Using AI and Collaborative Workflows to Predict and Prevent Clinical Deterioration

Stanford Medicine

Primary Contact Information:

Jenny Yiu
Senior Program Manager, Clinical Informatics
jyiu@stanfordhealthcare.org

Sabrina Veksler
Senior Manager, Information Technology
SVeksler@stanfordhealthcare.org

Lisa Tsering
TDS Communications Specialist
LTsering@stanfordhealthcare.org

Clinical Project Leads:

Lisa Shieh, MD, PhD
Associate Chief Quality Officer
Clinical Professor, Hospital Medicine
Lsheih@stanford.edu

Ron Li, MD
Clinical Assistant Professor, Hospital Medicine
Ronl@stanford.edu

Margaret Smith, MBA
Director of Operations
Stanford Healthcare AI Applied Research
Marsmith@stanford.edu

IT Project Leads:

Nigam Shah, MBBS, PhD
Chief Data Scientist
NigamShah@stanfordhealthcare.org

Nikesh Kotecha, PhD
Director, Data Science
NKotecha@stanfordhealthcare.org

Hamed Barahimi
Director, Inpatient Systems & Services
HBarahimi@stanfordhealthcare.org

Patrice Duhon
Director, Nursing Informatics
PDuhon@stanfordhealthcare.org

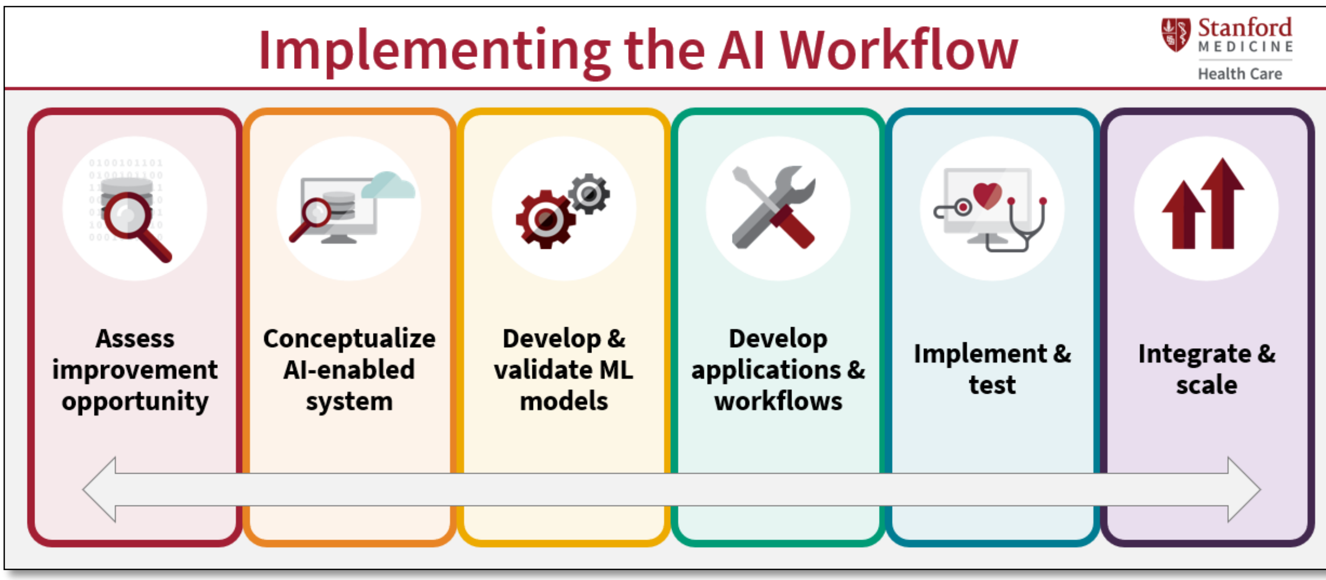
Lisa Flannery
Manager, Inpatient Clinical Applications
LFlannery@stanfordhealthcare.org

Jerri Westphal
Manager, Nursing Informatics
JWestphal@stanfordhealthcare.org

Brian Schenone
Lead, Clinical Informatics Analyst, Inpatient Clinical Applications
BSchenone@stanfordhealthcare.org

Executive Summary

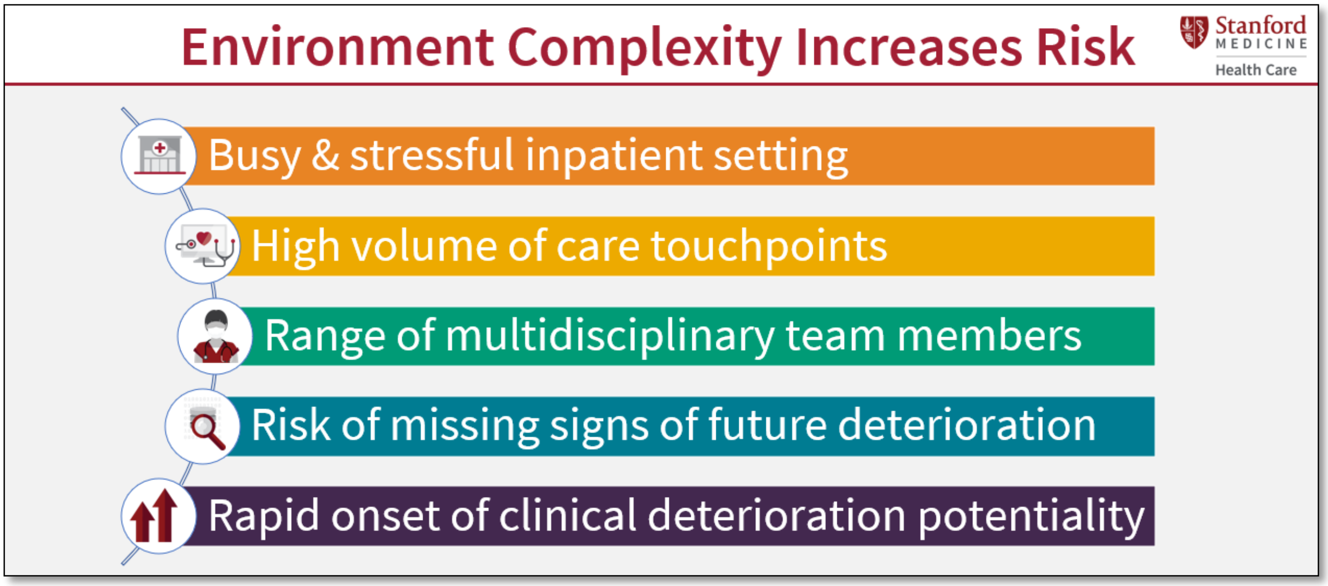
In order to facilitate the use of machine-learning (ML) models to improve care delivery, which remain poorly understood and executed, Stanford Medicine targeted an effort to address this implementation gap at the health system by addressing three key challenges: 1) developing a framework for designing integration of artificial intelligence (AI) into complex health care work systems; 2) identifying and building the teams of people, technologies, and processes to successfully develop and implement AI-enabled systems; and 3) executing in a manner that is sustainable and scalable for the health care enterprise.



This case study illustrates the pilot of a real-world implementation of AI into care delivery: clinical deterioration prediction to decrease unplanned escalations of care into the intensive care unit (ICU). We will describe how to apply the design principles to the health system, the barriers and facilitators we encountered, and how these experiences guided our collaborative approach to leveraging AI to improve patient outcomes and safety.

Define the Clinical Problem and Pre-Implementation Performance

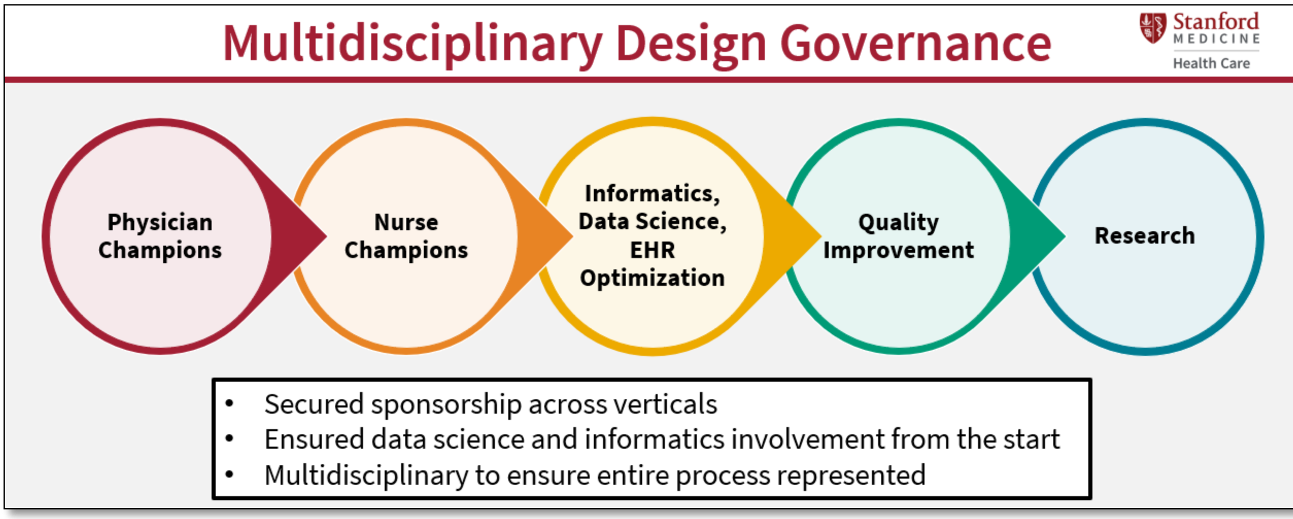
Stanford Health Care is a quaternary academic medical center with high-volume, often high-complexity inpatient services. Patients, particularly those at risk for deterioration, are cared for by multi-person care teams and require assessments of large amounts of data that change over time. These complexities can lead to barriers and gaps in care that may contribute to unanticipated clinical deterioration, resulting in the activation of rapid response teams [RRT], emergency resuscitation efforts on the wards and/or unplanned ICU transfers. The intent of this pilot is to identify patients with clinical deterioration so the care team can proactively intervene thereby avoiding an RRT, emergency resuscitation or emergent ICU transfer.



AI can facilitate alignment and coordination by acting as an objective assessor of risk. Patient care in the hospital, while supervised by the attending physician, is highly multidisciplinary, and patients interact with a variety of nonphysician clinical support services. One cause of process breakdowns was misalignment of risk perception and lack of coordination between physicians and nonphysician team members in performing needed clinical interventions. We found from stakeholder interviews that in times of disagreement, nonphysician team members frequently did not feel empowered to act, which may have led to missed opportunities for early identification of clinically deteriorating patients receiving a timely intervention.

Design and Implementation Model Practices and Governance

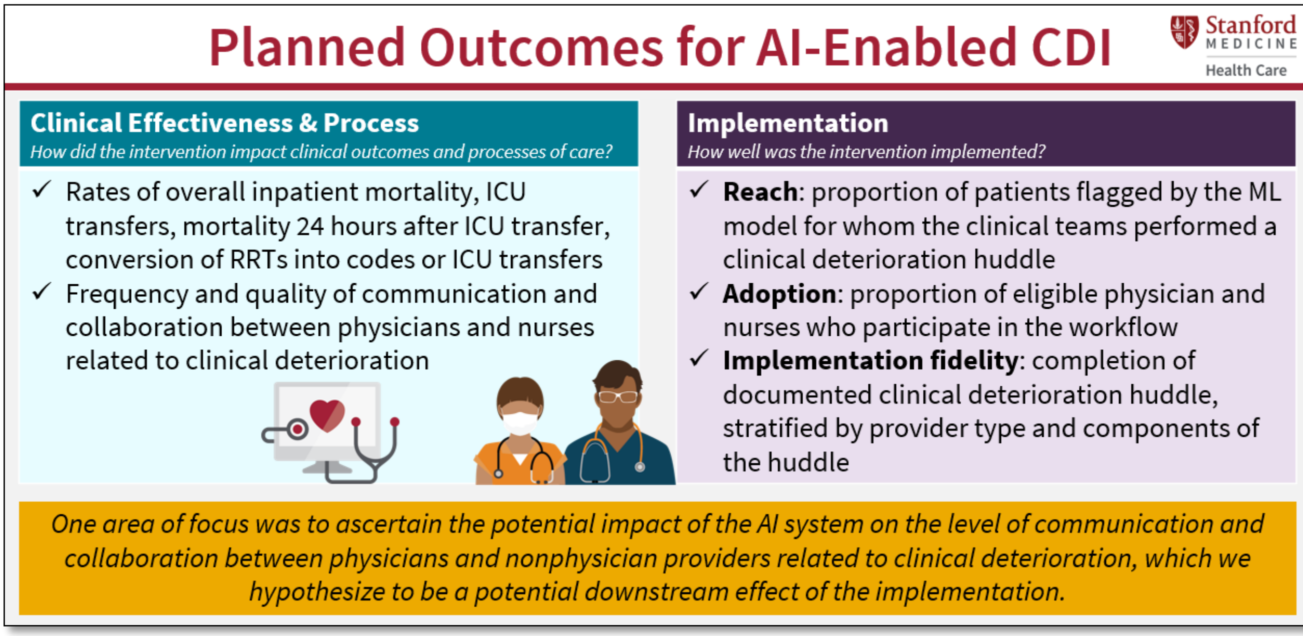
To successfully launch this project, we secured sponsorship across all stakeholders with collaboration across clinical, operational, data science, IT, and clinical informatics leadership. The multidisciplinary representatives included Bedside Nurses, Rapid Response Team Nurses, Attending Physicians, Residents, Medical Informaticists, Data Scientists, EHR optimization analysts, Quality improvement experts, and Researchers.



To align the care team on the appropriate early interventions, we determined that the ML model needed to identify patients with a high probability of a future clinical deterioration event (e.g., unplanned ICU transfer, RRT, or code), and that the predictions would have to be performed early enough to allow for enough time for the care team to respond. Predictions would also need to be updated in the EHR to reflect the frequent changes in the patients' clinical status, which enables the first key driver of providing a continuous assessment of risk.

We selected the Deterioration Index (DI), a model available through our EHR vendor, Epic Systems, because of the relative ease of technical integration while meeting most of these requirements. The DI model is a logistic regression that is capable of updating predictions on

hospitalized patients every 15 minutes using the most recent available clinical data on 31 physiological measures captured in the EHR; the DI tool also shows users the relative contributions of each physiological measure in generating the prediction. This last feature offers the additional benefit of providing a degree of model explainability, which can be useful for helping clinical users align around a shared mental model of risk.

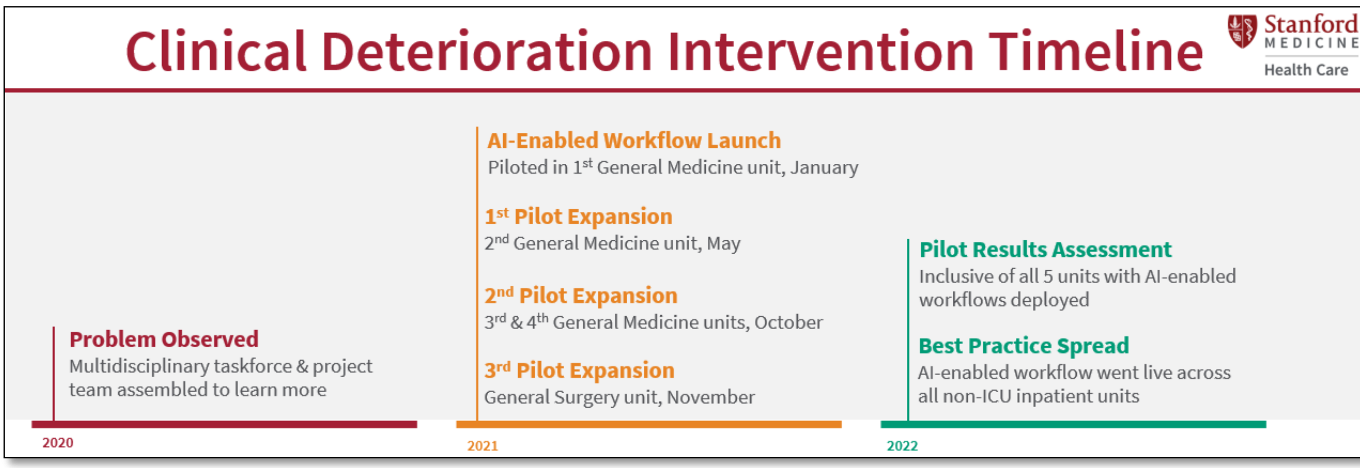


We then performed site-specific validation of the DI on a data set that we derived from a cohort of 6,232 non-ICU patient hospital encounters at our institution using a modified outcome definition that more closely reflected our product requirements: a composite outcome of RRT, code, or ICU transfer within 6 to 18 hours of the prediction. This validation strategy was modified from that of the vendor, which reported model accuracy in predicting the outcomes without the 6- to 18-hour time lag; this was thought not to be clinically meaningful because a model predicting an event within 6 hours of the event would not provide sufficient time for a clinical response. The area under the receiver operating characteristic (AUROC) (which is a performance metric for assessing ML models, in which 0.5 is the worst score and means the model is no better than random chance, and 1.0 is the best) calculated from our validation including these modified definitions was 0.71, which was lower than that reported by the vendor. Given this limited model discrimination, and to simplify the model output so that it could be more easily interpreted by the care team, we chose a binary classification threshold (high risk vs. not high risk), which was selected at a cutoff that maximized precision and recall, both of which were 20%.

We then validated with a focus group of clinicians that this level of accuracy would indeed be useful (i.e., most agreed they would want to be alerted if their patient had a “1 in 5 chances of experiencing an RRT or ICU transfer within the next 6–18 hours” while acknowledging that “four out of five patients who experience clinical deterioration would not be captured by the model”). While the low recall at this threshold (20%) would not make the DI an appropriate comprehensive screening tool for deterioration that would replace existing human-driven screening processes, there was consensus that, at a precision of 20%, it would still be useful to help align mental models and drive the desired physician–nurse team workflows for the patients whom the model does flag.

The digital applications embedded in the EHR incorporated ML predictions and enabled shared workflows between physician and nonphysician team members. The intent is to communicate and align risk across the Care Team with the following key product features:

- ML predictions had to be translated and displayed into usable information that is simple and avoids confusion that could lead to unintended consequences.
- Information had to be integrated into the clinicians' standard work in the EHR.
- Information had to be displayed transparently to all care team members to facilitate a shared mental model and collaborative work across the care team.

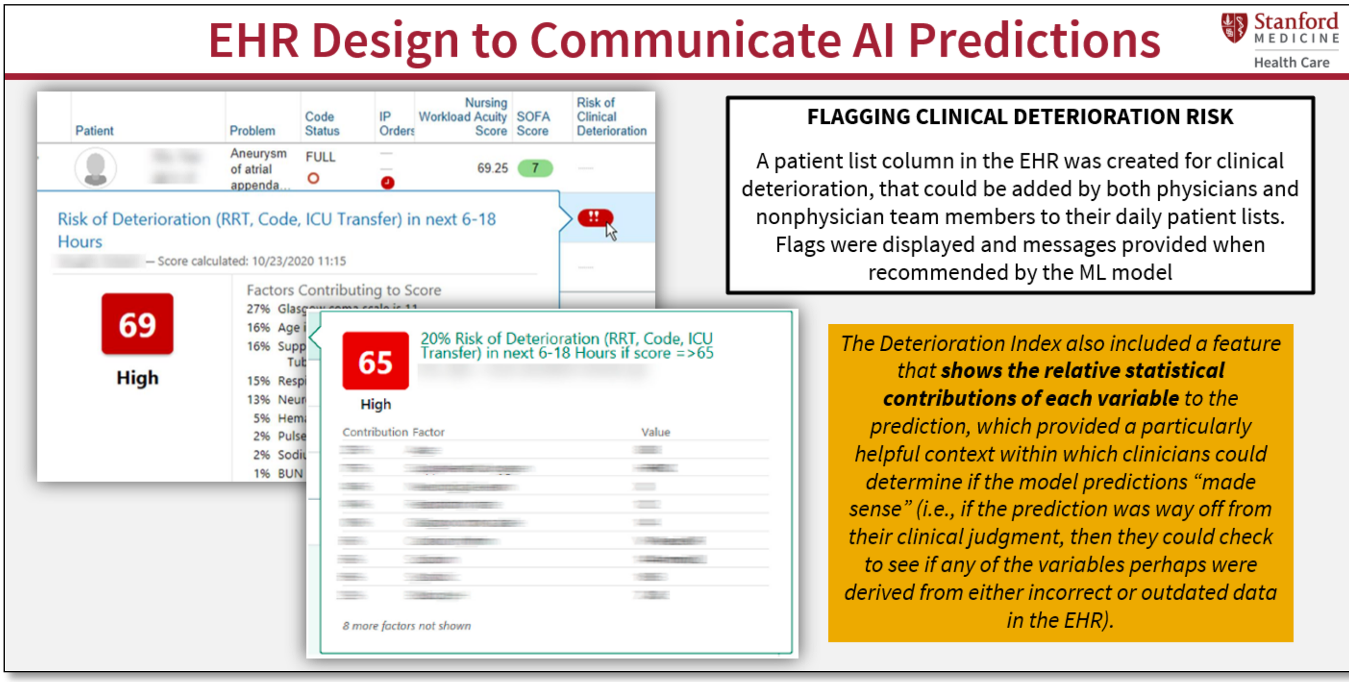


Clinical Transformation enabled through Information and Technology

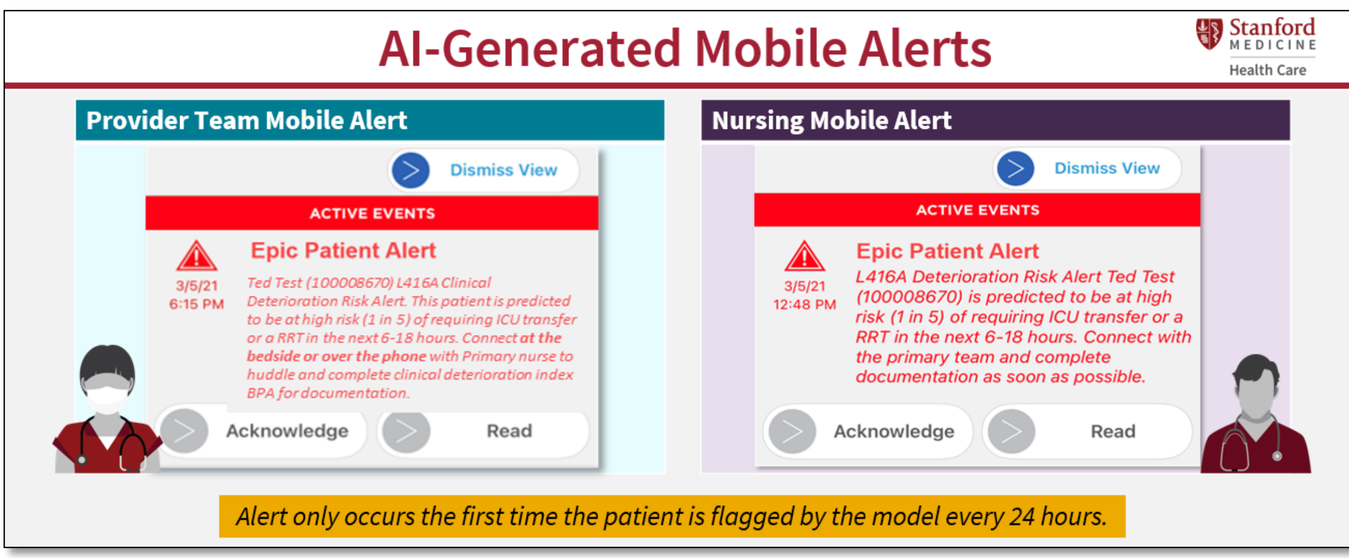
We needed to align the care team around a collaborative, standardized clinical response to patients flagged by the DI model. A key barrier to the adoption of AI systems in health care that we observed in our implementation is that clinicians disagree with the model predictions or believe that the AI system is not telling them anything that they do not already know. In our implementations, the emphasis was less on whether or not the model predictions were correct; rather, it was that for any given patient flagged by the model, physician and nonphysician care team members had to carry out a structured collaborative workflow to build a shared mental model of risk and a collaborative clinical response regardless of whether there is agreement with the model prediction. The role of the AI system was not necessarily to provide new information or to replace clinical decision-making, but to function as a dispassionate mediator for facilitating physician and nonphysician collaboration to assess the care plan in light of the new ML-generated information.

To promote consistency in this collaboration, we created the following structured workflows:

(1) Risk of clinical deterioration column flag and BPA when patient breaches model threshold (>20% chance of deterioration in 6-18 hours)

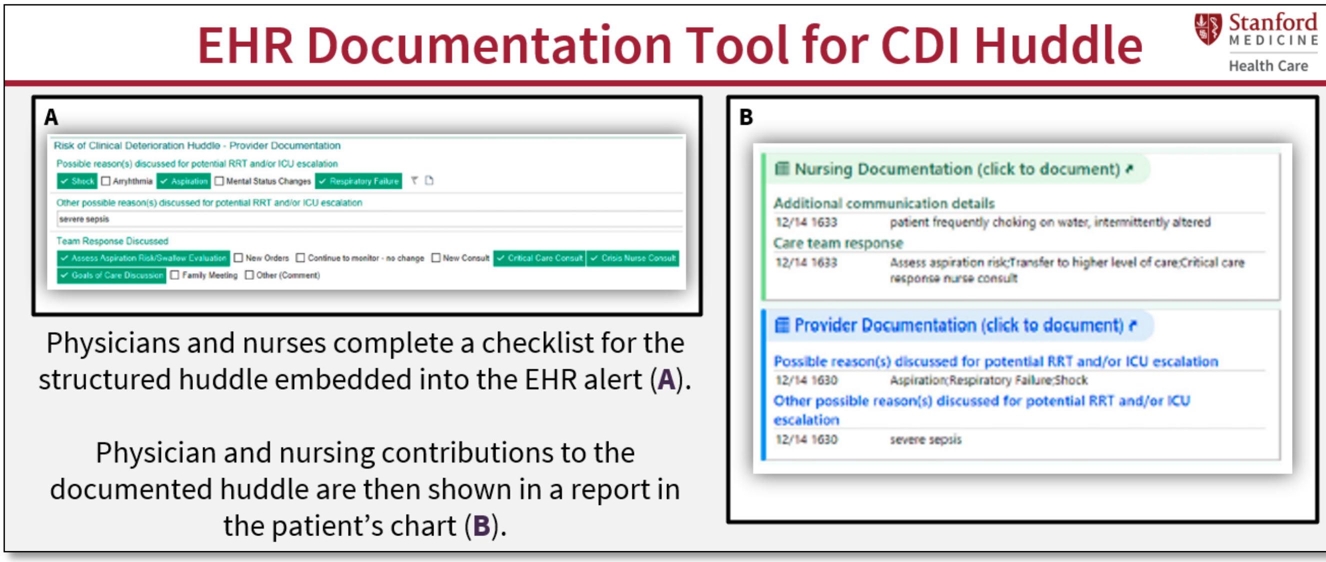


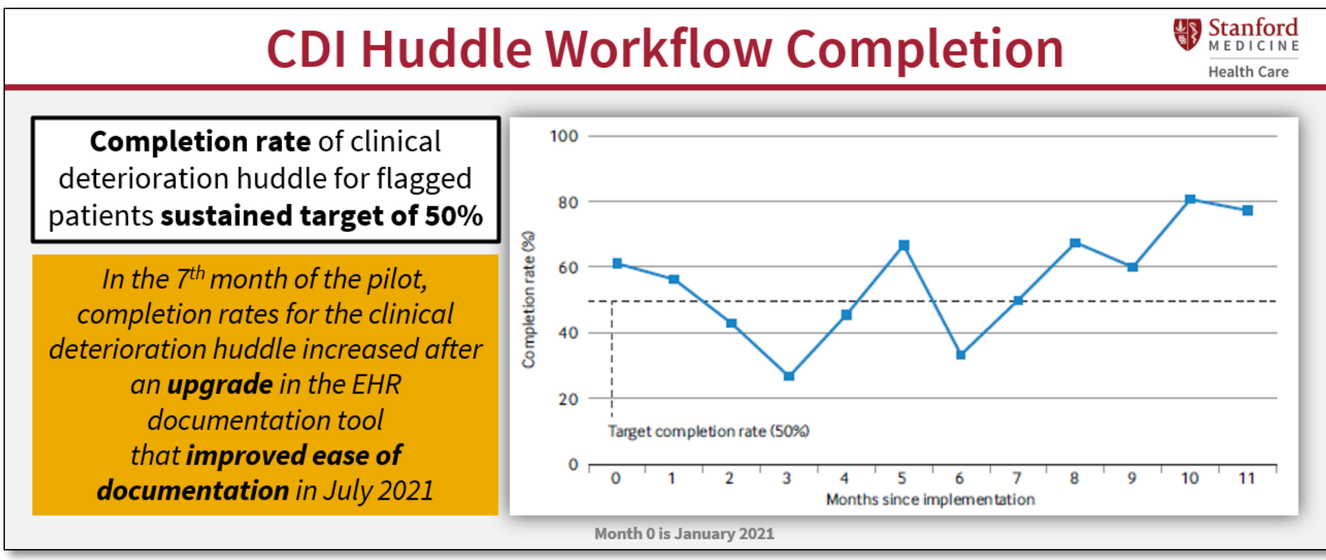
(2) Mobile alert to RN assigned to patient in EHR, Primary Resident/Intern, Cross Cover Resident/Intern.



(3) Primary Nurse and Charge Nurse connect to assess the patient and validate alert.

(4) Clinical Deterioration Huddle in person or on the phone within 2 hours, and communicate in SBAR format (Situation, Background/Assessment, Recommendation).

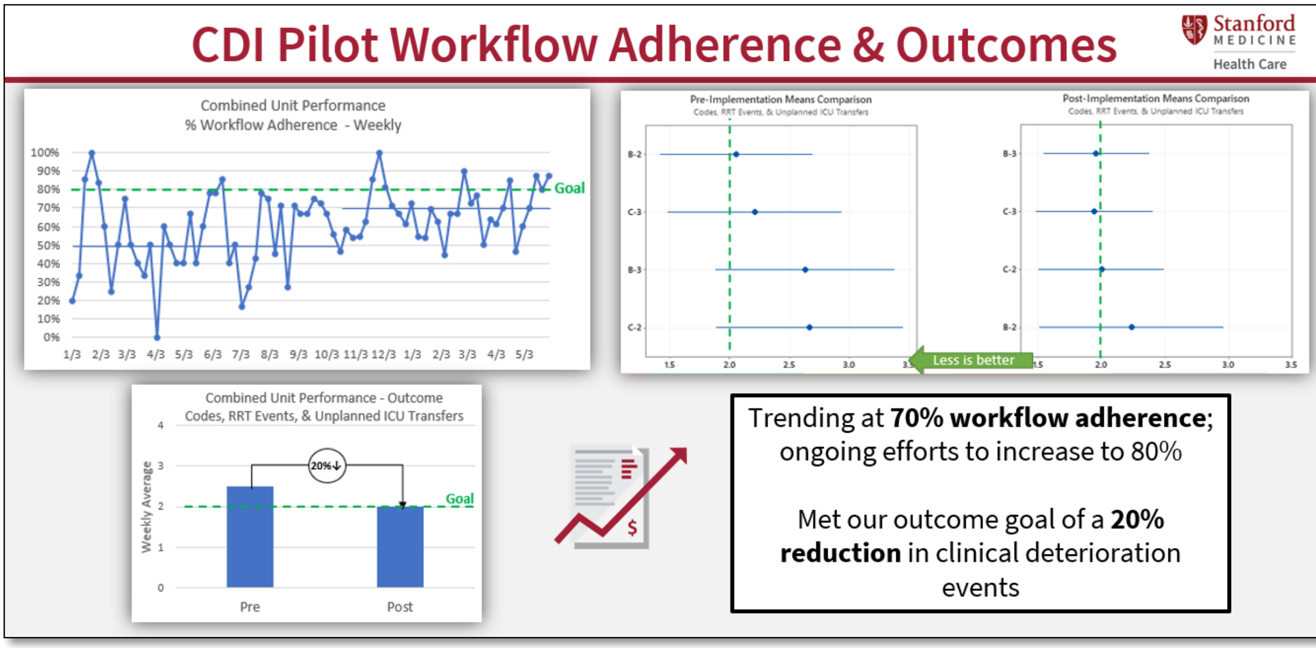




Improving Adherence to the Standard of Care

The clinical deterioration pilot was implemented in a stepwise fashion across two different nursing units for general medicine patients, which thus far has included 6,392 total patient encounters since the beginning of the implementation January 2021 to June 2022 (average of 355 encounters per month), with 601 total patient encounters experiencing at least one flag generated by the DI (average of 33 flags per month; 9.4% of total encounters).

We have yielded early promising results during the initial pilot phases, as measured by the documented workflow adherence rate and interviews with workflow participants. The workflow adherence rate is trending at 70% with ongoing efforts to increase to 80%. It met our outcome goal of a 20% reduction in clinical deterioration events.



Improving Patient Outcomes

The 20% reduction in clinical deterioration events is contributed by the observed sustained participation from nonphysician care team members with 100% of completed clinical deterioration huddles included contribution from a nurse. In a survey of nursing staff (57% response rate – 30/52 nurses responded). 96.5% reported that they felt the workflow was adding value to patient care. 89.6% indicated that the tool changes the way they care for their patients: charge nurses in the survey reported alternating patient assignments or ratios in anticipation of clinical changes with the flagging patient, and bedside nurses reported they rounded more frequently and/or completed a more in-depth patient assessment on their patients who were flagging.

While nurses have consistently documented completion of the huddles, physician documentation adherence has been minimal. However, survey results shed more light on physician participation and remaining challenges. In a survey among 19 medicine residents participating in the pilot:

- 50% indicated that they act on the alerts by calling the bedside nurse to huddle, messaging the bedside nurse, or going to the bedside to huddle with the nurse.
- 50% indicated that no personal action is taken on the alert; however, 64% said that after receiving an alert, the bedside nurse also reached out to them to discuss the patient's status.
- When asked about challenges to workflow adherence, 30% of physicians indicated that when they received the alert, they had recently assessed the patient, and, therefore, further action seemed redundant.

Accountability and Driving Resilient Care Redesign

Integrating novel workflows into health care is often challenging when there are competing demands for time and resources, especially with the record surges in patient volume our institution has experienced over the course of the implementation period (due to the Covid-19 pandemic and other factors). In particular, workflows involving AI can face a higher barrier of acceptance, because the mechanism triggering the workflow (the ML model) will, by definition, be wrong some percentage of the time (i.e., there is only a certain probability that the patient flagged by the ML model is, indeed, appropriate for the workflow).

Knowing that the model is not always a perfect prediction, we were thoughtful about where in the EHR to display the DI model risk score and how and when to alert members of the care team. That way, they could collaboratively review the score, interpret the risk scoring details, document their decision making and still have time to intervene earlier and before the patient's condition rapidly deteriorated. This workflow avoided data overload for busy clinicians as a cause of distraction and an inability to identify deteriorating patients.

It is important to conduct routine evaluations of the effectiveness of the ML generated alert. Our evaluation included workflow observations and system generated reports which can give better insights on user acceptance, and effective utilization. The results of the evaluation were shared with end-users, clinical teams, and clinical operational leaders for awareness and opportunities for workflow improvement.

Patient care teams need to continuously process large amounts of new information. If that information is ambiguous or not clearly actionable, it is at risk of being misinterpreted, misused, or not used at all. An important lesson we learned is that the ML prediction may not itself be necessarily informative, yet it still plays the important role of aligning clinical teams around a standard set of downstream actions that, on average for flagged patients, may lead to better outcomes. For example, a common piece of feedback we received from clinicians, particularly physicians, was that the model was "not telling [them] anything that [they] don't already know," in the sense that they often were already aware that a patient at risk of deteriorating. However, despite this prior awareness, physicians often did not actually perform the associated downstream tasks. Therefore, the true value.

Ultimately this AI system was not necessarily to provide new information, but rather to align the physicians with the rest of the care team around acting on an established workflow. To incorporate this concept early in each implementation, we pivoted from showing only model predictions to language that specifically outlines the appropriate interpretation and required action. For example, for patients flagged by the DI, nurses (and physicians) received an alert that concretely expressed the nature of the risk and next steps: "Clinical Deterioration Risk Alert — [insert patient name] is predicted to be at high risk (greater than 20%) of requiring ICU transfer or an RRT in the next 6–18 hours. Connect with the charge nurse and primary team as soon as possible and complete required documentation."

Building Clinician Trust and Buy-in for the Intervention is the key to success. The teams employed three strategies to build trust in the models and buy-in for the workflow designed in these implementations.

- First, site-specific quantitative model validation was conducted for clinical deterioration, and the results were shared with the clinical stakeholders during the participatory design sessions.
- Second, clinicians were directly involved in a parallel qualitative model validation process in which they indicated agreement or disagreement with the model predictions.
- Lastly, the team summarized and shared intervention success stories from early in the pilots to demonstrate patient-level benefit from the intervention. These stories included quotes from staff along with the case details and how the model output informed a different course of action and a favorable outcome.
- Forming a multidisciplinary team consisting of technical, operational, and clinical stakeholders, along with project management and quality improvement support, was convened. More specifically, the project team included about 15 members: data scientists, clinical informatics, enterprise analytics, nurse managers, frontline nurses, clinical nurse specialists, physicians, project managers, quality improvement experts, and social science researchers. Engaging all levels of the technical, operational, and clinical stack is a key facilitator of rapid and well-informed decision-making across all phases of the development and implementation of AI-enabled solutions.

References

1. Li R, Smith M, & Lu J, Avati A, Wang S, Teuteberg W, & Shum K, Hong G, Seevaratnam B, Westphal J, Dougherty M, Rao P, Asch Steven, Lin S, Sharp C, Shieh L, Shah NH. (2022). Using AI to Empower Collaborative Team Workflows: Two Implementations for Advance Care Planning and Care Escalation. NEJM Catalyst. 3. 10.1056/CAT.21.0457.

HIMSS Global Conference Audience – Main 3 Relevant Topics

1. Healthcare Applications and Technologies Enabling Care Delivery
2. Improving Quality Outcomes
3. Process Improvement, Workflow, and Change Management