2026 INC International Nursing Conference

Nursing Intelligence in the AI Era:
Strategic Empowerment of Research, Education, and Practice

AI and Decision-Making

Clinical Implications

HYEONEUI KIM

Seoul National University College of Nursing

1

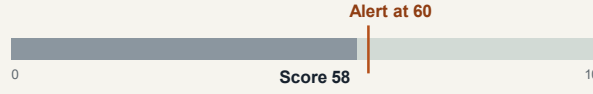
ILLUSTRATIVE CASE

A NIGHT SHIFT

03:12

Mr. Kim, 74, is on day 2 after abdominal surgery.

Sepsis risk score



Just below the threshold, so **no alert fires.** His vital signs are borderline.

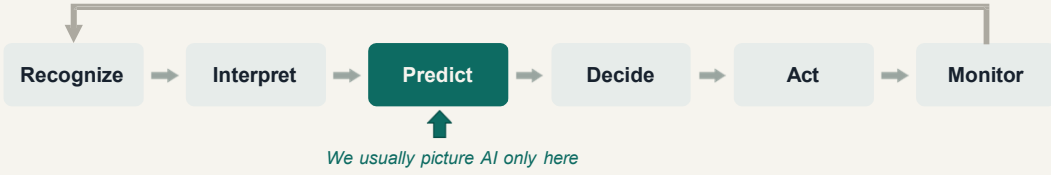
But his nurse noticed what no data field captures yet. During her night rounds, she found him awake. He seemed a little confused. Yesterday, he was clear. His daughter, staying the night, says, “He’s not himself.”

Should she call the doctor at 3 a.m.?
What can she show, besides a score below the threshold?

2

THE DECISION CYCLE

What does it actually mean for AI to support clinical decisions?



Clinical AI is not a one-time prediction tool, but a continuously learning decision-support system embedded in clinical workflow.

3

TODAY

Five parts

01	What AI can do: prediction and information synthesis
02	Hidden risks: how AI shapes our judgment and how models change over time
03	Redesigning human oversight
04	Nursing at the center
05	Closing: from better predictions to better decisions

4

01

What AI can do

Prediction and information synthesis

5

PREDICTION · A CAUTIONARY CASE

Even widely used prediction models can fail

Research

JAMA Internal Medicine | Original Investigation

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD; Erkin Otles, MEng; John P. Donnelly, PhD; Andrew Krumm, PhD; Jeffrey McCullough, PhD; Olivia DeTroyer-Cooley, BSE; Justin Pestrue, MEcon; Marie Phillips, BA; Judy Konye, MSN, RN; Carleen Penoza, MHSA, RN; Muhammad Ghous, MBBS; Karandeep Singh, MD, MMSc

A proprietary sepsis model used in hundreds of US hospitals

Tested outside its developer: 38,455 hospitalizations at one academic medical center

0.63

AUC for sepsis (0.5 = chance)

67%

of sepsis cases it missed

1,709 of 2,552

7%

of sepsis cases it caught before clinicians had already acted

183 of 2,552

18%

of all hospitalizations triggered an alert

8 patients evaluated to find 1 with sepsis

Missed patients, little added value, and alert fatigue, all at the same time.

Wong A, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med. 2021;181(8):1065–1070.

6

PREDICTION

From prediction to decision: four tests

1 The right question

Is the risk clinically meaningful?

2 The right data

Is the data available when the decision is made and does it reflect our patients??

3 Real added value

Does it tell clinicians something they don't already know?

4 A path to action

Who acts, by when, and what do they do?

A prediction is not yet a decision.

Most models pass the first three. Few are designed for the fourth.

7

OUR STUDY · THE FOUR TESTS IN PRACTICE

Predicting workplace violence in the ED

JOURNAL OF EMERGENCY NURSING VOLUME 49 • ISSUE 3 May 2023

PREDICTING WORKPLACE VIOLENCE IN THE EMERGENCY DEPARTMENT BASED ON ELECTRONIC HEALTH RECORD DATA

Authors: Hyungbok Lee, RN, Heeje Yun, MSN, RN, Minjin Choi, MSN, RN, and Hyeoneui Kim, PhD, MPH, RN, Seoul, Republic of Korea

Emergency department EHR data, 2016–2021

1,215 violence and 6,044 non-violence cases

1 The right question	Emergency nurses are frequently exposed to violence – a risk worth predicting and preventing
2 The right data	EHR data from the ED visit and stay; violence cases identified from nursing records
3 Real added value	Factors related to the ED stay (crowding, waiting, dissatisfaction) improved prediction (accuracy 0.90).
4 A path to action	**Not a flag on individual patients. Risk factors inform nurse education and ED operations**

Lee H, Yun H, Choi M, Kim H. Predicting workplace violence in the emergency department based on electronic health record data . J Emerg Nurs. 2023;49(3):415–424.

8

INFORMATION SYNTHESIS

Can AI reduce cognitive burden?

Information overload → **Extract what is relevant** → **Synthesize** → **Decide**

- The bottleneck in clinical decisions may not be prediction.
- It may be the *sheer volume of information* to process.
- AI can help with *decision preparation* by building the information state from which clinicians decide well.

9

OUR STUDY · ORGANIZE AND PRESENT THE DATA

Making wearable data usable for clinicians

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Development of a validation and Inspection Tool for Armband-based Lifelog data (VITAL) to facilitate the clinical use of wearable health data: A prototype and usability evaluation

Eunyoung Im [a], Sunghoon Kang [b], Hyeoneui Kim [a,c,d,*]

[a] College of Nursing, Seoul National University, Seoul, the Republic of Korea
[b] Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul National University Biomedical Informatics (SNUBI), Seoul, the Republic of Korea
[c] Center for World-leading Human-care Nurse Leaders for the Future by Brain Korea 21 (BK 21) Four Project, College of Nursing, Seoul National University, Seoul, the Republic of Korea
[d] The Research Institute of Nursing Science, Seoul National University, Seoul, the Republic of Korea

1. **Integrate:** Data from four device vendors in one view, at 10-minute intervals
2. **Set quality rules:** Completeness, recency, plausibility (e.g., minimum daily wear time)
3. **See what is valid:** Invalid days are filtered out before anyone draws conclusions

7 of 7 clinicians completed all tasks · **4.14 / 5** intention to use

Im E, Kang S, Kim H. Development of a Validation and Inspection Tool for Armband-based Lifelog data (VITAL) to facilitate the clinical use of wearable health data: a prototype and usability evaluation. Int J Med Inform. 2026;211:106322

10

EVIDENCE · AI SUMMARIZATION

AI summaries match experts, miss context

nature medicine

Article

https://doi.org/10.1038/s41591-024-02855-5

Adapted large language models can outperform medical experts in clinical text summarization

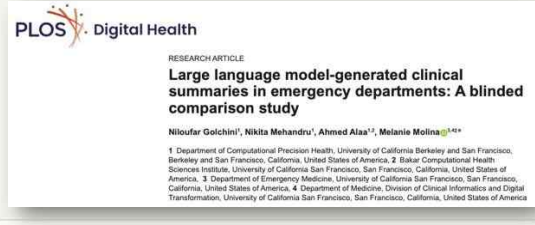
Received: 23 October 2023

Accepted: 2 February 2024

Published online: 27 February 2024

Check for updates

Dave Van Veen, Cara Van Uden, Louis Blankemeier, Jean-Benoit Delbrouck, Asad Aali, Christian Bluethgen, Anuj Pareek, Malgorzata Polacin, Eduardo Pontes Reis, Anna Seehofnerová, Nidhi Rohatgi, Poonam Hosamani, William Collins, Neera Ahuja, Curtis P. Langlotz, Jason Hom, Sergios Gatidis, John Pauly & Akshay S. Chaudhari



PLOS Digital Health

RESEARCH ARTICLE

Large language model-generated clinical summaries in emergency departments: A blinded comparison study

Niloufar Golchini, Nikita Mehandru, Ahmed Alaa, Melanie Molina

1 Department of Computational Precision Health, University of California Berkeley and San Francisco, Berkeley and San Francisco, California, United States of America, 2 Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, California, United States of America, 3 Department of Emergency Medicine, University of California San Francisco, San Francisco, California, United States of America, 4 Department of Medicine, Division of Clinical Informatics and Digital Transformation, University of California San Francisco, San Francisco, California, United States of America

CLINICAL TEXT SUMMARIZATION · 10 PHYSICIAN READERS

81% of LLM summaries rated as good as (45%) or better than (36%) experts'

But both made errors that could cause harm, and LLM summaries included fabricated information

ED CHART SUMMARIES · 21 PHYSICIANS, BLINDED

	LLM	Physician
Accuracy	**4.18**	3.40
Clinical utility	**3.88**	3.21

LLM: closer to the facts **Physicians:** more contextual nuance

Ratings on 5-point scales.

AI synthesizes facts; clinicians add context.

Van Veen D, Van Uden C, Blankemeier L, et al. Adapted large language models can outperform medical experts in clinical text summarization. Nat Med. 2024;30(4):1134–1142.
Golchini N, Mehandru N, Alaa A, Molina M. Large language model-generated clinical summaries in emergency departments: a blinded comparison study. PLOS Digit Health. 2026;5(7):e0001491..

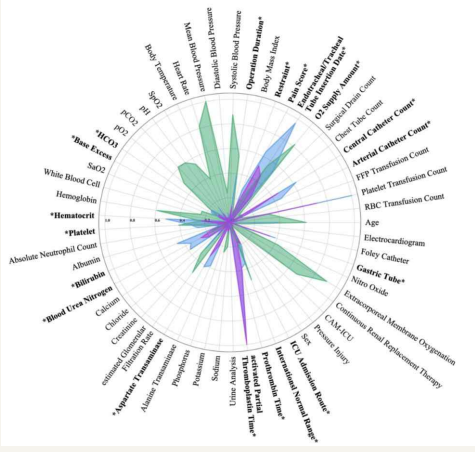
02

Hidden risks

How AI shapes our judgment and how models change quietly

OUR STUDY · EXPLAINABLE AI

Explanations changed how clinicians think



Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

ELSEVIER

Evaluating the impact of explainable AI on clinicians' decision-making: A study on ICU length of stay prediction

Jinsun Jung [a,d], Sunghoon Kang [b], Jeeyae Choi [c], Robert El-Kareh [d,e], Hyungbok Lee [a,b], Hyeoneui Kim [a,f,g,*]

AI predicting ICU length of stay · 15 ICU clinicians (13 nurses) ranked which patient factors matter, before and after seeing explanations

−0.15 → 0.87	3.9 → 3.1	2.2 → 3.7
Agreement with the model on key factors	"I am wary of the model"	"I can count on it to be correct all the time"

Trust items rated 1 to 5

Explanations pulled clinicians toward the model. Helpful if it is right. But what if it is wrong?

Jung J, Kang S, Choi J, El-Kareh R, Lee H, Kim H. Evaluating the impactof explainable AI on clinicians' decision-making: a study on ICU length of stay prediction. Int J Med Inform. 2025;201:105943.

EXPLANATION

Why should clinicians trust the prediction?

> ***"This patient is high risk because of these factors…"***

Explanation	≠	Understanding	≠	Appropriate decision

- ☞ Explainable AI does not automatically lead to good decisions.
- ☞ Instead, a plausible explanation can make a wrong answer more convincing.

EVIDENCE · RANDOMIZED VIGNETTE STUDY

Biased AI misled, explanations didn't help

Research

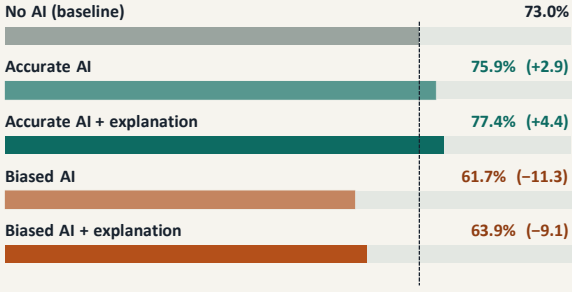
Jabbour S, et al. JAMA. 2023;330(23):2275–2284.

JAMA | Original Investigation | AI IN MEDICINE

Measuring the Impact of AI in the Diagnosis of Hospitalized Patients
A Randomized Clinical Vignette Survey Study

Sarah Jabbour, MSE; David Fouhey, PhD; Stephanie Shepard, PhD; Thomas S. Valley, MD; Ella A. Kazerooni, MD, MS; Nikola Banovic, PhD; Jenna Wiens, PhD; Michael W. Sjoding, MD

Diagnostic accuracy (change vs. no AI, percentage points)



457 hospitalist physicians, NPs, and PAs vignettes of acute respiratory failure.

A LOPSIDED TRADE-OFF

+4.4 best gain, with accurate AI

−11.3 loss, with biased AI

In this study, the harm from biased AI outweighed the gain from accurate AI.

EXPLANATIONS DIDN'T RESCUE CLINICIANS

−11.3 ➡ −9.1

With biased AI, adding an explanation barely reduced the harm

When AI was wrong, clinicians followed it even with an explanation.

Jabbour S, Fouhey D, Shepard S, et al. Measuring the impact of AI in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. JAMA. 2023;330(23):2275–2284.

16

15

HUMAN FACTORS · AUTOMATION BIAS AND DESKILLING

How clinicians use AI is itself a risk factor

NOW · EXPERTISE PROTECTS

Automation Bias in AI-Decision Support: Results from an Empirical Study

Florian KÜCKING[a,1], Ursula HÜBNER[a], Mareike PRZYSUCHA[a], Niels HANNEMANN[b], Jan-Oliver KUTZA[a,b], Maurice MOELLEKEN[a], Cornelia ERFURT-BERGE[c], Joachim DISSEMOND[d], Birgit BABITSCH[b], and Dorothee BUSCH[a,c]

210 nurses (63%) and physicians; AI wound advice (half of it wrong)

46% vs. **29%**

Wrong AI advice accepted
(without vs. with certified wound care training)

The more expertise, the less people followed wrong AI.

OVER TIME · AN EARLY WARNING SIGNAL

Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study

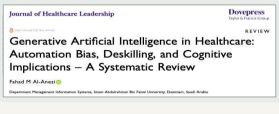
Krzysztof Budzyń, Marcin Romańczyk, Diana Kitala, Paweł Kołodziej, Marek Bugajski, Hans O Adami, Johannes Blom, Marek Buszkiewicz, Natalie Halvorsen, Cesare Hassan, Tomasz Romańczyk, Øyvind Holme, Krzysztof Janus, Shona Fielding, Melina Kunar, Maria Pellise, Nastazja Pilonis, Michał Filip Kamiński, Mette Kalager, Michael Bretthauer, Yuichi Mori

19 experienced endoscopists, 4 centres in Poland; 1,443 colonoscopies

28.4% vs. **22.4%**

Adenoma detection without AI
(3 months before vs. after AI was introduced)

Observational: a signal of possible deskilling, not proof.


Journal of Healthcare Leadership

Dovepress

REVIEW

Generative Artificial Intelligence in Healthcare: Automation Bias, Deskilling, and Cognitive Implications – A Systematic Review

Fahad M Al-Anezi

Not isolated findings

A 2026 systematic review of 29 studies found automation bias in 10 and deskilling in 9.

The less we can check AI, the more we rely on it. The more we rely on it, the less we practice checking

Kücking F, Hübner U, Przysucha M, et al. Automation bias in AI-decision support: results from an empirical study. Stud Health Technol Inform. 2024;317:298–304.
Budzyń K, Romańczyk M, Kitala D, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. Lancet Gastroenterol Hepatol. 2025;10:896–903.
Al-Anezi FM. Generative artificial intelligence in healthcare: automation bias, deskilling, and cognitive implications. A systematic review. J Healthc Leadersh. 2026;18:1–13.

17

16

WHAT CAN GO WRONG · WHEN THE SETTING CHANGES

A model can fail in a new place, or over time

JAMA Network | Open

Original Investigation | Health Informatics

Detecting and Remediating Harmful Data Shifts for the Responsible Deployment of Clinical AI Models

Vallijah Subasri, PhD; Amrit Krishnan, MSc; Ali Kore, BEng; Azra Dhalla, MBA; Deval Pandya, PhD; Bo Wang, PhD; David Malkin, MD; Fahad Razak, MD; Amol A. Verma, MD; Anna Goldenberg, PhD; Elham Dolatabadi, PhD

One early warning model for in-hospital death

7 hospitals in Toronto · 143,049 patient encounters

NEW PLACE

Same model, different hospital

Used in a hospital unlike the one it was built in

In this study: built at community hospitals, it got worse at large teaching hospitals

Like the sepsis model we saw earlier

OVER TIME · MODEL DRIFT

Same hospital, different time

Patients, practice, and data keep changing

In this study: when COVID-19 changed the patients, monitoring caught the shift and retraining kept the model working

The model stays the same. The world around it changes.

No alarm sounds when this happens. Someone has to watch and decide what to do.

Subasri V, et al. Detecting and remediating harmful data shifts for the responsible deployment of clinical AI models. JAMA Netw Open. 2025;8(6):e2513685.

WHAT CAN GO WRONG · SUMMARY

Failures can come from both sides

Human side

Trusting explanations too much
Explanation is not understanding

Following wrong AI advice
Biased AI misled clinicians; expertise protects

Losing the skill to check and decide
An early warning signal

A poor clinical decision

Machine side

Not validated where it is used
e.g., a widely used sepsis model missed 67% of cases

Facts without context
AI summaries can match experts but misses context

Quiet performance loss
Models drift after deployment

Both sides lead to the same question: who checks, and when?

03

Redesigning human oversight

Who should decide what, when AI performance changes?

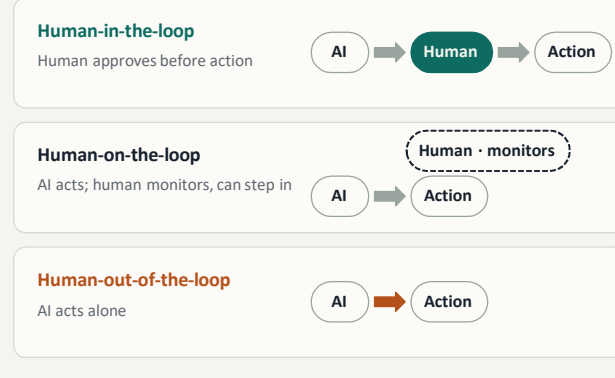
HUMAN OVERSIGHT

Human-in-the-loop: what it means

DEFINITION

"The structured, bidirectional integration of clinical expertise beyond passive use, enabling clinicians to validate, override, and provide feedback on AI outputs."

Not a final click to confirm, but a role built into the AI lifecycle.



Adaptive oversight moves between these modes as uncertainty, risk, and model drift.

Bakker M, et al. Human in the loop in AI-enabled clinical decision support: a systematic scoping review and reporting checklist for lifecycle governance. Int J Med Inform. 2026;222:106701..

MEANINGFUL OVERSIGHT

Presence is not oversight

A clinician in the workflow is not enough. Oversight is meaningful only when four conditions are designed in.

1 **Knowledge** (Epistemic capacity)
Know the model's intended use, validation population, limits, and where it is likely to fail

2 **Time to think** (Cognitive space)
Time in the workflow for independent appraisal — not reflexive acceptance

3 **Authority to disagree** (Decisional authority)
Disagreeing with the AI is expected, not deviant

4 **Power to act** (Intervention effectiveness)
Recommendations or actions can be paused, modified, or suspended

Without them, oversight becomes ***"procedural rather than protective."***

van de Sande D, Economou-Zavlanos N, van Genderen ME. Meaningful oversight of medical AI beyond human in the loop. npj Digit Med. 2026.

LIMITS OF STATIC HUMAN-IN-THE-LOOP

Clinician-in-the-loop shifts the risk

WHY THE BEDSIDE CANNOT BE THE FAILSAFE

Opaque models, limited training
Clinicians are asked to appraise systems they cannot see into

Alert fatigue
Accepting outputs without review becomes "a rational adaptation to excessive alarms"

Fluent generative AI
"Plausible rationales that mask factual errors" make checking harder, not easier

THE "MORAL CRUMPLE ZONE"

Clinicians absorb "the legal and reputational impact of system failure while shielding developers and implementers."

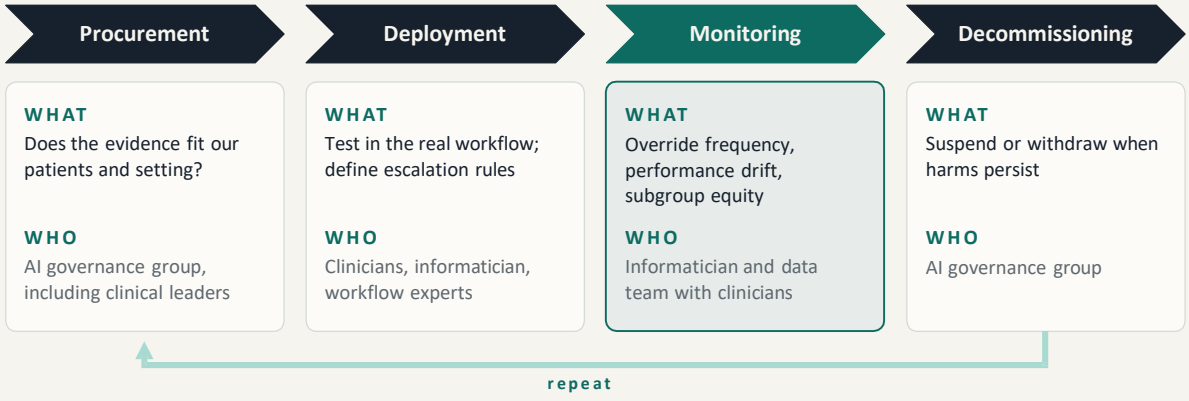
Clinicians are liable for overriding a correct AI verdict, and for following an incorrect one.

Proposed shift:
developer and implementer accountability | institutional AI governance with power to recalibrate or withdraw | continuous automated auditing for drift

Toro-Tobon D, Ponce Ponte O, Montori VM, Brito JP. Clinician in the loop: a flawed solution for AI oversight. BMJ. 2026;393:e089213. Term: Elish MC, 2019.

DISTRIBUTING DECISION AUTHORITY

Oversight is a cycle, not a checkpoint



Nurses belong at every stage as users, experts, and governance members.

Adapted from van de Sande D, Economou-Zavlanos N, van Genderen ME. Meaningful oversight of medical AI beyond human in the loop. npj Digit Med. 2026. doi:10.1038/s41746-026-02971-1 · Toro-Tobon D, Ponce Ponte O, Montori VM, Brito JP. Clinician in the loop: a flawed solution for AI oversight. BMJ. 2026;393:e089213.

THE KEY QUESTION

Does every AI decision require the same level of human oversight?

> **ADAPTIVE HUMAN OVERSIGHT**
>
> Dynamically changes the level and type of human involvement according to model uncertainty, system performance, clinical risk, and contextual change.

SELECTIVE HUMAN OVERSIGHT

When should humans step in?

UNCERTAINTY

Can we trust this output?

- The model signals low confidence (e.g., a score right at the threshold)
- The output doesn't fit what we see (e.g., low risk while the patient is visibly worsening)

Human role: Review this case against the patient

CLINICAL RISK

What if it is wrong?

- The decision has serious or irreversible consequences
- The patient is high-acuity

Human role: Confirm before acting, or escalate

MODEL DRIFT

Is the model still working?

- Performance or input data are shifting over time
- Performance drops for specific patient groups

Human role: Investigate, recalibrate, or suspend

No signals? AI proceeds. Humans watch the trend, not every output.

26

25

04

Nursing at the center

Nurses don't just consume AI outputs; they evaluate them

28

26

WHAT THE MODEL DOESN'T SEE

AI sees recorded data. Nurses also see context.

Subtle changes
"He's not himself today"

Family situation
Who is present, who decides, who worries

Treatment preferences
Goals of care, what matters to this patient

Coping
Fear, withdrawal, refusal of care

Ward events
A tense visit, roommate friction, a sleepless night

Unspoken concerns
What the patient has not said aloud

Much of this lives in verbal handoffs, not in the data a model learns from

Rossetti, S.C., Dykes, P.C., Knaplund, C. *et al.* Real-time surveillance system for patient deterioration: a pragmatic cluster-randomized controlled trial. *Nat Med* **31**, 1895–1902 (2025).

nature medicine

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

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

29

27

NURSING AT THE CENTER

Four roles nurses already play in AI-supported decisions

Context provider
Adds what the data do not capture

Interpreter
Translates AI outputs for this patient, this situation

Safety monitor
Notices and reports when AI behaves unexpectedly

Escalation decision-maker
Decides when a case needs more eyes, fast

These roles need time, authority, and support; not just goodwill!

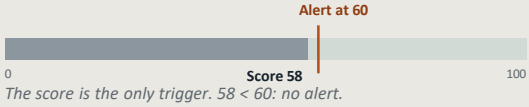
30

28

BACK TO MR. K

03:12, revisited

STATIC OVERSIGHT



The score is the only trigger. 58 < 60: no alert.

THE NURSE'S CONCERN

Stays in her notes; the system does not see it

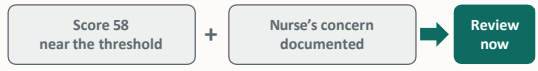
THE CALL AT 3 A.M.

Up to her confidence, with only a score below 60 to show

AFTERWARDS

If he gets worse, no one learns why the model missed him

ADAPTIVE OVERSIGHT



Uncertainty signal: the output doesn't fit what she sees

THE NURSE'S CONCERN

Counts as a signal, not just a note

THE CALL AT 3 A.M.

Expected. The doctor sees the score and her concern

AFTERWARDS

Her judgment and his outcome feed model monitoring

Same model, same nurse. The system made her judgment count.

05

Closing

From better predictions to better decisions

CLOSING

What clinical AI should become

FROM	TO	WHAT IT MEANS
Prediction	Decision support	Judged by the quality of decisions, not accuracy alone
Static models	Learning systems	Monitored and adapted long after deployment
Human vs. AI	Designed human–AI collaboration	Clear roles: who decides what, and when

The question is not AI or human. It is which decision belongs to whom.

TAKE-HOME MESSAGES

Three things to remember

1 **AI performance is not the endpoint. Clinical decision quality is.**
Recognizing patients in time, missing nothing important, acting appropriately

2 **Better AI does not make human judgment less important.**
Knowing when to trust, when to doubt, when to look further, when to override

3 **Oversight is an institution's job, not just the individual clinician's**
Adaptive, built into governance, with nurses at every level

We continuously reassess patients. We should also continuously reassess AI.

The future of clinical decision-making is not replacing human judgment with AI, but designing better decisions combining computational intelligence with clinical judgment, contextual knowledge, and continuous oversight.

Thank you

With thanks to Hungbok Lee, Eunyoung Im, Jinsun Jung, Jeong-ha Kim, and colleagues at SNU College of Nursing

Questions and discussion

Hyeoneui Kim | Seoul National University, College of Nursing

AI assistance (Claude, Anthropic) was used for slide design and editing; all content was developed and reviewed by the author