

Protocol

Clinical Effectiveness of AI-Based Cardiac Arrest Prediction System for General Ward Patients (VITAL-MD): Protocol for a Multicenter Pragmatic Stepped Wedge Cluster Randomized Controlled Trial

Mincheol Kim^{1,2}, BSc; Kyung Hoon Kim³, MD, PhD; Minsoo Kim¹, MSc; Sun Jung Kwon⁴, MD, PhD; Myung Soo Park⁵, MD; Ah Jin Kim⁶, MD, PhD; Man-Jong Lee⁶, MD, MSc; Mi Hwa Park⁶, MD, MSc; Hyung Keun Cha⁶, MD; Minhyeok Lee⁴, MD; In Beom Jeong⁴, MD; Soon Joo Wang⁷, MD; Sung Bak An⁸, MD; Jeon Hwa Lee⁵, MD, PhD; Kyung-Jae Cho¹, MSc; Jinhui Jang⁹, RN, MSc; HyoJin Jeong¹⁰, RN; Yoon Young Lee¹¹, RN; Woori Han¹², RN; VITAL-MD Study Group⁶; Jung Soo Kim⁶, MD, MSc; Dongjoon Yoo⁶, MD

¹VUNO, Seoul, Republic of Korea

²Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea

³Department of Internal Medicine, Division of Pulmonary and Critical Care Medicine, College of Medicine, Incheon St. Mary's Hospital, The Catholic University of Korea, Incheon, Republic of Korea

⁴Department of Internal Medicine, Division of Pulmonology and Critical Care Medicine, College of Medicine, Konyang University Hospital, Konyang University, Daejeon, Republic of Korea

⁵Department of Cardiology, College of Medicine, Dongtan Sacred Heart Hospital, Hallym University, Hwaseong, Gyeonggi-do, Republic of Korea

⁶Department of Critical Care Medicine, Inha University Hospital, Incheon, Republic of Korea

⁷Department of Emergency Medicine, College of Medicine, Dongtan Sacred Heart Hospital, Hallym University, Hwaseong, Republic of Korea

⁸Department of General Surgery, College of Medicine, Dongtan Sacred Heart Hospital, Hallym University, Hwaseong, Republic of Korea

⁹Department of Hospital Medicine, Critical Care Support Part, Inha University Hospital, Incheon, Republic of Korea

¹⁰Department of Nursing, Rapid Response Team, Konyang University Hospital, Daejeon, Republic of Korea

¹¹Department of Nursing, Incheon St. Mary's Hospital, The Catholic University of Korea, Incheon, Republic of Korea

¹²Rapid Response Team, Dongtan Sacred Heart Hospital, Hwaseong, Republic of Korea

Corresponding Author:

Dongjoon Yoo, MD
Department of Critical Care Medicine
Inha University Hospital
27 Inhang-ro, Jung-gu
Incheon 22332
Republic of Korea
Phone: 82 32-890-2114
Fax: 82 32-890-2114
Email: djinyoo@gmail.com

Abstract

Background: In-hospital cardiac arrest (IHCA) remains a frequent and devastating event, and conventional early warning scores (EWSs) coupled with rapid response systems (RRSs) have shown inconsistent effects on patient-centered outcomes, largely owing to limited predictive accuracy and high false-alarm burden. AI-based EWSs (AI-EWSs) have demonstrated superior discrimination retrospectively, but high-quality prospective evidence in true external validation environments remains scarce. Existing AI-EWS trials have either been confined to internal-development settings or bundled the algorithm with specific implementation infrastructure, leaving the stand-alone effect of AI-based software as a medical device (AI-SaMD) on patient outcomes uncharacterized.

Objective: This trial will evaluate whether implementation of AI-SaMD (VUNO Med-DeepCARS) as a stand-alone clinical decision-support tool in routine general-ward care reduces IHCA compared with usual care based on conventional EWSs.

Methods: The VITAL-MD trial (validation of an AI-based clinical deterioration prediction model on patient outcomes through a stepped-wedge cluster randomized trial [SW-CRT] multicenter evaluation of DeepCARS) is a pragmatic, multicenter, SW-CRT conducted at 4 tertiary academic hospitals in the Republic of Korea. Six tertiary academic hospitals were assessed for eligibility based on inpatient capacity, RRS operating hours aligned with the national RRS pilot program, and no prior AI-SaMD use. All adult patients aged 19 years or older admitted to general wards during the 12-month trial period are included; patients younger than 19 years, or with no documented vital signs throughout hospitalization, are excluded. The 4 sites are randomized to sequentially transition from preimplementation (usual care) to postimplementation (AI-SaMD-guided care) in 4 steps separated by 7-day washout periods. DeepCARS generates real-time risk scores from 4 vital signs and integrates into the existing electronic medical record interface without altering RRS workflows. The primary outcome is the incidence of general-ward IHCA, defined per the Utstein guidelines, among the target cohort (defined as patients exceeding the alert threshold at least once). Assuming a baseline IHCA incidence of 6.70%, a 40% relative risk reduction, an intracluster correlation coefficient of 0.01, and 350 patients per cluster-period, 7002 target-cohort patients (total enrollment approximately 72,695) provide 80% power. Analyses follow the intention-to-treat principle using generalized linear mixed models adjusting for prespecified covariates, with prespecified sensitivity, subgroup, and compliance analyses.

Results: The trial was registered with the Clinical Research Information Service (KCT0010243) on February 28, 2025. Recruitment was conducted from April 1, 2025, to March 31, 2026. All 4 sites completed transition to the postimplementation period, and data cleaning is currently underway before the final database lock, with primary results expected in 2027.

Conclusions: VITAL-MD will provide the first large-scale, multicenter randomized evidence evaluating whether an AI-EWS, deployed as a stand-alone device in external validation environments and benchmarked against real-world standard-of-care practice, improves patient-centered outcomes for general-ward patients, advancing the evaluation of AI-SaMD from retrospective diagnostic accuracy toward prospective, action-anchored clinical utility.

Trial Registration: Clinical Research Information Service KCT0010243; <https://tinyurl.com/4bd3prps>

International Registered Report Identifier (IRRID): DERR1-10.2196/104230

JMIR Res Protoc 2026;15:e104230; doi: [10.2196/104230](https://doi.org/10.2196/104230)

Keywords: rapid response system; rapid response team; early warning score; artificial intelligence; deep learning; clinical deterioration; clinical decision support systems; patient-centered outcome; randomized controlled trial; in-hospital cardiac arrest

Introduction

In-hospital cardiac arrest (IHCA) remains a frequent and devastating event among hospitalized patients. In the United States alone, approximately 292,000 patients experience IHCA annually [1]. Globally, the reported incidence in adults ranges from 1.2 to 10 cases per 1000 admissions, and although survival to hospital discharge has improved over time, long-term prognosis remains poor; a systematic review and meta-analysis estimated 1-year survival after IHCA at 13.4% [2].

Historically, hospital patient safety initiatives focused predominantly on the “efferent limb” of the response, optimizing code teams to react to arrests. Over the past 2 decades, the paradigm has shifted toward the “afferent limb,” emphasizing the proactive detection and prevention of clinical deterioration through early warning scores (EWS) and rapid response systems (RRSs) [3]. Multiple national and international bodies, including the American Heart Association (AHA) and the Society of Critical Care Medicine (SCCM), strongly recommend EWSs and hospital-wide surveillance to trigger RRS interventions in general wards [4-7].

Despite these broad recommendations, the evidence supporting the implementation of conventional EWSs and RRSs in improving patient-centered outcomes remains limited and inconsistent. The latest AHA guidelines highlight

a stark paucity of high-quality adult data, relying primarily on pediatric randomized controlled trials (RCTs) to substantiate EWS and RRS recommendations [4]. Furthermore, a recent Cochrane systematic review concluded that implementing conventional EWS and RRS on adult hospital wards may make little to no difference in key outcomes, such as IHCA incidence and in-hospital mortality [8]. A primary driver of this inconsistency is the suboptimal predictive accuracy of conventional track-and-trigger EWS tools, such as the National Early Warning Score 2 (NEWS2) and Modified Early Warning Score (MEWS). These scoring systems often have low sensitivity and positive predictive value, generating excessive numbers of nonactionable false-positive alerts. This overwhelms clinicians, causes alarm fatigue, and misdirects scarce critical care resources, while true episodes of atypical deterioration may still go unrecognized [9,10].

To overcome these critical limitations, AI and machine learning (ML) models have been actively developed. Retrospectively, AI-based EWSs consistently demonstrate superior discrimination capabilities (eg, area under the receiver operating characteristic curve [AUROC]) compared to conventional scores. However, exceptional algorithmic performance does not automatically translate into clinical utility, as emphasized by the Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by AI (DECIDE-AI) guidelines and recent statements from the AHA and the *Journal of the American Medical*

Association (JAMA), calling for evaluations using robust, patient-centered end points [11-13].

Indeed, the prospective clinical effectiveness of AI-based EWSs (AI-EWSs) in RCTs remains uncertain. For instance, the Continuous Monitoring of Event Trajectories (CoMET) pragmatic RCT reported negative results, with a passive visual display of AI-driven clinical deterioration risk failing to improve patient outcomes [14]. Conversely, the Communicating Narrative Concerns Entered by Registered Nurses (CONCERN) trial demonstrated significant reductions in mortality; however, its evaluation was conducted largely within the internal development environment [15]. As highlighted by recent analyses, deep learning models often experience significant cross-site performance degradation when deployed in external environments, raising substantial concerns regarding their generalizability and the need for additional local retraining to meet the promised performance [16,17]. For example, an external validation study of Pacmed (Pacmed B V)—a commercial AI device that obtained European regulatory clearance—revealed that the model experienced severe performance degradation when deployed at an independent site, ultimately requiring retraining to restore the intended predictive accuracy [18].

The AI-based software as a medical device (AI-SaMD) used in this study, VUNO Med-DeepCARS (version 1.3.2), continuously predicts the risk of IHCA within 24 hours using only 4 routinely measured vital signs. Unlike models that struggle with generalizability, DeepCARS has demonstrated robustness and discriminative accuracy across diverse clinical settings. Following its initial development [19], its clinical utility has been extensively validated through single-center and large-scale multicenter retrospective cohorts [20,21], as well as a prospective multicenter observational study [22]. Furthermore, real-world external validations have confirmed its efficacy during both RRS operating and nonoperating periods [23] and highlighted its practical utility during severe national health care workforce shortages [24]. Most notably, its implementation was associated with a 35% reduction in in-hospital mortality in a recent nonrandomized controlled trial (NRCT) [25].

We therefore designed the present study, the VITAL-MD trial (validation of an AI-based clinical deterioration prediction model on patient outcomes through a stepped-wedge cluster randomized trial [SW-CRT]: a multicenter evaluation of DeepCARS). To our knowledge, VITAL-MD is the first large-scale multicenter randomized trial designed to evaluate an AI-EWS as a stand-alone device in external validation environments, addressing the 2 principal evidence gaps in the existing AI-EWS literature. This study aims

to rigorously evaluate whether the implementation of an AI-SaMD into routine clinical workflows improves patient-centered outcomes, thereby establishing a milestone for the prospective validation of medical AI.

Methods

Overview

This study was structured according to the recommendations of the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2025 statement ([Checklist 1](#)).

Objectives

The primary objective of this trial is to evaluate whether implementation of an AI-SaMD as a stand-alone clinical decision-support tool in routine general-ward care reduces the incidence of IHCA compared with usual care based on conventional EWSs, especially among adult patients identified as high-risk (target cohort).

Secondary objectives include evaluating the effect of the intervention on all-cause in-hospital mortality, hospital and intensive care unit (ICU) length of stay (LOS), time to unplanned ICU transfer after the first alert, and neurologic outcome at discharge among patients who experienced IHCA. A further objective is to characterize the clinical actions (eg, critical care consultation, antibiotic escalation, early ICU transfer) that may mediate any observed outcome benefit in order to translate algorithmic risk prediction into action-anchored clinical utility.

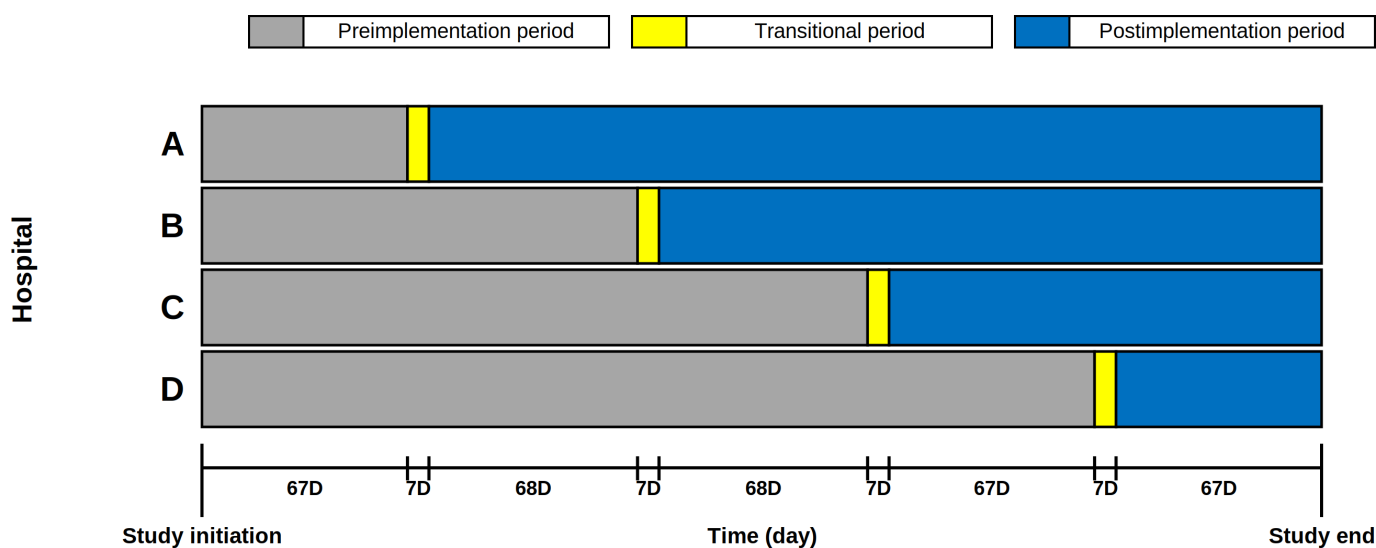
Estimand Framework

The primary estimand will follow a treatment-policy strategy: the population-level effect of being admitted to a cluster operating in the postimplementation (AI-SaMD-guided) phase versus the preimplementation (usual care) phase, regardless of individual clinician adherence to AI-SaMD alerts, with the outcome variable defined as the binary occurrence of general-ward IHCA during hospitalization.

Trial Design

This study is designed as a pragmatic, multicenter SW-CRT [26]. The unit of randomization is the hospital (cluster); the 4 participating hospitals were randomized to sequentially transition from control to intervention according to a prespecified schedule: preimplementation, transitional (washout), and postimplementation periods ([Figure 1](#)). This SW-CRT design allows each cluster to serve as its own control, accounting for intercluster variability. The total study period is 1 year to ensure coverage of all seasonal variations.

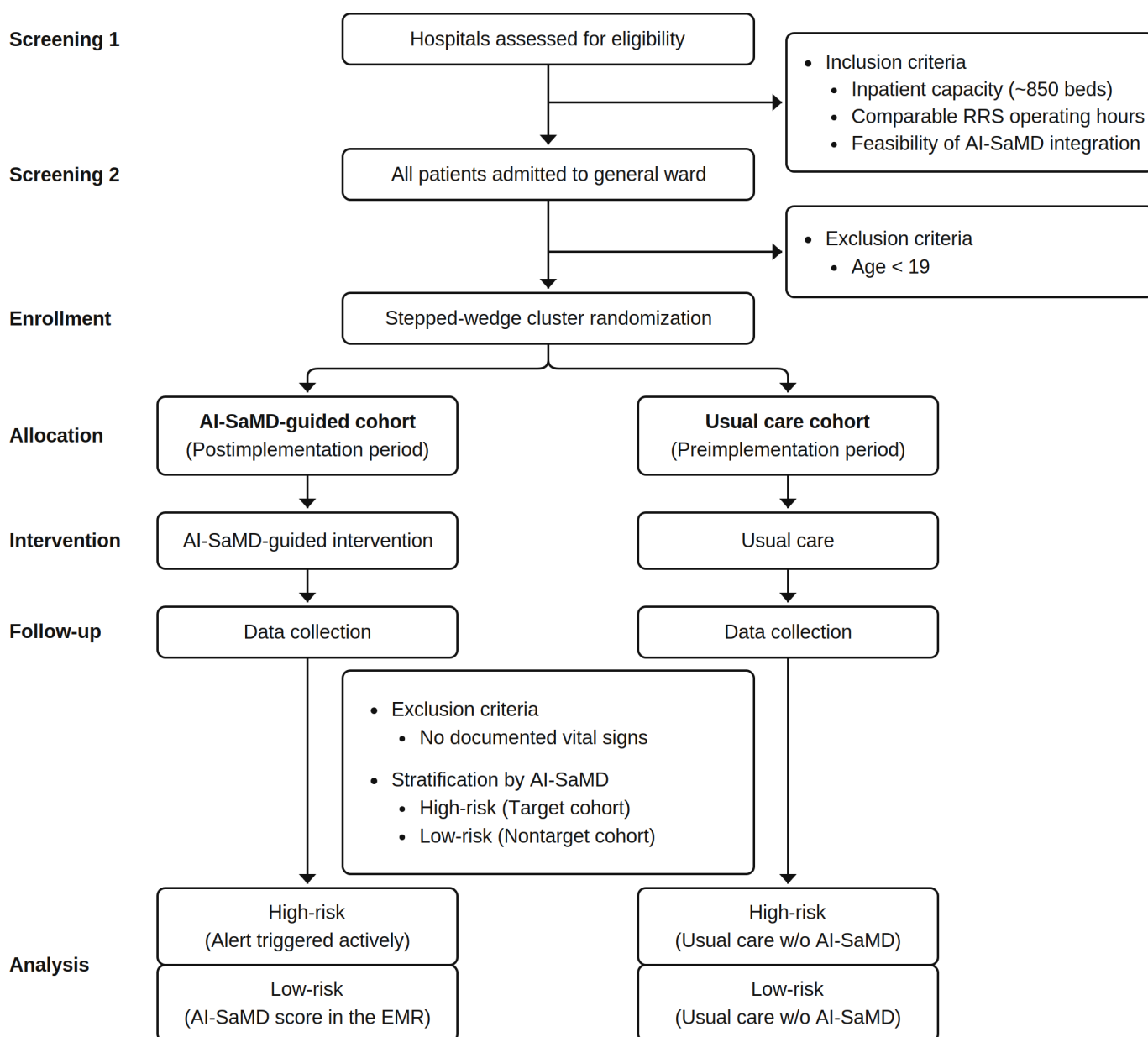
Figure 1. Schematic diagram of the stepped-wedge randomized design. The diagram illustrates the sequence of preimplementation, transitional (washout), and postimplementation periods. Each wedge spans 2.4 months and includes a 7-day transitional period.



The detailed study design is presented in [Figure 2](#). Intervention effectiveness will be evaluated by comparing patient outcomes between the AI-SaMD-guided (postimplementation) and usual care (preimplementation) phases. Within this overall comparison, the primary analysis will focus on the target cohort, defined as patients who exceed the AI-SaMD alert threshold at least once during general-ward hospitalization. This cohort aligns with the intended use of AI-SaMD (DeepCARS), which is designed to identify patients at high risk of clinical deterioration, and with the clinical estimand of interest, as an RRS is by definition intended to identify patients at risk of deterioration and facilitate intervention [3,19,21-23,25,27]. Patients who do not exceed

the alert threshold will be defined as the nontarget cohort. In prior studies, these non-high-risk patients demonstrated a very low incidence of the primary outcome [25,28-34]; the target cohort therefore represents the population in which the intervention is intended to act and in which a treatment effect can meaningfully be detected, supporting its designation as the primary analysis population. To validate the restriction of primary analysis to the target cohort, baseline characteristics, alert distribution, and patient outcomes will also be compared between the target and nontarget cohorts, as well as between the pre- and postimplementation phases, in a secondary analysis.

Figure 2. Trial flow diagram of the VITAL-MD stepped-wedge cluster randomized trial. AI-SaMD: AI-based software as a medical device; EMR: electronic medical record; RRS: rapid response system; VITAL-MD: validation of an AI-based clinical deterioration prediction model on patient outcomes through a stepped-wedge cluster randomized trial: multicenter evaluation of DeepCARS.



Participants, Interventions, and Outcomes

Trial Setting

This multicenter trial will be conducted at tertiary academic hospitals across different regions in the Republic of Korea: Inha University Hospital, Hallym University Dongtan Sacred Heart Hospital, Konyang University Hospital, and The Catholic University of Korea Incheon St. Mary's Hospital. All participating sites operate as inpatient tertiary care institutions with established RRS, and the AI-SaMD will be deployed within the general ward setting of each site.

Eligibility Criteria

Participating sites are selected based on the following criteria: (1) similar inpatient capacity (approximately 850

beds) to minimize intercluster heterogeneity; (2) comparable RRS operating hours, consistent with the national RRS pilot program [3,35-38]; and (3) no prior use of the AI-SaMD (DeepCARS), which will be newly integrated into the electronic medical record (EMR) system for the purpose of this trial. These criteria were selected to minimize structural heterogeneity across participating hospitals, thereby reducing unnecessary between-cluster variability. Of the 6 institutions assessed for eligibility, 4 were enrolled; 2 were excluded because RRSs were discontinued due to institutional resource reallocation in response to financial constraints.

All adult patients aged 19 years or older admitted to the general ward at participating hospitals during the trial period will be included. Consistent with the cluster-level allocation, patients are not individually recruited, sampled, or consented; the cohort reflects the full population of ward admissions

during each exposure period. Exclusion criteria are patients younger than 19 years at admission or no documented vital signs throughout hospitalization, resulting in no AI-SaMD score generation.

Intervention and Comparator

Intervention Description

VUNO Med-DeepCARS (version 1.3.2) is a deep learning-based cardiac arrest risk management system in general-ward patients [19,21-23,25,27]. It generates real-time 0-100 risk scores estimating the likelihood of cardiac arrest within 24 hours based on 4 routinely monitored vital signs (heart rate, respiratory rate, blood pressure, and body temperature), using the longitudinal measurements collected since hospital admission.

The AI-SaMD (DeepCARS) will be integrated into the EMR vital sign interface, replacing the conventional EWSs

for patient monitoring and risk stratification without altering existing workflows. The risk score is updated whenever new vital signs are recorded, and alerts are generated when a predefined threshold is exceeded. To minimize alarm fatigue, hospital-specific thresholds were selected to achieve an alert volume comparable to that of the existing RRS, based on historical data from each site, and repeated alerts were suppressed within the same calendar day [21-23,25,27]. Upon alert activation, the RRS team reviews the patient and health care professionals (HCPs) are recommended to reassess care; however, no mandatory response is required. RRS operations and reassessment criteria follow the national RRS pilot program guidelines (Table 1) [35,36]. The AI-SaMD functions autonomously within routine workflows, preserving physician discretion and requiring no additional staffing or protocol modifications.

Table 1. List of reasons and types of interventions implemented in response to the AI-SaMD^a alert.

Reasons for intervention	Types of intervention
Respiratory distress	• Treatment plan and consultation
Sepsis	• Advice for referral consultation
Septic shock	• Health care professional education
Hypotension	• Patient/guardian education
Cardiac arrest	• Consultation through telephone
Cardiogenic shock	• ABGA ^b
	• Blood test
	• POCUS ^c
Anaphylactic shock	• Chest X-ray
Hypovolemic shock	• EPOC (sort of ABGA)
Obstructive shock	• EtCO ₂ monitoring ^d
Arrhythmia	• Intubation
Metabolic acidosis	• Ventilator
Altered mental status	• Suction
For education/procedure	• Low/high-flow oxygen supply
For transfer	• Central line insertion
Others	• Peripheral line insertion
	• IV ^e fluid and medication
	• CPR ^f
	• DNR ^g recommend
	• UIT ^h recommend

^aAI-SaMD: AI-software as a medical device.

^bABGA: arterial blood gas analysis.

^cPOCUS: point-of-care ultrasound.

^dEtCO₂: end-tidal carbon dioxide.

^eIV: intravenous.

^fCPR: cardiopulmonary resuscitation.

^gDNR: do not resuscitate.

^hUIT: unplanned intensive care unit transfer.

Explanation for the Choice of Comparators

In the preimplementation period of the SW-CRT, usual care without the AI-SaMD serves as the comparator. Risk stratification relies on conventional EWSs as routinely deployed at each participating hospital, and RRS activation

and subsequent response follow the national RRS pilot program in the Republic of Korea [3,35-38].

Usual care and the recent standard of care, rather than a sham device or no-monitoring comparator, were chosen for both ethical and methodological reasons. Conventional EWS-based monitoring and RRS activation are the current

standard of care for detecting clinical deterioration in general-ward inpatients, as endorsed by the AHA and SCCM [4,5]. Withholding such monitoring from at-risk inpatients would violate clinical equipoise and the ethical principles of trials in acutely ill populations.

Methodologically, recent reviews and consensus statements have highlighted these considerations for the design of AI-EWS interventional studies. A systematic review of randomized clinical trials of ML interventions in *JAMA Network Open* reported that few such RCTs have been conducted, with high variability in adherence to reporting standards and substantial risk of bias across the existing evidence base [39]. The recent AHA Science Advisory on pragmatic approaches to health care AI evaluation has emphasized that AI tools should be evaluated in real-world contexts and against the care they would actually displace, rather than against artificially standardized comparators [40]. By using a routine established standard of care as the comparator, VITAL-MD estimates the added value of AI-SaMD over the care patients would otherwise receive.

This setting better reflects the question this trial seeks to answer, whether adding an AI-driven decision-support layer improves outcomes beyond what current EWS-anchored RRS systems already achieve. Although structural heterogeneity across participating hospitals was minimized through the site-selection criteria, real-world variation in day-to-day clinical workflows, staffing, and patient severity remained. By evaluating the AI-SaMD against usual care under these routine clinical conditions, the trial estimates the incremental benefit of the intervention in the setting in which it is intended to be used. These operational variations support the pragmatic generalizability of the trial.

Criteria for Discontinuing or Modifying Allocated Interventions

No predefined criteria exist for discontinuing or modifying the intervention, as the AI-SaMD will operate passively within routine workflows and poses minimal risk.

Patients admitted during the preimplementation period who remain hospitalized during the postimplementation period will be considered crossover cases and analyzed separately [26].

Strategies to Improve Adherence to Interventions

Adherence may be limited by delayed alert acknowledgment, noncompliance with alerts, patient refusal, alert fatigue, and staffing constraints [36,41-45].

To address these challenges, each participating hospital will undertake structured preimplementation preparation before AI-SaMD deployment. This will include (1) early coordination with hospital IT teams to enable seamless integration of the AI-SaMD into existing EMR and clinical workflows and (2) preimplementation education programs for relevant HCPs—including rapid response team members, ward nurses, and attending physicians—to standardize the

interpretation of AI-SaMD outputs and the corresponding alert response procedures.

Ongoing education throughout the trial will further reinforce appropriate use and timely response to alerts. However, because the AI-SaMD-guided intervention is nonmandatory in order to preserve physician autonomy, full compliance cannot be guaranteed and will be evaluated in a secondary analysis.

Relevant Concomitant Care Permitted or Prohibited During the Trial

All usual clinical practices are permitted. The AI-SaMD provides advisory alerts only and does not mandate clinical decisions.

Patient and Public Involvement

No patient or public involvement was incorporated into the design or conduct of this trial, given the cluster-level intervention, waiver of individual consent, and exclusive reliance on routinely collected data.

Outcomes

The primary outcome is the incidence of general-ward cardiac arrest, defined according to the in-hospital Utstein guidelines as cessation of cardiac activity with the absence of pulse, unresponsiveness, and apnea requiring resuscitation [46]. Secondary outcomes include all-cause in-hospital mortality, hospital and ICU LOS, time to unplanned ICU transfer (UIT) after the first alert, and Cerebral Performance Category (CPC) at discharge among general-ward patients with cardiac arrest. UIT is defined as ICU transfer within 24 hours for clinical deterioration, excluding planned postoperative ICU admissions [47].

Harms

This study poses minimal risk, as the intervention (AI-SaMD) functions as an additional safety measure without altering usual care. Adverse events are expected to be rare and likely reflect underlying illness rather than the intervention. Adverse events will be recorded in the case report form (CRF) and reviewed during interim analyses by the Data and Safety Monitoring Board (DSMB) and institutional review board (IRB) at each participating site.

In addition, the conduct of this trial will be reported with the routine mortality and cardiac arrest conferences held at each of the 4 participating hospitals. If any case is judged at such a conference to raise concern about potential harm related to the intervention—for example, workflow disruption associated with the intervention—the conference participants may independently notify the site principal investigator (PI) and the DSMB chair at any time by telephone and email, using the contact details provided in the DSMB charter (Multimedia Appendix 1). Similarly, any notable increase in mortality or IHCA at a participating site that is considered at such a conference to be associated with AI-SaMD implementation may also be reported through the same channel at any time.

Participant Timeline

The study schedule is shown in [Table 2](#).

Table 2. Schedule of procedures during study period.

Procedure	Study period		
	Admission	General ward stay	Discharge
Screening/enrollment/allocation			
Eligibility screening	✓	— ^a	—
Exclusion criteria ^b	✓	—	✓
Randomization	✓	—	—
Intervention			
Intervention (AI-SaMD) ^c	—	✓	—
Comparator (usual care)	—	✓	—
Data collection ^d			
Baseline characteristics	✓	—	—
Daily vital signs	—	✓	—
Intervention records	—	✓	—
General ward cardiac arrest	—	—	✓
All-cause in-hospital mortality	—	—	✓
Hospital length of stay	—	—	✓
Total ICU ^e length of stay	—	—	✓
Time to UIT ^f after the first alert	—	—	✓
Cerebral Performance Category	—	—	✓

^aNot applicable.

^bExclusion criteria: two-stage, patient-level age <19 years at admission; no vital-sign measurements recorded throughout hospitalization, resulting in no AI-based software as a medical device score being generated.

^cAI-SaMD: AI-based software as a medical device.

^dData collection: vital signs and intervention records captured daily during ward stay; outcomes recorded from electronic medical record after discharge.

^eICU: intensive care unit.

^fUIT: unplanned intensive care unit transfer.

Sample Size

The required sample size is estimated to detect a reduction in general ward cardiac arrest among patients identified as high risk by the AI-SaMD (target cohort). The baseline incidence in the control group is assumed to be 6.70% based on prior studies [23,25]. Assuming a 40% relative risk reduction (RRR), consistent with previous trials in comparable high-risk populations, the expected incidence in the intervention group is set at 4.00% [25,28,29,48].

With a 2-sided α of .05, 80% power, an intracluster correlation coefficient (ICC) of 0.01 (based on a meta-analysis of RRS studies) [49], and 350 patients per cluster-period (derived from prior institutional data) [21–23], the required sample size is calculated at 7002 patients in the target cohort [50]. Assuming a 10% high-risk alert rate and 5% dropout, total enrollment is estimated at 72,695 [21–23], including 65,693 patients in the nontarget cohort.

Recruitment

No specific recruitment strategy is required, as this study implements hospital-level intervention and includes all general ward admissions eligible for screening.

Assignment of Interventions

Randomization: Sequence Generation

The transition order of the 4 participating hospitals was determined before study initiation through a centralized randomization procedure performed by an independent statistician not involved in trial conduct. Given the small number of clusters and the prespecified eligibility criteria ensuring baseline comparability across hospitals, no stratification was applied. Full procedural details are provided in [Multimedia Appendix 2](#).

Allocation Concealment Mechanism

The randomization sequence for cluster transition order was generated by independent statisticians from a statistical contract research organization (CRO) and concealed from all site investigators, clinical staff, and the trial management group (TMG) until each cluster's prescheduled notification. Participating hospitals will be notified of their assigned transition time point one month before the scheduled transition to allow for technical preparation.

Implementation

In this SW-CRT, clusters will be randomly assigned to transition across 4 sequential steps, dividing the 12-month study period into five 2.4-month phases. Patients admitted in the preimplementation period will receive usual care, and those admitted in the postimplementation period will receive AI-SaMD-guided intervention. All adult patients admitted to general wards at each participating hospital during the study period will be identified through routine EMR data extraction; no individual-level sampling, selective recruitment, or consent-based enrollment will be performed, consistent with the cluster-level design of the trial.

Patients hospitalized across both periods will be considered crossover cases [26]. To minimize contamination at each cluster's phase transition, a 7-day transitional (washout) period, based on average general ward LOS, will be applied between the pre- and postimplementation periods [22,26,51].

Blinding

Given the stepped-wedge cluster deployment of an AI-SaMD, conventional blinding is not applicable, and detailed blinding is described here by role. Patients and caregivers are not informed of their site's group assignment, as AI-SaMD outputs are restricted to an EMR interface with no patient access during both the pre- and postimplementation periods. HCPs and RRS staff are not blinded due to direct interaction with the AI-SaMD alerts, and hospitals are informed of their assigned transition one month before implementation; however, HCPs and RRS staff cannot access the AI-SaMD scores during the preimplementation period.

Statistical analyses, including interim analyses, will be conducted by independent statisticians at the CRO who receive intervention/control labels in masked form, with unblinding only after the primary analysis dataset is locked and approved by the principal investigator.

Data Collection, Management, and Analysis

Data Collection Methods

Outcome data will be captured for all eligible patients admitted to participating general wards during the 12-month SW-CRT study period, irrespective of the cluster allocation phase.

Only routinely collected data and data from the national RRS pilot program will be used [35,36]. These include demographics (age, sex, admission and discharge dates, department, and ward information), routinely recorded vital signs (blood pressure, heart rate, respiratory rate, and body temperature), oxygen therapy (whether administered, type of equipment, and oxygen amount), oxygen saturation (SpO₂), and Sequential Organ Failure Assessment (SOFA) components, along with the time of measurement. Baseline illness severity will be quantified by the admission SOFA score,

computed from the first available measurements of each component during the index hospitalization, with per-component data sources adapted for general ward EMR availability, the preadmission carry-forward window for laboratory values, and handling of residual missingness operationalized consistently with prior work [52,53] and detailed in [Multi-media Appendix 3](#). In addition, ICU transfers, cardiopulmonary resuscitation (CPR), do-not-resuscitate (DNR) orders, deaths, interventions, staffing, bed occupancy, and Charlson Comorbidity Index (CCI) components will be collected.

At each hospital, a site-level clinical research coordinator (CRC) will perform data collection and quality control. Data elements will be extracted from each participating hospital's EMR and from its RRS module, which is built upon the national RRS pilot program [35,36]. The standardized structured data collection template, derived from the national RRS pilot program registry, will be deployed across all participating sites.

AI-SaMD scores will be stored in a separate EMR-based research module accessible only to the CRC for data collection and target/nontarget cohort allocation. The scores will not be visible to clinicians or guide clinical care during the preimplementation period. After transition to the postimplementation period, the AI-SaMD will be linked to the RRS module and integrated into the RRS workflow.

As the intervention is delivered at the cluster (hospital) level and outcomes are ascertained from routinely captured clinical data during the hospitalization period, individual-level retention strategies and follow-up adherence procedures are not applicable to this trial design.

Data Management

Data management will be conducted under the oversight of the PI in accordance with the Declaration of Helsinki and ICH-GCP guidelines.

Following site-level data entry and verification, each participating hospital will deidentify the captured records in conformity with the Health Insurance Portability and Accountability Act (HIPAA) deidentification standard and securely transfer them, via encrypted channels, to the designated trial server. The trial database is hosted on secure, access-controlled servers operated by VUNO Inc as the technical infrastructure provider, with audit logging of all data access events; VUNO Inc-affiliated personnel are granted no analytic or read-level access to individual-level outcome data or to aggregate treatment-allocation information before final database lock.

Access to trial data is restricted by prespecified role; site CRCs may access individual participant-level records at their respective sites, and the central PI may access the full trial dataset for trial oversight, while all other study personnel are granted access only to the data elements required to fulfill their predefined role.

At prespecified analytic time points, deidentified datasets will be exported from the trial server to a closed analytic environment maintained by the external statistical CRO via encrypted transfer. Outcome data will be analyzed by an independent statistician at the CRO, without involvement of the funders or clinicians directly providing care to enrolled participants, in accordance with the prespecified statistical analysis plan (SAP); interpretation of the analytic results will be the responsibility of the academic principal investigators, in consultation with the broader VITAL-MD study group. Research data will be retained for 3 years after completion of the trial and then securely destroyed and will not be used for secondary purposes beyond the prespecified objectives of this trial.

Statistical Methods

Statistical Methods for Primary and Secondary Outcomes

The primary outcome (general ward cardiac arrest) will be analyzed using a generalized linear mixed model (GLMM) with cluster as a random effect and time and intervention as fixed effects [26]. Secondary outcomes will be analyzed using mixed effects models appropriate to outcome type. To ensure valid inference given the small number of clusters, small-sample corrections (between-within degrees of freedom) will be applied [54]. Results will be reported as adjusted risk ratios (ARR) or adjusted risk differences (ARD) with 95% CIs [55]. Models will adjust for prespecified covariates, including age, sex, department, season, weekday, NEWS2, SOFA, CCI, number of staff physicians, and bed occupancy rate.

Baseline characteristics will be compared between the intervention and the control groups using appropriate parametric or nonparametric tests for continuous variables and chi-square tests for categorical variables.

All analyses will be performed using R (version 4.3 or higher; R Foundation for Statistical Computing) or Python (version 3.7 or higher; Python Software Foundation), with the specific software and package versions reported in the primary results publication.

Who Will Be Included in Each Analysis

The primary analysis will be conducted in the full analysis set (FAS), consistent with the intention-to-treat (ITT) principle. The primary analysis will focus on the target cohort, defined as patients who exceed the AI-SaMD alert threshold at least once during general ward hospitalization. Patients who do not exceed the AI-SaMD alert threshold (the nontarget cohort) will be included in secondary analyses. A per-protocol set (PPS) will be used for supportive analyses. Patients admitted during the washout period and exposed to both phases (crossover) will be excluded from the primary analysis but included in sensitivity analyses.

How Missing Data Will Be Handled in the Analysis

The AI-SaMD system internally applies last observation carried forward (LOCF) for missing vital signs in score calculation. Missing outcome data will be reviewed by clinicians.

Missing data regarding outcomes are expected to be limited, as they are ascertained from routinely collected EMR data. The primary outcome (IHCA) and secondary outcomes subject to mandatory hospital reporting, including in-hospital mortality, LOS, and ICU transfer, are expected to have minimal missing data. The AI-SaMD internally applies LOCF to missing vital-sign inputs during real-time score computation; this is a fixed device behavior and does not affect outcome ascertainment.

Methods for Additional Analyses

Sensitivity analyses will include crude, propensity score matching (PSM), e-value analysis, and crossover analyses to evaluate the robustness and consistency of the primary findings [56,57]. A nontarget cohort analysis will assess the robustness of restricting the primary analysis to high-risk patients by comparing baseline characteristics, alert distribution, and patient outcomes. In addition, a full-population analysis will be conducted using the same modeling framework as the primary analysis to estimate the hospital-wide intervention effect. Time-to-event outcomes will be descriptively summarized using Kaplan-Meier estimates to visualize event timing [58]. Subgroup analyses will evaluate the consistency of the intervention effect across clinically relevant subgroups (eg, sepsis, surgical vs nonsurgical admissions, and others), including patient characteristics, clinical contexts, and reasons for or types of interventions in response to the AI-SaMD alert. A prespecified sepsis subgroup, defined by the Centers for Disease Control and Prevention (CDC) Adult Sepsis Event (ASE) surveillance criteria, will be analyzed using the same modeling framework (Multimedia Appendix 3). Lead-time analyses will evaluate changes over time in alert response time and their association with patient outcomes to assess staff response patterns and intervention adherence after implementation.

Oversight and Monitoring

Trial Organization

Coordinating Site

The coordinating site at Inha University Hospital, which also serves as the trial sponsor, manages trial operations, including protocol implementation, data management coordination, and intersite communication. No separate external coordinating center has been established, given the relatively small number of clusters and the routinely collected data nature of this trial.

TMG

A TMG, composed of the principal investigators and RRS nurses at each of the 4 participating hospitals together with

the trial statistician, is responsible for the day-to-day conduct of the trial, including site coordination, protocol adherence, data quality oversight, and review of operational issues. The TMG holds scheduled meetings at each cluster transition time point. Independent oversight of the trial is provided separately by the DSMB (see below) rather than through a Trial Steering Committee, given the low-risk, pragmatic nature of the trial and the use of routinely collected clinical data only.

Data Monitoring Committee

DSMB

An independent DSMB, composed of independent experts in ethics, critical care, biostatistics, and clinical trials, oversees participant safety and data integrity. The DSMB reviews interim safety and efficacy data at each scheduled interim analysis. Further details are provided in the DSMB charter ([Multimedia Appendix 1](#)).

Interim Analyses and Stopping Guidelines

Three interim analyses will be performed at the second, third, and fourth cluster transition time points by an independent DSMB statistician. Each interim analysis will evaluate the IHCA rate (cases per 1000 admissions) after confounder adjustment and compare it with the corresponding calendar period of prior years. A >50% increase relative to this reference period will constitute a sentinel signal warranting closed-session DSMB review with clinical expert input; the DSMB will then report to the IRBs and the PI and may convene an open session to consider trial modification or termination. Further details are provided in the DSMB charter ([Multimedia Appendix 1](#)).

Trial Monitoring

Consistent with the risk-proportionate monitoring approach recommended by the International Council for Harmonisation (ICH) Good Clinical Practice Guideline E6(R3) and the SPIRIT 2025 statement (Item 29), and given that this is a minimal-risk pragmatic trial conducted with a waiver of individual consent and using only routinely collected EMR data, formal external on-site auditing has not been adopted. Instead, trial monitoring is conducted through a combination of (1) ongoing remote central data quality review by the CRO, including range checks, completeness checks, and consistency checks across sites; (2) periodic safety and data integrity oversight by the independent DSMB at each scheduled interim analysis; and (3) integration with the routine mortality and cardiac arrest conferences held at each of the 4 participating hospitals, as described in the Harms section, which allows independent institutional clinical judgment to flag any potential intervention-related harm directly to the site PI and the DSMB chair.

Ethical Considerations

The study protocol was approved by the IRBs of all 4 participating hospitals: Inha University Hospital (INHAUH 2024-09-034; approved January 20, 2025), Hallym University Dongtan Sacred Heart Hospital (HDT 2024-10-006;

approved March 24, 2025), Konyang University Hospital (KYUH 2024-12-004; approved December 18, 2024), and The Catholic University of Korea Incheon St. Mary's Hospital (OC24DIDS0141; approved February 17, 2025). The requirement for informed consent was independently waived by the IRBs of all 4 participating hospitals.

Important protocol amendments will be promptly communicated to relevant parties. The principal investigator will submit revised protocols to the IRBs at all sites for approval before implementation. Approved amendments will be updated in the trial registry.

This study requests the waiving of consent from individual patients for the following three reasons: (1) this study poses minimal risk, as the AI-SaMD supplements standard care by providing nonmandatory screening recommendations without altering clinical workflows or physician autonomy; (2) the waiver will not adversely affect participants' rights or welfare because the intervention does not restrict access to standard care; and (3) obtaining individual consent is not feasible in a large-scale hospital-based study and may introduce selection bias, as the study targets the entire inpatient population. Additional consent provisions are not applicable, as no additional participant data or biological specimens will be collected.

This study will ensure strict confidentiality throughout the data collection, analysis, and publication. All personal identifiers will be deidentified in accordance with the HIPAA requirements. Records will be retained for 3 years after completion of the trial, in accordance with the Korean Bioethics and Safety Act, and then securely disposed of. No secondary use beyond the study objectives is permitted. Ancillary and posttrial care is not applicable, as the study covers the entire general-ward hospitalization period.

Results

The trial was registered with the Clinical Research Information Service (CRIS; KCT0010243), a World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) primary registry in the Republic of Korea, on February 28, 2025. Of 6 tertiary academic hospitals assessed for eligibility, 4 were included; site-specific IRB approvals were obtained prior to study initiation. Patient recruitment ran from April 1, 2025, to March 31, 2026, covering the prespecified 12-month stepped-wedge schedule, and all 4 sites completed the transition to the postimplementation period. CRF finalization and data cleaning are currently underway prior to final database lock, after which statistical analysis will proceed, with the first results expected to be submitted for publication in 2027.

Discussion

Principal Anticipated Findings

The VITAL-MD trial will provide the first large-scale, multicenter randomized evidence on whether the deployment

of an AI-EWS, used as a stand-alone clinical decision-support device—without algorithm-specific human curation or mandated workflow modification—improves patient-centered outcomes among general-ward patients in external validation environments. Building on the demonstrated diagnostic superiority of DeepCARS over conventional EWS in retrospective and prospective observational studies [19,21-24] and on the suggestive but nonrandomized signal from the recent NRCT [25], we hypothesize that the intervention period will be associated with a reduction in ward IHCA and in in-hospital mortality compared with the control group in which routine conventional EWS-based care (eg, NEWS2 and MEWS) is used. We additionally anticipate a reduction in the rate of late or missed escalations of care, an increase in proactive interventions (eg, critical care consultation, antibiotic escalation, early ICU transfer), and no clinically meaningful increase in alert burden.

Strengths

This trial directly addresses the 2 principal limitations of the existing AI-EWS evidence base for ward-level deterioration. First, the 2 prior multicenter cluster RCTs for ward deterioration—the negative CoMET trial and the positive CONCERN trial, currently representing the highest-level randomized evidence in this field—were both conducted within their internal development environments [14,15]. Because AI model performance frequently drops substantially when transferred to external settings due to dataset shifts [59], whether these divergent outcomes generalize beyond their developing institutions remains unestablished. To our knowledge, VITAL-MD will be the first multicenter randomized trial to evaluate the patient outcome effect of an AI-EWS in true external validation environments. Second, previous interventional studies bundled algorithms with specific implementation infrastructures, such as off-site nurse curation in the Advance Alert Monitor deployment [28]. By deploying DeepCARS as a stand-alone device without algorithm-specific human curation or mandated workflow modification, VITAL-MD isolates the pure effect of the AI-SaMD itself from the surrounding human and workflow scaffolding. Furthermore, establishing causal outcome benefits for rapid response mechanisms has historically been challenging. Comprehensive Cochrane systematic reviews and the landmark Medical Early Response, Intervention, and Therapy (MERIT) trial demonstrated little or no significant reduction in unexpected mortality from conventional RRS and EWS interventions [8,60]. Despite this, major societies such as the AHA and SCCM continuously and strongly recommend hospital-wide surveillance systems [4, 5]. VITAL-MD aims to bridge this critical evidence gap by generating definitive causal evidence that replacing the inaccurate afferent limb with a highly accurate AI-EWS can validate the clinical rationale behind these ongoing recommendations. Evaluating the intervention directly against usual clinical practice and the standard of care is strongly emphasized by major guidelines as the most rigorous benchmark for establishing clinical utility [11]. The SW-CRT design robustly controls for secular trends and between-cluster

confounding, evaluating this real-world standard across a massive cohort of more than 70,000 anticipated admissions.

Shifting From Explainability to Actionability

Furthermore, this study seeks to propose a novel paradigm for understanding and evaluating AI outputs, pivoting the focus from algorithmic “explainability” to clinical “actionability.” The prevailing approach to mitigating the “black-box” nature of deep learning models involves post hoc mathematical interpretability techniques, such as Shapley Additive Explanations (SHAP) values. However, emerging literature clearly indicates that such mathematical explanations fail to translate into tangible clinical actions. As highlighted by Bienefeld et al [61], mathematical feature weightings do not provide clinical actionability and are frequently perceived by clinicians as unhelpful in acute care settings. We will therefore investigate the specific clinical mechanisms driving outcome improvements by prospectively analyzing lead times to targeted clinical interventions (eg, critical care consultation, antibiotic escalation, and early ICU consultation) and identifying which clinical subgroups benefit most from actions prompted by AI alerts. Providing HCPs with practical, action-oriented guidelines on how to respond to AI alerts will directly maximize the real-world utility of the AI-SaMD.

Limitations

Several limitations warrant explicit acknowledgment. First, while this study provides causal evidence for one specific algorithm, clinical efficacy is not a universal characteristic of all AI-EWS platforms. Emerging literature demonstrates that commercial medical AI devices exhibit vast disparities in their real-world performance [16]. Recent multicenter evaluations have shown that sensitivities, positive predictive values, and false-alarm burdens range widely across AI products even when evaluating the exact same clinical deterioration targets [9,62,63]. Consequently, the patient-centered benefits observed in this trial apply exclusively to the implemented AI-SaMD and do not validate the universal clinical utility of commercial early warning algorithms. Second, while all general ward admissions are included in the analysis, the primary evaluation focuses predominantly on the “target cohort,” patients who exceed the high-risk alert threshold. Although restricting the primary analysis to this subset limits inferences about the overall unselected hospital population, this targeted methodology perfectly aligns with the algorithm’s intended clinical use as a medical device to identify and trigger responses for high-risk patients. Furthermore, this approach is consistent with prior AI-EWS evaluations, such as the studies by Escobar et al [28] and Winslow et al [29], justifying its analytical validity for evaluating algorithmic rapid response interventions. Third, the study is conducted exclusively in tertiary academic hospitals within South Korea. The lack of racial and geographic diversity limits the direct generalizability of the findings to health care systems in other countries or nontertiary settings. Fourth, the planned number of randomized clusters was scaled down to 4 sites, which not only decreases statistical efficiency but also means the

between-cluster variance is imprecisely estimated; we address the resulting inferential limitations through the stepped-wedge structure and prespecified small-sample corrections [26,54]. Finally, the passive, nonmandatory nature of the AI intervention might lead to variable compliance and alert fatigue among clinicians over time. Potential cross-contamination during transition periods will be carefully monitored via strict washout period definitions and prespecified sensitivity analyses.

Implications for Research, Practice, and Policy

If the trial demonstrates a beneficial effect of DeepCARS on patient-centered outcomes, it will set a methodological benchmark for AI-EWS evaluation, complementing prior multicenter cluster RCTs by being the first conducted in external validation environments as a stand-alone device. For clinical practice, it will provide direct evidence on whether an AI-EWS can improve outcomes in resource-limited or workforce-constrained settings where additional curating personnel are not feasible. For regulatory science, it will inform the ongoing debate on whether AI-SaMDs—currently approved primarily based on retrospective diagnostic performance—should also be required to demonstrate prospective effects on patient-centered outcomes before or shortly after market authorization [13].

Conclusions

The VITAL-MD trial is designed to generate the first multicenter randomized evidence on whether an AI-EWS, deployed as stand-alone software as a medical device in external validation environments, can improve patient-centered outcomes for general-ward patients. By directly comparing the intervention with usual care and prospectively characterizing the clinical actions that mediate any observed benefit, the trial seeks to advance AI-SaMD evaluation from retrospective diagnostic accuracy toward prospective, action-oriented clinical utility, with anticipated implications for clinical practice, hospital policy, and regulatory science.

Acknowledgments

The authors are thankful to the institutional review boards of Inha University Hospital, Hallym University Dongtan Sacred Heart Hospital, Konyang University Hospital, and The Catholic University of Korea Incheon St. Mary's Hospital for approving the protocol, and to the rapid response team members at each participating hospital for their operational support during trial implementation.

The VITAL-MD Data and Safety Monitoring Board (DSMB) comprised Hyuna Bae, MD, PhD (Chairperson; Ewha Womans University School of Law); Jaehun Jung, MD, PhD (Korea University College of Medicine); Tai Sun Park, MD, PhD (Hanyang University College of Medicine); and Sooyeon Chung, MS (C&K INSIGHT Co., Ltd.).

The authors used Claude (Claude Opus 4.8; Anthropic) solely for language editing of author-written text. The tool did not generate any scientific content, data, analyses, results, interpretation, or references; all authors reviewed the AI-assisted text and take full responsibility for the manuscript.

Funding

This investigator-initiated trial is conducted with Inha University Hospital as the trial sponsor. The trial is funded primarily by an institutional clinical research grant from Inha University Hospital (Grant Number: INHAUH 2024-09-034), which constitutes the principal source of support for trial design, conduct, data management, and statistical analysis. Support was also provided by VUNO Inc. (Seoul, Republic of Korea), the manufacturer of the investigational AI-SaMD. VUNO Inc. provided (1) the investigational device, VUNO Med-DeepCARS (version 1.3.2), at no cost to the participating hospitals for the duration of the trial, and (2) support for local operational costs at the participating sites associated with deployment of the device into

Dissemination Policy

Dissemination of Trial Results

Trial results will be disseminated regardless of the direction or statistical significance of the findings. The primary results manuscript will be submitted to a peer-reviewed international journal within 12 months after final database lock and prepared in accordance with the CONSORT (Consolidated Standards of Reporting Trials) 2025 statement. Findings will be presented at relevant scientific conferences in critical care medicine, RRS, patient safety, cardiovascular disease, and biomedical informatics. Authorship of the primary results manuscript will follow the International Committee of Medical Journal Editors (ICMJE) criteria. The author list will be finalized before database lock, based on contributions to trial conception, design, conduct, analysis, or interpretation. Members of the VITAL-MD Study Group who do not meet ICMJE authorship criteria will be acknowledged as collaborators.

The TMG retains sole authority over the review and approval of all trial-related publications and presentations. The executive leadership of Inha University Hospital and all other funding organizations are strictly prohibited from editing, approving, or delaying the publication of the trial outcomes.

Access to the SAP, Participant-Level Data, and Statistical Code

The SAP is provided as [Multimedia Appendix 4](#). Deidentified data and analysis code will be made available from March 2027 upon reasonable request to JSK (Inha University Hospital), subject to ethical approval. The dataset will include each participant's vital signs, AI-SaMD scores, intervention type and reason, and corresponding clinical outcomes.

routine clinical workflows. Role of the Sponsor and Funders: Inha University Hospital, as the trial sponsor, holds institutional responsibility for trial conduct and data integrity at the sponsor level; however, the hospital administration has no role in the design of the trial, the enrollment or management of participants, the collection, management, analysis, or interpretation of the data, the review or approval of the manuscript, or the decision to submit the manuscript for publication. VUNO Inc., as one of the funders and the manufacturer of the investigational device, was permitted to provide technical commentary on the initial draft of the trial protocol and the statistical analysis plan, restricted to methodological accuracy regarding the investigational device. Beyond this clearly defined role, VUNO Inc. has no involvement in the final design of the trial; the enrollment, recruitment, or management of participants; the collection, management, analysis, or interpretation of the data; the review or approval of the manuscript reporting the trial results; or the decision to submit the manuscript for publication. Strictly adhering to the SPIRIT 2025 reporting guidelines, the academic investigators at the four participating hospitals retain absolute operational independence and exclusive authority over all trial operations and the ultimate dissemination of the results. The initial drafts of the trial protocol and the SAP were prepared with methodological and technical input from the VUNO Inc.-affiliated investigators (Mincheol Kim and Minsoo Kim) and the academic co-corresponding authors (DY and JSK). All draft documents were subsequently reviewed, independently modified, and approved by the academic principal investigators (JSK) at Inha University Hospital and all other site principal investigators (KHK, SJK, and MSP), who retain exclusive final decision-making authority over the trial design, conduct, and reporting. The SAP was further finalized in collaboration with an independent external statistical CRO with no affiliation to VUNO Inc. Data acquisition, patient management, and clinical decisions: All clinical data acquisition, patient management, and intervention-response decisions are conducted exclusively by the clinical site investigators (JSK, KHK, SJK, and MSP) and the rapid response team members at the four participating academic hospitals, none of whom hold any employment, consultancy, equity, advisory, or other financial relationship with VUNO Inc. VUNO Inc.-affiliated personnel have no role in identifying eligible participants, delivering or modifying clinical care, or determining the timing or content of any AI-SaMD-triggered response. Statistical analysis: To guarantee analytical integrity, the primary statistical analyses will be conducted by the independent external statistical CRO strictly in accordance with the prespecified SAP. All VUNO Inc.-affiliated personnel and analysts will remain blinded to aggregate treatment allocation until database lock, and no interim analysis results will be shared with VUNO Inc. prior to the final lock of the trial database.

Authors' Contributions

Conceptualization: Mincheol Kim, JSK, and DY

Methodology: Mincheol Kim, KHK, SJK, MSP, AJK, MJL, MHP, HKC, ML, IBJ, SJW, SBA, JHL, JJ, HJ, YYL, WH, JSK, and DY

Software: Mincheol Kim and KJC

Resources: Mincheol Kim, KHK, SJK, MSP, AJK, MJL, MHP, HKC, ML, IBJ, SJW, SBA, JHL, and KJC

Investigation: KHK, SJK, MSP, AJK, MJL, MHP, HKC, ML, IBJ, SJW, SBA, JHL, JJ, HJ, YYL, and WH

Data curation: KHK, SJK, MSP, AJK, MJL, MHP, HKC, ML, IBJ, SJW, SBA, JHL, JJ, HJ, YYL, WH, JSK, and DY

Project administration: KHK, SJK, MSP, JSK, and DY

Supervision: KHK, SJK, MSP, JSK, and DY

Funding acquisition: JSK

Writing – original draft: Mincheol Kim, Minsoo Kim, and DY

Writing – review & editing: KHK, SJK, MSP, JSK, and DY

All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

Mincheol Kim, Minsoo Kim, and KJC are employees of VUNO Inc, the manufacturer of the AI-SaMD (DeepCARS) evaluated in this trial. No other authors report financial or nonfinancial conflicts of interest related to the subject of this work.

To mitigate the potential influence of these industry relationships on trial integrity, structural firewalls were implemented in advance and will be maintained throughout the trial, as detailed above in the Role of the sponsor and funders, Data acquisition, patient management, and clinical decisions, and Statistical analysis sections.

Collaborators

VITAL-MD Study Group

The VITAL-MD Study Group additionally comprises the rapid response team members and clinical research coordinators who supported trial conduct at the four participating sites:

Inha University Hospital, Incheon, Republic of Korea

Jung Soo Kim, MD, MSc (Department of Critical Care Medicine); Dongjoon Yoo, MD (Department of Critical Care Medicine); Ah Jin Kim, MD, PhD (Department of Critical Care Medicine); Man-Jong Lee, MD, MSc (Department of Critical Care Medicine); Mi Hwa Park, MD, MSc (Department of Critical Care Medicine); Hyung Keun Cha, MD (Department of Critical Care Medicine); Jinhui Jang, RN, MSc (Critical Care Support Part, Department of Hospital Medicine); Seolhee Kim, RN (Critical Care Support Part, Department of Hospital Medicine); Joohee Lee, RN (Critical Care Support Part, Department of

Hospital Medicine); Keunhye Park, RN (Critical Care Support Part, Department of Hospital Medicine); Suhyun Kwon, RN (Critical Care Support Part, Department of Hospital Medicine); Inhwa Jeon, RN (Research Support Department) Hallym University Dongtan Sacred Heart Hospital, Hwaseong, Republic of Korea
Myung Soo Park, MD (Department of Cardiology); Jeon Hwa Lee, MD, PhD (Department of Cardiology); Soon Joo Wang, MD (Department of Emergency Medicine); Sung Bak An, MD (Department of General Surgery); Woori Han, RN (Rapid Response Team); Hye Ran Lee, RN (Rapid Response Team); Han Sol Lee, RN (Rapid Response Team); Ji Soo Hong, RN (Rapid Response Team); Jung Sik Shin, RN (Rapid Response Team); JiYeon Woo, RN (Clinical Research Nurse, Department of Cardiology)
Konyang University Hospital, Daejeon, Republic of Korea
Sun Jung Kwon, MD, PhD (Division of Pulmonology and Critical Care Medicine, Department of Internal Medicine); Minhyeok Lee, MD (Division of Pulmonology and Critical Care Medicine, Department of Internal Medicine); In Beom Jeong, MD (Division of Pulmonology and Critical Care Medicine, Department of Internal Medicine); Hyojin Jeong, RN (Rapid Response Team, Department of Nursing)
Incheon St. Mary's Hospital, The Catholic University of Korea, Incheon, Republic of Korea
Kyung Hoon Kim, MD (Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine); Yoon Young Lee, RN (Department of Nursing)

Multimedia Appendix 1

Data and Safety Monitoring Board (DSMB) charter.

[\[DOCX File \(Microsoft Word File\), 36 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Randomization procedure and intervention time point allocation in a stepped-wedge cluster randomized trial (SW-CRT) for the 4 participating sites.

[\[PDF File \(Adobe File\), 6544 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Supplementary methods.

[\[DOCX File \(Microsoft Word File\), 42 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Statistical analysis plan (SAP).

[\[DOCX File \(Microsoft Word File\), 231 KB-Multimedia Appendix 4\]](#)

Checklist 1

SPIRIT 2025 checklist.

[\[DOCX File \(Microsoft Word File\), 35 KB-Checklist 1\]](#)

References

1. Andersen LW, Holmberg MJ, Berg KM, Donnino MW, Granfeldt A. In-hospital cardiac arrest: a review. *JAMA*. Mar 26, 2019;321(12):1200-1210. [doi: [10.1001/jama.2019.1696](#)] [Medline: [30912843](#)]
2. Schluep M, Gravesteyn BY, Stolker RJ, Endeman H, Hoeks SE. One-year survival after in-hospital cardiac arrest: a systematic review and meta-analysis. *Resuscitation*. Nov 2018;132:90-100. [doi: [10.1016/j.resuscitation.2018.09.001](#)] [Medline: [30213495](#)]
3. DeVita MA. Textbook of Rapid Response Systems: Concept and Implementation. Springer; 2024. [doi: [10.1007/978-3-031-67951-3](#)] ISBN: 978-3-031-67950-6
4. Dezfulian C, Cabañas JG, Buckley JR, et al. Part 4: systems of care: 2025 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. *Circulation*. Oct 21, 2025;152(16_suppl_2):S353-S384. [doi: [10.1161/CIR.0000000000001378](#)] [Medline: [41122886](#)]
5. Honarmand K, Wax RS, Penoyer D, et al. Society of critical care medicine guidelines on recognizing and responding to clinical deterioration outside the ICU: 2023. *Crit Care Med*. Feb 1, 2024;52(2):314-330. [doi: [10.1097/CCM.0000000000006072](#)] [Medline: [38240510](#)]
6. Perkins GD, Neumar R, Monsieurs KG, et al. The International Liaison Committee on resuscitation-review of the last 25 years and vision for the future. *Resuscitation*. Dec 2017;121:104-116. [doi: [10.1016/j.resuscitation.2017.09.029](#)] [Medline: [28993179](#)]
7. Berg KM, Bray JE, Djäv T, et al. Executive summary: 2025 International Liaison Committee on resuscitation consensus on science with treatment recommendations. *Circulation*. Oct 21, 2025;152(16_suppl_1):S2-S22. [doi: [10.1161/CIR.0000000000001361](#)] [Medline: [41122844](#)]

8. McGaughey J, Fergusson DA, Van Bogaert P, Rose L. Early warning systems and rapid response systems for the prevention of patient deterioration on acute adult hospital wards. *Cochrane Database Syst Rev*. Nov 22, 2021;11(11):CD005529. [doi: [10.1002/14651858.CD005529.pub3](https://doi.org/10.1002/14651858.CD005529.pub3)] [Medline: [34808700](https://pubmed.ncbi.nlm.nih.gov/34808700/)]
9. Edelson DP, Churpek MM, Carey KA, et al. Early warning scores with and without artificial intelligence. *JAMA Netw Open*. Oct 1, 2024;7(10):e2438986. [doi: [10.1001/jamanetworkopen.2024.38986](https://doi.org/10.1001/jamanetworkopen.2024.38986)] [Medline: [39405061](https://pubmed.ncbi.nlm.nih.gov/39405061/)]
10. Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA. Minimal impact of implemented early warning score and best practice alert for patient deterioration. *Crit Care Med*. Jan 2019;47(1):49-55. [doi: [10.1097/CCM.0000000000003439](https://doi.org/10.1097/CCM.0000000000003439)] [Medline: [30247239](https://pubmed.ncbi.nlm.nih.gov/30247239/)]
11. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ*. May 18, 2022;377:e070904. [doi: [10.1136/bmj-2022-070904](https://doi.org/10.1136/bmj-2022-070904)] [Medline: [35584845](https://pubmed.ncbi.nlm.nih.gov/35584845/)]
12. Aroundas AA, Narayan SM, Arnett DK, et al. Use of artificial intelligence in improving outcomes in heart disease: a scientific statement from the American Heart Association. *Circulation*. Apr 2, 2024;149(14):e1028-e1050. [doi: [10.1161/CIR.0000000000001201](https://doi.org/10.1161/CIR.0000000000001201)] [Medline: [38415358](https://pubmed.ncbi.nlm.nih.gov/38415358/)]
13. Liu TYA, Hswen Y, Bonow RO, et al. Artificial intelligence across the JAMA network: a call for submissions. *JAMA*. Nov 11, 2025;334(18):1626-1627. [doi: [10.1001/jama.2025.15074](https://doi.org/10.1001/jama.2025.15074)]
14. Keim-Malpass J, Ratcliffe SJ, Clark MT, et al. A randomized controlled trial of artificial intelligence-based analytics for clinical deterioration. *Sci Rep*. Feb 5, 2026;16(1):7345. [doi: [10.1038/s41598-026-39051-z](https://doi.org/10.1038/s41598-026-39051-z)] [Medline: [41644641](https://pubmed.ncbi.nlm.nih.gov/41644641/)]
15. Rossetti SC, Dykes PC, Knaplund C, et al. Real-time surveillance system for patient deterioration: a pragmatic cluster-randomized controlled trial. *Nat Med*. Jun 2025;31(6):1895-1902. [doi: [10.1038/s41591-025-03609-7](https://doi.org/10.1038/s41591-025-03609-7)] [Medline: [40175738](https://pubmed.ncbi.nlm.nih.gov/40175738/)]
16. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med*. Apr 2021;27(4):582-584. [doi: [10.1038/s41591-021-01312-x](https://doi.org/10.1038/s41591-021-01312-x)] [Medline: [33820998](https://pubmed.ncbi.nlm.nih.gov/33820998/)]
17. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. *Radiol Artif Intell*. May 2022;4(3):e210064. [doi: [10.1148/ryai.210064](https://doi.org/10.1148/ryai.210064)] [Medline: [35652114](https://pubmed.ncbi.nlm.nih.gov/35652114/)]
18. de Hond AAH, Kant IMJ, Fornasa M, et al. Predicting readmission or death after discharge from the ICU: external validation and retraining of a machine learning model. *Crit Care Med*. Feb 1, 2023;51(2):291-300. [doi: [10.1097/CCM.0000000000005758](https://doi.org/10.1097/CCM.0000000000005758)] [Medline: [36524820](https://pubmed.ncbi.nlm.nih.gov/36524820/)]
19. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. *J Am Heart Assoc*. Jun 26, 2018;7(13):e008678. [doi: [10.1161/JAHA.118.008678](https://doi.org/10.1161/JAHA.118.008678)] [Medline: [29945914](https://pubmed.ncbi.nlm.nih.gov/29945914/)]
20. Cho KJ, Kwon O, Kwon JM, et al. Detecting patient deterioration using artificial intelligence in a rapid response system. *Crit Care Med*. Apr 2020;48(4):e285-e289. [doi: [10.1097/CCM.0000000000004236](https://doi.org/10.1097/CCM.0000000000004236)] [Medline: [32205618](https://pubmed.ncbi.nlm.nih.gov/32205618/)]
21. Lee YJ, Cho KJ, Kwon O, et al. A multicentre validation study of the deep learning-based early warning score for predicting in-hospital cardiac arrest in patients admitted to general wards. *Resuscitation*. Jun 2021;163:78-85. [doi: [10.1016/j.resuscitation.2021.04.013](https://doi.org/10.1016/j.resuscitation.2021.04.013)] [Medline: [33895236](https://pubmed.ncbi.nlm.nih.gov/33895236/)]
22. Cho KJ, Kim JS, Lee DH, et al. Prospective, multicenter validation of the deep learning-based cardiac arrest risk management system for predicting in-hospital cardiac arrest or unplanned intensive care unit transfer in patients admitted to general wards. *Crit Care*. Sep 5, 2023;27(1):346. [doi: [10.1186/s13054-023-04609-0](https://doi.org/10.1186/s13054-023-04609-0)] [Medline: [37670324](https://pubmed.ncbi.nlm.nih.gov/37670324/)]
23. Cho KJ, Kim KH, Choi J, Yoo D, Kim J. External validation of deep learning-based cardiac arrest risk management system for predicting in-hospital cardiac arrest in patients admitted to general wards based on rapid response system operating and nonoperating periods: a single-center study. *Crit Care Med*. Mar 1, 2024;52(3):e110-e120. [doi: [10.1097/CCM.0000000000006137](https://doi.org/10.1097/CCM.0000000000006137)] [Medline: [38381018](https://pubmed.ncbi.nlm.nih.gov/38381018/)]
24. Jang H, Yoo W, Hwang S, Lee K. DeepCARS-identified high-risk patients: clinical interventions and outcomes during the Korean healthcare crisis. *Medicina (Kaunas)*. Oct 22, 2025;61(11):1896. [doi: [10.3390/medicina61111896](https://doi.org/10.3390/medicina61111896)] [Medline: [41303733](https://pubmed.ncbi.nlm.nih.gov/41303733/)]
25. Park MH, Kim M, Lee MJ, et al. Clinical effectiveness of an artificial intelligence-based prediction model for cardiac arrest in general ward-admitted patients: a non-randomized controlled trial. *Diagnostics (Basel)*. Jan 20, 2026;16(2):335. [doi: [10.3390/diagnostics16020335](https://doi.org/10.3390/diagnostics16020335)] [Medline: [41594312](https://pubmed.ncbi.nlm.nih.gov/41594312/)]
26. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. Feb 6, 2015;350:h391. [doi: [10.1136/bmj.h391](https://doi.org/10.1136/bmj.h391)] [Medline: [25662947](https://pubmed.ncbi.nlm.nih.gov/25662947/)]
27. Shin Y, Cho KJ, Lee Y, et al. Multicenter validation of a deep-learning-based pediatric early-warning system for prediction of deterioration events. *Acute Crit Care*. Nov 2022;37(4):654-666. [doi: [10.4266/acc.2022.00976](https://doi.org/10.4266/acc.2022.00976)] [Medline: [36442471](https://pubmed.ncbi.nlm.nih.gov/36442471/)]

28. Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P. Automated identification of adults at risk for in-hospital clinical deterioration. *N Engl J Med*. Nov 12, 2020;383(20):1951-1960. [doi: [10.1056/NEJMsa2001090](https://doi.org/10.1056/NEJMsa2001090)] [Medline: [33176085](https://pubmed.ncbi.nlm.nih.gov/33176085/)]
29. Winslow CJ, Edelson DP, Churpek MM, et al. The impact of a machine learning early warning score on hospital mortality: a multicenter clinical intervention trial. *Crit Care Med*. Sep 1, 2022;50(9):1339-1347. [doi: [10.1097/CCM.0000000000005492](https://doi.org/10.1097/CCM.0000000000005492)] [Medline: [35452010](https://pubmed.ncbi.nlm.nih.gov/35452010/)]
30. Mayampurath A, Carey K, Palama B, et al. Machine learning-based pediatric early warning score: patient outcomes in a pre- versus post-implementation study, 2019-2023. *Pediatr Crit Care Med*. Feb 1, 2025;26(2):e146-e154. [doi: [10.1097/PCC.0000000000003656](https://doi.org/10.1097/PCC.0000000000003656)] [Medline: [39982152](https://pubmed.ncbi.nlm.nih.gov/39982152/)]
31. Kollef MH, Chen Y, Heard K, et al. A randomized trial of real-time automated clinical deterioration alerts sent to a rapid response team. *J Hosp Med*. Jul 2014;9(7):424-429. [doi: [10.1002/jhm.2193](https://doi.org/10.1002/jhm.2193)] [Medline: [24706596](https://pubmed.ncbi.nlm.nih.gov/24706596/)]
32. Bailey TC, Chen Y, Mao Y, et al. A trial of a real-time alert for clinical deterioration in patients hospitalized on general medical wards. *J Hosp Med*. May 2013;8(5):236-242. [doi: [10.1002/jhm.2009](https://doi.org/10.1002/jhm.2009)] [Medline: [23440923](https://pubmed.ncbi.nlm.nih.gov/23440923/)]
33. Adams R, Henry KE, Sridharan A, et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. *Nat Med*. Jul 2022;28(7):1455-1460. [doi: [10.1038/s41591-022-01894-0](https://doi.org/10.1038/s41591-022-01894-0)] [Medline: [35864252](https://pubmed.ncbi.nlm.nih.gov/35864252/)]
34. Henry KE, Adams R, Parent C, et al. Factors driving provider adoption of the TREWS machine learning-based early warning system and its effects on sepsis treatment timing. *Nat Med*. Jul 2022;28(7):1447-1454. [doi: [10.1038/s41591-022-01895-z](https://doi.org/10.1038/s41591-022-01895-z)] [Medline: [35864251](https://pubmed.ncbi.nlm.nih.gov/35864251/)]
35. Lee BY, Hong SB. Rapid response systems in Korea. *Acute Crit Care*. May 2019;34(2):108-116. [doi: [10.4266/acc.2019.00535](https://doi.org/10.4266/acc.2019.00535)] [Medline: [31723915](https://pubmed.ncbi.nlm.nih.gov/31723915/)]
36. Song MJ, Lee YJ. Strategies for successful implementation and permanent maintenance of a rapid response system. *Korean J Intern Med*. Sep 2021;36(5):1031-1039. [doi: [10.3904/kjim.2020.693](https://doi.org/10.3904/kjim.2020.693)] [Medline: [34399572](https://pubmed.ncbi.nlm.nih.gov/34399572/)]
37. Subbe CP, Bannard-Smith J, Bunch J, et al. Quality metrics for the evaluation of rapid response systems: Proceedings from the third international consensus conference on rapid response systems. *Resuscitation*. Aug 2019;141:1-12. [doi: [10.1016/j.resuscitation.2019.05.012](https://doi.org/10.1016/j.resuscitation.2019.05.012)] [Medline: [31129229](https://pubmed.ncbi.nlm.nih.gov/31129229/)]
38. Haegdorens F, Van Bogaert P, Roelant E, et al. The introduction of a rapid response system in acute hospitals: a pragmatic stepped wedge cluster randomised controlled trial. *Resuscitation*. Aug 2018;129:127-134. [doi: [10.1016/j.resuscitation.2018.04.018](https://doi.org/10.1016/j.resuscitation.2018.04.018)] [Medline: [29679694](https://pubmed.ncbi.nlm.nih.gov/29679694/)]
39. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized clinical trials of machine learning interventions in health care: a systematic review. *JAMA Netw Open*. Sep 1, 2022;5(9):e2233946. [doi: [10.1001/jamanetworkopen.2022.33946](https://doi.org/10.1001/jamanetworkopen.2022.33946)] [Medline: [36173632](https://pubmed.ncbi.nlm.nih.gov/36173632/)]
40. Jain SS, Goto S, Hall JL, et al. Pragmatic approaches to the evaluation and monitoring of artificial intelligence in health care: a science advisory from the American Heart Association. *Circulation*. Dec 9, 2025;152(23):e433-e442. [doi: [10.1161/CIR.0000000000001400](https://doi.org/10.1161/CIR.0000000000001400)] [Medline: [41208719](https://pubmed.ncbi.nlm.nih.gov/41208719/)]
41. Difonzo M. Performance of the afferent limb of rapid response systems in managing deteriorating patients: a systematic review. *Crit Care Res Pract*. 2019;2019:6902420. [doi: [10.1155/2019/6902420](https://doi.org/10.1155/2019/6902420)] [Medline: [31781390](https://pubmed.ncbi.nlm.nih.gov/31781390/)]
42. Alhmoud B, Bonnici T, Patel R, Melley D, Williams B, Banerjee A. Performance of universal early warning scores in different patient subgroups and clinical settings: a systematic review. *BMJ Open*. Apr 8, 2021;11(4):e045849. [doi: [10.1136/bmjopen-2020-045849](https://doi.org/10.1136/bmjopen-2020-045849)] [Medline: [36044371](https://pubmed.ncbi.nlm.nih.gov/36044371/)]
43. Baker K, Rodger J. Assessing causes of alarm fatigue in long-term acute care and its impact on identifying clinical changes in patient conditions. *Inform Med Unlocked*. 2020;18:100300. [doi: [10.1016/j.imu.2020.100300](https://doi.org/10.1016/j.imu.2020.100300)]
44. Cvach M. Monitor alarm fatigue: an integrative review. *Biomed Instrum Technol*. 2012;46(4):268-277. [doi: [10.2345/0899-8205-46.4.268](https://doi.org/10.2345/0899-8205-46.4.268)] [Medline: [22839984](https://pubmed.ncbi.nlm.nih.gov/22839984/)]
45. Ruskin KJ, Hueske-Kraus D. Alarm fatigue: impacts on patient safety. *Curr Opin Anaesthesiol*. Dec 2015;28(6):685-690. [doi: [10.1097/ACO.0000000000000260](https://doi.org/10.1097/ACO.0000000000000260)] [Medline: [26539788](https://pubmed.ncbi.nlm.nih.gov/26539788/)]
46. Jacobs I, Nadkarni V, Bahr J, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: update and simplification of the Utstein templates for resuscitation registries. *Resuscitation*. Dec 2004;63(3):233-249. [doi: [10.1016/j.resuscitation.2004.09.008](https://doi.org/10.1016/j.resuscitation.2004.09.008)] [Medline: [15582757](https://pubmed.ncbi.nlm.nih.gov/15582757/)]
47. Orosz J, Bailey M, Udy A, Pilcher D, Bellomo R, Jones D. Unplanned ICU admission from hospital wards after rapid response team review in Australia and New Zealand. *Crit Care Med*. Jul 2020;48(7):e550-e556. [doi: [10.1097/CCM.0000000000004353](https://doi.org/10.1097/CCM.0000000000004353)] [Medline: [32304417](https://pubmed.ncbi.nlm.nih.gov/32304417/)]
48. Lin CS, Liu WT, Tsai DJ, et al. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. *Nat Med*. May 2024;30(5):1461-1470. [doi: [10.1038/s41591-024-02961-4](https://doi.org/10.1038/s41591-024-02961-4)] [Medline: [38684860](https://pubmed.ncbi.nlm.nih.gov/38684860/)]

49. Maharaj R, Raffaele I, Wendon J. Rapid response systems: a systematic review and meta-analysis. *Crit Care*. Jun 12, 2015;19(1):254. [doi: [10.1186/s13054-015-0973-y](https://doi.org/10.1186/s13054-015-0973-y)] [Medline: [26070457](#)]
50. Woertman W, de Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S. Stepped wedge designs could reduce the required sample size in cluster randomized trials. *J Clin Epidemiol*. Jul 2013;66(7):752-758. [doi: [10.1016/j.jclinepi.2013.01.009](https://doi.org/10.1016/j.jclinepi.2013.01.009)] [Medline: [23523551](#)]
51. Baek H, Cho M, Kim S, Hwang H, Song M, Yoo S. Analysis of length of hospital stay using electronic health records: a statistical and data mining approach. *PLoS One*. 2018;13(4):e0195901. [doi: [10.1371/journal.pone.0195901](https://doi.org/10.1371/journal.pone.0195901)] [Medline: [29652932](#)]
52. Lamperti M, Romero CS, Guarracino F, et al. Preoperative assessment of adults undergoing elective noncardiac surgery: updated guidelines from the European Society of Anaesthesiology and intensive care. *Eur J Anaesthesiol*. Jan 1, 2025;42(1):1-35. [doi: [10.1097/EJA.0000000000002069](https://doi.org/10.1097/EJA.0000000000002069)] [Medline: [39492705](#)]
53. Kim M, Yoo D, Noh E, et al. Implementation of an AI-based clinical decision support system predicting in-hospital cardiac arrest in general wards: a multicenter staggered-implementation study in secondary hospitals in Korea. *Diagnostics (Basel)*. May 29, 2026;16(11):1682. [doi: [10.3390/diagnostics16111682](https://doi.org/10.3390/diagnostics16111682)] [Medline: [42279550](#)]
54. Li F, Hughes JP, Hemming K, Taljaard M, Melnick ER, Heagerty PJ. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: an overview. *Stat Methods Med Res*. Feb 2021;30(2):612-639. [doi: [10.1177/0962280220932962](https://doi.org/10.1177/0962280220932962)] [Medline: [32631142](#)]
55. Norton EC, Miller MM, Kleinman LC. Computing adjusted risk ratios and risk differences in Stata. *Stata J*. Sep 2013;13(3):492-509. [doi: [10.1177/1536867X1301300304](https://doi.org/10.1177/1536867X1301300304)]
56. Haukoos JS, Lewis RJ. The propensity score. *JAMA*. Oct 20, 2015;314(15):1637-1638. [doi: [10.1001/jama.2015.13480](https://doi.org/10.1001/jama.2015.13480)] [Medline: [26501539](#)]
57. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the e-value. *Ann Intern Med*. Aug 15, 2017;167(4):268-274. [doi: [10.7326/M16-2607](https://doi.org/10.7326/M16-2607)] [Medline: [28693043](#)]
58. Wang P, Li Y, Reddy CK. Machine learning for survival analysis. *ACM Comput Surv*. Nov 30, 2019;51(6):1-36. [doi: [10.1145/3214306](https://doi.org/10.1145/3214306)]
59. Finlayson SG, Subbaswamy A, Singh K, et al. The clinician and dataset shift in artificial intelligence. *N Engl J Med*. Jul 15, 2021;385(3):283-286. [doi: [10.1056/NEJMc2104626](https://doi.org/10.1056/NEJMc2104626)] [Medline: [34260843](#)]
60. Hillman K, Chen J, Cretikos M, et al. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. *Lancet*. 2005;365(9477):2091-2097. [doi: [10.1016/S0140-6736\(05\)66733-5](https://doi.org/10.1016/S0140-6736(05)66733-5)] [Medline: [15964445](#)]
61. Bienefeld N, Boss JM, Lüthy R, et al. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. *NPJ Digit Med*. May 22, 2023;6(1):94. [doi: [10.1038/s41746-023-00837-4](https://doi.org/10.1038/s41746-023-00837-4)] [Medline: [37217779](#)]
62. Lee AY, Yanagihara RT, Lee CS, et al. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. *Diabetes Care*. May 2021;44(5):1168-1175. [doi: [10.2337/dc20-1877](https://doi.org/10.2337/dc20-1877)] [Medline: [33402366](#)]
63. Qin ZZ, Van der Walt M, Moyo S, et al. Computer-aided detection of tuberculosis from chest radiographs in a tuberculosis prevalence survey in South Africa: external validation and modelled impacts of commercially available artificial intelligence software. *Lancet Digit Health*. Sep 2024;6(9):e605-e613. [doi: [10.1016/S2589-7500\(24\)00118-3](https://doi.org/10.1016/S2589-7500(24)00118-3)] [Medline: [39033067](#)]

Abbreviations

AHA: American Heart Association
AI-EWS: AI-based Early Warning Score
AI-SaMD: AI-based software as a medical device
ARD: adjusted risk difference
ARR: adjusted risk ratio
ASE: Adult Sepsis Event
AUROC: area under the receiver operating characteristic curve
CCI: Charlson Comorbidity Index
CDC: Centers for Disease Control and Prevention
CoMET: Continuous Monitoring of Event Trajectories
CONCERN: Communicating Narrative Concerns Entered by Registered Nurses
CONSORT: Consolidated Standards of Reporting Trials
CPC: Cerebral Performance Category
CPR: cardiopulmonary resuscitation
CRC: clinical research coordinator
CRF: case report form
CRIS: Clinical Research Information Service

CRO: contract research organization
DECIDE-AI: Developmental and Exploratory Clinical Investigations of Decision Support systems driven by AI
DNR: do-not-resuscitate
DSMB: Data and Safety Monitoring Board
EMR: electronic medical record
EWS: early warning score
FAS: full analysis set
GLMM: generalized linear mixed model
HCP: health care professional
HIPAA: Health Insurance Portability and Accountability Act
ICC: intracluster correlation coefficient
ICH: International Council for Harmonisation
ICMJE: International Committee of Medical Journal Editors
ICTRP: International Clinical Trials Registry Platform
ICU: intensive care unit
IHCA: in-hospital cardiac arrest
IRB: institutional review board
ITT: intention-to-treat
JAMA: *Journal of the American Medical Association*
LOCF: last observation carried forward
LOS: length of stay
MERIT: Medical Early Response, Intervention and Therapy
MEWS: Modified Early Warning Score
ML: machine learning
NEWS2: National Early Warning Score 2
NRCT: nonrandomized controlled trial
PI: principal investigator
PPS: per-protocol set
PSM: propensity score matching
RCT: randomized controlled trial
RRR: relative risk reduction
RRS: rapid response system
SAP: statistical analysis plan
SCCM: Society of Critical Care Medicine
SHAP: Shapley Additive Explanations
SOFA: Sequential Organ Failure Assessment
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
SpO₂: oxygen saturation
SW-CRT: stepped-wedge cluster randomized trial
TMG: trial management group
UIT: unplanned intensive care unit transfer
VITAL-MD: validation of an AI-based clinical deterioration prediction model on patient outcomes through a stepped-wedge cluster randomized trial: multicenter evaluation of DeepCARS
WHO: World Health Organization

Edited by Javad Sarvestan; peer-reviewed by Gyu Chong Cho, Heather Brom; submitted 11 Jun.2026; final revised version received 11 Aug.2026; accepted 31 Aug.2026; published 19 Oct.2026

Please cite as:

Kim M, Kim KH, Kim M, Kwon SJ, Park MS, Kim AJ, Lee MJ, Park MH, Cha HK, Lee M, Jeong IB, Wang SJ, An SB, Lee JH, Cho KJ, Jang J, Jeong H, Lee YY, Han W, VITAL-MD Study Group, Kim JS, Yoo D

Clinical Effectiveness of AI-Based Cardiac Arrest Prediction System for General Ward Patients (VITAL-MD): Protocol for a Multicenter Pragmatic Stepped Wedge Cluster Randomized Controlled Trial

JMIR Res Protoc 2026;15:e104230

URL: <https://www.researchprotocols.org/2026/1/e104230>

doi: [10.2196/104230](https://doi.org/10.2196/104230)

© Mincheol Kim, Kyung Hoon Kim, Minsoo Kim, Sun Jung Kwon, Myung Soo Park, Ah Jin Kim, Man-Jong Lee, Mi Hwa Park, Hyung Keun Cha, Minhyeok Lee, In Beom Jeong, Soon Joo Wang, Sung Bak An, Jeon Hwa Lee, Kyung-Jae Cho, Jinhui Jang, HyoJin Jeong, Yoon Young Lee, Woori Han, VITAL-MD Study Group, Jung Soo Kim, Dongjoon Yoo. Originally published in JMIR Research Protocols (<https://www.researchprotocols.org>), 19.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Research Protocols, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.researchprotocols.org>, as well as this copyright and license information must be included.