secondary outcome of 30-day hospital readmission was defined as an unplanned hospital readmission of at least 24 hours (for multiple events, the earliest was counted) for any cause to an acute care hospital (trackable in the same health system) within 30 days of discharged alive from a hospital–patient encounter in our study, excluding planned, same-day and other specific types of readmissions⁵². We aimed to detect the OR of readmission from the CONCERN group to usual-care group. Among patients discharged alive, we examined whether a patient was readmitted to a hospital in the same health system within 30 days of discharge by calculating the OR using a GLM with logit-link function (logistic regression). The study group assignment for each patient was determined based on where the majority of time was spent during that hospital–patient encounter (intervention versus usual-care unit). Consistent with all analyses being performed at the hospital encounter level, a 30-day hospital readmission was treated as an outcome of the original hospital encounter and not a new hospital encounter. No outcomes were measured in this study across a patient’s multiple hospital encounters.

Transfers between intervention and usual-care units. Hospitalized patients are routinely transferred between units due to changing care needs. In this pragmatic trial, we employed several methods to control for hospital–patient encounters in which the patient was transferred between units that were randomized as intervention and usual care⁵¹. Survival analysis defined the event location as the clinical unit the patient was on for the majority of time during that 12-hour time window (that is, clinical shift) when an event outcome occurred. GLM analysis for LOS labeled the hospital–patient encounter as either intervention or usual-care group according to which unit the patient spent the most

time in during their hospital encounter. Visualizations of these analyses are presented in Extended Data Fig. 2 and Supplementary Figure 1 and demonstrate that a low percentage of crossover occurred for hospital–patient encounters between intervention and usual-care-assigned units. We were unable to draw Kaplan–Meier survival curves by group because the study assignment group was a time-dependent variable and ICU status, a covariate, is also a time-dependent variable.

Ethics and inclusion statement. Roles and responsibilities were agreed amongst collaborators ahead of the research and the research included local researchers/collaborators throughout the research process and as coauthors when authorship criteria was met. Collaborators for whom authorship criteria was not met are acknowledged in additional contributions. Clinical relevance was determined in collaboration with local stakeholders throughout the research process. All processes were approved by local ethics review committees.

Reporting summary

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

Data availability

We will make our de-identified outcomes data publicly available within 6 months of publication with no anticipated end date. This will include data dictionaries, study protocol, statistical analysis plan and de-identified individual shift-level and encounter-level data for our primary and secondary outcomes. We will make the data available under the Creative Commons Attribution 4.0 International License on a public repository, such as PhysioNet, that provides functionality for users to register, agree to terms of use and provide evidence of human subjects research training before downloading. Users will be able to access the data upon agreeing to the Creative Commons Attribution 4.0 International License and the public repository terms of use. Access will be granted according to the time frame for processing requests provided by the public repository. Our study website (www.concernearlywarningscore.org) will provide study team contact information and a link to the public repository to support findability. Prior to our data being uploaded on a public repository please contact the corresponding author for data requests. The CONCERN study team can also be contacted directly at CONCERN-EWS@cumc.columbia.edu. The conditions and purposes for sharing include that requesting individuals provide proof of a training program in human research subject protections and HIPAA regulations, not being from a for-profit organization and agreeing to criteria consistent with the PhysioNet Credentialled Health Data Use Agreement 1.5.0. The CONCERN study team will process data requests within 1 month of receipt of a data sharing request that meets the above stated criteria. The CONCERN EWS algorithm is considered intellectual property that will not be shared but is described in our online supplement.

Code availability

We will share our SAS statistical analysis code, along with our data as mentioned above, within 6 months of publication on a public repository such as PhysioNet that provides functionality for users to register, agree to terms of use and provide evidence of human subjects research training before downloading. Prior to our code being uploaded on a public repository please contact the corresponding author for data requests. The conditions and purposes for sharing include that requesting individuals provide proof of a training program in human research subject protections and HIPAA regulations, not being from a for-profit organization and agreeing to criteria consistent with the PhysioNet Credentialled Health Data Use Agreement 1.5.0. The CONCERN study team will process code requests within 1 month of receipt of a code request that meets the above stated criteria.

References

53. Shcherbatykh, I., Holbrook, A., Thabane, L., Dolovich, L. & COMPETE III investigators. Methodologic issues in health informatics trials: the complexities of complex interventions. *J. Am. Med. Inform. Assoc.* **15**, 575–580 (2008).
54. Rossetti, S. C. et al. The Communicating Narrative Concerns Entered by Registered Nurses (CONCERN) clinical decision support early warning system: protocol for a cluster randomized pragmatic clinical trial. *JMIR Res. Protoc.* **10**, e30238 (2021).
55. Rossetti S. C. et al. Leveraging clinical expertise as a feature - not an outcome - of predictive models: evaluation of an early warning system use case. In *AMIA Annual Symposium Proc.* 323–332 (AMIA, 2019).
56. Collett D. *Modelling Survival Data in Medical Research* 3rd edn (Chapman and Hall/CRC, 2014).
57. Jolliffe I. T. & Stephenson D. B. *Forecast Verification: A Practitioner's Guide in Atmospheric Science* (Wiley, 2012).
58. Fu, L.-H. et al. Utilizing timestamps of longitudinal electronic health record data to classify clinical deterioration events. *J. Am. Med. Inform. Assoc.* **28**, 1955–1963 (2021).
59. Abbey, E. J., Mammen, J. S. R., Soghoian, S. E., Cadorette, M. A. F. & Ariyo, P. In-hospital mortality and the predictive ability of the modified early warning score in Ghana: single-center, retrospective study. *JMIRx Med.* **2**, e24645 (2021).
60. Delgado-Hurtado, J. J., Berger, A. & Bansal, A. B. Emergency department modified early warning score association with admission, admission disposition, mortality, and length of stay. *J. Community Hosp. Intern. Med. Perspect.* **6**, 31456 (2016).
61. Kao, C.-C. et al. Prognostic significance of emergency department modified early warning score trend in critical ill elderly patients. *Am. J. Emerg. Med.* **44**, 14–19 (2021).
62. Kim W. Y. et al. Modified early warning score changes prior to cardiac arrest in general wards. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0130523> (2015).
63. Knoery, C. et al. Modified early warning score and risk of mortality after acute stroke. *Clin. Neurol. Neurosurg.* **202**, 106547 (2021).
64. Lee, J.-R. et al. Derivation and validation of modified early warning score plus SpO₂/FiO₂ score for predicting acute deterioration of patients with hematological malignancies. *Korean J. Intern. Med.* **35**, 1477–1488 (2020).
65. Martín-Rodríguez, F. et al. Early warning scores in patients with suspected COVID-19 infection in emergency departments. *J. Pers. Med.* <https://doi.org/10.3390/jpm1030170> (2021).
66. Monzon, L. D. R. & Boniatti, M. M. Use of the modified early warning score in intrahospital transfer of patients. *Rev. Bras. Ter. Intensiva* **32**, 439–443 (2020).
67. Takada, K. et al. Association between Intraoperative early warning score and mortality and in-hospital stay in lower gastrointestinal spontaneous perforation. *Anesthesiol. Res. Pract.* **2023**, 8910198 (2023).
68. Quan, H. et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med. Care.* **43**, 1130–1139 (2005).
69. Figueiredo, S. in *Stroke Engine* (eds Zeltzer, L. et al.) (Canadian Partnership for Stroke Recovery Heart & Stroke Foundation, 2009); <https://strokedatabase.ca/en/assessments/charlson-comorbidity-index-cci>
70. Hosmer, D. W., Lemeshow, S. & May, S. *Applied Survival Analysis: Regression Modeling of Time-To-Event Data* (Wiley, 2008).
71. Austin, P., Rothwell, D. & Tu, J. A comparison of statistical modeling strategies for analyzing length of stay after CABG surgery. *Health Serv. Outcomes Res. Methodol.* **3**, 107–133 (2002).

Acknowledgements

This study was funded by the National Institute of Nursing Research (NINR R01NR016941, COmmunicating Narrative Concerns Entered by RNs (CONCERN): Clinical Decision Support Communication for Risky Patient States) (S.C.R., P.C.D., C.K., S.C., J.W., G.L., D.A., R.Y.L., H.J., S.B., M.J.K., F.Y.C., L.Z., T.D., F.L., M.T., S.M.A.B., J.T., K.D.C.) and Reducing Health Disparities Through Informatics (T32NR007969) (J.W., R.Y.L., S.B., J.S.D.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank all the nurses, prescribing providers and patients who participated in this study. We also acknowledge B. Westra for serving on our Advisory Board and the American Nurses Foundation Reimagining Nursing Initiative funding for sharing and evaluating the implementation of CONCERN EWS to additional study sites. We thank C. Stillwell for her contribution to the design of tables and figures and L. Schweiß for editorial review.

Author contributions

S.C.R. and K.D.C. had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. More than one author (K.D.C., S.C.R., G.L., H.J., M.T., J.W. and R.Y.L.) have directly accessed and verified the underlying data reported in the paper. The research concept was designed by S.C.R. and K.D.C. System development and study design was provided by S.C.R., K.D.C., P.C.D., C.K., S.C., D.A., G.L., S.B., M.J.K., F.Y.C., L.Z., J.W., H.J., F.L. and J.S.D. Acquisition, analysis or interpretation of data was performed by S.C.R., K.D.C., P.C.D., C.K., S.C., D.A., G.L., S.B., J.W., H.J., F.L., R.Y.L., T.D., M.T., S.M.A.B. and J.T. The paper was drafted by S.C.R., R.Y.L., K.D.C. and P.C.D. All authors contributed to the critical review of the paper for important intellectual content. Statistical analysis was carried out by H.J., K.D.C. and D.A. Funding was obtained by S.C.R. and K.D.C. Administrative, technical or material support was provided by S.C.R., K.D.C., P.C.D., R.Y.L., T.D. and F.L. Supervision was carried out by S.C.R., K.D.C., P.C.D. and S.B.

Competing interests

Columbia University has filed a US nonprovisional patent application (US Patent Application No. 18/814,823) related to the technology that is the subject of this article. S.C.R., C.K. and K.D.C. are named inventors on the patent application and are entitled to revenue sharing with the university per the terms of the university's patent policy. The university and the named inventors are committed to making the technology freely available upon request for academic noncommercial research purposes. Any entity interested in obtaining a license to practice the technology for commercial purposes may contact Columbia Technology Ventures at techventures@columbia.edu. D.W.B. reports grants and personal fees from EarlySense, personal fees from CDI Negev, equity from ValeraHealth, equity from Clew, equity from MDClone, personal fees and equity from AESOP, personal fees and equity from FeelBetter, personal fees and equity from Guided Clinical Solutions and grants from IBM Watson Health, outside the submitted work. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41591-025-03609-7>.

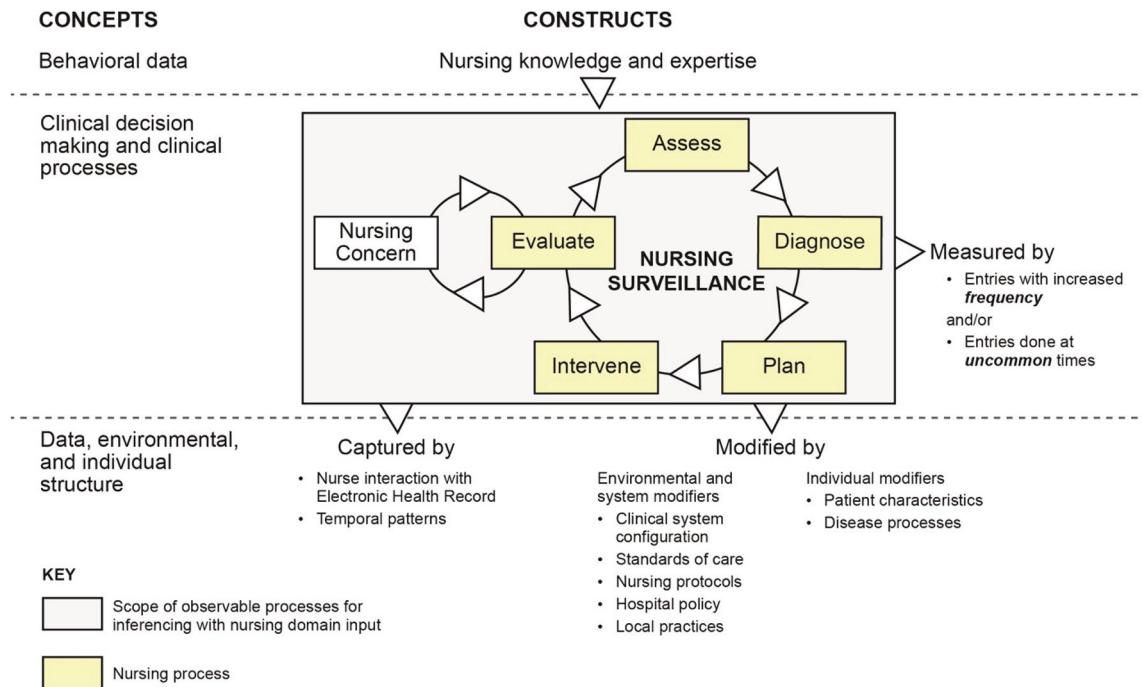
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-025-03609-7>.

Correspondence and requests for materials should be addressed to Sarah C. Rossetti.

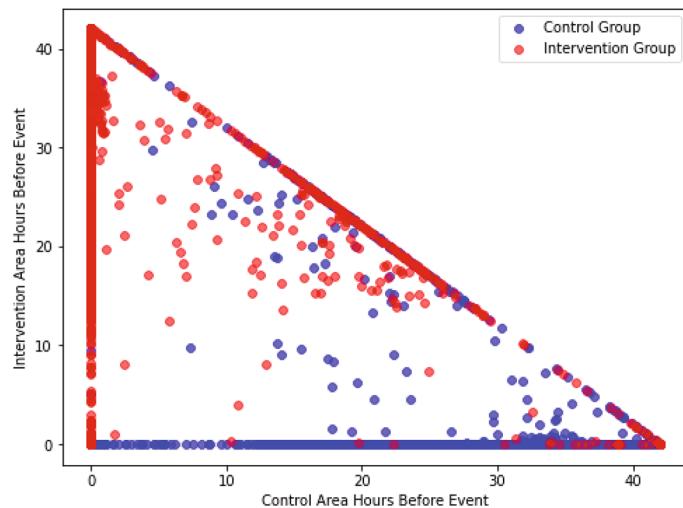
Peer review information *Nature Medicine* thanks Alessio Crippa, Rupert Pearse, Jean-Louis Vincent and the other, anonymous, reviewer(s)

for their contribution to the peer review of this work. Primary Handling Editor: Lorenzo Righetto, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | CONCERN Predictive Model Conceptual Modeling Approach. Figure adapted from ref. 10 under a Creative Commons license CC BY 4.0.



Extended Data Fig. 2 | Percentage of Time Intervention and Usual Care Hospital Encounters Spent in Each Area 42 Hours Before Deterioration Event or Discharged Alive (N = 60,893). This scatter plot visualizes the distribution of

hours spent in the usual care (control) area (x-axis) and intervention area (y-axis) 42 h prior to an in-hospital deterioration event or being discharged alive for patients in the Control Group (blue) and Intervention Group (red).

Extended Data Table 1 | Characteristics of Hospital Encounters During the Trial (N=60,893) by Study Site

	Site A		Site B	
	CONCERN Intervention	Usual Care	CONCERN Intervention	Usual Care
	N = 16,838	N = 15,630	N = 16,186	N = 12,239
Age	63.03 ± 16.86	63.72 ± 16.67	62.19 ± 18.26	63.60 ± 17.62
Male Sex – no. (%)	8,276 (49.15)	7,543 (48.26)	7,780 (48.07)	6,423 (52.48)
Race – no. (%)				
White	13,483 (80.07)	12,327 (78.87)	5,440 (33.61)	3,951 (32.28)
Black	1,564 (9.29)	1,540 (9.85)	3,314 (20.47)	2,684 (21.93)
Asian	455 (2.70)	467 (2.99)	248 (1.53)	167 (1.36)
Other or missing	1,336 (7.93)	1,296 (8.29)	7,184 (44.38)	5,437 (44.42)
Ethnic group – no. (%)				
Not Hispanic or Latino	15,298 (90.85)	14,200 (90.85)	8,320 (51.40)	6,104 (49.87)
Hispanic or Latino	1,172 (6.96)	1,151 (7.36)	6,270 (38.74)	4,752 (38.83)
Unknown or not reported	368 (2.19)	279 (1.79)	1,596 (9.86)	1,383 (11.30)
Primary Language English	15,662 (93.02)	14,378 (91.99)	11,051 (68.28)	8,206 (67.05)
Charlson Comorbidity Index	3.56 ± 3.34	4.01 ± 3.53	2.64 ± 2.89	2.87 ± 2.97
Discharge Disposition ^b				
Home	13,667	12,860	13,106	9,346
Other	3,171	2,770	3,080	2,893

^a Plus-minus values are mean±SD *Race and ethnic group were reported in the EHR \$Discharge Disposition: Home includes home with services; Other includes any disposition not to home.

Extended Data Table 2 | Top 10 Admission and Principal Diagnoses^a Across Study Sites

Rank	Site A Top 10 Admission Diagnoses	Site B Top 10 Admission Diagnoses
1	Shortness of breath	Shortness of breath
2	Unspecified abdominal pain	Unspecified abdominal pain
3	Chest pain, unspecified	Chest pain, unspecified
4	Fever, unspecified	Sepsis, unspecified organism
5	Altered mental status, unspecified	Altered mental status, unspecified
6	Weakness	COVID-19
7	COVID-19	Syncope and collapse
8	Morbid (severe) obesity due to excess calories	Weakness
9	Solitary pulmonary nodule	Nonrheumatic aortic (valve) stenosis
10	Nausea with vomiting, unspecified	Fever, unspecified
Rank	Site A Top 10 Principal Diagnoses	Site B Top 10 Principal Diagnoses
1	COVID-19	Sepsis, unspecified organism
2	Hypertensive heart and chronic kidney disease with heart failure and stage 1 through stage 4 chronic kidney disease, or unspecified chronic kidney disease	COVID-19
3	Hypertensive heart disease with heart failure	Hypertensive heart and chronic kidney disease with heart failure and stage 1 through stage 4 chronic kidney disease, or unspecified chronic kidney disease
4	Non-ST elevation (NSTEMI) myocardial infarction	Hypertensive heart disease with heart failure
5	Spinal stenosis, lumbar region with neurogenic claudication	Nonrheumatic aortic (valve) stenosis
6	Morbid (severe) obesity due to excess calories	Non-ST elevation (NSTEMI) myocardial infarction
7	Acute kidney failure, unspecified	Morbid (severe) obesity due to excess calories
8	Sepsis, unspecified organism	Atherosclerotic heart disease of native coronary artery without angina pectoris
9	Nonrheumatic aortic (valve) stenosis	Acute kidney failure, unspecified
10	Spinal stenosis, cervical region	Hypertensive heart and chronic kidney disease with heart failure and with stage 5 chronic kidney disease, or end stage renal disease

^aICD10 Codes; Gray shading indicates present in Top 10 list for both study sites.

Extended Data Table 3 | Sensitivity Analysis for Imbalance in COVID-19 Diagnoses for Hospital Encounters Across Study Arms

Imbalance Analysis		
	Intervention	Usual Care
Mortality rate for patients with COVID-19 Diagnosis	23/1,247 (1.84%)	61/2,167 (2.81%)
Mortality rate for patients without COVID-19 Diagnosis	158/31,777 (0.49%)	204/25,702 (0.79%)
Sensitivity Analysis		
Excluded patients with any COVID-19 Diagnoses	adjusted hazard ratio [HR], 0.70; 95% confidence interval [CI], 0.57 - 0.86; p=0.0008	
Excluded patients with COVID-19 Primary Diagnosis	adjusted hazard ratio [HR], 0.71; 95% confidence interval [CI], 0.59 - 0.87; p=0.0008	

Extended Data Table 4 | Count of Study Units by Type and Intervention versus Usual Care Group Across Study Sites

Unit Type	Site A		Site B		Totals
	CONCERN Intervention	Usual Care	CONCERN Intervention	Usual Care	
Acute Care Unit (ACU)	19	17	9	8	53
Intensive Care Unit (ICU)	5	7	4	5	21
Totals	24	24	13	13	74

Extended Data Table 5 | Performance Metrics for CONCERN Predictive Model^a

Setting	Accuracy	Precision	Recall	Log Loss	AUC
ICU	0.970938	0.431373	0.594595	0.073695	0.934683
ACU	0.973341	0.813559	0.643935	0.089369	0.955982

^aMultinomial Gradient Boosted Machine (GBM) model built on random 12-hour time slices to predict (over the next 24 h) whether a patient is discharged alive, will still be in the hospital or has a hospital event (in-hospital mortality, cardiopulmonary arrest, sepsis, unanticipated ICU transfer). Modeling was trained on 70% of the dataset, with 30% used for 10-fold cross-validation. Average performance reported for ensemble models²².

Extended Data Table 6 | CONCERN Model Factors and Features

Nursing Note Content	Abdominal pain* Abnormal heart rhythm* Abnormal mental state* Abnormal rate, rhythm, depth and effort of respirations* Abnormal temperature* Back pain* Chest pain* Communication problem* Diagnosis related with infection* Deficit of circulation* Fall risk* Fluid volume alteration* General concern* Headache* Improper renal function* Medication related with infection* Monitoring* Mood disorder* Musculoskeletal pain* Pain level* Violence gesture*
Vital Sign Frequency	Heart rate measurement* Respiratory rate measurement* Blood pressure measurement* Temperature measurement* SpO ₂ measurement* All 5 vital measurements taken at same time* Only 1 vital measurement taken*
Nursing Note Frequency	Nursing note written*
Vital Sign Comment Frequency	Heart rate comment* Respiratory rate comment* Blood pressure comment* Temperature comment* SpO ₂ comment*
Medication Administration	PRN medication administered* Scheduled medication withheld*

*Feature is aggregated over the past 12h.

Extended Data Table 7 | Statistical Power Analysis

Original analysis plan	<ul style="list-style-type: none">Generalized linear mixed models for the comparison of mortality rates between those randomized to CONCERN and usual careBased on cluster randomized trial using intra-cluster correlation coefficient (ICC) equal to 0.22-sided tests with alpha = 0.05Expected sample:<ul style="list-style-type: none">2,000 total admissions per month50,000 total inpatient days based on average occupancyMean mortality rate of 37.5 deaths per 10,000 inpatient days (ranging from 11.9 to 48.8 across our study sites)80% statistical power to detect a difference of less than 1% relative difference in mortality rates
Parameters	
Estimated effect size	

Corresponding author(s): Rossetti

Last updated by author(s): Feb 3, 2025

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	SQL queries to EHR data warehouse
Data analysis	SAS 9.4 Copyright(c) 2002-2012 by SAS Institute Inc, Cary, NC, USA. We will share our SAS  statistical analysis code, along with our data as mentioned below, within 6 months of publication on a public repository such as PhysioNet that provides functionality for users to register, agree to terms of use, and provide evidence of human subjects research training prior to downloading. Site A used MS SQL server 2012 and site B used MS SQL server 2014.

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](#)

Data Availability: We will make our de-identified outcomes data publicly available within 6 months of publication with no anticipated end date. This will include data dictionaries and de-identified individual shift level and encounter level data for our primary and secondary outcomes. We will make the data available on a public repository, such as PhysioNet, that provides functionality for users to register, agree to terms of use, and provide evidence of human subjects research training prior to downloading. The CONCERN EWS algorithm is considered intellectual property that will not be shared but is described in our online supplement.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Sex is reported per intervention arm in Table 1 Characteristics of Hospital Encounters During the Trial. Patients' sex data was derived from each patient's EHR record data on our EHR data warehouse (and therefore the provenance of sex data includes both self-reported and assigned) and was used as a covariate in the analysis. These data are also reported in extended data eTable1 broken down per site and intervention arm. Gender was not used in the analysis due to the inconsistent unavailability of data in the EHR for the patient population.

Reporting on race, ethnicity, or other socially relevant groupings

Race, ethnicity, and primary language (English or not English) are reported per intervention arm in Table 1 Characteristics of Hospital Encounters During the Trial. These data were derived from each patients' EHR record data on our EHR data warehouse. These data are also reported in extended data eTable1 broken down per site and intervention arm.

Population characteristics

Age, Charlson comorbidity scores and discharge disposition are reported in Table1 Characteristics of Hospital Encounters During the Trial. These data are also reported in extended data eTable1 broken down per site and intervention arm. We also report extended data eTable2 the Top Admission and Principal Diagnoses across sites.

Recruitment

Individual study units across the 4 hospitals included all non-specialty acute care units (ACUs) and ICUs. Specialty ACUs, such as oncology, psychiatric, and rehabilitation, were excluded. Randomization was performed at the unit level and stratified by site prior to trial initiation. Each study unit (ACU or ICU) was randomly allocated to one of 2 groups (CONCERN EWS intervention or usual-care) using a computer-generated randomization scheme. The trial included 74 clinical units (37 intervention, 37 usual-care) with the following distribution: Site A intervention-19 ACUs and 5 ICUs; Site A usual-care-17 ACUs and 7 ICUs; Site B intervention-9 ACUs and 4 ICUs; Site B usual-care-8ACUs and 5 ICUs (eTable1).

Ethics oversight

Columbia University and Brigham and Women's Hospital

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

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

Behavioural & social sciences study design

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

Study description

1-year multi site, pragmatic, cluster-randomized clinical trial at 2 large health systems in the Northeastern United States. The objective of this study was to determine at the individual patient level whether CONCERN EWS led to a decrease in primary outcomes (in-hospital mortality and LOS) and influenced secondary outcomes (cardiopulmonary arrest, sepsis, unanticipated ICU transfers, and 30-day readmission). We hypothesized a priori that patients whose care team received CONCERN EWS risk scores would have lower mortality and LOS than a control group of patients with non-EWS informed teams.

Research sample

This clinical trial took place at 2 large health systems in the Northeastern United States. Each system was a study site (A and B), and sites A and B each comprise 1 academic medical center and 1 community hospital. Individual study units across 4 hospitals included all non-specialty acute care units (ACUs) and ICUs. Specialty ACUs, such as oncology, psychiatric, and rehabilitation, were excluded.

The research sample consisted of 60,893 hospital encounters from two large U.S. health systems, including 33,024 encounters in the CONCERN intervention group and 27,869 in the usual-care group. Participants were adult patients (aged ≥18 years) admitted to

acute and intensive care units, with exclusions for those receiving hospice, palliative care, or with pre-existing Do Not Resuscitate/Do Not Intubate orders

- Age: Mean age was 62.62 years (SD = 17.56) in the intervention group and 63.67 years (SD = 17.10) in the usual-care group.
- Sex: The sample was nearly balanced by sex, with 48.62% male in the intervention group and 50.11% male in the usual-care group.
- Race: Race of patients included White (57.30% intervention; 58.41% usual-care), Black (14.77% intervention; 15.16% usual-care), Asian (2.13% intervention; 2.27% usual-care), and Other or Missing (25.80% intervention; 24.16% usual-care).
- Ethnicity: Not Hispanic or Latino (71.52% intervention; 72.86% usual-care), Hispanic or Latino (22.54% intervention; 21.18% usual-care), and Unknown/Not Reported (5.95% intervention; 5.96% usual-care).
- Primary Language: Approximately 81% of patients in both groups spoke English as their primary language.

The sample is broadly representative of hospitalized adult inpatients in urban U.S. healthcare settings but may not be fully generalizable to rural hospitals or smaller community health systems due to potential differences in patient demographics.

Sampling strategy

Randomization was performed at the unit level and stratified by site prior to trial initiation. Power analysis comparing mortality rates between groups estimated a 1-year trial on the targeted units at our 2 sites would result in sufficient sample size for at least 80% statistical power to detect a difference of less than 1% relative difference in mortality rates (2-sided; alpha=.05)(eTable4).

Data collection

Primary outcomes were in-hospital mortality rate and LOS. Secondary outcomes were rates of cardiopulmonary arrest, sepsis, unanticipated ICU transfer, and 30-day hospital readmission. See Table1 for definitions. All data were collected from the EHR and evaluated and reported across study sites. More than one author (Cato, Rossetti, Lowenthal, Jia, Tran, Withall, Lee) have directly accessed and verified the underlying data reported in the manuscript. In addition, we performed randomly selected chart reviews on 10% of the sample per outcome per site to validate our outcome definitions and extractions. The researchers were not blinded to the experimental condition or the study hypothesis. As a pragmatic cluster-randomized controlled trial, the allocation of units to either the intervention or usual care group was known to the study team.

Timing

Outcomes data were collected for all patients admitted to our study units during the trial dates: At Site A, the first inpatient hospital admission for the cohort was October 1, 2020, the last inpatient hospital admission of the cohort was October 31, 2021, and the last inpatient hospital discharge of the cohort was March 30, 2022. At Site B, the first inpatient hospital admission was October 1, 2021, the last inpatient hospital admission was October 30, 2022, and the last inpatient hospital discharge was December 26, 2022.

Data exclusions

Patients on study units were included in the analyses if they were 18 years of age and older, hospitalized for greater than 24 hours (EWS score begins displaying after 24 hours), admitted to a study unit for a minimum of 12 hours, and free from any in-hospital event (including discharge) until at least 6 hours after study unit admission. Hospice and palliative care patients and patients with do not resuscitate/do not intubate and comfort care orders activated prior to any trial outcome event were excluded.

Non-participation

N/A; this field is not applicable to the study because the research was conducted as a pragmatic cluster-randomized controlled trial with waived individual patient consent.

Randomization

Randomization was performed at the unit level and stratified by site prior to trial initiation. Each study unit (ACU or ICU) was randomly allocated to one of the 2 groups (CONCERN EWS Intervention or usual-care) using a computer-generated randomization scheme. The trial included 74 clinical units (37 intervention, 37 usual-care) with the following distribution: Site A intervention-19 ACUs and 5 ICUs; Site A usual-care-17 ACUs and 7 ICUs; Site B intervention-9 ACUs and 4 ICUs; Site B usual-care-8ACUs and 5 ICUs (eTable1). The intervention (CONCERN EWS prediction risk score) was displayed to a patient's care team consisting of nurses and prescribing providers (i.e., physician, nurse practitioners, physician assistants) only if that patient was admitted to an intervention-assigned unit. The intervention was not displayed to care teams of patients admitted to usual-care assigned units.

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
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

NCT03911687

Study protocol

Please see in our Supplement materials submitted online section sMethods1 which provides a summary of changes in our final analysis plan. Our original study protocol was published here: <https://pubmed.ncbi.nlm.nih.gov/34889766/>

Data collection

Primary outcomes were in-hospital mortality rate and LOS. Secondary outcomes were rates of cardiopulmonary arrest, sepsis, unanticipated ICU transfer, and 30-day hospital readmission. See Table 1 for definitions. All data were collected from the EHR and evaluated and reported across study sites. More than one author (Cato, Rossetti, Lowenthal, Jia, Tran, Withall, Lee) have directly accessed and verified the underlying data reported in the manuscript. In addition, we performed randomly selected chart reviews on 10% of the sample per outcome per site to validate our outcome definitions and extractions.

Outcomes data were collected for all patients admitted to our study units during the trial dates: At Site A, the first inpatient hospital admission for the cohort was October 1, 2020, the last inpatient hospital admission of the cohort was October 31, 2021, and the last inpatient hospital discharge of the cohort was March 30, 2022. At Site B, the first inpatient hospital admission was October 1, 2021, the last inpatient hospital admission was October 30, 2022, and the last inpatient hospital discharge was December 26, 2022.

Outcomes

Primary outcomes were in-hospital mortality rate and length of stay. Secondary outcomes were rates of cardiopulmonary arrest, sepsis, unanticipated ICU transfer, and 30-day hospital readmission. See Table 1 for definitions. All data were collected from the EHR and evaluated and reported across study sites.

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A