Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Clinicians' perspectives on the design of early warning systems presenting AI-based multiple outputs for clinical deterioration: A qualitative analysis

Monica Noselli

m.noselli@uq.edu.au

The University of Queensland https://orcid.org/0009-0003-5756-9214

Anton Van Der Vegt

The University of Queensland https://orcid.org/0000-0001-5642-5188

Maxime Cordeil

The University of Queensland https://orcid.org/0000-0002-9732-4874

Victoria Campbell

Clinical Excellence Queensland https://orcid.org/0000-0001-9846-9410

Ian A. Scott

4. Metro South Digital Health and Informatics, Princess Alexandra Hospital https://orcid.org/0000-0002-7596-0837

Research Article

Keywords: Artificial Intelligence, Clinical decision support system, Human Factors, Qualitative analysis

Posted Date: November 11th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-8071546/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background

AI-enabled clinical decision support systems (CDSS) can help detect patient deterioration, but their value depends on outputs that fit workflows, support reasoning, and avoid alert fatigue.

Objective.

Little is known about clinician perspectives on multi-output Early Warning Systems (MO-EWS), which provide simultaneous predictions (e.g., overall deterioration, organ- or diagnosis-specific). This study explores clinician views on future MO-EWS and offers user-centred recommendations to improve design, usability, and adoption.

Methods

Reflexive thematic analysis was used to examine clinicians' perceptions of existing early warning systems (EWS) and envisioned AI-based tools for predicting clinical deterioration. Between January and June 2025, 22 clinicians from Australian hospitals with experience in managing clinical deterioration were interviewed using a semi-structured questionnaire. Resulting themes were mapped to system components (user interface [UI], AI model, clinical workflow) to inform practical design recommendations which were then assessed for their feasibility.

Results

Participants found existing EWS distracting and sometimes unsafe. They envisioned AI-based systems that provide multiple, contextualised outputs (e.g., causes of deterioration, organ dysfunction, clinical severity) as layered, adaptable tools to support reasoning, prioritisation, and communication. While many UI improvements are readily achievable, model development is more complex, and workflow integration will depend on infrastructure, policies, and clinician preferences.

Conclusion

Clinicians perceive potential benefits in an AI-enabled system that provides more insightful information when predicting clinical deterioration in individual patients, but various design and implementation challenges remain.

1. Introduction

As artificial intelligence (AI) continues to be integrated into clinical decision support systems (CDSS) [1], there is growing interest in how AI-generated predictions can assist clinicians to identify patient deterioration thereby enhancing decision-making, facilitating early intervention and supporting resource allocation [2]. While risk prediction algorithms are getting more accurate [3], their impact on clinical decision making and care outcomes depends on the actionable content of the CDSS outputs and how

these are presented to clinicians [4]. Poorly designed CDSS user interfaces can lead to poor utility, misinterpretation, alert fatigue and non-adoption of the system [5].

Previous AI-based CDSS (AI-CDSS) user interface research focused mostly on single condition or single alert scores (e.g., sepsis or general deterioration only) [6] [7] [8] [9] [10]. These studies identified important design factors. For example, Blythe et al [8] explored how early warning scores (EWS) influence clinicians' decision-making, highlighting the importance of models that facilitate but not replace critical thinking, provide transparency regarding included and excluded variables, and support decision-making processes. Owoyemi et al. [9] examined user experiences with the Epic Sepsis System (ESS) and identified challenges with workflow fit and usability. A second study of the ESS [10] proposed that supporting clinicians in hypothesis generation and information gathering was more important than providing a final diagnosis or decision. AI-CDSS user-interface investigators have also found that successful implementation of AI-based tools requires understanding clinicians' cognitive needs and framing AI as complementary support [11]. In all these investigations, a singular AI-based CDSS tool was considered.

However, as more AI-based predictive models are researched and developed, it is likely that future AI-CDSS deterioration systems will encompass multiple predictive models working in concert. Currently, there is very limited research and understanding of how multiple concurrent AI-based models should present their outputs within a single AI-CDSS to support clinician decision-making.

The purpose of this study is to investigate clinicians' views on how multiple AI-based predictions could and should be presented within a deterioration focused AI-CDSS. The purpose of this investigation is to better understand (i) the issues that clinicians have with their current deterioration early warning tools that they would like to see resolved in future solutions; and (ii) clinician preferences with respect to the scope and design of future multi-prediction AI-based deterioration CDSS. This study applies a clinician-centric lens to identify actionable insights for system developers, co-development teams and researchers seeking to build multiple AI-based prediction CDSS that are both more effective with high levels of clinical adoption.

2. Methods

1.1. Participants and recruitment

Between January and June 2025, clinicians experienced in managing deterioration were recruited via email containing a flyer and participant information sheet. These were distributed through professional networks including two clinical research groups (AI-KEYS: Artificial Intelligence – Keeping You Safer, and Queensland Health Sepsis group). Contacted participants were invited to forward the call for participation email to extend recruitment reach. Recruitment was initially aimed at ward-based staff but expanded to include clinicians from Intensive Care Unit (ICU), Emergency Department (ED), and Post-

Anaesthesia Care Unit (PACU) in order to capture a broader range of perspectives across settings where early warning systems are used.

Participation was voluntary with consent. Recruitment was guided by the principle of informational power [12], given the narrow and clearly defined study aim, the participants' high level of clinical expertise, and the anticipated richness of dialogue, a relatively modest sample (22 participants) was considered sufficient. Recruitment concluded once the available networks had been fully approached and no additional eligible participants could be identified. No prior relationship existed between the researcher and participants. Ethical approval was obtained from the [name deleted to maintain the integrity of the review process].

1.2. Data collection

Researcher [name deleted to maintain the integrity of the review process], a female PhD student, conducted 45-minute one-on-one (single sessions only, without follow-up) Microsoft Teams interviews, using a semi-structured questionnaire, constructed de novo with all authors, that consistently captured: (a) participants' clinical roles and settings; (b) challenges in using current EWT for identifying clinical deterioration; and (c) perceptions of a novel MO-EWS. Two pilot interviews with two clinician co-authors helped refine clarity, relevance and suitability of the questions. Participants gave verbal consent to have interviews recorded for analysis using Microsoft Teams. Participants did not comment or correct the transcripts, they were invited to provide feedback on the findings, however, no responses were received.

1.3. Analysis

The study was conducted in three phases (Fig. I). First, following Braun and Clarke's approach to using Reflexive Thematic Analysis (RTA) (see Box) [16, 17]. Author [name deleted to maintain the integrity of the review process] engaged iteratively with transcripts, moving recursively between coding, theme development, review, and writing. Using NVivo 14 software [15], [name deleted to maintain the integrity of the review process] coded semantic and latent content, initially distinguishing 'actual' and 'envisioned' tools. However, since participants often discussed envisioned tool through the lens of their experience with existing tools, this overlap shaped the final coding. While some codes aligned with frameworks like technology acceptance model (TAM) [16], these limited reflexive analysis, so different themes were developed. Second, the final themes were then mapped to system components comprising (a) UI: what clinicians see and interact with; (b) AI algorithms that analyse data, make predictions, or generate alerts; (c) Workflow: the sequence of tasks and processes clinicians undertake when using the system in practice. Third, this mapping was used to derive practical design recommendations which were then assessed for their feasibility by consensus of all authors who comprised the main author [name deleted to maintain the integrity of the review process], two senior clinicians and system analysts [name deleted to maintain the integrity of the review process], a senior data scientist [name deleted to maintain the integrity of the review process] and a senior social scientist [name deleted to maintain the integrity of the review process]. In contrast to prior studies. we did not employ or demonstrate any tool prototypes to avoid cueing or biasing effects on clinicians in exploring these issues.

This study is reported according to the COREQ (COnsolidated criteria for REporting Qualitative research) Checklist [17].

Figure I Study goals and processes

3. Results

2.1. Participants

Twenty-two clinicians participated in interviews, comprising 11 nurses and 11 doctors, working in different departments across hospitals (ED = 8, ICU = 6, Ward = 5, PACU = 2, ICU + ED = 1) throughout Australia and all having a minimum 4 years of postgraduate experience.

Participants mentioned various currently used clinical deterioration alert tools, including Queensland-Adult Deterioration Detection System (Q-ADDS) and Between the Flags (BTF) [18], and downstream systems that report these alerts (Compass). They also worked with different electronic medical record (EMR) systems, such as Sunrise, Oracle Cerner, and the Cerner module FirstNet used in EDs.

2.2. Themes

The themes derived from interviews (Fig. II) are discussed in more detail in the following sections. Consistency between data and findings was ensured and illustrative participant quotes, identified using numbers and letters in parentheses in the following text, are provided in the supplementary material (Online Resource 1).

Figure II Themes and subthemes derived from interviews on clinicians' perceptions and expectations

2.2.1. Theme 1 - It cried wolf too many times and sometime not at all

Participants described how frequent, redundant, and poorly calibrated alerts (Table S1: Q-1 to Q-4) lead to alert burden and fatigue and eventual rejection of the alerting system through desensitisation, including when clinical intuition contradicts algorithmic outputs (Q-11).

The effective use of EWTs is shaped by a complex interplay of data quality, system design, clinical context and human judgment. As P10-D cautions, even the most sophisticated AI model is "only as good as the data you feed into it", which can be subjective (Q-5), inconsistent across settings (Q-6), have long time gaps between nursing observations (Q-7), and contain omissions or entry errors (Q-8, Q-9). Among design factors, rigid criteria embedded into systems can introduce bias and cognitive error (Q-10). Threshold-based triggers can miss deteriorations happening below the cut-off point (Q-11) and patient-specific thresholds and delta calculations, not incorporated into EWTs, also influence decisions (Q-12, Q-13). Human factors such as both over-reliant and overly cautious response to outputs can lead to missed deterioration in patients requiring urgent review (Q-14). Vital signs automatically plotted by the

computer may cause clinicians to forego the cognitive act of processing data that makes explicit their understanding of the trends (Q-15). We grouped both false positives (Q-1/10) and false negatives (Q-11/15) under this theme. Although opposite in form, participants perceived them as part of the same reliability problem: a system that either alarms unnecessarily or fails to alarm when needed ultimately reduces clinicians' trust in its outputs

2.2.2. Theme 2 - Hunting for answers in a fractured system

Poor UI layouts, fragmented workflows, or excessive alerts increase cognitive load and slow responses. Participants reported frustration from information being often dispersed across multiple pages or tabs, obscuring temporal trends and forcing them to mentally integrate disparate data sources (Table S2. Q-16). While some graphical displays offer trend visualization, inability to cross-reference trends directly with related clinical documentation disrupts workflow. Accessibility is insufficient if information is not presented in ways that align with clinicians' mental models (Q-17). Icon-heavy dashboards and copy-pasted documentation bring noise, information overload and reduced visibility of critical clinical indicators (Q-18).

Even with interactive functions, information may be updated in different pages or tabs, diminishing their utility for dynamic patient monitoring (Q-19). Clinicians disliked systems that are not easily searchable, wanting functionality that retrieves pertinent information with a minimal number of steps (Q-20).

Moreover, multiple, non-integrated platforms and data sources, combined with limited ability to extract actionable meaning from disjointed or cluttered interfaces, all foster redundancy, inefficiency, and cognitive strain, diverting time from patient care (Q-21, Q-22).

2.2.3. Theme 3 - Ghost in the Interface

This theme captures clinicians' vision of AI as an intelligent collaborator delivering actionable, context-aware outputs aligned with reasoning modes, workflow, and expertise. Interfaces should be layered, drillable, and adaptive, with alerts providing timely, user-sensitive decision support integrated seamlessly into workflows.

Sub-Theme 1 – Navigating Complexity: Clinician perceptions of single vs multi-output predictions

This theme highlights a tension between potential benefits and risks of MO-EWS.

In general, diagnosis-specific outputs are well regarded as they mimic the reasoning process of differential diagnosis (Table S3, Q-23), allowing clinicians to pursue their own hypotheses (Q-24). Participants see value in multiple outputs that provide richer detail and help identify diagnoses that may be overlooked or prioritise investigations. AI augments, not replaces, clinician reasoning, offering a proactive, exploratory approach to patient assessment (Q-25).

On the other hand, a general deterioration score is often sufficient to initiate investigation into underlying causes (Q-26), and where offering specific causative conditions could lead clinicians to prematurely

close on certain diagnoses without considering alternative explanations (Q-27). Multiple, simultaneous outputs could also overwhelm clinicians during busy shifts, further impairing situational awareness and timely responses (Q-28).

Overall, clinicians desired a layered or multi-functional system with drill-down interactions (Q-29 and Q-30).

Sub-theme 2 - From chasing problems to catching them early

Clinicians envision a tool that prompts action before deterioration becomes clinically overt (Q-31). A novel, proactive MO-EWS should surface relevant, context-aware information before the alert triggers, enabling clinicians to efficiently initiate preventive actions (Q-32). One clinician actively searches the data for early signs of deterioration prior to receiving alerts, supported by hospital resources (staff, hardware), centralised alert system design, and timely communication with front-line nurses (Q-33). Proactive intervention could potentially free up time in caring for more patients (Q-34), as well as potentially saving lives when clinicians' gestalt sees patients as worsening prior to triggering of alerts (Q-35). Observations and trends should remain visible in the UI as a first indicator of possible deterioration (Q-36).

Sub-theme 3 – It is only useful if I can act on it

Clinicians emphasised reactive alert systems should be action-oriented, regardless of the output, whether a deterioration score, or diagnosis-specific, or organ-specific outputs. For example, one participant pointed to actionable alerts tied to severity, expressing willingness to receive prompts suggesting appropriate actions based on the level of urgency (Q-37). Others gave practical examples of outputs tied to lab results (Q-38), not just showing the raw result (e.g. microurine analysis) but also giving diagnosis-specific guidance or suggested next step (Q-30.1). The system not only considers sepsis as a cause, for example, but also identifies a missing action (no fluids charted) and suggests next actions. Prioritising alerts according to type and severity of organ failure (Q-31.1) aids in ranking the urgency of response and suggesting remedial actions. Other clinicians considered time to act as necessary (Q-32.1), presenting both the required response and the urgency to act.

Participants noted specific diagnoses with a pre-determined response within established protocols reduced cognitive burden, while general deterioration was more ambiguous as to cause and appropriate response. They also stressed the need for role-appropriate recommendations, particularly for junior clinicians (Q-33.1) but also for specific staff groups (Q-34.1) and different levels of expertise (Q-35.1).

Also important was the desire for on-demand, flexible querying, describing a model that monitors and notifies using an interactive approach according to pre-specified rules (Q-36.1).

Sub-theme 4 – More Than a Number: Supporting Interpretation of Deterioration Alerts

Clinicians wanted to understand the alerts given by AI models (Q-37.1), wanting clinically meaningful AI explanations. Simply reporting a probability of deterioration was insufficient, as clinicians often trust their own judgement unless information is provided that complements or challenges their reasoning (Q-38.1).

While some clinicians valued patient-specific details (Q-39), others considered cohort-level outputs as providing more useful insights and context (Q-40). Preferences varied between those seeking simplicity (e.g., clear trends) (Q-41), and those wanting more in-depth, feature-based information (Q-42). Metrics of model confidence in its predictions were considered important (Q-43) for gauging reliability, and senior clinicians in particular desired supporting evidence about model performance from validation studies (Q-44).

Finally, explanations were seen as necessary for both real-time decision support and training and adoption (Q-45).

Sub-theme 5 – Suggesting priority of actions but not mandating them

Ranking patient risk can help separate high-risk from low-risk patients, thereby lessening alert fatigue and failure to recognise critical deterioration (Q-46). High-risk patients with life threatening conditions (Q-47), for which established treatments are available, receive priority (Q-48). However, the system should not mandate particular responses because different alerts may have similar levels of criticality (Q-49), and clinicians must decide which alert warrants immediate action (Q-50).

Sub-theme 6 - Responsive systems for real-world clinical complexity

Alert systems should facilitate model refinement and improved accuracy through continuous clinician feedback (Q-51), particularly when clinicians disagree with system outputs (Q-52 - Q-54). Any new AI-based tool should include adaptability to clinicians' intuitive, experience-based gestalt (Q-55), rather than relying solely on quantitative data from EMRs (Q-56) with their inherent technical and infrastructure limitations (Q-57).

Clinicians expressed contrary positions regarding personal customisation of the system (Q-58, Q-59), depending on their roles, desiring critical patient deterioration data calibrated to the level of expertise, while also optimising communication among different staff (Q-60), again highlighting the importance of adaptative UIs.

Sub-theme 7 – Single Watchtower vs Local Patrols

In a centralised system, alerts are directed to a dedicated team responsible for monitoring and escalation. In a distributed system, alerts go directly to clinicians caring for individual patients, with escalation managed locally. Clinicians were undecided about which system to adopt. Some nurses envisioned a centralised overview of patient alerts (Q-61), while some doctors preferred a mixed approach (Q-62). Doctors valued seeing outputs specific to their patients which facilitated response

coordination, streamlined workflows and time savings (Q-63). One doctor described a mixed system (Q-64) in which both nurses and doctors shared access to the same interface and could independently update alert status (e.g., from active to deactivated). A nurse (Q-65) cautioned that alerts sent only to the clinician on duty, who had to interpret prediction outputs, could create a single point of failure. In contrast, a mixed approach allows alerts to be managed collaboratively by a team with junior and more senior clinicians, an important safeguard given the “fear of speaking up” often experienced by junior clinicians. The method for delivering an alert can vary (Q-66) to a device carried by clinicians (Q-67) as one option, especially one with a dedicated UI (Q-68). However, staff will not actively interact with a device throughout all working hours, as duties vary, so the device would have to be shared or monitored by others. This requires the new system to support individual logins and logouts for each clinician, so that urgent calls can be managed more efficiently (Q-69). One nurse noted that having alerts delivered at the bedside using an audible sound made her more likely to respond (Q-70).

Sub-theme 8 - More than just accurate: a system that works for us!

This sub-theme covers UI features raised by participants not covered in previous sub-themes.

Alerts should be visible but not distracting (Q-71), understandable (Q-72), convey information clearly (Q-73), consider the patient’s clinical status (Q-74), be pertinent to the user (Q-75), and be amenable to customisation (Q-76). Moreover, each alert should include a ‘status’ indicator (e.g., ‘active’, ‘acknowledged’) to help clinicians track and manage them effectively (Q-77). In ward settings, multiple modalities (e.g., text, sounds, visual cues) (Q-78) could help nurses prioritise urgent cases and provide reassurance to patients, when they hear or see these alerts, that their own, less urgent needs will be addressed (Q-79).

Participants noted clinical deterioration tools can serve as a safety net (Q-80), particularly for junior staff (Q-81), while maintaining clinician control to select issues of most concern (Q-82).

The need for rigorous evaluation of system utility (Q-83) in building trust and supporting adoption (Q-84) was emphasised, with evidence of effectiveness preceding implementation (Q-85) in reducing resistance to adoption.

2.2.4. Theme 4 - The Ghost in the Machine

MO-EWS brings a fundamental tension between AI’s intended role as a clinical aid versus becoming an intrusive or unreliable agent. For some participants, MO-EWS risk fragmenting attention and creating cognitive overload (Q-86), alert fatigue (Q-87), desensitisation and eventual dismissal of potentially critical outputs (Q-88).

Overreliance on system outputs with little or no critical appraisal (Q-89) (i.e. automation bias), could inadvertently erode critical thinking skills, particularly among junior staff (Q-90). How clinicians with different expertise levels respond to AI recommendations raises concerns in that system performance is

only as effective in guiding care as its user's ability to consider and apply its outputs appropriately. This is more so if outdated or incomplete input data also risks generating flawed recommendations (Q-91).

Other concerns relate to how AI outputs might unintentionally distort priorities in care delivery, whether due to technical design limitations (Q-92), constraints in the wider healthcare system (Q-93), or the AI's inability to dynamically adapt to evolving clinical exigencies (Q-94).

2.2.5. Theme 5 – Ghost in the loop: The Role of User Training in Shaping the Machine

Successful adoption of AI in clinical settings depends on comprehensive, well-timed staff training (Q-95), involvement of clinical champions or 'super users' in tool design and training (Q-96), attention to generational differences in AI familiarity, from paper-based tools to emerging generative AI systems (Q-97), and positioning AI as an always-available support tool for less experienced clinicians dealing with urgent situations (Q-98). Training should occur prior to AI system implementation and be reinforced through in-service education, ensuring staff understand the system's purpose and benefits (Q-99).

2.2.6. Theme 6 - Beyond the ghost: the ecosystem matters

Beyond the AI algorithm, factors such as digital infrastructure, organisational policies, communication channels, and social dynamics critically influence the success of an AI system. Changed workflows due to system implementation must be standardised and accepted by clinicians, making this a clinical governance issue (Q-100, Q-105). In addition, a new clinical deterioration tool should be integrated into the current EMR platform in avoiding time consuming dispersal of outputs across different systems (Q-101) and ensuring easy and quick clinician access to information (Q-102). The AI system should aim to ameliorate communication breakdown between different staff (Q-103), and between staff and patients and their families (Q-104), while cognisant of constraints in the wider healthcare system (Q-106).

3. Mapping of themes to system components and deriving recommendations

Mapping these clinician-identified themes to system components of UI, AI model, and workflow [3] allowed derivation of recommendations for designing MO-EWS, summarised in Table 1 (acknowledging most of them can be useful for single output systems). In Table 2, we propose a categorisation of the feasibility of these different design options for each of the three system components based on consensus following discussion among all authors.

Table I Recommendations according to theme. Abbreviations: UI, User Interface, AI, Artificial Intelligence

UI	AI	Workflow
It cried wolf too many times but sometimes not at all		
Reduce alert noise through intuitive visualization of trends, patient-specific thresholds, and contextual cues that support clinician judgment or multiple AI-based prediction outputs.	Use adaptive thresholds, delta changes, precision medicine, and calibrated predictions to minimise false positives and better capture subtle deterioration.	Deliver timely, role-appropriate alerts while preserving opportunities for hands-on scrutiny of patient data.
Hunting for answers in a fractured system		
Integrate vitals, labs, documentation, and multiple predictions into a clean, role-adapted dashboard with trends and quick search functions related to the multiple prediction outputs	Enable cross-referencing of predictions with clinical data, provide search and filtering functions, and align predictions with clinicians' mental models to support rapid interpretation.	Embed predictions seamlessly into core EMR workflows, timed and tailored to clinicians' roles and responsibilities.
Navigating complexity: single vs multi-predictions		
Design layered, drill-down interfaces that allow clinicians to explore multiple condition-specific alerts without overwhelming the main view; support proactive exploration while minimising cognitive overload.	Provide multiple, detailed AI-based predictions that support differential diagnosis and augment clinician reasoning without priming towards specific conclusions; allow clinicians to filter, prioritise, or dismiss alerts as needed.	Integrate alert presentation into workflow in ways that prevent simultaneous alert overload, ensuring clinicians can respond effectively during busy shifts and maintain situational awareness.
From chasing the problems to catching them early		
Keep observations and trends visible in the interface to allow early recognition of deterioration; use dashboards that highlight subtle changes before alerts trigger.	Provide proactive, context-aware predictions to identify early signs of deterioration and enable preventive actions.	Support clinicians in initiating preventive care before deterioration occurs; integrate proactive alerts with centralised communication and hospital resources to improve efficiency and patient safety.
It is only useful if I can act of it		
Present alerts in actionable formats with clear severity ranking, prioritization, and role-specific guidance; allow on-demand querying and interactive exploration of patient data.	Generate predictions with actionable recommendations, urgency indicators, and clinical guidance tailored to existing protocols and staff expertise.	Integrate alerts into workflow to ensure timely responses, support role-appropriate actions (especially for juniors), and highlight priorities to prevent missed or delayed interventions.
More than a number: supporting interpretation of deterioration alerts		

UI	AI	Workflow
It cried wolf too many times but sometimes not at all		
Display AI explanations in a clear, interpretable way, offering both simple trends and deeper feature-based insights; indicate model confidence and provide supporting evidence to enhance trust.	Provide meaningful, patient-specific and cohort-level explanations that complement clinician reasoning; include metrics of prediction reliability and links to validation studies.	Integrate explanations into the clinical workflow to support real-time decisions, training, and adoption without slowing urgent actions.
Suggesting priority of actions but not mandating them		
Display alerts with clear prioritization cues for high-risk patients without forcing actions; use visual hierarchy to highlight urgency.	Prioritise high-risk predictions based on severity and likelihood of life-threatening conditions while allowing clinician discretion.	Support clinician decision-making by filtering and ranking patients according to risk, avoiding alert fatigue, and keeping clinicians in control of prioritization and actions.
Responsive systems for real clinical complexity		
Implement adaptive, role-specific user interfaces that allow clinicians to personalise alert displays and visualizations; enable intuitive interaction to provide feedback and additional data.	Incorporate clinician gestalt and experience-based judgment into AI predictions; allow AI models to learn from continuous clinician feedback to refine accuracy and relevance.	Maintain clinician control over choice of response; support communication among staff and integrate feedback loops to improve system performance and coordination in patient care.
Single watchtower vs local patrol		
Design interfaces that support both centralised and distributed alert viewing; provide login/logout flexibility and device-agnostic delivery; use auditory or visual cues to increase responsiveness.	Ensure AI alerts can be routed to multiple recipients, track acknowledgement of receipt, and support collaborative management; adapt alert prioritisation based on real-time context and user roles.	Enable collaborative alert management across teams to prevent single points of failure; integrate mixed centralised/distributed approaches for efficiency, coordination, and safeguarding junior clinicians.
More than just accurate: system that works for us!		
Design alerts that are visible but non-distracting, clear, understandable, and use multiple modalities (text, sound, visuals); include status indicators (active, acknowledged, dismissed); allow user customisation.	Prioritise alerts based on patient status and context; ensure relevance to the clinician receiving it; support safety-net functionality for new or junior staff.	Integrate alerts into clinical workflow as supportive tools rather than mandates; provide evidence of effectiveness to build trust and facilitate adoption; allow clinicians to select which alerts to act on while maintaining oversight of critical cases.
Theme 4 – The ghost in the machine		

UI	AI	Workflow
It cried wolf too many times but sometimes not at all		
Design interfaces that provide alert visibility and prioritisation but avoid overwhelming the user.	AI should provide interpretable, context-aware predictions, indicate confidence levels, and dynamically adapt to evolving patient status while mitigating alarm fatigue.	Workflow should preserve clinician critical thinking; system should support appropriate verification of AI predictions, prevent desensitisation, and account for varying expertise levels and real-world clinical constraints.
Theme 5 - Ghost in the loop: the role of user training in shaping the machine		
Interface should be intuitive and user-friendly to support learning; include guided tutorials, contextual help, and clear visualizations to reinforce understanding.	AI should act as an on-demand expert, offering real-time guidance, explanations, and actionable insights to bridge experience gaps.	Workflow should incorporate structured training programs, engage clinical champions, and provide ongoing reinforcement; consider generational differences in AI familiarity and ensure alignment with clinical reasoning.
Theme 6 – Beyond the ghost: the ecosystem matters		
The interface should integrate information from multiple sources into a single, coherent view, reducing fragmentation and enabling quick access.	AI can support communication by highlighting relevant alerts or updates to appropriate staff, while considering system constraints.	Standardised workflows, policies, and governance should be established before system design; integration into existing systems is crucial to streamline care and avoid inefficiencies.

Feasibility took into consideration aspects of UI, AI, and workflow that are technically moderately or highly feasible versus those constrained by hospital infrastructure and organisational factors. These recommendations serve as a checklist for co-design and prototyping sessions, ensuring tools are grounded in clinician experience while effectively leveraging AI capabilities.

Table 2
Assessment of feasibility of different design options for system components. Abbreviation: UI, User Interface, AI, Artificial Intelligence.

UI	AI	Workflow
High feasibility: Layered dashboards, status indicators, hierarchical structures, trend visualizations, search/filter functions, multi-modal alerts (text, sound, visual). These are mostly standard UI/UX practices and achievable with modern software development. **Moderate feasibility**: Fully adaptive and role-specific interfaces, drill-down views for multiple AI predictions, integrating explanations and model confidence dynamically. Requires significant development effort and close collaboration with clinicians. **Low feasibility**: Real-time integration of multiple AI prediction outputs while avoiding cognitive overload; presenting all relevant patient data without clutter for proactive monitoring. Technically complex, and financially challenging, especially for legacy/proprietary EMR systems.	**High feasibility**: Generating multiple condition-specific outputs, providing patient-specific and cohort-level insights, integrating confidence scores, supporting prioritization based on severity. Current ML/AI methods can achieve this. **Moderate feasibility**: Adaptive AI that learns continuously from clinician feedback, accounts for clinician gestalt, and dynamically updates predictions in real time. Conceptually possible but requires extensive validation and regulatory approval. **Low feasibility**: AI that perfectly mimics clinician reasoning, dynamically balances multiple alerts to prevent fatigue, and customises predictions per staff experience level. This is still cutting-edge and challenging to deploy safely.	**High feasibility**: Delivering alerts in centralised or distributed systems, supporting collaborative alert management, providing actionable recommendations and role-appropriate guidance, integrating alerts into existing workflows. **Moderate feasibility**: Seamless proactive monitoring before alerts trigger, integrating multi-source data into a single coherent workflow view, supporting on-demand querying. Requires workflow redesign and clinician input. **Low feasibility**: Personalising alert prioritisation across staff while maintaining safety and efficiency, dynamically adapting workflows in real time based on evolving patient states. Complex due to human, policy, and infrastructural constraints.

4. Discussion

This study explored clinicians' experiences with current and envisioned clinical deterioration tools, revealing a disconnect between the intended CDSS function and its real-world implementation [2, 8, 12, 18, 19]. A MO-EWS can help bridge this gap by offering context-sensitive, adaptive outputs rather than a single fixed deterioration score. By learning patient-specific baselines and trajectories, such systems can support nuanced decision-making while reducing false alarms. When paired with transparent UIs, co-designed workflows, and customisable levels of detail, they can minimise alert fatigue, avoid tool overreliance, and preserve critical thinking [20]. Continuous feedback loops [23, 24] further ensure predictions evolve with additional quantitative data and clinician input.

4.1. Single vs. Multiple AI-based Prediction Outputs

AI predictions with multiple versus single outputs are not inherently better. Senior clinicians (P01-D, 30y; P04-D, 37y) disagreed in their choice of either, showing experience shapes views on usefulness, risk, and cognitive load. Accordingly, MO-EWS should adopt a customisable interface supporting both single and multiple outputs as desired by clinicians. This can be achieved through progressive disclosure (providing a key prediction by default, with the option to expand into multiple outputs), adaptive presentation (tailoring detail to patient risk level or clinician role), and co-design with diverse clinical groups in ensuring alignment with varying workflows, cognitive needs, and decision-making styles.

4.2. Proactive systems

Doctors desire proactive systems, not just reactive alerts, that monitor trends, respond to queries, and disclose context-aware information before alerts trigger, supporting clinicians to act preventively. Displaying patient-specific trends and trajectories across multiple conditions and offering layered outputs beginning with a synthesised summary but allowing drill-down into detailed predictions when required, can all help manage informational complexity without adding cognitive burden [11]. Such systems can foster human–AI teaming [8], whereby AI is most effective when it reflects how clinicians think, not just delivers facts they need [23]. The “layered” interface metaphor aligns with iterative reasoning and differential diagnosis, especially when coupled with explanations that make clinical sense [8].

4.3. Situational awareness or tailored user experience

Combining centralised and distributed methods of alert delivery can enhance team communication and decision-making, user-centred personalisation of output presentation tailored to cognitive style, role, expertise, or setting [10] may further support workflow [11] but could create information silos and undermine multidisciplinary collaboration and shared situational awareness. A more standardised approach would support consistency, coordination, and safety in acute care. This tension reveals a design dilemma, reflecting differing assumptions about hierarchy, role boundaries, and knowledge parity. Effective CDSS design should try to accommodate both perspectives with role-aware and modifiable interfaces, adapted to user preferences where possible.

4.4. Authority or safety net

Some clinicians value AI as a safety net or second opinion, while others resist systems that dictate decisions or priorities and threaten clinician autonomy and accountability [24]. Safety concerns around clinicians overriding alerts highlights the need to balance automation with human judgment (“human-in-the-loop AI”) and take account of medico-legal liability if patients are harmed by inaccurate or misinterpreted AI systems. MO-EWS should offer context-aware recommendations while preserving clinicians’ autonomy to interpret, adapt, and act.

4.5. Reliability, integration and interpretability

System reliability and integration are critical since out-of-date data and fragmented systems can delay or distort decision-making. Cognitive ergonomics and system design aligned with how clinicians think and

act under pressure are required. Clear, timely and easily interpretable predictions delivered on the appropriate platform [25] are also essential for trust and adoption [8]. Failing to prioritise cognitive usability over computational precision can lead to alert fatigue, overload, and mistrust [22].

Regarding feasibility, most UI and AI features are technically achievable with existing software development and machine learning methods. Major bottlenecks lie in workflow integration within hospital settings. High-level UI/AI capabilities risk being undermined if alerts, data, and predictions cannot be seamlessly integrated into clinical workflows. Grading of feasibility depends not only on technical UI/AI advances but also on the level of hospital digital infrastructure, organisational policies, and clinician preferences. Legacy EMRs and proprietary systems, associated with poor interoperability and organisational inertia, make MO-EWS implementation particularly challenging.

5. Study implications

Clinician adoption of EWS rests on meeting diverse, often subjective expectations [26]. Even well-designed tools may fail if expectations are unrealistic and not addressed through iterative, user-centred design and evaluation. Preserving autonomy and shared mental models requires careful design trade-offs, adaptive UIs, and frameworks centred on clinical reasoning needs rather than accuracy alone. Designers may underestimate workflow disruption and alert burden; clinician–system dynamics should be understood, not treated as a straightforward technical problem to fix. This study proposes an opportunity for clinicians to compare perspectives, invites developers to collaborate with medical staff and organisations, and encourages researchers to use the identified themes as an investigative checklist that ensures EWS accuracy is not prioritised at the expense of usability and workflow integration.

6. Limitations

This study is limited by reliance on self-reported experiences from volunteer clinicians in specific hospitals with particular EMR systems, which brings selection bias and may not generalise to other contexts. The 22 participants constrained subgroup exploration, limiting transferability of themes, although this is not a methodological weakness in RTA. Negative experiences may be overrepresented, introducing potential bias against beneficial redesign. Finally, conducting RTA from a single positional perspective may limit interpretive diversity, although clinician authors [name deleted to maintain the integrity of the review process] with expertise in deterioration reviewed the analyses.

5. Conclusion

This study of clinician perspectives on current EWTs and future MO-EWS highlights persistent challenges in optimising CDSS. Effective solutions require real-time monitoring, responsive queries, user-centred modelling, and calibrated predictions, without violating cognitive usability principles[19]. More AI does not guarantee better support; interfaces must fit clinicians' mental models and workflows so the system,

not the user, carries the burden of sense-making. Even small design–practice mismatches risk unsafe decisions, missed detections, or delayed responses.

References

1. Tun HM, Rahman HA, Naing L, Malik OA (Dec. 2024) Trust in AI-Based Clinical Decision Support Systems Among Healthcare Workers: A Systematic Review (Preprint). J Med Internet Res. 10.2196/69678
2. Yuan S, Yang Z, Li J, Wu C, Liu S (June 2025) AI-Powered early warning systems for clinical deterioration significantly improve patient outcomes: a meta-analysis. BMC Med Inf Decis Mak 25(1):203. 10.1186/s12911-025-03048-x
3. Van Der Vegt AH et al (Jan. 2024) Systematic review and longitudinal analysis of implementing Artificial Intelligence to predict clinical deterioration in adult hospitals: what is known and what remains uncertain. J Am Med Inf Assoc 31(2):509–524. 10.1093/jamia/ocad220
4. Noselli M et al (2025) Presenting Artificial Intelligence Predictions Based on Electronic Medical Records to Clinicians in Hospitals: A Systematic Review, *IEEE J. Biomed. Health Inform.*, vol. 29, no. 10, pp. 7619–7632, Oct. 10.1109/JBHI.2025.3572494
5. Jayawardena T, Baysari M, Bamgboje-Ayodele A (2025 Apr) Interface design features of clinical decision support systems for real-time detection of deterioration: A scoping review. Int J Med Informatics 22:105946
6. Verma AA (Oct. 2024) Toward the Rigorous Evaluation of Early Warning Scores. JAMA Netw Open 7(10):e2438966. 10.1001/jamanetworkopen.2024.38966
7. Cánovas-Segura B, Morales A, Juarez JM, Campos M (July 2023) Meaningful time-related aspects of alerts in Clinical Decision Support Systems. A unified framework. J Biomed Inf 143:104397. 10.1016/j.jbi.2023.104397
8. Blythe R et al (Sept. 2024) Clinician perspectives and recommendations regarding design of clinical prediction models for deteriorating patients in acute care. BMC Med Inf Decis Mak 24(1):241. 10.1186/s12911-024-02647-4
9. Owoyemi A, Okpara E, Salwei M, Boyd A (Oct. 2024) End user experience of a widely used artificial intelligence based sepsis system. JAMIA Open 7(4):ooae096. 10.1093/jamiaopen/ooae096
10. Zhang S et al (2024) Rethinking Human-AI Collaboration in Complex Medical Decision Making: A Case Study in Sepsis Diagnosis, in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, May pp. 1–18. 10.1145/3613904.3642343
11. Butler JM et al (2025) 'Be Really Careful about That': Clinicians' Perceptions of an Intelligence Augmentation Tool for In-Hospital Deterioration Detection, *Appl. Clin. Inform.*, vol. 16, no. 02, pp. 377–392, Mar. 10.1055/a-2505-7743
12. Malterud K, Siersma VD, Guassora AD (2016) Sample Size in Qualitative Interview Studies: Guided by Information Power, *Qual. Health Res.*, vol. 26, no. 13, pp. 1753–1760, Nov.

10.1177/10497323315617444

13. Braun V, Clarke V (2024) Supporting best practice in reflexive thematic analysis reporting in *Palliative Medicine* : A review of published research and introduction to the *Reflexive Thematic Analysis Reporting Guidelines* (RTARG), *Palliat. Med.*, vol. 38, no. 6, pp. 608–616, June 10.1177/02692163241234800
14. Byrne D (June 2022) A worked example of Braun and Clarke's approach to reflexive thematic analysis. Qual Quant 56(3):1391–1412. 10.1007/s11135-021-01182-y
15. Azeem M, Salfi NA, USAGE OF NVIVO SOFTWARE FOR QUALITATIVE DATA ANALYSIS (2012), vol. 2, no. 1
16. Ammenwerth E (2019) Technology Acceptance Models in ealth nformatics: TAM and UTAUT. Stud Health Technol Inf 263:64–71. 10.3233/SHTI190111
17. Tong A, Sainsbury P, Craig J (2007) Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups, *Int. J. Qual. Health Care*, vol. 19, no. 6, pp. 349–357, Sept. 10.1093/intqhc/mzm042
18. Campbell V et al (Aug. 2020) Predicting clinical deterioration with Q-ADDS compared to NEWS, Between the Flags, and eCART track and trigger tools. Resuscitation 153:28–34. 10.1016/j.resuscitation.2020.05.027
19. Cho H et al (May 2022) Assessing the Usability of a Clinical Decision Support System: Heuristic Evaluation. JMIR Hum Factors 9:e31758. no. 210.2196/31758
20. Abdulnour R-EE, Gin B, Boscardin CK (2025) Educational Strategies for Clinical Supervision of Artificial Intelligence Use, *N. Engl. J. Med.*, vol. 393, no. 8, pp. 786–797, Aug. 10.1056/NEJMra2503232
21. Weigand L, Viers T, Tipton E (2025) Integration of Rapid Response Teams and Early Warning Systems to Reduce Cardiac Arrests and Intensive Care Unit Readmissions, *Crit. Care Nurse*, vol. 45, no. 4, pp. 49–56, Aug. 10.4037/ccn2025131
22. Gandhi TK et al (July 2023) How can artificial intelligence decrease cognitive and work burden for front line practitioners? JAMIA Open 6(3):ooad079. 10.1093/jamiaopen/ooad079
23. Wang L et al (June 2023) Human-centered design and evaluation of AI-empowered clinical decision support systems: a systemtic review. Front Comput Sci 5:1187299. 10.3389/fcomp.2023.1187299
24. Elgin CY, Elgin C (Dec. 2024) Ethical implications of AI-driven clinical decision support systems on healthcare resource allocation: a qualitative study of healthcare professionals' perspectives. BMC Med Ethics 25(1):148. 10.1186/s12910-024-01151-8
25. Sirajuddin AM, Osheroff JA, Sittig DF, Chuo J, Velasco F, Collins DA (2012) Implementation Pearls from a New Guidebook on Improving Medication Use and Outcomes with Clinical Decision Support:
26. Raita E, Oulasvirta A (July 2011) Too good to be bad: Favorable product expectations boost subjective usability ratings. Interact Comput 23(4):363–371. 10.1016/j.intcom.2011.04.002

Box

Box is available in the Supplementary Files section.

Figures

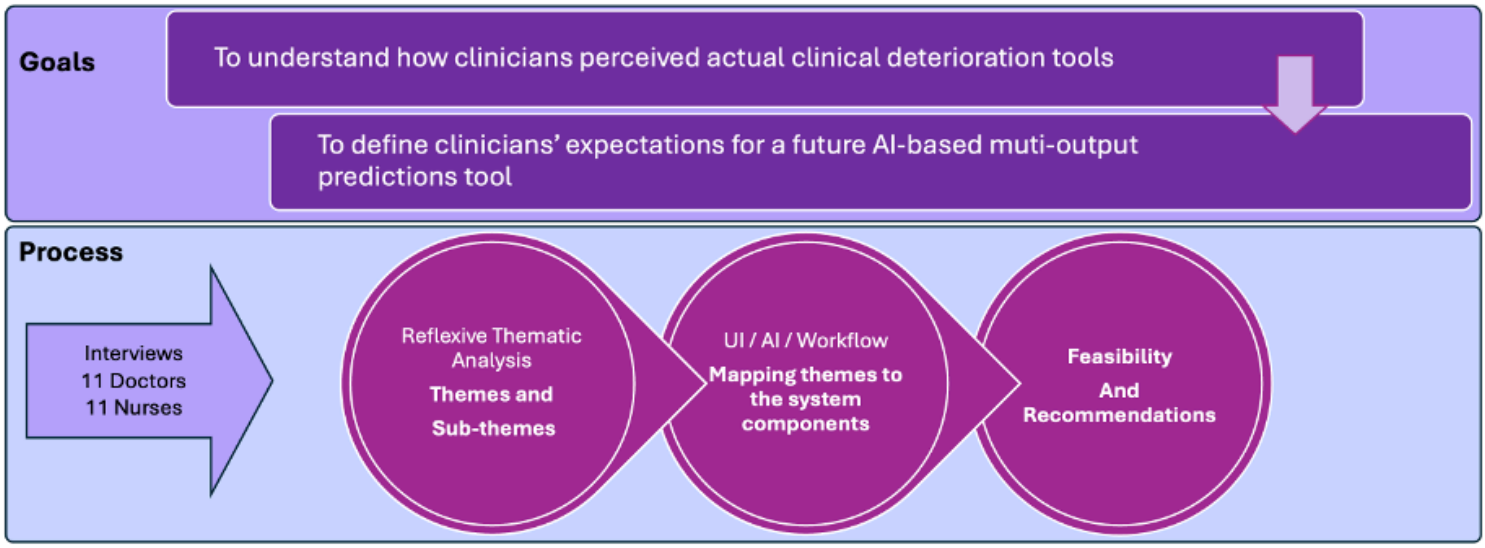

Figure 1

Study goals and processes

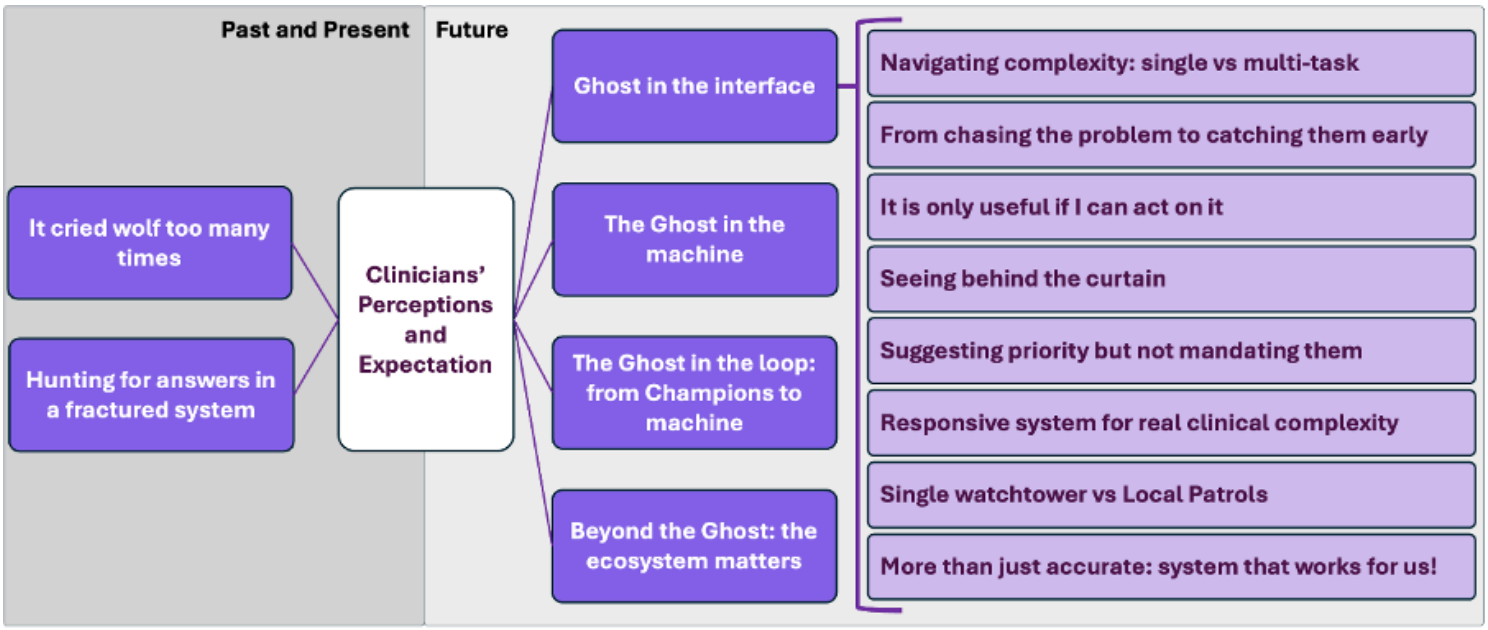

Figure 2

Themes and subthemes derived from interviews on clinicians' perceptions and expectations

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RTACOREQChecklist.pdf
- SupplementaryTableofParticipantsQuotesAnonymised.docx
- box1.png