Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 02 Aug 2026
Revised 16 Sep 2026
Accepted 04 Oct 2026

Natural Resources for Human Health 2026, Vol. 6 (23s): 198–210

KEYWORDS:
Artificial Intelligence; Remote Patient Monitoring; Hospitalization; Emergency Department; Heart Failure; Clinical Decision Support; Narrative Review; Implementation Science.

Artificial Intelligence–Enabled Remote Patient Monitoring and Acute-Care Utilization: A Critical Narrative Review of Comparative Clinical Evidence

M.B. Habeeb[1], Fatima Habib Allah Salama[2], Gawahir Murad Abdelrahman Alzein[3], Mayada Farah Obaid[4], Malaz Taglwy[5], Maysaa Ahmed[6], Ashruf Ishag Mohammed Elhassan[7], Abdelazim Ahmed Idris[8], Aseel Alswairky[9] , Aiyda Mohammed Mohammed Ahmed[10], Osama Mohammed Abdelnabee Ali[11], Nisreen Abbas Mohammed Ali Osman[12], Adl Hagelarabi[13], Awadalla Abdelwahid[14*]

[1]Department of obstetrics and Gynecology, Albutana University, Geziera, Sudan
Mohamedbakry31817@gmail.com
[2]Consultant of Obstetrics and Gynecology, Adele Medical Clinics Complex, Riyadh, Saudi Arabia.
fatima19801111@gmail.com
[3]Obstetrics and Gynecology Consultant, Maternity and Children Hospital Alkharj, Alkharj, Saudi Arabia.
totaelzein2017@gmail.com
[4]Pulmonolgy Senior Registrar, Altawiin Medical Group, Alkharj, Saudi Arabia.
Email: mayada_farah@yahoo.com
ORCID ID: 0009-0005-2095-5044
[5]Consultant of Obstetrics and Gynecology, Wellness Experts Medical Center, Riyadh, Saudi Arabia.
malazbabekir990@gmail.com
[6]Obstetrics and Gynecology Consultant, Dallah Hospital, Riyadh, Saudi Arabia.
Email: maysaaahmed00@gmail.com
[7]Community Medicine Specialist, Infection Prevention and Control director, Namerah General Hospital ,Al Qunfudah, Saudi Arabia.
ashruf.ishag@gmail.com
[8]Obstetrics&Gynecology Specialist, Jazan Health Cluster, Jazan, Saudi Arabia.
d.abazim@gmail.com
[9]Department of Obstetrics and Gynecology, Specialized Medical Center (SMC1), Riyadh, Saudi Arabia.
alswairkyaseel@gmail.com
[10]Family Medicine Registrar, Home Health care Department, King Khalid Hospital, Najran
Email: draiydamohd@gmail. com
[11]Registrar Of Family Medicine, Armed Forces Hospital Southern Region, Ministry of Defense Health Services, Khamis Mushait, Saudi Arabia.
oma_dr@yahoo.com
ORCID 0009-0007-3413-9798

Natural Resources for Human Health
eISSN: 2583-1194

[12]Obstetrics and Gynecology Specialist, MOH,Almokhwa Hospital, Albaha,Saudi Arabia.
Email: nsrynbasmhmd@gmail.comn
[13] Department of Paediatrics and Child Health, University of Medical Sciences and Technology, Khartoum, Sudan
Email: Dradlhight13@gmail.com
ORCID ID: 0009-0000-2431-0239
[14]Department of Obstetrics and Gynecology, Alneelain University, Khartoum, Sudan.
Email: awadallaabdelwahid@neelain.edu.sd
*Corresponding Author: Awadalla Abdelwahid
Department of Obstetrics and Gynecology, Alneelain University, Khartoum, Sudan.
Email: awadallaabdelwahid@neelain.edu.sd
ORCID ID : 0009-0008-3102-2786

ABSTRACT:

Background: AI is increasingly embedded in RPM (remote patient monitoring) to detect deterioration, prioritize alerts and support outpatient intervention. Does AI add clinical benefit beyond monitoring infrastructure and clinical staffing?

Objective: To critically appraise comparative evidence linking AI-enabled RPM to hospitalization, ED utilization and other patient important outcomes; to separate algorithm specific effects from those of broader care programmes; and to distinguish algorithmspecific effects from those of broad care programmes.

Evidence Synthesis: This focused, non-systematic narrative review critically summarizes five principal comparative evaluations and five contextual studies of related alert-driven or hybrid telemedicine pathways (mostly heart failure and postoperative care). In randomized LINK-HF2 feasibility trial, rehospitalization for heart failure occurred in 13/80 AI-notification recipients vs 17/91 controls (HR 0.90; 95% CI 0.43–1.84). In a nonrandomized conversational AI pilot death or hospitalization for heart failure was an exploratory composite favorable (unadjusted HR 0.39; 95% CI 0.16–0.96) but hospitalization alone inconclusive. ED attendance occurred in 2/62 (3/58) patients in a 120-patient randomized postoperative wound-monitoring feasibility trial in the same condition. Signals of reduced utilization from bundled care programmes cannot be attributed uniquely to AI.

Conclusions: None of the comparisons we selected show fewer admissions or ED visits due to AI alone. Clinical effect depends on data completeness, alerts that clinicians can act upon, treatment available to clinicians and implementation equity. Future trials should isolate incremental algorithmic value and patient safety alongside utilization.

1. INTRODUCTION

Remote patient monitoring (RPM) is collection, transmission and clinical interpretation of measurements or symptoms outside encounters: connected blood pressure devices, pulse oximeters, wearable biosensors implanted devices, mobile applications and telephone symptom collection. It's supposed to improve longitudinal awareness of disease and enable earlier intervention (following hospitalization or in people with chronic illness). But RPM is a service architecture rather than a treatment: monitoring frequency, completeness of data capture staffing alert escalation and therapies available to respond can vary widely [1,2]. Admissions and ED visits are consequential outcomes because they are patient important (expensive), financially costly (patient important) and responsive to early clinical action (but also hard to modify). Heart-failure decompensation, postoperative infection and respiratory deterioration follow different clinical pathways; some admissions are appropriate care

in response to warning not failure of intervention. So fewer alerts, better risk prediction and more nursing contact should not be taken as interchangeable proxies for fewer hospital visits. Any clinical evaluation of a monitoring pathway should specify an explicit path from prediction to clinician decision to treatment and outcome. The wider RPM literature is helpful context but does not quantify AI-specific effect: Taylor et al. [1] showed variation in acute care outcomes among RPM interventions; Smedslund et al. [3] found fewer participants hospitalized among noncommunicable disease patients in synthesis of conventional and digital RPM while effects on ED attendance were less clear; De Lathauwer et al. [4] demonstrated education, self-management and video interaction may modify heart failure RPM outcomes. None of these results can be attributed solely to algorithmic decision making. As RPM matures ML, NLP and computer vision will move monitoring from recording isolated measurements to anticipating clinically actionable deterioration (a wearable sensor identifies your personalised physiological pattern; automated telephone system detects concerning change in symptoms; postoperative image classifier prioritises wound assessment). These functions are fundamentally different from basic data transmission but "AI-enabled" is often used inconsistently for predictive models, fixed threshold rules and proprietary multiparameter indices. These need to be distinguished when interpreting reported effectiveness [5,6] . The clinical question is also different from the traditional one asked of diagnostic algorithms: even a model that correctly distinguishes between stable and deteriorating patients may not have an effect on admissions if alerts are delayed/ignored/misleading or disconnected from effective treatment; remote care programme may reduce hospitalization through increased nursing contact and medication adjustment but cannot show any unique contribution from AI. This distinction between algorithm, deployment and care system motivates the present narrative review. This review asks a narrower question than RPM reviews: what comparative clinical evidence shows incremental benefit from AI generated interpretation (not sensors alone, follow-up intensity alone or staffing alone)? It sorts selected evidence by causal contrast each study can support, path from detection to treatment and reporting/implementation priorities. Hospitalization, ED use, mortality, quality of life and safety are regarded as separate rather than interchangeable endpoints.

2. SCOPE AND APPROACH OF THE NARRATIVE REVIEW

This is a critical narrative review (not systematic review) article. The evidence base consists of ten comparative investigations identified during preparation for an earlier evidence synthesis: five main evaluations of AI-containing RPM pathways and five contextual evaluations of automated alerting, proprietary indices or hybrid digital care. Additional systematic reviews and reporting guideline publications were used to guide interpretation. The current revision checked the named clinical reports against their published bibliographies and available primary source abstracts or full texts (studies released online by November 2026). This targeted verification was not a fresh exhaustive database search. Studies were selected on the basis of (i) deployment of remotely collected physiological, symptom or image data in a clinical service; (ii) explicit comparator (usual care, AI on versus AI off or historical cohort); (iii) hospitalisation, ED utilisation, mortality, patient reported outcomes or safety reporting; (iv) intervention description sufficient to determine whether algorithm was different from care pathway. Randomized feasibility trials, nonrandomized controlled pilots, matched historical comparisons and difference-in-differences analyses are described separately. The three interpretive contrasts are incremental AI effect, total AI-containing programme effect and contextual algorithm-assisted effects. No pooled clinical estimate or formal risk-of-bias score is reported because this is a selective synthesis of clinically heterogeneous studies. Selection was interpretive not exhaustive: no prespecified registration history, reproducible multi-database search history, independent duplicate screening history or complete exclusion log is available; so the ten studies are the papers discussed here not an assertion about how many were eligible in the field. SANRA quality principles guided structure of this critical synthesis [7]. Limitations of study designs are judged descriptively (randomisation, baseline severity, completeness of follow-up, concurrent clinical contact, endpoint precision and uncertainty about AI component). These distinctions prevent selective evidence from being misread as systematic estimate of effectiveness.

3.1 From observation to personalized clinical prioritization: Conventional RPM collects information that repeats outside an encounter but care is still done by someone who reviews those data and decides what to do with them (anomaly detection in high frequency biosignals; individualized risk estimation; NLP extraction of patient-reported symptoms or image classification of postoperative findings). The expected advantage is not just faster computation — ideally the system will identify change sooner, avoid avoidable overload of information and direct limited clinician time toward patients who would benefit from intervention. Value of an AI feature depends on use in clinical practice. A model that forecasts well but doesn't change triage cannot plausibly change treatment. Nor does a fixed rule become machine learning intervention because it is embedded within a sophisticated digital platform. Clinical reports should specify inputs, training and deployment settings, thresholds,

model updates, notification recipients, human oversight and resulting actions [8,9]. where prediction-model evidence is considered TRIPOD+AI adds further standards for transparent model reporting [10].

3.2 Clinical comparator

An AI-on versus AI-off design (same devices and staff support) is informative about how much extra value algorithmic interpretation adds. When the experimental arm instead gets sensors, frequent calls, nurse review and early treatment while controls get usual care then the trial tests the whole care package. Both questions matter to health systems but attributing an outcome improvement at programme level solely to AI can mislead procurement and implementation decisions.

Table 1. Interpretive categories for comparative AI-enabled RPM evidence.

Evidence class	Comparison	What the comparison estimates	Key interpretation constraint
Incremental AI effect	AI-on versus AI-off; similar devices, staffing and care access	Added value of algorithmic output or notification	Masking, clinician spillover and differential response protocols may remain
AI-containing programme	AI-assisted RPM pathway versus usual care or another service	Total effect of a clinical technology-and-staffing bundle	Device access, contact frequency and therapies may explain part of effect
Algorithm-assisted contextual evidence	Proprietary indices, automated rules, or response-based observational comparisons	Effect of an alert programme or observed response	Technology may not meet defined AI/ML criteria; confounding may be substantial

These categories distinguish causal questions rather than formally ranking study quality. The operational classifications in Table 2 are based on descriptions in the source reports; no independent algorithm audit was performed.

Table 2. Operational classification of the ten comparative evaluations discussed in this review.

Study	Reported AI or alert function	Clinical comparator	Interpretation for AI attribution
Sideris et al. (2026) [5]	Personalized physiological anomaly detection with AI alerts	Same wearable infrastructure; AI notification versus control	Most direct incremental-AI comparison; exploratory clinical endpoints
Olivella et al. (2026) [11]	NLP-enabled conversational symptom acquisition and alerts	AI-supported calls versus usual care	NLP effect inseparable from contact frequency and nursing actions
Sultan et al. (2026) [12]	Wearable data, fixed rules and machine-learning alerts	Nurse-led programme versus matched historical care	Bundle effect; severity and clinician-contact imbalances
Jamal et al. (2026) [13]	AI-supported remote patient care; component details limited	Programme adoption versus matched nonadoption	Service-level effect; AI-specific contribution unidentified
Rochon et al. (2026) [6]	AI-assisted postoperative image prioritization	Wound-platform pathway versus usual care	Platform feasibility; not an isolated algorithm comparison

Study	Reported AI or alert function	Clinical comparator	Interpretation for AI attribution
Alcoceba-Herrero et al. (2024) [14]	Automatic alert management; model architecture incompletely specified	Wearables plus app versus app-based follow-up	Device and monitoring-intensity differences; feasibility
Yu et al. (2025) [15]	Hybrid telemedicine with AI-assisted consultations	Multicomponent telemedical care versus usual care	AI effect cannot be separated from telemedicine package
Assad et al. (2026) [16]	HeartLogic proprietary multiparameter alert index	Patient-facing alert pathway versus standard care	Alert-programme benefit; prospective ML status not established
Feijen et al. (2023) [17]	HeartLogic-enhanced device-alert care	Propensity-matched standard care	Observational pathway effect; no isolated ML contrast
Oka et al. (2026) [18]	HeartLogic alerts and observed clinical response	Alert response versus nonresponse	Post-exposure selection; confounding by response and care access

Classification is based on each publication's description, not an independent source-code or model audit. "AI-supported" does not imply that a trained machine-learning model was independently validated. NLP, natural-language processing.

4. CLINICAL EVIDENCE ACROSS MONITORING APPLICATIONS

4.1 Physiological AI notification after heart-failure hospitalization

In the multicentre LINK-HF2 randomized feasibility trial 176 individuals were randomised (171 in the reported analyses, 80 to AI notification and 91 to control). Both groups wore a continuous multisensory patch but clinicians only received structured AI notifications from intervention group. During 90 days, rehospitalisation for HF occurred in 13/80 versus 17/91 patients (16.3% vs 18.7%; HR 0.90, 95% CI 0.43–1.84; p=0.76) and exploratory composite of all cause unplanned rehospitalization, ED visit or death occurred in 29/80 versus 34/91 patients (HR 1.03, 95% CI 0.63–1.70). A post hoc physiological outcome favoured intervention: AI-detected instability was lower (HR 0.76, 95% CI 0.60–0.97) but intermediate signal did not translate into clear clinical benefit. Monitoring adherence covered a median 72.4% of follow-up days; clinicians responded to notifications within a median one day. The trial was designed for feasibility and exploratory efficacy only [5] . LINK-HF2 is particularly relevant because the intervention and control groups used the same wearable monitoring infrastructure. Its reported results show that more sophisticated physiological notification might change monitoring signals but not clearly change near term admissions. The trial wasn't powered to exclude modest clinical benefit and its short follow up limits interpretation but it challenges the assumption that improved detection automatically translates into hospitalization prevention. For health systems this means how quickly a clinician reviews an alert, whether the warning is trusted and if outpatient therapies can be deployed before deterioration becomes irreversible [5].

4.2 Conversational AI and structured symptom surveillance

Olivella et al. [11] compared 86 heart-failure outpatients managed weekly using NLP-enabled telephone calls with 40 contemporaneous usual-care patients. The system completed 3,919 of 4,272 scheduled calls (91.7%) and produced 1,962 alerts followed by 648 actions. At 12 months the unadjusted exploratory composite of all cause death or heart-failure hospitalization was AI arm HR 0.39 (95% CI 0.16–0.96); heart-failure hospitalization alone was not statistically significant (sub distribution HR 0.43; 95% CI 0.16–1.13). Cardiovascular death occurred in approximately 1.2% versus 10.0%; intervention group had within group improvement in KCCQ-12 score from baseline reported as 7.13 points. The between-group clinical signal is hypothesis generating because baseline severity of illness, diuretic use and contact with clinicians cannot be separated in a small nonrandomized study. The strength of conversational systems is that people who may not reliably use wearables get the same symptom acquisition. Weekly automated calls can standardize questions and send abnormal reports to nurses but changes information gathering as well as how much clinical attention patients receive. The positive composite outcome in Olivella pilot

cannot tell us whether it was NLP or more surveillance, proactive nursing action or baseline differences between groups; evidence for reduced heart-failure hospitalization itself is unclear [11].

4.3 Wearables embedded in nurse led post discharge care

In Heart Failure Cascade evaluation Sultan et al. [12] reported 39 patients completing post discharge wearable monitoring with rules-based and machine learning alerts reviewed by home health nurses compared with matched retrospective patients. The reported adjusted relative estimate for 30-day readmission was 0.31 (95% CI 0.06–1.38; p=0.138) with confidence interval spanning null and no statistically significant reduction. Baseline NYHA class was still imbalanced and intervention patients had more clinical contacts and diuretics escalated so attribution was limited to algorithm. This model is pragmatic: clinical staff interpret alerts and change treatment; makes biological pathway credible (recognition of worsening congestion leads early medication adjustment or assessment); historical controls, differential clinical contacts make attribution difficult. Confidence interval is also consistent with benefit as well as no benefit so point estimate should not be promoted as established effectiveness [12].

4.4 Integrated AI-supported care programmes

Jamal et al. [13] performed a retrospective difference-in-differences evaluation of an AI supported remote-care programme on 500 patients against propensity score matched controls over 12 months after programme activation. Coefficient for 30-day readmission (−0.0381) (p<0.0001) was reported; ED visits (−0.0043), length of stay (−0.2044) were also negative. These model coefficients should not be translated into percentage-point reductions universally applicable without the original variable coding and model specification. Parallel-trends assumption, timing of intervention and simultaneous health-system changes need to be appraised before causal interpretation can be made. This was a multicomponent care programme not an AI on versus AI off comparison [13]. The distinction between an algorithm and its service context is particularly important for large programmes where deployment occurs at the same time as new staffing models, follow up or other digital infrastructure are put in place. Difference-in-differences analysis can help address stable differences between groups but still depends on parallel trends assumption, comparable outcome capture and absence of other change linked to rollout of programme. A favourable programme coefficient means something about redesign of service not that the AI model caused fewer admissions [13].

4.5 AI-assisted postoperative wound monitoring

Rochon et al [6] randomized 120 postoperative cardiac surgery patients to an AI prioritized digital wound platform (n=62) vs usual care (n=58). Exploratory ED attendance was reported in 2/62 vs 3/58; of those with wound assessment surgical site infections occurred in 7/61 vs 9/55. Intervention supported postop wound-image capture, automated prioritization and clinician review; study was explicitly feasibility, acceptability and safety trial rather than powered clinical effectiveness trial. Clinically it's different from physiologic heart failure monitoring and should not be pooled indiscriminately with it.

Postoperative imaging has a different risk profile. Maybe we want to speed up appropriate wound review rather than minimize every ED visit. A low ED event count in feasibility trial is reassuring about apparent utilization but cannot prove treatment effect. The pathway raises questions too about image quality, skin-tone performance, missed infection, patient anxiety and clinical responsibility for flagging images [6].

Table 3. Selected comparative evidence on AI-containing RPM pathways and acute-care utilization. Effect measures, follow-up periods and statistical estimands differ and are not pooled.

Source and design	Participants / AI application	Reported clinical result	Important limitation
Sideris et al. (2026); LINK-HF2 randomized feasibility trial [5]	80 AI notification vs 91 control analysed; wearable anomaly detection	HF readmission 13/80 vs 17/91; HR 0.90 (95% CI 0.43–1.84), 90 days	Feasibility focus; exploratory, imprecise clinical endpoints

Natural Resources for Human Health
eISSN: 2583-1194

Source and design	Participants / AI application	Reported clinical result	Important limitation
Olivella et al. (2026); prospective controlled pilot [11]	86 NLP-assisted phone follow-up vs 40 usual care	Death/HF hospitalization: unadjusted HR 0.39 (0.16–0.96)	Nonrandomized; treatment and baseline severity differences
Sultan et al. (2026); matched historical comparison [12]	39 monitoring completers; rules and ML alerts, nurse escalation	Adjusted 30-day readmission relative estimate 0.31 (0.06–1.38)	Small sample; NYHA imbalance; intensified clinical contacts
Jamal et al. (2026); retrospective DiD [13]	500-patient AI-supported remote-care programme; matched controls	Readmission DiD coefficient −0.0381 ($p<0.0001$)	Assumptions and scale require full-model appraisal; bundled programme
Rochon et al. (2026); randomized feasibility trial [6]	62 AI-enabled wound platform vs 58 usual care	ED attendance 2/62 vs 3/58	Different clinical indication; not powered for event reduction

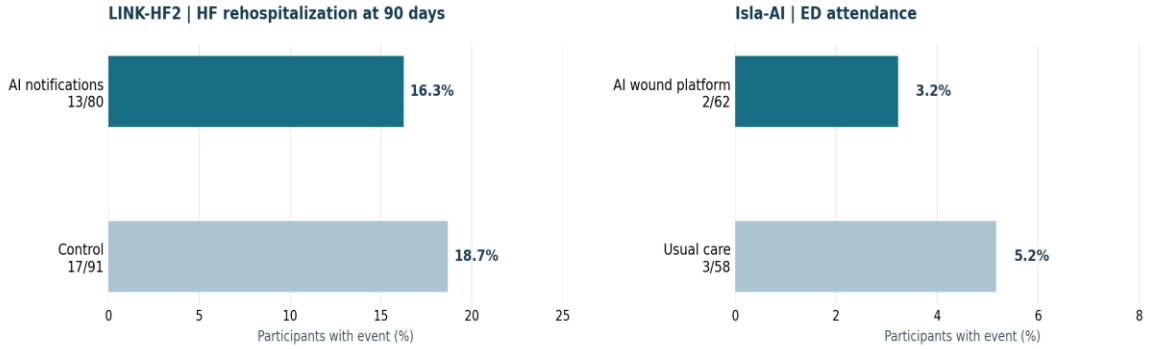


Figure 1. Descriptive event proportions in two distinct randomized feasibility studies. LINK-HF2: heart-failure rehospitalization within 90 days (13/80 versus 17/91); Isla-AI: exploratory ED attendance following cardiac surgery (2/62 versus 3/58). Percentages are rounded to one decimal place. Outcomes and follow-up differ; these are not pooled effect estimates.

4.6 Related but distinct algorithm-assisted and hybrid telemedicine evidence

Other comparative trials broaden the clinical context but make AI attribution less certain: Alcoceba-Herrero et al. [14] randomized 342 COVID-19 home monitor patients; per-protocol analysis used 247 (103 intervention, 144 control); groups had access to wearable devices and monitoring intensity different; feasibility and alerts were emphasized rather than clearly isolated hospitalization estimand in trial. Yu et al. [15] tested telemedical package post PCI with AI consultation/monitoring (not algorithm assistance). In this latter randomized trial major adverse cardiovascular and cerebrovascular events occurred in 36/1,040 intervention patient's vs 55/1,046 controls; neither trial isolates incremental effect of AI on admission rates.

NOTIFY-HF randomly assigned 160 individuals with implanted cardiac devices to HeartLogic-based patient-facing alerts or standard care; at least one hospitalization was reported in 19/81 intervention patients and 34/79 controls (exploratory analysis HR all cause hospitalization 0.51 [95% CI 0.29–0.90]). These are about a proprietary multiparameter alert pathway; they are

contextual or sensitivity evidence unless qualifying prospective AI functionality is shown [16]. Feijen et al. [17] compared 127 propensity-matched pairs and found fewer worsening heart-failure events with HeartLogic-enhanced care whereas Oka et al. [18] compared alert responsive vs nonresponsive patients in a retrospective cohort which is especially prone to confounding. All are separately reported from the strictly classified AI interventions.

These studies show how proactive alert pathways might be beneficial but they cannot be used as interchangeable demonstrations of ML efficacy. For instance, an intervention based on whether clinicians responded to an alert is subject to treatment-selection bias: responsiveness may depend on staffing, severity of clinical presentation, engagement of the patient and availability of effective therapy [16,17,18].

Table 4. Contextual algorithm-assisted and multicomponent monitoring studies.

Study	Selected observation	Reason for separate handling
Alcoceba-Herrero et al. (2024) [14]	342 randomized; app plus wearable monitoring / automatic alerts	The AI component and hospitalization estimand are not clearly isolated
Yu et al. (2025) [15]	PCI telemedicine RCT (1,040 vs 1,046); major cardiovascular/cerebrovascular events 36 vs 55	Broad consultation and monitoring package rather than isolated AI alerts
Assad et al. (2026); NOTIFY-HF [16]	81 vs 79; any hospitalization reported in 19 vs 34	HeartLogic proprietary multiparameter index; exploratory clinical outcomes
Feijen et al. (2023) [17]	127 propensity-matched pairs; fewer worsening-HF events reported	Observational effect of an enhanced algorithm-assisted care programme
Oka et al. (2026) [18]	246 patients; observed associations by alert response	Response determined after exposure; high risk of selection/confounding

Studies are shown for clinical context; AI architecture, comparators or primary outcomes differ from direct AI-specific notification evaluations.

5. WHY BETTER PREDICTION DOESN'T NECESSARILY REDUCE ACUTE CARE USE

The evidence has to be interpreted in terms of how a prediction might prevent an outcome. There is a pathway from reliable remote information acquisition to algorithmic signal, actionable triage decision, timely clinical intervention and finally patient-important result. A breakdown at any step can negate the proposed benefit. Figure 2 shows this as an explanatory model not a validated causal mechanism.



Figure 2. The proposed signal-to-action pathway for AI-enabled remote monitoring. Clinical effect requires successful transitions from data capture to model interpretation, triage, intervention and patient outcome; predictive discrimination alone is insufficient.

5.1 Data completeness and representativeness

Connectivity loss, sensor wear rate, device failure and symptom reporting irregularity can cause missingness when risk is highest. If training and validation data are skewed towards patients who use technology consistently then reported model

performance may overestimate real-world effectiveness. Language barriers, disability, older age and unequal access to connected devices add further limitations. Implementation studies should therefore report proportion of monitoring time covered, characteristics of nonusers and alternatives to digital only care.

5.2 Alert quality, calibration and workload

Sensitivity is not calibration (clinically useful calibration). A sensitive model might generate irrelevant or duplicated alerts; identify worsening illness too late to change its course; a conservative threshold may instead give false reassurance. In the deployment population investigators should measure false-positive burden, missed high-risk events, calibration, threshold changes, subgroup performance and model drift. Explanatory graphics are no substitute for external validation and safety evaluation [19,10]. Alert fatigue also erodes benefit if nominally accurate outputs compete with other clinical work.

5.3 Response ownership and treatment opportunity

Alert recipients, acknowledgement procedures, response deadlines and interventions available determine whether a risk signal becomes meaningful care action (a patient facing alert has a different causal pathway than a nurse reviewed notification; the former depends on patient understanding and access to care while the latter depends on staffing and escalation protocols). For heart failure an early warning is useful only when fluid status can be assessed and appropriate treatment started. For postoperative infection urgent review may increase ED use but prevent more serious harm so decrease in ED visits is not an unconditional marker of safer care).

6. MEASURING OUTCOMES SAFETY AND EQUITY

6.1 Patient-important outcomes and utilization interpretation: All-cause admissions, condition specific readmissions, time to first hospitalization and recurrent events are different outcomes with different mechanisms (an alert system could reduce severity or duration of admissions without changing proportion admitted; or it could increase early ED attendance and lower later emergency deterioration). Trials should therefore pair utilization with mortality, clinical severity, patient reported outcomes and adverse events. Composite outcomes have to be decomposed into their components (mortality and admission trends can move in different directions).

The selected studies show this. LINK-HF2 reported a composite of all-cause unplanned rehospitalization, ED attendance or death; its heart failure readmission endpoint was narrower. Olivella study had favourable composite but inconclusive heart failure hospitalization alone; wound-monitoring trial few ED events so that even numerically positive difference imprecise. Comparisons between different follow-up periods and outcome definitions should not be interpreted as if they estimated the same intervention effect [5,11,6].

6.2 Patient safety, fairness and unintended effects

Safety concerns go beyond algorithmic error rates. False alarms increase testing, anxiety, inappropriate escalation of clinical care and opportunity costs; a missed alert or data stream delays treatment or promotes reassurance. Studies should report human oversight, procedures for technical failure, alert acknowledgement, clinician override and downstream harms. Patients need to know what is being monitored (which abnormal findings will be reviewed), how urgently a response can be expected and when it might occur.

Equity should be measured as a feature of the deployed service not just the prediction model. Digital literacy, language, disability, device access, broadband coverage and caregiver support may determine who gets the intervention. If sample sizes are large then subgroup evaluation should account for calibration and outcome differences but small exploratory strata should not be overinterpreted — a clinically successful programme that excludes people with greatest need widens disparities even though average outcomes are good.

6.3 Workforce, cost and governance

An economic appraisal needs to take into account monitoring hardware, software licensing, integration with EHRs, clinician time, training, technical support, cybersecurity and costs of additional consultations. Avoided admissions cannot be translated into net savings without these inputs (and potentially urgent assessments). Implementers must assign alerts responsibility; audit changes in model version; establish incident-response and drift-monitoring procedures. The translational challenges are technical as well as organizational [20]; DECIDE-AI and CONSORT-AI define reporting standards for clinical deployment

and human–AI interaction (Va7. Clinical practice and future research implications. Cautionary adoption by indication is preferable to a blanket statement that AI-RPM prevents hospitalisation; health services may pilot implementation when clinical deterioration pathway is defined, alerts actionable, response capacity available and safety prospectively monitored. Evaluation should distinguish whether the system improves organisation of follow-up, overall effectiveness of care programme or incremental accuracy/utility of AI compared with simpler monitoring. Future comparative research should use sufficiently detailed intervention descriptions for replication, clear comparator pathways and patient important primary endpoints. Where possible trials can randomize availability of algorithm but keep sensors, staffing, clinical contact and response resources equivalent; where randomization is not possible investigators specify explicit target trial and account for baseline severity, prior utilisation, staffing, calendar time and cointerventions. Recurrent admissions, mortality, ED attendance, adverse events, quality of life, false negative deterioration, subgroup access and fidelity of response should be measured. Reporting should separate algorithm performance from treatment change and downstream clinical benefit [9,10].sey et al., 2022; Liu et al., 2020).

Table 5. Design priorities for future AI-enabled RPM evaluations.

Design problem	Preferred approach	Reason
Attribution of AI effect	Compare AI-on versus AI-off within otherwise similar RPM where ethical and feasible	Separates algorithmic value from device and staff contact
Inconsistent endpoints	Prespecify all-cause and disease-specific utilization, mortality, safety and patient-reported measures	Prevents selective emphasis on alert counts or post hoc composites
Sparse clinical events	Power for clinically meaningful differences and sufficient follow-up; collect recurrent events	Avoids interpreting non-significance as equivalence
Human factors and fidelity	Report alert receipt, acknowledgement, response latency, therapy, nonwear and version changes	Tests the complete detection-to-treatment pathway
Risk of confounding	Randomize where possible; otherwise emulate a target trial and justify adjustment/parallel trends	Strengthens causal interpretation
Generalisability and equity	Prespecify subgroup calibration and access measures; account for language, connectivity and literacy	Reduces unequal benefit and silent noncoverage

8. DISCUSSION

The main finding of this narrative synthesis is not that AI-enabled RPM is ineffective but that positive predictive and operational signals have caught up with clinical evidence to establish fewer hospitalizations specifically because of AI (the best isolated notification comparison in the literature was LINK-HF2 but no clear reduction in acute care events occurred during feasibility follow-up). In contrast favourable observational and quasi-experimental results often involve more care contact and larger service changes so these are implementation signals but they cannot settle how much benefit comes from algorithm rather than clinical programme around it. That distinction helps reconcile apparently conflicting reports. Sensor model flagging physiological instability and conversational system increasing contact with patients both improve surveillance but act at different parts of care pathway; effectiveness will depend on disease, baseline risk, frequency of measurements, signal quality, staffing,

treatments and outcome timing. Separating these dimensions gives more clinical insight than treating every digital intervention as the same exposure. Evidence also warns against making ED use reduced a universal success criterion. For suspected infection or acute heart failure deterioration earlier, urgent review can be an appropriate outcome. Future evaluation should ask whether monitoring changes avoidable escalation, prevents severe complications, reduces mortality or improves patient experience rather than just rewarding fewer contacts regardless of safety. Better outcomes may require coordinated system redesign even when AI plays a meaningful role. Finally, transparency in technology classification matters to clinicians and researchers too. Proprietary scores, fixed rules and trained ML models can all be clinically useful but have different mechanisms. A narrative review preserves that diversity not the implication that good experience with multiparameter alert index validates every implementation of AI. More convincing evidence will emerge when algorithm function, human response and patient outcome are evaluated as linked but distinguishable parts of care.

9. STRENGTHS AND LIMITATIONS

This review juxtaposes selected randomized and nonrandomized comparative evaluations, keeps source effect measures, separates incremental algorithm effects from full programme effects and proposes a clinically actionable signal-to-treatment framework. The main weakness is study selection bias. Ten comparative studies were not assembled by reproducible comprehensive search so publication and selection bias are possible; relevant evidence may be missing. Studies differ in populations, outcome definitions, follow-up periods and algorithm architectures so no pooled effectiveness estimate can be defensible; several reports do not describe algorithms, clinical response protocols or endpoint ascertainment; some compare treatment completers or historical controls so confounding is substantial. No duplicate screening has been done, independent extraction nor formal certainty-of-evidence assessment carried out on this synthesis. Numerical findings are descriptive results from individual publications not estimates of one overall treatment effect. Narrative-review design was chosen intentionally to be different from PRISMA 2020 systematic review/meta-analysis.

10. CONCLUSION

AI-enabled RPM provides a plausible way to detect deterioration earlier and intervene first but the comparative evaluations we selected don't consistently show reductions in hospitalization or ED use due to the algorithm alone. Small feasibility trials and multicomponent observational programmes ask different causal questions. Patient safety, timely clinical response, equity of access and utility must be measured alongside predictive accuracy. More powerful comparisons need to isolate incremental value of AI from surrounding care pathway before claims of admission prevention can be made generally.

Declarations

Ethics approval and consent to participate

Not applicable: this narrative review discusses published aggregate results and does not involve new human-participant data collection.

Funding

This research received no external funding.

Competing interests

Author confirmed no conflict of interest.

Data availability

No original participant-level data or new quantitative dataset were generated for this review.

Review design

Focused critical narrative review; no PRISMA systematic review or meta-analysis is claimed.

Author contributions

M.B. Habeeb: Conceptualization, Methodology, Investigation, Literature Review, Writing – Original Draft Preparation, and Project Administration.

Fatima Habib Allah Salama: Investigation, Literature Search, Data Curation, and Writing – Review & Editing.

Gawahir Murad Abdelrahman Alzein: Methodology, Investigation, Evidence Synthesis, and Writing – Review & Editing.

Mayada Farah Obaid: Investigation, Literature Screening, Data Extraction, and Writing – Review & Editing.

Malaz Taglwy: Investigation, Data Extraction, Data Curation, and Writing – Review & Editing.

Maysaa Ahmed: Investigation, Evidence Synthesis, Visualization, and Writing – Review & Editing.

Ashruf Ishag Mohammed Elhassan: Methodology, Validation, Critical Appraisal, and Writing – Review & Editing.

Abdelazim Ahmed Idris: Investigation, Literature Review, Data Interpretation, and Writing – Review & Editing.

Aseel Alswairky: Investigation, Data Curation, Validation, and Writing – Review & Editing.

Aiyda Mohammed Mohammed Ahmed: Investigation, Literature Screening, Data Extraction, and Writing – Review & Editing.

Osama Mohammed Abdelnabee Ali: Methodology, Evidence Interpretation, Critical Appraisal, and Writing – Review & Editing.

Nisreen Abbas Mohammed Ali Osman: Investigation, Literature Screening, Data Extraction, and Writing – Review & Editing.

Adl Hagelarabi: Investigation, Data Curation, Validation, and Writing – Review & Editing.

Awadalla Abdelwahid: Conceptualization, Methodology, Supervision, Project Administration, Validation, Writing – Original Draft Preparation, and Writing – Review & Editing.

All authors reviewed the manuscript critically for important intellectual content, approved the final version for publication, and agreed to be accountable for their respective contributions to the work.

REFERENCES

[1] Taylor, M.L., Thomas, E.E., Snoswell, C.L., Smith, A.C. and Caffery, L.J. (2021) 'Does remote patient monitoring reduce acute care use? A systematic review', BMJ Open, 11, e040232. https://doi.org/10.1136/bmjopen-2020-040232

[2] Iqbal, F.M. et al. (2021) 'Clinical outcomes of digital sensor alerting systems in remote monitoring: a systematic review and meta-analysis', npj Digital Medicine, 4, 7. https://doi.org/10.1038/s41746-020-00378-0

[3] Smedslund, G., Østerås, N. and Hestevik, C.H. (2025) 'Effects of remote patient monitoring on health care utilization in patients with noncommunicable diseases: systematic review and meta-analysis', JMIR mHealth and uHealth, 13, e68464. https://doi.org/10.2196/68464

[4] De Lathauwer, I.L.J. et al. (2025) 'Remote patient monitoring in heart failure: a comprehensive meta-analysis of effective programme components for hospitalization and mortality reduction', European Journal of Heart Failure, 27(9), pp. 1670–1685. https://doi.org/10.1002/ejhf.3568

[5] Sideris, K. et al. (2026) 'Artificial intelligence–enabled detection and management of physiological instability in heart failure: results of the randomized LINK-HF2 trial', JACC: Heart Failure, article 103422 (published online 9 October 2026). https://doi.org/10.1016/j.jchf.2026.103422

[6] Rochon, M. et al. (2026) 'AI-enabled digital wound monitoring after cardiac surgery: a randomised controlled feasibility, safety, and acceptability trial', Journal of Hospital Infection, 174, pp. 428–435. https://doi.org/10.1016/j.jhin.2026.04.020

[7] Baethge, C., Goldbeck-Wood, S. and Mertens, S. (2019) 'SANRA—a scale for the quality assessment of narrative review articles', Research Integrity and Peer Review, 4, 5. https://doi.org/10.1186/s41073-019-0064-8

[8] Vasey, B. et al. (2022) 'Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI', BMJ, 377, e070904. https://doi.org/10.1136/bmj-2022-070904

[9] Liu, X. et al. (2020) 'Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension', The Lancet Digital Health, 2(10), pp. e537–e548. https://doi.org/10.1016/S2589-7500(20)30218-1

[10] Collins, G.S. et al. (2024) 'TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods', BMJ, 385, e078378. https://doi.org/10.1136/bmj-2023-078378

[11] Olivella, A. et al. (2026) 'Conversational AI for remote monitoring in heart failure: a prospective controlled pilot study', European Heart Journal – Digital Health, 7(3), ztag032. https://doi.org/10.1093/ehjdh/ztag032

[12] Sultan, S. et al. (2026) 'Continuous remote patient monitoring in heart failure patients: the Heart Failure Cascade Study: phase II and III outcomes', Applied Clinical Informatics, 17(4), pp. 789–801. https://doi.org/10.1055/a-2946-7234

[13] Jamal, A., Tauseef, F., Naseer, F. and Nasim, F. (2026) 'Do AI-enabled remote patient monitoring programs reduce readmissions? A quasi-experimental evaluation of adoption timing and outcomes', International Journal of Medical Informatics, 222, 106728 (published online 24 September 2026). https://doi.org/10.1016/j.ijmedinf.2026.106728

[14] Alcoceba-Herrero, I. et al. (2024) 'Randomized controlled trial to assess the feasibility of a novel clinical decision support system based on the automatic generation of alerts through remote patient monitoring', Journal of Clinical Medicine, 13(19), 5974. https://doi.org/10.3390/jcm13195974

[15] Yu, X. et al. (2025) 'Efficacy of telemedical interventional management in patients with coronary heart disease undergoing percutaneous coronary intervention: randomized controlled trial', Journal of Medical Internet Research, 27, e63350. https://doi.org/10.2196/63350

[16] Assad, J. et al. (2026) 'NOTIFY-HF: a randomized pilot trial of patient-facing alerts for worsening heart failure using multiparameter monitoring', JACC: Heart Failure, 14(9), 103184. https://doi.org/10.1016/j.jchf.2026.103184

[17] Feijen, M. et al. (2023) 'Fewer worsening heart failure events with HeartLogic on top of standard care: a propensity-matched cohort analysis', Journal of Cardiac Failure, 29(11), pp. 1522–1530. https://doi.org/10.1016/j.cardfail.2023.04.012

[18] Oka, T. et al. (2026) 'Reduction in heart failure hospitalizations through real-world HeartLogic alert response: a multicenter study in Japan', Circulation Journal, 90(11), pp. 1666–1674. https://doi.org/10.1253/circj.CJ-26-0182

[19] Ghassemi, M., Oakden-Rayner, L. and Beam, A.L. (2021) 'The false hope of current approaches to explainable artificial intelligence in health care', The Lancet Digital Health, 3(11), pp. e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9

[20] Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G. and King, D. (2019) 'Key challenges for delivering clinical impact with artificial intelligence', BMC Medicine, 17, 195. https://doi.org/10.1186/s12916-019-1426-2