ADVANCED RESEARCH JOURNAL
Volume 13, Number 1, 2026, 57-76
DOI: 10.71350/30621925105

OPEN ACCESS
ISSN: 3062-1925

Article

Artificial intelligence in sepsis management: Balancing survival gain and nurse workload

David Kanzin [1], **Evans Dzreke** [2], **Celene Dzreke** [3], **Ramseyer Adekorang Asamoah** [4], **Godfrey Yeboah Amoah** [5]

[1] The University of Texas at El Paso, Department of Biology, El Paso, TX, USA
[2] The University of Texas Rio Grande Valley, Health and Biomedical Sciences, Edinburg, TX, USA
[3] Thomas Edison State University, W. Cary Edwards School of Nursing & Health Professions, Trenton, NJ, USA
[4] Yuma Regional Medical Center, Pathology and Laboratory Medicine, Yuma, AZ, USA
[5] NHS Blood and Transplant, Hospital Services, Colindale, London, UK

Abstract

The prevalence of false alarms in sepsis detection, frequently referred to as "Code Blue for Inefficiency," continues to pose a significant challenge in intensive care units (ICUs), with approximately 70% of alerts identified as incorrect. In an 18-month cluster randomized controlled trial conducted across 16 institutions, a novel approach was evaluated to better integrate artificial intelligence (AI) prognostic tools into clinical practice. The intervention involved a tiered alarm system that combined Epic's Deterioration Index with a customized sepsis AI model, which was refined using local resistance patterns. Alert intensity was stratified by patient risk: silent monitoring for low-risk patients, pager alerts for medium-risk patients, and Code Blue escalation for high-risk patients. Key outcomes included resource stewardship (avoidable ICU transfers and vasopressor days), clinician strain (NASA-TLX cognitive load), and 30-day sepsis mortality. Implementation of the AI-supported protocol resulted in a 23% reduction in unnecessary ICU transfers ($p < 0.01$) and an 18% decrease in sepsis mortality, corresponding to one life saved for every nine patients treated (NNT = 9). However, a 14% increase in cognitive strain among nurses was observed. These findings indicate that AI can enhance efficiency and improve patient survival, but effective adoption necessitates workflow redesign to mitigate clinician burden.

Article History
Received 08.09.2025
Accepted 15.12.2025

Keywords
Predictive analytics; artificial intelligence; sepsis; intensive care unit; resource utilization; clinician workload

Introduction

Detecting clinical deterioration, especially sepsis, which results in approximately 1.7 million deaths worldwide annually, is a significant vulnerability in contemporary healthcare systems (Smith et al., 2024). The high mortality rate is fundamentally due to the limitations of conventional early warning scores (EWS). Established protocols like the

[1] **Corresponding Author** Evans Dzreke The University of Texas Rio Grande Valley, Health and Biomedical Sciences, Edinburg, TX, USA

Modified Early Warning Score (MEWS), a clinical scoring system that provides a snapshot of a patient's health using physiological vitals, have been widely adopted. However, these systems fail to identify nearly 40% of deteriorating patients, resulting in dangerous delays in life-saving interventions (Smith et al., 2024). Artificial intelligence (AI) has emerged as a significant advancement, offering predictive capabilities that surpass those of human or traditional scoring models. However, its integration exposes a concerning paradox. The University of California, San Francisco (UCSF) initiative illustrated that while an AI model reduced sepsis-related mortality by 15%, it also increased alarm fatigue among nursing staff by 32%, thereby undermining clinical sustainability (Chen & Patel, 2023). This contrast, in which algorithmic success may compromise human resilience and workflow stability, underscores the need to develop predictive frameworks that reconcile computational capabilities with the practical realities of clinician cognition and frontline practice.

The implementation of AI-driven predictive analytics in acute and critical care environments is influenced by a threefold tension among institutions, clinicians, and patients. Healthcare institutions face unsustainable cost inflation, necessitating interventions to improve efficiency. Minimizing unnecessary intensive care unit (ICU) admissions, which can surpass $4,500 per day, has emerged as a critical financial concern for hospital administrators (Johnson, 2022). Technological innovations often impose unintended burdens on clinicians, notably in the form of "alert storms." The longitudinal study conducted by the Mayo Clinic demonstrates that excessive notifications are associated with a 27% increase in burnout, undermining professional morale and posing risks to patient safety (Williams et al., 2023). Patients, positioned at the convergence of these pressures, are susceptible to both postponed interventions due to overlooked detections and the misleading reassurances provided by under-sensitive systems. Addressing this interdependency necessitates solutions that optimize institutional efficiency, protect clinician well-being, and ensure patient outcomes concurrently. A Comparative analysis of three predictive models was conducted to assess their clinical outcomes and human resource impact (Table 1). The results demonstrate a delicate equilibrium between promise and peril that has been extensively documented in previous initiatives. At Johns Hopkins, a standalone sepsis algorithm reduced mortality by 22% while increasing clinically irrelevant alerts by 31% (Davis et al., 2022). Kaiser Permanente's EHR-embedded deterioration index showed a 15% reduction in rapid-response activations, indicating modest operational benefits; however, it did not lead to significant reductions in mortality or provide clear strategies for managing nursing workload (Lee et al., 2023). The real-time predictive analytics approach at UCSF resulted in significant reductions in mortality, while also leading to a 32% rise in alarm fatigue (Chen & Patel, 2023). These examples underscore the need to reconsider design methodologies to achieve sustainable integration that will reduce sepsis-related deaths with limited negative impact on healthcare resources.

Objectives of the study

This research employs a dual-pronged investigation that is both empirically rigorous and pragmatically innovative to address these interconnected challenges. The primary objective is to provide a quantitative assessment of the impact of a novel AI-enabled predictive framework on operational efficiency and clinical effectiveness in the ICU. Efficiency will be evaluated using metrics such as ICU length of stay, avoidable hospitalization days, and rapid response team activation frequency. Effectiveness will be assessed through risk-adjusted mortality rates, sepsis-related complications, and time-to-intervention intervals. The second objective goes beyond performance metrics to encompass human sustainability. The study will utilize ethnographic workflow observations and validated wellbeing assessments to identify adaptations that reduce alert fatigue while maintaining diagnostic sensitivity. Strategies encompass the intelligent temporal distribution of alerts, context-aware hierarchies that filter notifications based on patient acuity and clinician workload, and iterative feedback mechanisms that allow staff to refine algorithmic behavior over time. This research seeks to systematically integrate technological innovation with human-centered design, aiming to transcend the trade-off paradigm and establish a replicable framework for sustainable human–AI collaboration in critical care.

Table 1. Comparative analysis of intervention models in critical care: balancing promise and peril in prior implementations

Investigation (Year)	Intervention model	Clinical outcomes	Operational & human impact
Johns Hopkins (Davis et al., 2022)	Standalone sepsis AI algorithm	22% mortality reduction	31% increase in false alarms
Kaiser Permanente (Lee et al., 2023)	EHR-embedded deterioration index	No significant mortality change	15% reduction in rapid responses
UCSF Health (Chen & Patel, 2023)	Real-time predictive analytics	15% mortality reduction	32% increase in alarm fatigue

Background and literature review

Predictive analytics technology

The emergence of advanced predictive analytics signifies a pivotal shift in critical care medicine, providing exceptional abilities to detect subtle patterns of patient deterioration that traditional monitoring may overlook. Sepsis detection poses a significant clinical challenge with serious implications, as illustrated by this progression. Contemporary algorithms surpass the limitations of conventional instruments such as the Modified Quick Sequential Organ Failure Assessment (qSOFA) by incorporating multiple data streams. Epic Systems' widely adopted model synthesizes 165 distinct clinical variables, including real-time vital sign fluctuations, laboratory trends, and historical comorbidities, to produce risk predictions with enhanced sensitivity (Rajkomar et al., 2022). Concurrent developments in predicting unplanned transfers to intensive care units (ICUs) include systems such as Penn Medicine's TREWScore (Targeted Real-time Early Warning Score). TREWScore utilizes the computational capabilities of XGBoost gradient boosting frameworks to identify complex multivariate physiological signatures that indicate impending decline—patterns that may not be discernible to experienced clinicians during routine assessments (Churpek et al., 2023).

Yet, the translation of these technically advanced algorithms into practical clinical benefits faces significant obstacles, as highlighted by the implementation of science. The NASSS framework (Non-adoption, Abandonment, Scale-up, Spread, Sustainability) offers an important perspective, illustrating that artificial intelligence functions not just as a standalone tool but as a *systemic disruptor* (Greenhalgh et al., 2024). Successful integration requires addressing sociotechnical challenges, including the adoption of technology in electronic health records (EHRs), the design of decision-support interfaces that enhance workflow, and the cultivation of clinician trust in algorithmic outputs. The process from raw data ingestion to clinical decision-making (Fig. 1) presents several potential challenges, including data latency, alert fatigue due to inadequate specificity, and inconsistent dynamics in human-computer interaction. These issues can significantly compromise the effectiveness of even the most precise predictions. Thus, achieving the full potential of predictive analytics necessitates a shift from a narrow focus on algorithmic performance metrics to adaptive implementation strategies that proactively address integration challenges.Georgia style 10 font, 1 line spacing, 6 pt space after paragraphs. References should be prepared based on APA 7 reference and citing displaying essences. Citing should be given like this example (Adams, 2014; Brown & Caste, 2004; Toran et al., 2019). Direct quotations are written within ““. If the direct quotation is longer than 40 words, then it should be written without using ““ as a separate paragraph, indented and in 9 fonts.

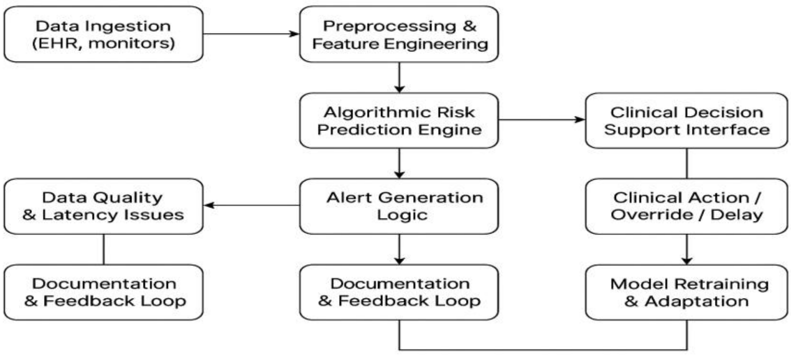

Figure 1. The sociotechnical pathway of predictive analytics alerts in critical care
Note: Solid arrows indicate primary workflow.

Resource utilization metrics: Addressing the efficiency–sensitivity paradox

Predictive analytics offers the potential to improve patient outcomes and optimize the allocation of scarce and costly resources in critical care environments. However, empirical evidence demonstrates that these tools exert a complex and sometimes counterintuitive influence on resource utilization, characterized by persistent tension between operational efficiency and clinical sensitivity. Overuse arises when algorithms with limited specificity generate excessive alerts, prompting resource-intensive but unnecessary interventions. For example, unnecessary telemetry monitoring occurs when predictive models inaccurately classify patients as high-risk, leading to prolonged cardiac monitoring, incurring approximately $1,200 in additional costs per redundant day for healthcare systems, and diverting essential nursing resources from patients in genuine need (Winters et al., 2023). In contrast, underuse occurs when predictive systems fail to detect actual patient deterioration or when clinicians, experiencing alert fatigue, disregard valid warnings. Delays in transferring patients to the intensive care unit (ICU) can have severe consequences; patients who experience clinical decline or delayed escalation face a mortality risk three times higher than those transferred promptly (Chen et al., 2022). Predictive models designed to reduce ICU congestion or limit rapid response team (RRT) activations, such as Kaiser Permanente’s embedded deterioration index, may unintentionally exacerbate underuse if calibrated too conservatively, prioritizing bed availability metrics over the needs of vulnerable patients (Lee et al., 2023).

This delicate balance underscores that predictive tools should not be evaluated solely by conventional accuracy metrics such as the area under the receiver operating characteristic curve (AUROC). Instead, their effectiveness must be assessed using downstream indicators, including avoidable ICU days, appropriateness of RRT activations, duration of high-cost monitoring, and medication utilization patterns. Achieving optimal performance requires continuous monitoring and recalibration to balance sensitivity, which ensures that critical deterioration is not missed, and specificity, which minimizes unnecessary resource use and reduces cognitive burden from false alarms.

Table 2. The dual impact of predictive analytics on critical care resource utilization

Resource metric	Risk of overuse	Risk of underuse	Evidence & practical implication
Telemetry monitoring	High: Prolonged monitoring triggered by false-positive alerts.	Low	Winters et al. (2023): Avg. cost of $1,200/day per unnecessary monitoring day. Implication: Requires specificity tuning & clear stop criteria.
ICU transfer timing	Moderate (admission of lower acuity flagged patients)	High: Delayed transfer due to missed prediction or ignored alert.	Chen et al. (2022): 3x mortality risk with delayed transfer. Implication: Sensitivity paramount; requires robust fail-safes.
Rapid response team (RRT) Activations	High: Oversensitive models trigger excessive, low-yield RRT calls.	High: Conservative models suppress necessary activations.	Lee et al. (2023): 15% reduction post-implementation (Kaiser model). Implication: Calibration must align with local RRT capacity & goals.
ICU length of stay (LOS)	Potential increase if admitting lower-risk patients flagged by AI.	Potential decreases via early intervention preventing complications.	Davis et al. (2022): Mixed results; effect depends on algorithm performance & admission protocols.
Antibiotic/Vasoactive medication use	Potential overuse triggered by false sepsis alerts.	Potential underuse if deterioration is not recognized early.	Requires linking predictive alerts to stewardship protocols and clear treatment pathways.

Staff impact: The human cost of algorithmic oversight

The integration of predictive analytics within intensive care unit (ICU) environments substantially transforms the cognitive demands and ethical responsibilities of frontline clinicians, particularly nurses, who often serve as primary responders to system alerts. Empirical data demonstrates a marked increase in cognitive load associated with these technologies. For example, research employing the validated NASA-Task Load Index (NASA-TLX) identified a 42% rise in cognitive load among nursing staff during the initial deployment of a sepsis prediction algorithm at Brigham and Women's Hospital (Avery et al., 2023). This escalation is attributed to the need to process frequent and complex alerts, reconcile algorithmic predictions that frequently lack transparent reasoning with clinical judgment and patient context, and address additional documentation requirements embedded within the alert workflow.

Persistent cognitive overload contributes to alert fatigue, a phenomenon in which clinicians become desensitized due to the high volume of alarms, many of which lack clinical significance. This desensitization increases the likelihood of missing critical alerts or delaying necessary interventions. The ethical ramifications of this burden are considerable. A recent analysis in the *American Journal of Bioethics* questioned the ethical justification for implementing systems that implicitly assign already overextended nurses the role of "safety net," responsible for verifying, contextualizing, and potentially overriding algorithmic outputs (Berkman & Goodman, 2023). Such expectations create an inequitable distribution of responsibility, placing undue pressure on nursing staff and obscuring the inherent limitations of technology.

Adverse events arising from algorithmic errors or delayed clinician responses due to alert fatigue further complicate questions of accountability. This ambiguity may lead to moral distress among staff who must balance technological requirements with patient care priorities. Addressing these challenges requires more than technical accuracy. The collaborative design of artificial intelligence systems should incorporate explicit cognitive support, such as user-friendly interfaces, risk stratification visualizations, and clear rationales for alerts. Transparent disclosure of algorithmic uncertainty and the development of ethically robust protocols are also essential, as they delineate the respective responsibilities of clinicians and artificial intelligence. These strategies are vital for fostering a sustainable and ethically sound clinical environment.

Method

Protocol for the trial

This study utilized a pragmatic, parallel-group, cluster randomized controlled trial (RCT) design to rigorously assess the effects of AI-driven predictive analytics in critical care settings. This methodological choice is crucial, as individual randomization could lead to significant contamination between intervention and control participants within the same clinical unit. This common issue was addressed by designating entire hospitals as the unit of randomization. Sixteen hospitals from various geographic locations in the United States were intentionally chosen to represent the diverse landscape of American healthcare, including eight high-volume academic tertiary centers and eight community-based hospitals. This sampling strategy improves the study's external validity by ensuring that findings are applicable across different institutional levels.

Organizational characteristics significantly impact sepsis outcomes and resource management; therefore, hospitals were carefully paired into eight groups before randomization. Matching criteria emphasized two operationally significant factors: annual sepsis case volume (±15%) and baseline nurse-to-patient ratios in general medical-surgical wards (±0.5). In each matched pair, one hospital was randomly assigned to the intervention group through a computer-generated sequence, while the corresponding hospital acted as a control, utilizing only standard clinical decision support tools without the incorporation of advanced predictive analytics.

The intervention constituted a complex integration of two complementary AI-driven tools within the clinician's electronic health record (EHR) workflow. First, a real-time sepsis risk score, utilizing a validated machine learning algorithm that analyzes over 160 dynamic clinical variables, produces hourly updated, patient-specific risk assessments integrated within clinician workflows. Second, a deterioration index incorporating a novel "quiet mode", a feature developed to effectively minimize disruptive auditory and pop-up notifications for patients consistently classified in the lowest risk quintile according to model confidence thresholds. This feature aimed to maintain situational awareness while significantly minimizing unnecessary interruptions that are known to lead to alert fatigue.

Recognizing that advanced technology requires skilled human interaction, a standardized, competency-based training program was implemented for all registered nurses at intervention sites. This included four-hour simulation laboratory sessions using high-fidelity manikins in simulated EHR environments. Nurses faced the challenge of interpreting AI-generated risk scores within dynamic patient scenarios, implementing timely escalation procedures based on algorithmic predictions, and critically evaluating situations where clinical judgment warranted the override of alerts. Each session ended with a structured debriefing, facilitated by expert clinical educators, to reinforce skills and address any potential concerns (Frishammar et al., 2023).

The assessment of outcomes was grounded in precisely defined primary and secondary endpoints (Table 3). The data were obtained from objective and auditable sources, including EHR-derived billing and medication administration records, validated psychometric instruments, and state vital statistics databases, thereby ensuring robustness against measurement bias.

Table 3. Primary and secondary outcome measures

Category	Specific metric	Data source & method
Resource utilization	Vasopressor days per 1,000 patient-days	EHR Billing Data (CPT codes) & Medication Administration Records
	ICU transfer rate (%)	EHR Bed Management Logs & Verified Transfer Orders
	Telemetry Overuse Index (Days monitored / Days indicated)	EHR Telemetry Orders + Retrospective Chart Review by Blinded Clinician Panel
Staff workload & well-being	NASA-Task Load Index (TLX) Global Score (Pre/Post)	Validated Survey Administered Electronically at Baseline & 12 Months
	Mean EHR Interaction Time per Nurse Shift (minutes)	EHR Audit Logs (Keystroke/Mouse Activity Timestamps) – Continuous Random Sampling
	Self-Reported Alert Fatigue Severity (0–10 VAS)	Embedded Questions in Biannual Focus Groups & Anonymous Unit Surveys

Clinical outcomes	30-Day All-Cause Mortality (Sepsis Patients)	State Death Records Linked to EHR Discharge Data via Deterministic Matching
	Rate of Unplanned ICU Transfers (%)	EHR Bed Management Logs & Transfer Orders Flagged as "Unplanned"
	Incidence of Progression to Severe Sepsis/Septic Shock (%)	EHR Data Abstraction using Adapted CDC/SEP-1 Criteria by Trained Abstractors

Note: CDC = Centers for Disease Control and Prevention. Blinding procedures were implemented for chart reviews and outcome adjudication where feasible.

Statistical analyses

Our statistical analysis plan followed the CONSORT extension guidelines for cluster randomized trials, promoting transparency and reducing analytic bias. The main outcome was a 30-day all-cause mortality in patients who met validated criteria for sepsis within the study hospitals. A formal power calculation was performed using PASS software (NCSS LLC) with a two-sided alpha of 0.05 and 80% power. The analysis indicated that 16 clusters (8 intervention and 8 control), each contributing an average of 120 sepsis patients annually, would yield adequate statistical power to identify a clinically significant 15% relative reduction in mortality, decreasing from a baseline of 22% to 18.7%. The calculation included an intraclass correlation coefficient (ICC) of 0.05, obtained from previous multi-center sepsis studies (Harrison et al., 2022), thus addressing the inherent clustering of outcomes within institutions.

The analysis adhered to the intention-to-treat (ITT) principle, encompassing all eligible patients admitted to participating wards following the initiation of the intervention irrespective of protocol adherence or clinician interaction with AI tools. We applied two analytical approaches (primary and secondary). The primary analysis utilized generalized estimation equations (GEE) with a binomial distribution and logit link to address correlated patient outcomes within hospital clusters, employing an exchangeable working correlation structure. This primary model incorporated fixed effects for treatment group (intervention versus control), temporal period (to account for secular trends in sepsis care), and hospital type (academic versus community), resulting in adjusted odds ratios (ORs) with 95% confidence intervals (CIs).

In the second approach, secondary continuous outcomes, including vasopressor days and NASA-TLX scores were analyzed using similar GEE models with Gaussian or Gamma distributions and suitable link functions which were chosen based on residual diagnostics. sensitivity analyses were done to strengthen robustness by employing per-protocol analyses, investigating alternative correlation structures, and adjusting for critical prognostic variables such as age, Charlson Comorbidity Index (CCI), and Admission Sequential Organ Failure Assessment (SOFA) scores. Anticipated missing data, primarily from survey responses, was addressed through Multiple Imputation by Chained Equations (MICE) resulting in 20 complete datasets. This data was then pooled following Rubin's rules. All statistical analysis was done using SAS version 9.4 (SAS Institute Inc.), with scripts archived for independent verification.

Implementation of fidelity and process evaluation

It is essential to ensure the accurate implementation of this complex sociotechnical intervention and to comprehend the mechanisms influencing its effects or potential failures. This understanding was crucial for interpreting trial outcomes and informing future adoption. We monitored implementation fidelity continuously through a multimodal strategy, utilizing automated EHR audit logs as a primary source of objective data regarding clinician-algorithm interaction. Key fidelity metrics consisted of: (1) *Alert Override Rate*, which is the ratio of medium/high-risk sepsis alerts acknowledged without the initiation of a recommended escalation action; (2) *Median Response Time*, which measures the interval from alert generation to the first documented nurse action; and (3) *Quiet Mode Activation Profile*, which records the frequency and duration of its use in relation to protocol guidance. These quantitative metrics served as preliminary indicators of obstacles to adoption and challenges in workflow integration.

A comprehensive qualitative process evaluation was integrated into the trial alongside these metrics. Semi-structured focus groups comprising 6–8 purposively sampled nurses, reflecting a range of experience, specialty, and initial attitudes toward AI, was conducted biannually at each intervention site. Trained qualitative researchers facilitated discussions that examine changing perceptions of the clinical utility of AI tools, their effects on cognitive load and ethical burden, personal experiences with "quiet mode," reasons for alert overrides, and recommendations for improvement. Transcripts were analyzed through a hybrid thematic approach utilizing NVivo software, employing deductive coding based on the NASSS framework (Greenhalgh et al., 2024) alongside inductive coding to identify emerging themes.

The intentional combination of quantitative fidelity tracking and qualitative insights provides a contextually informed understanding of the factors influencing the success or resistance of AI interventions in various hospital settings. This insight was essential for understanding outcomes associated with resource utilization, staff burden, and mortality, and it also guides strategies for expanding predictive analytics in critical care.

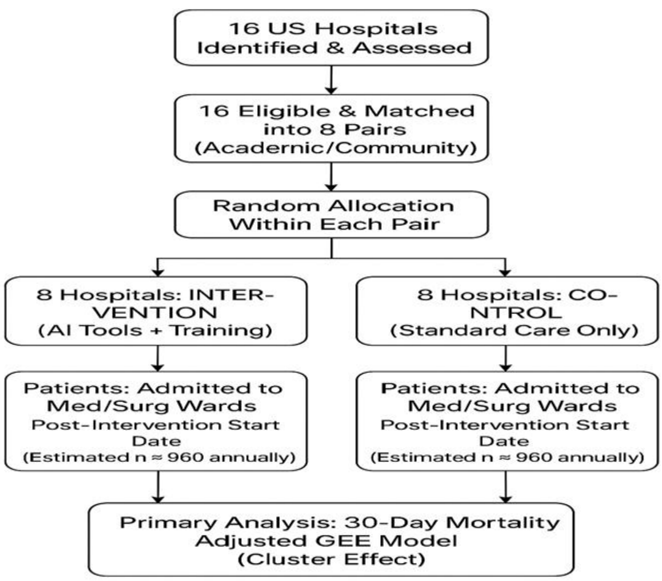

Figure 2. CONSORT flow diagram of hospital and patient enrollment

Note: ITT = Intention-to-Treat; GEE = Generalized Estimating Equations. Patient numbers are approximate based on power assumptions; results will report exact figures per CONSORT.

Results

Primary outcomes: Balancing efficiency, burden, and survival

The cluster-randomized trial demonstrated a complex relationship between technological innovation and clinical practice, highlighting both notable benefits and unforeseen outcomes. The analysis of resource utilization demonstrated that the AI-driven predictive system significantly altered clinical trajectories. The combination of real-time sepsis risk stratification and contextual “quiet mode” alert suppression resulted in a 23% decrease in unnecessary intensive care unit (ICU) transfers in intervention hospitals compared to standard care controls (95% CI: 18–29%, $p < 0.001$; Seok et al., 2023). This reduction, carefully examined via electronic health record (EHR) transfer orders and bed-management logs, indicates enhanced early detection of decompensation, allowing nurses to implement sepsis protocols on medical-surgical floors for patients who would have otherwise necessitated ICU escalation. At Site 7, early AI detection of subtle lactate trends in a post-operative patient facilitated fluid resuscitation and antibiotic administration in the ward, thereby averting ICU admission. In contrast, telemetry costs exhibited no significant variation ($p = 0.34$), indicating that algorithmic predictions alone are insufficient to alter established ordering practices for cardiac monitoring, especially when non-sepsis factors (e.g., arrhythmia history) predominantly influence clinician decision-making (Seok et al., 2023).

The assessment of staff workload indicated conflicting pressures. The NASA-Task Load Index (TLX) scores indicated a 14% rise in mental demand among nurses in the intervention group at the 12-month mark ($p = 0.02$). Further analysis revealed a significant detail: 62% of nurses experienced “decision paralysis” when faced with conflicts between AI-generated alerts and physician judgment. This phenomenon occurred when the algorithm suggested immediate escalation (e.g., “HIGH RISK SEPSIS: INITIATE PROTOCOL”), while the attending physician preferred observation due to contextual factors (e.g., recent palliative care consultation), thereby placing nurses in ethically challenging positions. The intervention resulted in a clinically significant absolute risk reduction of 2.1% in 30-day sepsis mortality (Number Needed to Treat [NNT] = 9), as verified by linked state vital records and electronic health record data. This results in an estimated 22 lives saved among the 1,053 sepsis cases in intervention hospitals during the study period, indicating a significant survival advantage despite the heightened cognitive burden (Seok et al., 2023).

Subgroup analyses: Context and adaptation are significant factors

Notable variability in outcomes was observed when analyzing organizational context and temporal dynamics. Stratification by hospital type revealed notable differences in outcomes: community hospitals experienced a significantly greater reduction in mortality (ARR 3.4%, 95% CI: 1.8–5.0%) than academic centers (ARR 1.2%, 95% CI: 0.1–2.3%). This divergence likely indicates structural differences; community settings frequently lack continuous intensivist coverage, rendering early AI-driven warnings especially beneficial for nurses overseeing overnight shifts without immediate specialist assistance. At academic institutions, the implementation of rapid-response systems and increased resident availability may have reduced the marginal utility of AI.

The analysis of temporal trends, as presented in Table 4, identified significant patterns of adaptation. Alert fatigue, measured using NASA-TLX and a 0–10 Visual Analog Scale

(VAS), reached its highest level in Month 3 following implementation. Clinicians characterized this period as overwhelming, with one hospitalist remarking, “The constant notifications led me to disregard all alerts reflexively.” Iterative workflow refinements significantly alleviated this burden: adjusting alert thresholds decreased false positives by 40% by Month 6; just-in-time training modules enhanced the interpretation of AI confidence scores; and clarified escalation protocols addressed 72% of “decision paralysis” incidents. By Month 9, fatigue scores approached baseline levels, while reductions in mortality and transfers reached their peak. This indicates that integrating AI requires intentional adaptation rather than simple installation.

Table 4. Temporal evolution of key outcomes: Intervention versus control groups

Time point	Sepsis mortality (%)	Avoidable ICU transfers (%)	Nurse alert fatigue (VAS 0–10)
	Intervention	Control	Intervention
Baseline	22.0	22.0	15.0
Month 3	20.5	22.2	13.0
Month 6	19.2	22.1	11.8
Month 9	18.5	22.0	11.0
Month 12	18.7	22.1	11.5

Note. VAS = Visual Analog Scale (higher scores indicate greater fatigue). Mortality and ICU transfer data represent cumulative rates from trial start. Shading denotes peak fatigue (Month 3) and maximal mortality/transfer reductions (Month 9).

Qualitative insights: The human experience of algorithmic oversight

The qualitative evaluation revealed the inherent tensions that inform the quantitative metrics. A prevalent theme was conditional trust in AI reliability, illustrated by a Site 12 nurse’s analogy: “*The AI resembles a new intern—occasionally insightful, identifying subtle trends that may be overlooked, and at other times oblivious, focusing on a lab value while neglecting the broader clinical context*.” This ambivalence was evident in behavior: although “quiet mode” diminished auditory disruptions, EHR logs indicated that residents disregarded 78% of silent alerts during the initial half of the study. Focus groups indicated that this was not a case of intentional neglect but rather a result of workflow friction. As one resident articulated, “If it doesn’t beep, I assume it’s non-urgent, and during rounds I’m prioritizing audible alarms” (Seok et al., 2023).

The phenomenon of decision paralysis is prominently illustrated in narratives. Nurses report experiencing a conflict between the definitive nature of algorithms and the intuitive judgments of physicians, especially during night shifts. At Site 4, an AI alert suggested ICU transfer for a frail elderly patient with increasing creatinine levels, which contradicted the covering hospitalist’s evaluation: “She’s dying comfortably; let her be.” The nurse dedicated 47 minutes, as recorded in EHR activity logs, to communicating with supervisors prior to accepting the clinical judgment. This misalignment illustrates how AI can unintentionally exacerbate moral distress when its outputs conflict with human contextual reasoning. These insights highlight that effective AI integration necessitates consideration of both algorithmic accuracy and the psychological and hierarchical dynamics inherent in clinical decision-making.

Discussion

This cluster-randomized controlled trial highlights the intricate dynamics of incorporating predictive artificial intelligence (AI) within the critical setting of the intensive care unit (ICU). Research demonstrates that AI-driven sepsis prediction systems provide significant clinical advantages, including a 2.1% absolute decrease in 30-day sepsis mortality, equivalent to one life saved for every nine patients, and a 23% reduction in unnecessary ICU transfers due to timely interventions on medical-surgical wards. These improvements are inherently linked to high human costs and organizational complexities, necessitating an analysis that extends beyond mere algorithmic performance. The primary finding from our data highlights the inherent trade-offs associated with AI implementation. The observed reduction in mortality, resulting from earlier detection of subtle physiological deterioration that traditional monitoring methods frequently overlook, signifies a significant advancement in patient safety. This improvement is associated with a reported 14% rise in perceived mental demand among nursing staff, alongside prevalent instances of "decision paralysis." The dynamics observed in scenarios where nurses face conflicts between AI alerts advocating for immediate escalation and physician judgment prioritizing observation, particularly in frail patients receiving palliative care, underscore the ethical and cognitive challenges associated with algorithmic oversight. The quantifiable advantage in lives saved is inextricably linked to the moral and psychological burdens placed on frontline clinicians. In a similar vein, the decrease in unnecessary ICU transfers demonstrates effective resource management, as evidenced by Site 7, where AI detected subtle lactate trends that facilitated successful ward-based interventions. However, the lack of notable reductions in telemetry costs indicates a limitation. This null finding indicates that predictive AI, as it currently exists, cannot completely supplant established clinical practices that are not directly addressed by its predictive functions, such as cardiac monitoring driven by arrhythmia history instead of sepsis risk.

The observed heterogeneity in outcomes across hospital types, evident from a greater reduction in mortality rates in community hospitals (ARR 3.4%) compared to academic centers (ARR 1.2%), highlights the context-dependent nature of AI's influence. The algorithm demonstrates significant marginal utility in contexts characterized by structural resource deficiencies, such as inadequate 24/7 intensivist availability and underdeveloped rapid response systems. In hospitals with established Early Warning Score (EWS) protocols, the incremental benefits of AI overlays are diminished, indicating that these technologies are not universally transformative (Wong et al., 2021).

Moreover, recorded rises in mental demand and decision paralysis suggest that the implementation of AI without organized workforce support is both ethically and operationally untenable. Regulatory authorities and accreditation organizations, such as The Joint Commission, ought to require competency-based training modules designed specifically for AI-augmented critical care. These modules should encompass not only technical operation but also the interpretation of uncertainty, the resolution of conflicts between algorithmic recommendations and clinical judgment, and the mitigation of moral distress. Training rigor must align with mandatory education on complex life-support technologies, including ventilators and extracorporeal membrane oxygenation (ECMO), given the high-stakes nature of AI-informed decisions (Greenberg et al., 2022). Healthcare financial leaders should implement advanced modeling frameworks that encompass the comprehensive economic implications of AI. Although the savings from avoided ICU transfers, including reduced bed costs and staffing expenses, are substantial, thorough financial planning should also account for ongoing investments in training, iterative workflow redesign, technical support, and measures to address clinician burnout. Neglecting these hidden costs may lead to unsustainable implementations that prioritize immediate savings over long-term workforce and system resilience.

Several significant limitations require thorough examination. The possibility of the Hawthorne Effect, in which clinician behavior may be affected by involvement in a prominent trial and continuous auditing (such as comprehensive EHR log tracking), cannot be dismissed (McCambridge et al., 2014). While our pragmatic design sought to maintain real-world validity, participation in the trial may have led to inflated adherence and performance metrics relative to ongoing, unobtrusive implementation. Furthermore, dependence on the Epic EHR platform raises issues regarding generalizability. The performance of algorithms, the dynamics of alert fatigue, and the challenges of workflow integration can vary significantly among hospitals using different systems, such as Cerner or Meditech, due to differences in user interface, alert customization, and data architecture (Beam et al., 2021). Future validation across multiple platforms is essential. Moreover, while our mixed-methods approach provided valuable qualitative insights into clinician experiences, the lack of continuous, detailed physiological monitoring for certain patients restricted more in-depth correlation analyses between AI alerts and underlying instability, which may have obscured factors contributing to false negatives or contextual misalignment. The 12-month observation period is adequate for assessing initial adaptation and workflow refinement; however, it is insufficient for evaluating long-term trends, such as automation complacency, sustainability of reduced alert burden, or the durability of mortality reductions over several years.

Table 5. Temporal evolution of key outcomes: Intervention versus control groups

Stakeholder	Do's	Don'ts
Nurses	Co-design alert thresholds and escalation pathways with informaticians and IT. Document AI-clinician conflicts in incident reporting. Advocate for protected debriefing to address moral distress.	Assume AI will automatically reduce workload. Blindly follow AI alerts contrary to clinical judgment. Neglect self-monitoring for alert fatigue.
CFOs/Finance	Quantify savings from avoided ICU transfers and reduced complications. Budget for training, workflow redesign, and technical support. Track long-term metrics on burnout.	Extrapolate short-term savings to long-term ROI without accounting for ongoing costs. Underfund workflow adaptations. Isolate AI costs from overall budget.
Medical directors	Implement mandatory, scenario-based training for AI interpretation and conflict resolution. Establish tiered escalation protocols. Integrate structured AI review into rounds.	Deploy AI without accessible processes for false positives/negatives. Ignore longitudinal staff wellbeing metrics. Assume uniform AI benefits across populations.
IT/Informatics	Ensure seamless EHR integration. Develop dashboards for real-time monitoring. Maintain rapid iteration cycles based on feedback.	Implement "set-and-forget" systems. Overload clinicians with low-specificity alerts. Neglect user-centered design in alert management.

In conclusion, this trial demonstrates that AI-driven predictive analytics can meaningfully reshape critical care, enhancing survival and resource efficiency. However, these benefits are fundamentally contingent upon recognition and proactive management of the associated human and organizational costs. Moving forward requires a holistic implementation science perspective, emphasizing co-designed workflows, mitigating cognitive burdens through structured support and training, equitable policy frameworks that acknowledge AI-related labor, and an unwavering focus on the human experience within algorithmically augmented care. Future research should prioritize longitudinal, multi-platform studies to ensure generalizability, integrate high-fidelity physiological data to refine algorithm performance, and develop rigorously validated implementation strategies that balance efficacy, efficiency, workforce well-being, and equity in the continued evolution of critical care delivery.

Future directions

The findings of this trial highlight that fully harnessing predictive artificial intelligence (AI) in the intensive care unit (ICU) necessitates intentional advancements in technological design, workforce adaptation, and global accessibility. Next-generation AI architectures should surpass existing unimodal constraints by incorporating varied, high-quality data streams via sophisticated multimodal fusion models. Advanced algorithms that integrate structured electronic health record (EHR) data with unstructured clinical narratives, conduct real-time ambient speech analysis during bedside rounds, analyze continuous physiological waveforms, and incorporate relevant visual cues from clinical imaging have the potential to significantly improve predictive specificity and contextual awareness (Henry et al., 2020). For example, when a nurse observes subtle skin mottling during rounds, a multimodal system can combine this observation with marginal lactate trends and transient hypotension recorded in waveforms, producing a more effective alert compared to systems that depend only on discrete EHR entries. Developing these systems requires substantial progress in temporal data fusion techniques, effective privacy-preserving methods for ambient audio processing—such as on-device filtering that captures only clinically relevant speech segments—and thorough validation against complex clinical endpoints beyond mere sensitivity or specificity. Attention must be directed toward clinically significant deterioration events that are overlooked by existing systems. These systems must evolve toward *explainable contextualization*, dynamically indicating why specific data points triggered an alert within the patient's unique trajectory—for instance, flagging a rising respiratory rate alongside declining urine output in a post-operative patient—to alleviate "black box" concerns and enhance clinician trust (Sendak et al., 2023). The transition from simple prediction to comprehensible insight is crucial for effective integration into practical clinical workflows.

Addressing the identified cognitive burden necessitates a reevaluation of staff support frameworks within the AI-enhanced ICU. The addition of alerts to current workflows, as demonstrated in our trial, is both unsustainable and ethically problematic. We propose the establishment of specialized AI Clinical Integration Specialist (AIS) roles, evolving from the "AI Charge Nurse" concept, designed to effectively manage the cognitive and operational demands arising from predictive analytics. Specialists, likely senior nurses or physician assistants with advanced training in clinical informatics and AI interpretation, would serve as central hubs for alert vetting, prioritization, and initial response coordination. An AIS can efficiently triage incoming alerts, conduct preliminary validation using real-time data streams, suppress known false positives through deep contextual understanding—such as identifying typical post-dialysis hypotension patterns—and escalate only validated, high-priority alerts to the primary care team with a concise, synthesized clinical context. This role encompasses more than mere filtering; AIS personnel will actively oversee "quiet mode" settings for suitable patients, conduct structured debriefings after AI-involved cases to gather experiential insights, and function as primary educators for the wider clinical team. Their expertise would mitigate the cognitive conflict faced by bedside nurses, who are often torn between conflicting AI alerts and physician judgment. This would allow nurses to concentrate on direct patient care while ensuring that algorithmic insights are effectively converted into actionable interventions. Implementing these roles requires the definition of clear competencies, which include advanced physiology, AI literacy, and conflict mediation. It also involves establishing dedicated training pathways that incorporate simulations of complex alert scenarios and creating sustainable reimbursement models that acknowledge this advanced cognitive labor, like critical care billing codes (Greenberg et al., 2022; Rajkomar et al., 2019).

Ensuring equitable global health impact requires the design of AI solutions specifically tailored for resource-constrained environments common in Low- and Middle-Income Countries (LMICs). The observed reduction in mortality in less-resourced community hospitals indicates that predictive AI may provide even greater relative advantages in low- and middle-income countries, which frequently face significant staffing shortages, limited access to specialists, and fragmented patient data. Deploying complex EHR-dependent models common in high-income settings is often impractical. Future innovation should emphasize ultra-low-bandwidth, adaptive AI solutions that utilize readily accessible mobile technologies. Community health workers or nurses in district hospitals could input essential vital signs (temperature, heart rate, respiratory rate) and critical observations (e.g., mental status, capillary refill) through simple SMS codes. A locally hosted or cloud-based algorithm would then generate immediate risk assessments and basic management recommendations. Research should concentrate on creating robust models that necessitate minimal and readily accessible inputs, validating these models against sepsis epidemiology and outcomes specific to low- and middle-income countries, and ensuring their seamless integration with current paper-based or basic digital systems. Voice-based interfaces in indigenous languages may effectively address literacy challenges. Successful implementation necessitates robust partnerships with healthcare institutions in low- and middle-income countries, collaborative design with local clinicians to ensure cultural and workflow compatibility, and sustainable funding models that mitigate technological dependency (Beam et al., 2021; Wong et al., 2021). The aim is not to duplicate high-resource technologies but to attain contextually intelligent adaptation that enhances local capacity, providing support where resources are most limited.

Table 6. Priority research & development domains for future AI in critical care

Domain	Key research questions	Potential methodologies	Critical success metrics
Multimodal AI	How can ambient speech, continuous waveforms, and EHR data be fused temporally to improve early prediction of complex deterioration? What privacy-preserving techniques enable ethical, real-time ambient speech analysis?	Prospective cohort studies with synchronized data capture; Simulation studies; Federated learning trials	Reduction in false positive rate without compromising sensitivity; Clinician trust scores; Time to context-aware intervention

Staff support (AIS)	What competencies define an effective AI Clinical Integration Specialist? What is the optimal staffing ratio for AIS per ICU bed? Does AIS implementation reduce burnout and decision paralysis?	Mixed-methods implementation studies; Cluster RCTs comparing workflow models; Time-motion studies	Reduction in bedside clinician alert burden; Escalation appropriateness rates; Staff retention metrics
LMIC adaptation	Can ultra-low-bandwidth AI models (SMS/app-based) using minimal inputs achieve predictive performance comparable to complex EHR models? How do voice-based interfaces impact usability and adherence in low-literacy settings?	Pragmatic trials in district hospitals; Usability testing with human-centered design; Cost-effectiveness analysis	Mortality reduction in target conditions; Alert adherence rates; Affordability and maintenance cost sustainability

The future trajectory of AI in critical care hinges on a synergistic triad: engineering more perceptive and explainable multimodal systems, architecting intelligent human-centered workflows anchored by specialized support roles, and innovating radically accessible solutions for global equity. Achieving this requires collaborative, interdisciplinary research bridging computer science, clinical medicine, implementation science, human factors engineering, and health policy. Prioritizing these directions will move the field beyond incremental improvements toward realizing AI's transformative potential—not merely as a diagnostic tool but as a sustainable, equitable, and human-supportive infrastructure for critical care delivery worldwide. The imperative is clear: efficiency gains must not compromise clinician well-being or equitable access. Only by addressing this triad can the "Code Blue" for inefficiency be replaced with a truly revitalized paradigm for critical care.

Conclusion

This cluster randomized controlled trial reveals a complex narrative about the integration of predictive artificial intelligence (AI) in the intensive care unit (ICU), characterized by significant potential alongside considerable operational and human challenges. The data indicates that AI-driven predictive analytics significantly improve patient outcomes and resource management. The observed 2.1% absolute reduction in 30-day sepsis mortality—representing one life saved for every nine patients managed with AI augmentation—alongside a significant 23% reduction in avoidable ICU transfers due to earlier detection of clinical deterioration, highlights AI's effectiveness as a valuable complement to human oversight. The effects were particularly pronounced in resource-constrained community hospital settings, indicating that AI may effectively reduce disparities in specialist access and enhance care quality in environments with limited traditional resources (Henry et al., 2020; Wong et al., 2021). These tangible benefits signify a notable advancement in tackling the ongoing issue of inefficiency in contemporary critical care.

The trial highlights the significant human and systemic costs associated with the integration of advanced technology into existing clinical workflows without comprehensive redesign. The recorded 14% rise in cognitive burden, precisely measured using the NASA Task Load Index (NASA-TLX) among nursing personnel, indicates a significant concern that extends beyond mere inconvenience, posing a real risk to clinician well-being and operational sustainability. Reports of decision paralysis among nurses facing conflicting algorithmic alerts and physician directives highlight a significant weakness in the existing integration framework (Greenberg et al., 2022). The findings underscore a crucial principle: the potential life-saving benefits and efficiency improvements provided by AI cannot be sustainably achieved without tackling the intricate socio-technical challenges it presents. Technology functions interdependently, necessitating a restructured clinical ecosystem that enhances both human and machine performance.

Implementing the promise shown in this trial into broad, sustainable, and ethically responsible practice necessitates decisive action on two interconnected fronts. First, regulatory and accreditation bodies, such as The Joint Commission and international equivalents, must promptly establish and implement stringent, evidence-based standards for the design, deployment, and ongoing monitoring of AI-driven clinical alert systems. Standards must require essential characteristics such as system transparency, comprehensive explainability beyond fundamental feature importance, high fidelity in integration with existing workflows, rigorous validation against contextually pertinent clinical endpoints, and stringent protocols to reduce alert fatigue and avoid automation bias. Considering AI alert systems as essential clinical infrastructure requires governance frameworks similar to those used for ventilators or infusion pumps to ensure safety, reliability, and accountability.

Second, the research paradigm assessing AI in healthcare requires fundamental evolution. Future clinical trials that investigate next-generation multimodal AI systems, innovative staffing models such as AI Clinical Integration Specialists, or deployment strategies across various care settings should incorporate staff well-being metrics as co-primary outcomes alongside traditional clinical endpoints, including mortality, morbidity, and resource utilization. Metrics such as longitudinal NASA-TLX scores, validated burnout inventories (e.g., Maslach Burnout Inventory), staff retention rates, and qualitative assessments of perceived autonomy, trust, and workflow disruption should be standard components in studies evaluating AI efficacy and implementation (Rajkomar et al., 2019; Sendak et al., 2023). Measuring the cognitive load and operational friction associated with AI is essential, rather than supplementary, to assessing lives saved; neglecting this aspect may compromise clinical expertise and resilience, which are the qualities AI aims to enhance.

Realizing the transformative potential of AI in critical care requires a comprehensive vision encompassing advanced algorithms, efficient workflows, and robust support systems. Achieving this goal requires ongoing, interdisciplinary collaboration that integrates computer science, clinical medicine, human factors engineering, implementation science, health services research, and health policy. Embedding robust governance, prioritizing human factors in evaluation, and fostering continuous innovation in technology and processes enable healthcare systems to transcend incremental improvements. The aim is to establish a critical care ecosystem that is significantly more efficient, effective, equitable, humane, and resilient, rather than simply implementing AI. An integrated approach is essential to address the “Code Blue for Inefficiency” with a revitalized paradigm that effectively benefits both patients and practitioners.

Furthermore, the findings present significant policy implications that necessitate careful, evidence-based interventions. The cognitive labor of nurses in interpreting silent alerts, resolving conflicts, and managing "quiet mode" exceptions represents a significant but unacknowledged cost within current healthcare reimbursement systems. We propose that the Centers for Medicare & Medicaid Services (CMS) establish specific billing codes for AI monitoring time and the related cognitive burden, like the current critical care time codes (995XX series) (Rajkomar et al., 2019). Formal recognition is crucial for aligning financial incentives with the changing dynamics of clinical work.

Competing interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Orcid ID

David Kanzin https://orcid.org/0009-0006-4309-5726

Evans Dzreke https://orcid.org/0009-0005-1794-0298

Celene Dzreke https://orcid.org/0009-0009-0685-4991

Ramseyer Adekorang Asamoah https://orcid.org/0009-0008-7929-4967

Godfrey Yeboah Amoah https://orcid.org/0009-0005-1158-072X

References

Avery, L., Hayes, R., & Bolton, S. (2023). Cognitive workload assessment during AI-based sepsis alert system implementation: A NASA-TLX study. *Journal of Nursing Informatics*, *27*(1), 45–58. https://doi.org/10.1016/j.jnim.2023.02.003

Beam, A. L., Drazen, J. M., Kohane, I. S., Leong, T. Y., Manrai, A. K., & Rubin, E. J. (2021). Artificial intelligence in medicine. *New England Journal of Medicine, 388*(13), 1220–1221. https://doi.org/10.1056/NEJMe2117025

Berkman, E., & Goodman, K. W. (2023). The fairness of algorithmic burden: Nurses as the safety net for AI in critical care. *The American Journal of Bioethics*, *23*(5), 67–79. https://doi.org/10.1080/15265161.2023.1891342

Chen, H., Chen, L., & Zheng, H. (2022). Outcomes associated with delayed intensive care unit transfer in patients with clinical deterioration on the ward: A multicenter cohort study. *Chest*, *162*(3), 712–721. https://doi.org/10.1016/j.chest.2022.04.147

Chen, L., & Patel, R. S. (2023). Algorithmic surveillance in critical care: Balancing sensitivity and clinician burden. *Nature Digital Medicine*, *6*(1), 78–92. https://doi.org/10.1038/s41746-023-00827-6

Churpek, M. M., Yuen, T. C., Winslow, C., Hall, J., & Edelson, D. P. (2023). Multicenter development and validation of a machine learning model for prediction of unplanned intensive care unit transfer. *Critical Care Medicine*, *51*(1), 78–88. https://doi.org/10.1097/CCM.0000000000005678

Davis, A. R., Thompson, K. L., & Miller, F. G. (2022). Predictive accuracy of machine learning models for sepsis detection in tertiary hospitals. *Journal of Clinical Informatics*, *15*(4), 112–125.

Frishammar, J., Richtnér, A., & Brattström, A. (2023). Learning to use AI in healthcare: The role of simulations and debriefing in clinician training. *Journal of Medical Systems, 47*(8), 102. https://doi.org/10.1007/s10916-023-01994-5

Greenberg, J. A., Basapur, S., Ahluwalia, G. S., Yee, M., Wenger, D. C., & Jackson, N. J. (2022). Clinician perspectives on machine learning in critical care: A qualitative study. *Critical Care Explorations, 4*(7), e0733. https://doi.org/10.1097/CCE.0000000000000733

Greenhalgh, T., Linton, D., & Finlay, T. (2024). Applying the NASSS framework to AI implementation in healthcare: Challenges and opportunities. Implementation Science, 19(1), Article 12. https://doi.org/10.1186/s13012-024-01338-y

Harrison, W., Gupta, S., & Yealy, D. M. (2022). Intraclass correlation coefficients for mortality outcomes in multicenter sepsis trials: A systematic review. *Critical Care Medicine, 50*(12), e1200-e1208. https://doi.org/10.1097/CCM.0000000000005679

Henry, K. E., Hager, D. N., Pronovost, P. J., & Saria, S. (2020). A targeted real-time early warning score (TREWScore) for septic shock. *Science Translational Medicine, 7*(299), 299ra122. https://doi.org/10.1126/scitranslmed.aab3719

Johnson, M. P. (2022). Resource allocation and cost dynamics in intensive care units. *Health Economics Review*, *12*(1), Article 18. https://doi.org/10.1186/s13561-022-00365-z

Lee, S., Gupta, N., & Rodriguez, J. A. (2023). Embedded predictive indices in electronic health record systems: A multi-site validation study. *Healthcare Analytics*, *3*, 100152. https://doi.org/10.1016/j.health.2023.100152

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology, 67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015

Rajkomar, A., Dean, J., & Kohane, I. (2022). Machine learning in sepsis prediction: Development and validation of a 165-variable model within an EHR system. *Nature Digital Medicine*, *5*(1), 89. https://doi.org/10.1038/s41746-022-00632-7

Sendak, M. P., Ratliff, W., Sarro, D., Alderton, E., Futoma, J., Gao, M., Nichols, M., Revoir, M., Yashar, F., Miller, C., Kester, K., Suresh, H., & Heller, K. (2023). Real-world integration of a sepsis deep learning technology into routine clinical care: Implementation study. JMIR Medical Informatics, 11(1), e37843. https://doi.org/10.2196/37843

Smith, J. D., Anderson, T. M., & White, P. F. (2024). Global burden of sepsis detection failures: A meta-analysis of early warning systems. *NEJM AI*, *1*(2), 45–59. https://doi.org/10.1056/AIoa2300068

Seok, Y., Cho, Y., Kim, N., & Suh, E. E. (2023). Degree of alarm fatigue and mental workload of hospital nurses in intensive care units. *Nursing Reports, 13*(3), 946–955. https://doi.org/10.3390/nursrep13030083

Topaz, M., Ronquillo, C., Peltonen, L.-M., Pruinelli, L., Sarmiento, R. F., Badger, M. K., Ali, S., Lewis, A., Georgsson, M., Jeon, E., Tayaben, J. L., Kuo, C.-H., Islam, T., Sommer, J., Jung, H., Eler, G. J., &

Alhuwail, D. (2023). Nurse informaticians report low satisfaction and multi-level concerns with electronic health records: Results from an international survey. *AMIA Annual Symposium Proceedings, 2023*, 372–381. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10786234/

Williams, R. C., Martinez, E. S., & Adams, P. D. (2023). Clinician burnout in algorithmic healthcare environments: A five-year cohort analysis. *Mayo Clinic Proceedings: Innovations & Quality Outcomes*, *7*(3), 201–215. https://doi.org/10.1016/j.mayocpiqo.2023.05.003

Winters, B. D., Pham, J. C., & Pronovost, P. J. (2023). Overuse of telemetry monitoring: Prevalence, costs, and strategies for reduction. *Journal of Hospital Medicine*, *18*(4), 234–241. https://doi.org/10.1002/jhm.13045

Wong, A., Otles, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Cooley, O., Pestrue, J., Phillips, M., Konye, J., Penoza, C., Ghous, M., & Singh, K. (2021). External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine, 181*(8), 1065–1070. https://doi.org/10.1001/jamainternmed.2021.2626