How to use artificial intelligence tools to predict and prevent care-related harm

Context

In the Australian setting, risk prediction aligns with the goals of the National Safety and Quality Health Service Standards (especially Clinical Governance and Preventing and Controlling Healthcare-Associated Infection Standards), as well as the aims of the RACP Evolve initiative, which seeks to reduce low-value care and promote evidence-based practice.

Low-value care tests, treatments, or procedures that provide little or no clinical benefit remains a significant challenge in the Australian health system. Research suggests that up to one in three healthcare services delivered may be of low or uncertain value, contributing to avoidable patient harm, inefficiencies, and escalating costs.[1]

Artificial intelligence (AI)-driven risk prediction tools are being explored across healthcare to identify patients at risk of harm before it occurs – such as deterioration, sepsis, falls, or unplanned readmissions – and to target such individuals with evidence-based preventive strategies and avoid low-value practices that provide limited clinical benefit.

Ensuring physicians evaluate and effectively use AI tools is essential to addressing low-value care and improving patient outcomes and system sustainability.

Ideal future

AI risk prediction tools are being used in the hospital environment to flag a patient as being at higher risk of disease by analysing subtle trends in vital signs, lab results, and clinical notes contained in electronic medical records (EMRs). The tools are typically trained on large datasets obtained from EMRs containing clinical, demographic, and administrative data to identify patterns associated with risk of future harm and enable early detection.

Clinical scenario

Risk prediction dashboard

With a busy inpatient ward round ahead, Dr Patel, a general physician, logs into the hospital's AI dashboard that contains a risk prediction tool for sepsis. Her first patient is a 72-year-old man admitted overnight with community-acquired pneumonia and is treated with a penicillin.

Alert

The tool flags the patient as being at high risk for becoming septic based on early changes in vital signs and laboratory markers.

Consultation and decision

This prompts Dr Patel, after discussion with the team, to order closer monitoring of vital signs and lab results as well as a repeat set of blood cultures.

Prevention

Dr Patel's registrar suggests changing the patient's intravenous antibiotics to a broad-spectrum cephalosporin. However, Dr Patel feels this is not necessary unless the patient deteriorates further or blood cultures return an organism resistant to the penicillin the patient is currently receiving. She reminds her registrar of the need for antibiotic stewardship to counter the growing problem of antibiotic resistance due to overuse of broad-spectrum agents.[2]

Pause and reflect:
What factors would you prioritise here: the AI alert or antibiotic stewardship?

evolve
Better care. Better decision-making. Better use of resources.

[1] Braithwaite J, Glasziou P, Westbrook J. The three numbers you need to know about healthcare: the 60-30-10 Challenge. BMC Med. 2020 May 4;18(1):102. doi: 10.1186/s12916-020-01563-4. PMID: 32362273; PMCID: PMC7197142.
[2] Australian Commission on Safety and Quality in Health Care. Antimicrobial Stewardship Clinical Care Standards. November 2020. https://www.safetyandquality.gov.au/sites/default/files/2020-11/Antimicrobial%20Stewardship%20Clinical%20Care%20Standard%20%20Clinician%20fact%20sheet.pdf

RACP Specialists. Together
EDUCATE ADVOCATE INNOVATE

How to use artificial intelligence tools to predict and prevent care-related harm

Benefits	Risks
Earlier intervention and prevention: Clinicians are alerted to patients at risk of certain harms who warrant early initiation of preventive measures before harms occur.	**Data quality and bias:** AI tools trained on flawed or incomplete data may misclassify risks or perpetuate inequities (e.g. among Indigenous or rural patients), negatively affecting the AI tool's external validity.
Efficiency in resource allocation: High-risk patients can be identified, enabling interventions to be targeted to them while avoiding potentially unnecessary and wasteful care directed at those with lower risk.	**Alert fatigue:** Excessive or non-specific alerts may reduce clinician engagement or even lead to dismissal of critical warnings.
Scalability and automation: AI tools in their training and deployment can process vast amounts of data, enabling rapid and more accurate risk prediction than current rule-based algorithms which often require manual data entry and human calculations.	**Over-reliance:** Clinicians may defer too readily to the tool rather than apply clinical reasoning, particularly in high-pressure settings.
Supporting clinical decision-making: AI tools can provide an additional insight to complement clinical judgement.	**False sense of objectivity:** AI outputs may appear neutral and evidence-based, but underlying biases in training data and design such as encoding historical racial or socioeconomic disparities, underrepresenting rare conditions, or favouring billing-driven care patterns can perpetuate existing inefficiencies or low-value practices.
Quality improvement: Analyses of aggregated data from risk prediction tools can inform policy, guideline development, and education.	

To find out more about the use of AI in medicine, see the Evolve *AI in Healthcare* webinar series on the Evolve website: https://www.racp.edu.au/evolve/home

evolve
Better care. Better decision-making. Better use of resources.

RACP
Specialists. Together
EDUCATE ADVOCATE INNOVATE

How to use artificial intelligence tools to predict and prevent care-related harm

Key considerations for physicians using AI tools to predict and prevent care-related harm

Understanding development and evaluation:

- Check whether the tool has undergone external validation in comparable Australian settings and patient groups. Use frameworks such as SALIENT[3] and DECIDE-AI[4] for early-phase clinical evaluation.
- Assess tool performance using metrics such as AUROC, AUPRC, sensitivity, and specificity, and question whether use of a seemingly high performing tool translates into meaningful improvement in patient outcomes.
- Determine whether the tool's predictive accuracy changes across different patient subgroups or in different clinical settings.

Maintain clinical accountability:

- AI tools should augment, not replace clinical decision-making.
- Physicians remain accountable and must weigh AI outputs against patient context and preferences.
- Discuss the reasons for considering AI-generated risk predictions with patients and carers to ensure trust and acceptability.

Governance and monitoring:

- Ensure compliance of AI tool use with the Privacy Act 1988[5] and local data governance frameworks.
- Ensure compliance with regulatory standards that apply to AI tools which influence clinical decision-making.
- Advocate for continuous monitoring and recalibration of tools to maintain safety and equity.

Balancing effectiveness, efficiency and utility[6]:

- Evaluate whether the tool integrates seamlessly into existing workflows without creating extra administrative burden. Use frameworks such as SALIENT[3] to assess end-to-end implementation.
- Consider trade-offs in selecting the risk threshold for generating alerts: a tool optimised for sensitivity (i.e. lower risk threshold) may capture more at-risk patients however creates inefficiencies through generating more false positives in patients who have no or little risk, thus worsening alert fatigue. Conversely, a tool prioritising specificity (i.e. higher risk threshold) will reduce false alerts but will miss more at-risk patients.
- Determine if risk prediction tools actually alert clinicians to patients at risk of harm at an earlier time than clinicians would have recognised based on clinical judgement alone.
- Examine whether it supports stewardship of care by not causing overuse of care in many patients at low risk as the tool is too sensitive, while simultaneously maintaining patient safety by not missing many patients at high risk as the tool is too specific.

Glossary

AUROC:
Area under the receiver operator characteristic curve

AUPRC:
Area under the precision-recall curve.

evolve
Better care. Better decision-making. Better use of resources.

RACP Specialists. Together
EDUCATE ADVOCATE INNOVATE

[3] Van der Vegt, Scott IA, Dermawan K, et al. Implementation frameworks for end-to-end clinical AI: derivation of the SALIENT framework. J Am Med Inform Assoc 2023; 30(9):1503-1515.
[4] Vasey B, Nagendran M, Campbell B, et al.; DECIDE-AI expert group. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ 2022;377:e070904.
[5] Office of the Australian Information Commissioner. The Privacy Act. Australian Goverment, https://www.oaic.gov.au/privacy/privacy-legislation/the-privacy-act
[6] Van der Vegt AH, Campbell VK, Webb R, et al. A novel, standardised approach to balancing effectiveness, efficiency and utility of surveillance AI prediction models for hospitalised patients using sepsis prediction as an exemplar. J Am Med Inform Assoc 2025 Nov 11:ocaf192. doi: 10.1093/jamia/ocaf192. Epub ahead of print.