nature medicine

Supplementary information

https://doi.org/10.1038/s41591-025-03609-7

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

In the format provided by the authors and unedited

Supplementary Table 1. Reporting of trial findings according to the CONSORT guidelines and extensions for AI, cluster-randomized trials, and pragmatic trials

Section/Topic	Item No.	CONSORT 2010 item	Extension for AI studies	Extension for cluster randomized trials	Extension for pragmatic trials	Page No. (Section)
Title and abstract						
Title and abstract	1a	Identification as a randomised trial in the title	CONSORT-AI 1a,b Elaboration: Indicate that the intervention involves artificial intelligence/machine learning in the title and/or abstract and specify the type of model	Identification as a cluster randomised trial in the title		Title page; Abstract
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	State the intended use of the AI intervention within the trial in the title and/or abstract	See extension for CONSORT-abstracts		Abstract
Introduction						
Background and objectives	2a	Scientific background and explanation of rationale	CONSORT-AI 2a (i) Extension Explain the intended use of the AI intervention in the context of the clinical pathway, including its purpose and its intended users (for example, healthcare professionals, patients, public).	Rationale for using a cluster design	Describe the health or health service problem that the intervention is intended to address and other interventions that may commonly be aimed at this problem	1.Main; sMethods1

	2b	Specific objectives or hypotheses		Whether objectives pertain to the cluster level, the individual participant level, or both		1.Main
Methods						
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio		Definition of clusters and description of how the design features apply to the clusters		4.1 Study Design, Trial Sites, and Randomization
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons				4.1 Study Design, Trial Sites, and Randomization; sMethods1
Participants	4a	Eligibility criteria for participants	CONSORT-AI 4a (i) Elaboration : State the inclusion and exclusion criteria at the level of participants. CONSORT-AI 4a (ii) Extension State the inclusion and exclusion criteria at the level of the input data.	Eligibility criteria for clusters		4.1 Study Design, Trial Sites, and Randomization; 4.2 Trial Participants and Outcomes
	4b	Settings and locations where the data were collected	CONSORT-AI 4b Extension Describe how the AI intervention was integrated into the trial setting, including any onsite or offsite requirements.			4.1 Study Design, Trial Sites, and Randomization; 4.2 Trial Participants and Outcomes
Interventions	5	The interventions for each group with sufficient details to	CONSORT-AI 5 (i) Extension	Whether interventions pertain to the cluster level,	Describe extra resources added to (or resources	4.1 Study Design, Trial Sites, and Randomization;

		allow replication, including how and when they were actually administered	State which version of the AI algorithm was used. CONSORT-AI 5 (ii) Extension Describe how the input data were acquired and selected for the AI intervention. CONSORT-AI 5 (iii) Extension Describe how poor quality or unavailable input data were assessed and handled. CONSORT-AI 5 (iv) Extension Specify whether there was human–AI interaction in the handling of the input data, and what level of expertise was required of users. CONSORT-AI 5 (v) Extension Specify the output of the AI intervention CONSORT-AI 5 (vi) Extension Explain how the AI intervention's outputs contributed to decision-making or other elements of clinical practice.	the individual, participant level, or both	removed from) usual settings in order to implement intervention. Indicate if efforts were made to standardise the intervention or if the intervention and its delivery were allowed to vary between participants, practitioners, or study sites Describe the comparator in similar detail to the intervention	4.3 CONCERN EWS Intervention) eFigure 1, eTable2, eTable 3, sMethods2

Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed		Whether outcome measures pertain to the cluster level, the individual participant level, or both	Explain why the chosen outcomes and, when relevant, the length of follow-up are considered important to those who will use the results of the trial	4.2 Trial Participants and Outcomes; 4.4 Statistical Analysis
	6b	Any changes to trial outcomes after the trial commenced, with reasons				sMethods1
Sample size	7a	How sample size was determined		Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intracluster correlation (ICC or k), and an indication of its uncertainty	If calculated using the smallest difference considered important by the target decision maker audience (the minimally important difference) then report where this difference was obtained	4.4 Statistical Analysis; eTable4
	7b	When applicable, explanation of any interim analyses and stopping guidelines				N/A
Randomization						
Sequence generation	8a	Method used to generate the random allocation sequence				4.1 Study Design, Trial Sites, and Randomization

	8b	Type of randomisation; details of any restriction (such as blocking and block size)		Details of stratification or matching if used		
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned		Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level, or both		4.1 Study Design, Trial Sites, and Randomization
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions		Replaced by 10a, 10b, and 10c		4.1 Study Design, Trial Sites, and Randomization
	10a			Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions		41 Study Design, Trial Sites, and Randomization
	10b			Mechanism by which individual participants were included in clusters for the purposes of the trial (such as		4.1 Study Design, Trial Sites, and Randomization

				complete enumeration, random sampling)		
	10c			From whom consent was sought (representatives of the cluster, or individual cluster members, or both) and whether consent was sought before or after randomisation		4.1 Study Design, Trial Sites, and Randomization
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how			If blinding was not done, or was not possible, explain why	4.1 Study Design, Trial Sites, and Randomization
	11b	If relevant, description of the similarity of interventions				N/A
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes		How clustering was taken into account		4.4 Statistical Analysis; sMethods1
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses				N/A

Results						
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome		For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome	The number of participants or units approached to take part in the trial, the number which were eligible, and reasons for non-participation should be reported	2.1 Trial Participants and Hospital Encounters; eTable1
	13b	For each group, losses, and exclusions after randomisation, together with reasons		For each group, losses and exclusions for both clusters and individual cluster members		2.1 Trial Participants and Hospital Encounters
Recruitment	14a	Dates defining the periods of recruitment and follow-up				4.1 Study Design, Trial Sites, and Randomization sMethods1
	14b	Why the trial ended or was stopped				N/A
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group		Baseline characteristics for the individual and cluster levels as applicable for each group		2.1 Trial Participants and Hospital Encounters; eTable1, eTable 5, eTable 6
Numbers analyzed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups		For each group, number of clusters included in each analysis		2.1 Trial Participants and Hospital Encounters

Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)		Results at the individual or cluster level as applicable and a coefficient of intracluster correlation (ICC or k) for each primary outcome		2.2 Co-Primary Outcomes; 2.3 Secondary Outcomes
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended				2.2 Co-Primary Outcomes; 2.3 Secondary Outcomes
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory				2.2 Co-Primary Outcomes; eTable7
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	CONSORT-AI 19 Extension Describe results of any analysis of performance errors and how errors were identified, where applicable. If no such analysis was planned or done, justify why not.			eMethods2
Discussion						
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant,				3. Discussion

		multiplicity of analyses				
Generalizability	21	Generalisability (external validity, applicability) of the trial findings		Generalisability to clusters and/or individual participants (as relevant)	Describe key aspects of the setting which determined the trial results. Discuss possible differences in other settings where clinical traditions, health service organisation, staffing, or resources may vary from those of the trial	3. Discussion
Interpretation	22	Interpretation consistent with results, balancing benefits, and harms, and considering other relevant evidence				3. Discussion
Other Information						
Registration	23	Registration number and name of trial registry				Title page
Protocol	24	Where the full trial protocol can be accessed, if available				Rossetti SC, Dykes PC, Knaplund C, et al. The Communicating Narrative Concerns Entered by Registered Nurses (CONCERN) Clinical Decision Support Early Warning System: Protocol for a Cluster Randomized Pragmatic Clinical Trial.

						JMIR Res Protoc. 2021;10(12):e30238. doi:10.2196/30238
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	CONSORT-AI 25 Extension State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use.			Title Page; Funding/Support; Data Sharing Statement

Supplementary Methods 1: Summary of Modifications from Original to Final Analysis Plan

Our original study protocol is published here[3]. Below we describe all changes made in our study, all of which were modifications within our analysis plan, and rationale for those changes. We note that the trial's cluster randomized design, two study sites, primary and secondary outcomes, and enrollment eligibility remained the same as originally proposed.

We originally planned to use generalized linear mixed models for an interrupted time series analysis with six-month pre-intervention data at the clinical unit level of analysis. However, in our original statistical plan we also specified analyses that we would conduct at the individual patient encounter level. As noted in the main manuscript, since a significant proportion of patients moved across different units during their hospital stay, it was not possible to conduct analyses at the clinical unit level. Instead, we analyzed all outcomes at the individual patient level per hospital encounter. While we did remove the unit level analysis (interrupted time series) from the final analyses, the rest of our final analytical plan is not different from that in the original proposal. Because we analyzed at the individual patient encounter level, inclusion of pre-intervention data was not needed in the analysis. The original analysis plan specified use of generalized linear mixed model for LOS (Poisson models) and readmission (logit link function) and survival models for other outcomes, which is consistent with our final analytical plan. Our trial sample size was based on our power analysis for a cluster-randomized controlled clinical trial, which was not changed; details of our power analysis are provided in the main manuscript. The power calculation was still appropriate for our final analysis plan at the individual patient encounter level.

The original published protocol also allowed for continuous refinement of our machine learning algorithm, pushing out a new version every 3 months and tracking outcomes for each version analyzed using a multiple time-series analysis. This was proposed due to the possibility that nursing documentation patterns may change overtime in response to the use of a novel algorithm that is based on real-time documentation patterns. Our Advisory Board recommended that we perform an interim analysis at our Study Site A (go-live 2020) to confirm model performance before our Study site B go-live in 2021 to inform if model versioning was needed or appropriate. Based on the interim analysis we concluded that no updates were needed to the model, therefore the same version of the model was used during the entirety of the trial at both study sites. Given the same version of the model was tested during the entirety of the trial a multiple time-series intervention analysis was no longer necessary.

Supplementary Methods 2: Detailed Description of CONCERN Early Warning System (EWS) Intervention

CONCERN EWS Prediction: Predictive Model Output as Early Warning Score of Patient Risk

Our risk prediction model (version 1.5) is nonlinear, with time-varying variables (e.g., vital signs recorded at variable times within a clinical shift). Every hour, the CONCERN EWS processes EHR data from the past 24 hours through the ensemble models that best fit the patient characteristics, calculating a risk score (green, yellow, or red). See eFigure1 for conceptual modeling approach and eTable3 for model factors and features. CONCERN EWS' conceptual modeling approach and model factors can be summarized as reflecting patterns of increased nurses' surveillance above and beyond the standard of care and resultant nursing interventions consistent with an observed change in the patient's clinical state. Increased nurses' surveillance in our model, for example, may be measured when a nurse assesses vital signs every 2 hours for a specific patient displaying subtle changes even though the clinical unit policy only requires assessment every 6 hours. As an example of a nursing intervention consistent with observed but subtle changes in the patient's clinical state, a nurse may decide not to administer a scheduled medication, such as metoprolol, because the patient's heart rate is hovering around 60 beats per minute and their usual baseline is around 90 beats per minute. While medications are ordered by a prescribing provider in the hospital setting, the administration is completed by a nurse and requires the nurse to assess the appropriateness of that medication given the patient's condition prior to administration. Finally, our time-varying modeling approach also considers if nursing assessments or interventions are performed at non-common times, based on our data-driven analyses. The CONCERN EWS includes a monitoring function to check that all variables are available to the model and that the volume of data for each variable is consistent with historical volumes, with automated emailed notification of issues. The CONCERN EWS has robust error reporting and logging functionalities for each step in our data pipeline, including data inputs, pre-processing, prediction generation, and CONCERN score writing to the EHR. A system performance error was defined as the automated prediction score not being generated every hour for eligible patients. The few times that automated prediction did not occur during our trial were due to a web service error. The research team was alerted and able to manually trigger the prediction processes for all patients.

Approaches to evaluate the predictive accuracy of a nonlinear, time-varying model must consider the clinical problem, data inputs, and model features. See eTable2 for Performance Metrics for CONCERN Predictive Model. Through experimentation on retrospective data sets of hospitalized patients' encounters in the EHR, we refined our machine learning modeling approach (see eFigure1), which is focused on temporal patterns of nursing observations, using ensemble learning to identify the best performing models to predict clinical deterioration in the next 24 hours, for a patient's specific clinical situation. For our model building, patient deterioration was defined as the first occurrence of unanticipated transfer to the ICU, sepsis, mortality, or cardiopulmonary arrest. For example, ensemble learning allowed us to identify the best model for a patient who has been in an ICU for 5 hours on a Friday night and to make our EWS transparent and explainable. We are not aware of any other decision-support implementations of EWSs that use a similar multi-factorial, advanced computational, *and* explainable modeling approach.[2]

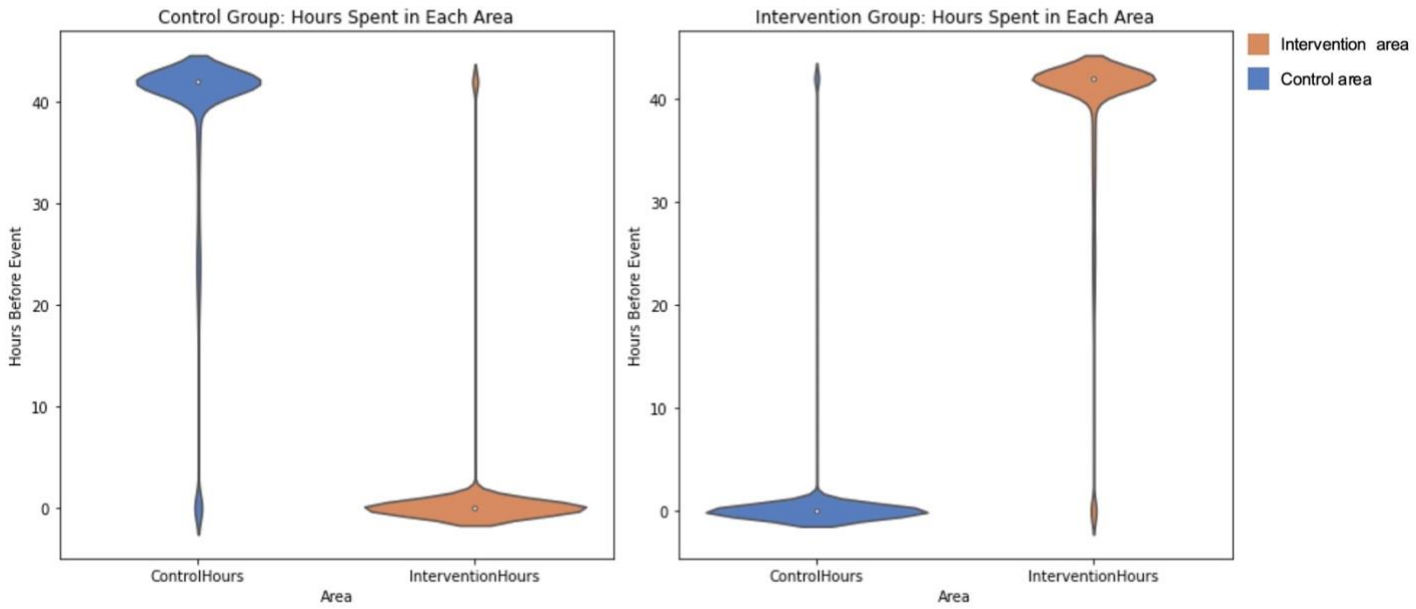
CONCERN EWS Display: Clinical Decision Support in the EHR to Increase Team Situational Awareness of Patient Risk

Clinical decision support functionality was developed based on substantial user-centered design sessions and simulation evaluations and was integrated into the EHR using the FHIR (Fast Healthcare Interoperability Resources) standard for exchanging healthcare information electronically.[2] The Clinical Decision Support displayed the early warning score (green, yellow, red) in the EHR patient list for all nurses and prescribing providers (physicians and advanced practice providers) on the clinical unit. Displaying the score in the EHR patient list meant that the score was visible on nurses' and prescribing providers' log-in screen and part of the typical EHR workflow when selecting a new patient chart to view. Nurses and prescribing providers could double-click the score icon linked to a screen with detailed information about the prediction (Figure 1)[3], including the factors that determined a patient's score, a score trendline adjustable up to 72 hours, and a scale

displaying the patient's score relative to other hospitalized patients. Users could also click on the factors that determined the score to view the clinical data associated with that factor. The factors are comprised of one or more granular CONCERN EWS features (e.g., vital sign frequency factor includes measurement patterns related to each type of vital sign).

CONCERN EWS Intervention Implementation Activities

As part of the intervention to ensure hospital-level support and end-user awareness of the score and information displayed within the EHR, we created robust stakeholder engagement and end-user training processes and materials. As with most information technology implementations in the clinical setting, early and frequent engagement sessions were held with stakeholders at all levels within the hospital, including nurse and physician leadership, registered nurses, prescribing providers, clinical champions, analysts, technicians, and other hospital leaders. We also created training materials and training processes for super-users and end-users (nurses and prescribing providers) such as presentations, pocket cards, and posters that conveyed information about the EWS score and its display within the EHR.



Supplementary Figure 1. Violin plot of intervention vs usual care patient's time spent in each area 42 hours before deterioration event (N=60,893)

References Cited in Supplement

1. Opitz D, Maclin R. Popular ensemble methods: an empirical study. *J Artif Intell Res*. July 1999. doi:10.5555/3013545.3013549
2. Rossetti S, Knaplund C, Albers A, et al. Healthcare Process Modeling to Phenotype Clinician Behaviors for Exploiting the Signal Gain of Clinical Expertise (HPM-ExpertSignals): Development and Evaluation of a Conceptual Framework. *JAMIA*. 2021;28(6):1242-1251.
3. Rossetti SC, Dykes PC, Knaplund C, et al. The Communicating Narrative Concerns Entered by Registered Nurses (CONCERN) Clinical Decision Support Early Warning System: Protocol for a Cluster Randomized Pragmatic Clinical Trial. *JMIR Res Protoc*. 2021;10(12):e30238. doi:10.2196/30238
4. Shcherbatykh I, Holbrook A, Thabane L, Dolovich L, COMPETE III investigators. Methodologic issues in health informatics trials: the complexities of complex interventions. *J Am Med Inform Assoc*. 2008;15(5):575-580. doi:10.1197/jamia.M2518