Canadian Journal of Emergency Medicine (2025) 27:165–169
https://doi.org/10.1007/s43678-025-00870-0

CJEM DEBATE SERIES

CJEM debate: clinical decision rules–thinking beyond the algorithm

Justin Morgenstern[1] · Ryan Radecki[2] · Lauren Westafer[3,6] · Joshua D. Niforatos[4] · Paul Atkinson[5]

Received: 6 January 2025 / Accepted: 16 January 2025 / Published online: 3 February 2025

Introduction

Paul Atkinson (@PaulAtkinsonEM)

The typical emergency physician, after rigorous academic selection, spends 11 to 13 years in post-secondary training, shaping and honing their mind, a product of over 80-billion neurons, with up to 1000-trillion synaptic connections. Yet, when it comes to complex decision-making about serious illness and injury, we increasingly try to distill the data points down to 5 or 6 questions, providing a score upon which we proceed. Clinical decision rules, or instruments, have been shown to be as accurate as a trained physician for certain decisions, they provide a consistent approach, they have been researched and adapted, and are popular for learning and standardizing practice. But do they also oversimplify decision-making? Do they neglect to factor in individual patient variability? Who is the average patient for whom they are designed anyway?

Justin Morgenstern of First10EM.com and the University of Toronto and Ryan Radecki of Christchurch, New Zealand propose that clinical decision rules are ruining medicine, proliferating without proper scrutiny and testing, and distracting physicians from applying evidence-based medicine in an individualized manner. Lauren Westafer of foamcast.org and the University of Massachusetts Medical School along with Joshua Niforatos of Wisconsin counter that knowing the inherent bias and flaws we all have, never mind the impact of stress, tiredness and a hectic environment, it is better to embrace these tools to help us make better decisions, in line with the Choosing Wisely philosophy. This series of editorials provides CJEM readers with the opportunity to hear differing perspectives on topics pertinent to the practice of Emergency Medicine. The debaters have been allocated opposing arguments on topics where there is some controversy or perhaps scientific equipoise.

Readers can follow the debate on X.com or bsky.social and vote for either perspective by going to @CJEMonline or by searching #CJEMdebate and #CDR.

Proposal: Clinical decision instruments may hinder clinical decision-making

Justin Morgenstern (@First10EM)
Ryan Radecki (@emlitofnote)

Clinical decision instruments (CDIs) have become ubiquitous in medicine. Given the cognitive complexity of clinical medicine and the fatiguing number of decisions made each workday, it is not surprising that clinical decision instruments are adopted widely. Unfortunately, with few exceptions, the available evidence rarely demonstrates either superiority to clinical judgment or durable, favorable impact on clinical practice. Despite the aim of decreased testing, an emphasis on high sensitivity results in tools with low specificity, the impact of which may result in paradoxical increases in testing. Decision instruments serve to simplify complex clinical questions, but evidence suggests clinicians lack proficiency with their probabilistic nature [1]. Therefore, it is possible, or even probable, that the widespread

✉ Paul Atkinson
paul.atkinson@dal.ca

1 Department of Emergency Medicine, Saint Joseph's Healthcare, Toronto, ON, Canada

2 Department of Emergency Medicine, Christchurch Hospital, Christchurch, New Zealand

3 Department of Emergency Medicine, UMass Chan Medical School - Baystate, Springfield, MA, USA

4 BerbeeWalsh Department of Emergency Medicine, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA

5 Department of Emergency Medicine, Dalhousie Medicine New Brunswick, Saint John Regional Hospital, Saint John, NB, Canada

6 Department of Healthcare Delivery and Population Science, UMass Chan Medical School - Baystate, Springfield, MA, USA

adoption of clinical decision instruments in medicine has been deleterious to clinical decision-making.

CDIs are rarely compared with clinical judgment, and when such comparisons exist, clinical decision instruments often fail to improve upon baseline clinical practice [2, 3]. For example, despite extensive use, in the only study available, the PECARN minor traumatic brain injury instrument did not improve upon clinical judgment [4]. The PECARN instrument has similar sensitivity to clinicians, but diminished specificity, suggesting premature adoption of this rule may be increasing CT utilization among children without a corresponding reduction in missed clinically important traumatic brain injury. For many other widely used clinical decision instruments, we simply do not know how they compare to clinical judgment, as such comparisons have never been made.

The long-term impact of clinical decision instruments is rarely observed. The gold standard research step for clinical decision instruments is high quality implementation research demonstrating sustained improvement in patient outcomes [5]. Ideally, this implementation work occurs prior to widespread clinical adoption. Proponents of clinical decision instruments will point to the very small handful of rules with positive implementation studies. However, these studies often suffer threats to internal validity such as the Hawthorne Effect and other practice-changing confounders, and do not reliably demonstrate long-term effects on practice. Other clinical decision instruments, such as the Canadian CT Head Rule, are widely used despite high quality implementation research failing to demonstrate measurable effects on imaging [6]. Other observational data describing changes in behavior associated with the Canadian CT Head Rule cannot be reliably attributed to the clinical decision instrument, as the study designs lack control groups and are confounded by additional cultural changex [7, 8]. The absence of well-conducted implementation work does not exclude the possibility of benefit, but reduces the level of confidence with which clinicians ought rely upon clinical decision instruments.

Similar to outcomes of other clinical research, clinical decision instrument validation studies suffer the malaise of imperfect external validity. Only a small minority of decision instruments are validated outside the settings in which they are derived [5]. When validation studies are performed, the settings are often very similar in health system, culture, or capabilities. Furthermore, clinical decision instruments are frequently developed in settings with pre-existing practice patterns of overuse, such as the United States, and are unlikely to affect outcomes to a similar magnitude in other clinical cultures. In fact, promotion of these clinical decision instruments may serve to export the culture of overuse in which they were developed and validated.

Like other clinical practices, clinical decision instruments are prone to "indication creep": a propensity to be applied in inappropriate clinical scenarios outside of the original research populations. For example, the Canadian CT Head Rule is frequently applied to patients who have "minimal" rather than "minor" head injury, a demonstration of the unintended "real-world" use of the instrument [9]. (Although it is an incorrect interpretation of the rule, this rule is almost certainly responsible for the widespread belief that every patient over the age of 65 needs a CT, no matter how trivial their head injury.) Even when "validated" in an implementation trial, apparently effective clinical decision instruments may give rise to unanticipated deviations when adopted widely.

Special attention should be paid to decision instruments with low specificity and their impact on practice. Simple, widespread tools including the Pulmonary Embolism Rule-Out Criteria (PERC) and Ottawa Subarachnoid Hemorrhage (SAH) rules are designed to be used as "one-way" decision instruments. That is, their test characteristics have such poor positive likelihood ratios it is proposed they be used solely for their negative likelihood ratios. Unfortunately, due to clinician acclimation to typical clinical decision instruments featuring rule-in and rule-out components, these "one-way" clinical decision instruments are easily misinterpreted as "two way" tools, allowing their low positive likelihood ratios to inappropriately influence clinician behavior [10].

The shortcomings in the evidence for clinical decision instruments are likely to be amplified as decision-support is more frequently embedded into electronic health records. The presence of decision-support may guide clinical practice, but in routine use produces decision errors when decision-support fails to activate in the correct context, or produces guidance based on imperfect information or processing [11]. "Automation bias", a term used to describe the reliance on decision-support, introduces errors via "heuristic replacement for vigilant information seeking and processing" [12]. Thus, the structure and features of clinical decision instruments, similar to electronic decision-support, introduces cognitive biases relating to gathering and processing of information.

Finally, although more difficult to quantify, we must consider the impacts of clinical decision instruments on the culture of medicine. Does the emphasis on high sensitivity clinical decision instruments amplify a "zero miss" culture in medicine? Can clinical decision instruments with completely binary outputs ("image" or "do not image") actually be used to improve judgment or shared decision-making? Are we judging clinicians based on the apparently objective answers these tools provide, ignoring the inherent subjectivity of most clinical decisions? As clinicians become dependent on clinical decision instruments, their repeated use fundamentally alters their underlying "gestalt" and risk estimation. This may result in a paradoxical situation in which the features of a clinical decision instrument may be

the primary inputs into a pretest probability estimate before application of that same clinical decision instrument. Trainees witnessing these processes may mimic a practice pattern based on clinical decision instrument features, rather than a wider base of clinical knowledge that was used when the clinical decision instrument was validated. It is not yet known whether these shifts would be net beneficial or harmful to the practice of medicine.

CDIs are not intrinsically problematic in principle. They are, however, as Goodman et al. describe, limited tools whose use and application are not sufficiently understood by those who use them [1]. Existing clinical decision instruments ought to be regarded more critically than at present, particularly by those whose practice setting is unlike the environment in which the rules were derived. Then, to use these decision-making adjuncts, clinicians need to have the statistical knowledge to consciously apply the clinical decision instruments like a test, with pre- and post-test odds. Finally, clinical decision instruments need to be compared to clinical judgment and evaluated in prospective, experimental and quasi-experimental designs capable of providing true insight in their long-term beneficial effects and unanticipated harms.

Counter: Clinical decision instruments can improve and standardize clinical decision-making

Joshua D. Niforatos @reverendofdoubt
Lauren Westafer @LWestafer

In the United States and Canada there are more than 150-million emergency department (ED) visits annually. [13]. Emergency physicians are tasked with rapid clinical decision-making in a time-limiting environment where the risk tolerance for a missed diagnosis is exceedingly low [14]. Given these constraints, a liberal approach to hospital admissions and diagnostic testing, such as the use of advanced imaging, can be viewed as an attempt to mitigate uncertainty and liability in this fast-paced clinical environment [15, 16].

The accessibility of advanced imaging such as computed tomography (CT) over the past few decades, which can aid in efficient and effective diagnostic strategies, has led to widespread use. Although values in healthcare have shifted toward more testing, there is an increased recognition, within the house of medicine, of the need to identify and eliminate low-value care that does not improve patient-oriented outcomes, that may lead to overdiagnosis and downstream patient harm, and may be driven by implicit bias [17, 18].

In a prospective multicenter study of the Pregnancy-Adapted YEARS algorithm, radiation-associated imaging was avoided in 39% of pregnant patients [19]. Although not a direct comparison to gestalt, the counterfactual is that prior to this study, evaluation of pulmonary embolism in pregnant patients relied almost exclusively on chest imaging, and all of these patients would have likely received imaging. Finally, in an international multicenter, cluster-randomized trial of the PERC rule in patients at very low risk for pulmonary embolism, randomization to PERC versus gestalt resulted in an absolute difference in reduction of imaging of 9.7% (95% CI 6.1–13.2%) and reduction of hospital admission of 3.3% (95% CI, 0.1–6.6%) with a non-significant miss rate for pulmonary embolism [20].

Critics of clinical decision instruments cite minimal difference in performance against clinician gestalt, though rarely specify what that minimal difference entails. A real-world outcome metric of interest when comparing clinical decision instruments with clinician gestalt is how gestalt correlates to clinician action. As in the above example, clinical decision instruments have meaningfully and safely decreased the rate of imaging for patients with suspected pulmonary embolism, compared to the counterfactual and usual care. In contrast, the PECARN brain injury instrument, in one paper from EDs in Australia and New Zealand, did not improve upon clinical judgment [4]. Despite impressive clinical gestalt and a low rate of contrast-enhanced imaging (8.3%), what is worrisome in this study was the unnecessarily high rate of hospitalization (22%), which was significantly higher than the 9% hospitalization rate in the primary PECARN head trauma study, calling into question the utility of gestalt when it does not correlate to clinician action [21].

Another study comparing physician gestalt and diagnosis with clinical decision instruments of acute pediatric appendicitis revealed a favorable ROC curve for gestalt, particularly in low-risk patients. In that group, 1.1% had appendicitis compared to gestalt group rate of 1%-10%. Despite an exceedingly low pretest probability for appendicitis, in this low risk group, 22.8% of patients underwent imaging where the clinical decision instrument recommended observation for 6 h with serial abdominal exams and discharge with 24 h follow-up, and no imaging required in those with improving symptoms [22]. Syncope risk stratification is even more illuminating. The Canadian Syncope Risk Score stratifies patients presenting with syncope who are at risk for 30-day serious adverse events. An external validation study revealed that clinician gestalt, which is part of the risk score, has a similar area under the curve compared to the clinical decision instrument for predicting serious adverse events [23]. Nevertheless, in one robust multicenter validation study, 25% of those in the very-low-risk syncope group were still admitted unnecessarily to the hospital for syncope [24]. It is not the clinical decision instruments that drive low-resource care, but likely our risk-intolerant culture.

Another criticism is the need for robust external validation of clinical decision instruments and the lack of 'real-world' evidence while neglecting to acknowledge the

significant clinician practice pattern variation for similar disease processes. The idea that a patient presenting with chest pain may receive significantly divergent evaluations depending on which ED they might present to is ethically and financially problematic. Research has shown that emergency physician practice varies greatly with regard to rates of imaging and patient disposition for similar conditions [25, 26]. Emerging qualitative research suggests that what emergency clinicians believe and are highly motivated by is what they perceive and call "standard of care" without realizing it is just a local practice pattern [27]. Despite some of the limitations of clinical decision instruments, their integration in electronic health records or through quality improvement initiatives has been shown to decrease unnecessary imaging. One study of 13 EDs in California revealed that clinician education of the Canadian CT Head Rule, along with clinical decision support in the electronic health record, was associated with a 5.3% (95% CI 2.5–8.1%) reduction in CT use with an associated 2.3% (95% CI 1.5%-3.1%) in diagnostic yield of intracranial injuries [7]. In another large study of 305 EDs in the US participating in the Emergency Quality Network (E-QUAL) Avoidable Imaging Initiative, researchers found that EDs with sustained participation in the initiative showed significant decreases in imaging utilization rates for syncope and minor head injury, the latter of which relied on the Canadian CT Head Rule [8, 22]. Utilizing the Canadian CT Head Rule for adults patients and the PECARN brain injury instrument for pediatric patients, the authors estimated that implementation of these rules would decrease inappropriate imaging by an estimated 1,083 to 3,643, and 44 to 51 CTs for adult and pediatric patients respectively, with significant costs savings. Although this study did not directly look at implementation of the clinical decision instruments, it demonstrates the potential to decrease what is otherwise liberal imaging use in low-risk patients.

CDIs emerged to assist with targeting the right intervention to the right patient. These instruments are supplements to clinical gestalt and reasoning, which alone led to decades of binge-testing. Although the plethora of clinical decision instruments are heterogenous regarding rigor and validation, several have been validated and have demonstrated real-world impact, such as decreased imaging rates. CDIs are only as good as the clinicians using the data. Medical education must emphasize the ability to critically appraise research and underscore that clinicians understand the appropriate patient and scenario for clinical decision instrument use, just like any diagnostic test. It is time to recalibrate our risk tolerance and discover how to get quality evidence-based care to the bedside.

Declarations

Conflict of interest The authors have no conflicts of interest to declare.

References

1. Goodman KE, Rodman AM, Morgan DJ. Preparing physicians for the clinical algorithm era. N Engl J Med. 2023;389(6):483–7.
2. Schriger DL, Elder JW, Cooper RJ. Structured clinical decision aids are seldom compared with subjective physician judgment, and are seldom superior. Ann Emerg Med. 2017;70(3):338-344.e3.
3. Sanders S, Doust J, Glasziou P. A systematic review of studies comparing diagnostic clinical prediction rules with clinical judgment. PLoS ONE. 2015;10(6): e0128233.
4. Babl FE, Oakley E, Dalziel SR, et al. Accuracy of clinician practice compared with three head injury decision rules in children: a prospective cohort study. Ann Emerg Med. 2018;71(6):703–10.
5. Finnerty NM, Rodriguez RM, Carpenter CR, et al. Clinical decision rules for diagnostic imaging in the emergency department: a research agenda. Acad Emerg Med. 2015;22(12):1406–16.
6. Stiell IG, Clement CM, Grimshaw JM, et al. A prospective cluster-randomized trial to implement the Canadian CT Head Rule in emergency departments. CMAJ. 2010;182(14):1527–32.
7. Sharp AL, Huang BZ, Tang T, et al. Implementation of the Canadian CT head rule and its association with use of computed tomography among patients with head injury. Ann Emerg Med. 2018;71(1):54-63.e2.
8. Venkatesh AK, Scofi JE, Rothenberg C, et al. Choosing wisely in emergency medicine: Early results and insights from the ACEP emergency quality network (E-QUAL). Am J Emerg Med. 2021;39:102–8.
9. Davey K, Saul T, Russel G, Wassermann J, Quaas J. Application of the Canadian computed tomography head rule to patients with minimal head injury. Ann Emerg Med. 2018;72(4):342–50.
10. Green SM, Schriger DL, Yealy DM. The case for 1-way clinical decision rules in emergency medicine reply. Ann Emerg Med. 2015;66(6):690.
11. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. J Am Med Inform Assoc. 2017;24(2):423–31.
12. Mosier K, Skitka L. Chapter 10: human decision makers and automated decision aids: made for each other? In: Parasuraman R, Mouloua M, editors. Automation and human performance: theory and applications. Mahwah, NJ: Lawrence Erlbaum Associates; 1996. p. 201–20.
13. Cairns C, Ashman. J, King JM. Emergency Department Visit Rates by Selected Characteristics: United States, 2020 [Internet]. National Center for Health Statistics (U.S.); 2022 [cited 2024 Apr 11]. Available from: https://stacks.cdc.gov/view/cdc/121837
14. Kelen G, Wolfe R, D'Onofrio G, et al. Emergency department crowding: the canary in the health care system. NEJM Catalyst 2021;5(2)
15. Carpenter CR, Raja AS, Brown MD. Overtesting and the downstream consequences of overtreatment: implications of "preventing overdiagnosis" for emergency medicine. Acad Emerg Med. 2015;22(12):1484–92.
16. Platts-Mills TF, Nagurney JM, Melnick ER. Tolerance of uncertainty and the practice of emergency medicine. Ann Emerg Med. 2020;75(6):715–20.
17. Kwee TC, Almaghrabi MT, Kwee RM. Diagnostic radiology and its future: what do clinicians need and think? Eur Radiol. 2023;33(12):9401–10.
18. Freund Y, Chauvin A, Jimenez S, et al. Effect of a diagnostic strategy using an elevated and age-adjusted d-dimer threshold on thromboembolic events in emergency department patients with suspected pulmonary embolism: a randomized clinical trial. JAMA. 2021;326(21):2141–9.

19. van der Pol LM, Tromeur C, Bistervels IM, et al. Pregnancy-adapted years algorithm for diagnosis of suspected pulmonary embolism. N Engl J Med. 2019;380(12):1139–49.
20. Freund Y, Cachanado M, Aubry A, et al. Effect of the pulmonary embolism rule-out criteria on subsequent thromboembolic events among low-risk emergency department patients: the Proper randomized clinical trial. JAMA. 2018;319(6):559.
21. Vinson DR, Ballard DW, Dayan PS, Kuppermann N. Clinical judgment and the pediatric emergency care applied research network head trauma prediction rules. Ann Emerg Med. 2018;72(3):323–5.
22. Simon LE, Kene MV, Warton EM, et al. Diagnostic performance of emergency physician gestalt for predicting acute appendicitis in patients age 5 to 20 years. Acad Emerg Med. 2020;27(9):821–31. https://doi.org/10.1111/acem.13931.
23. Solbiati M, Talerico G, Villa P, et al. Multicentre external validation of the Canadian Syncope Risk Score to predict adverse events and comparison with clinical judgment. Emerg Med J. 2021;38(9):701–6.
24. Zimmermann T, du Fay J, de Lavallaz NT, et al. International validation of the canadian syncope risk score : a cohort study. Ann Intern Med. 2022;175(6):783–94.
25. Valtchinov VI, Ip IK, Khorasani R, et al. Use of imaging in the emergency department: do individual physicians contribute to variation? AJR Am J Roentgenol. 2019;213(3):637–43.
26. Westafer LM, Shieh M-S, Pekow PS, Stefan MS, Lindenauer PK. Outpatient management of patients following diagnosis of acute pulmonary embolism. Acad Emerg Med. 2021;28(3):336–45.
27. Westafer LM, Jessen E, Zampi M, et al. Barriers and facilitators to the outpatient management of low-risk pulmonary embolism from the emergency department. Ann Emerg Med. 2023;82(3):381–93.