



Hippocrates and prophecies: the unfulfilled promise of prediction rules

Claudio M. Martin, MSc, MD, FRCPC · Fran Priestap, MSc

Received: 1 November 2021 / Revised: 1 November 2021 / Accepted: 1 November 2021 / Published online: 31 January 2022
© Canadian Anesthesiologists' Society 2022

Around 2500 BC (Before COVID-19), Hippocrates stated that it is unwise to prophesy either death or recovery in acute disease.¹ Despite realms of data and computing power available to generate predictions, this still seems valid today, as evidenced by the huge number of publications reporting new prediction models. An admittedly crude search on PubMed with “prognosis OR prediction” produces over 3.3 million papers and more than 20,000 when combined with “COVID-19 OR SARS-CoV-2”. A systematic reviewed published in March 2020, just as the virus was beginning to spread globally, screened 2,696 titles to identify 27 studies that developed or validated a multivariable COVID-19 prediction model.² These authors conclude that the proposed models are “poorly reported, at high risk of bias, and their reported performance is probably optimistic”.

Statistical models that relate patient characteristics to outcomes can serve three general purposes. First, the least contentious purpose is consistent and standard reporting and risk adjustment of clinical trial results, which was purportedly the primary reason for the *Sepsis-3* definition.³ The second purpose is controlling for heterogeneity when reporting quality metrics, although caution must be

exercised when applying and interpreting these results, which are typically aggregated at an institutional level.⁴ The third purpose is using these results for individual prognostication. Often implemented as clinical prediction rules, these are frequently found wanting.^{5,6} In critical care, an early approach to determining futility was based on three or more organ failures for three or more days.⁷ Nevertheless, improvements in outcomes, for undetermined reasons, soon outdated that clinical prediction rule.⁸

“Unreliable predictions could cause more harm than benefit in guiding clinical decisions”² is a statement that we strongly endorse. We are particularly concerned with proposals to use scores for triage. For example, a framework to guide resource allocation for critically ill patients with COVID-19 included prediction of survival *without specifying how that should be performed*.⁹ Nevertheless, a retrospective evaluation of two triage scoring guidelines for the allocation of mechanical ventilators identified few patients as low priority and there was poor agreement between the two triage scoring guidelines.¹⁰ In fact, two well-established mortality prediction models were compared at the level of the individual patient using a threshold mortality prediction of 50% as proposed in triage models, and the resulting graph was a cloud.¹¹

Prediction may be improved through longitudinal observations and modelling.^{8,12} In this issue of the *Journal*, Bartoszko *et al.* report such an approach using dynamic modelling based on three-day intervals.¹³ Their population was highly selected at a single quaternary centre where nearly one in three patients in the cohort received extracorporeal membrane oxygenation. They rightly concluded that external validation is required, *but also suggest that the tool can be used to inform decision-making and resource allocation and allow population level comparisons across institutions.*

C. M. Martin, MSc, MD, FRCPC (✉)
Division of Critical Care, Department of Medicine, Schulich School of Medicine and Dentistry, Western University and Lawson Health Research Institute, London, ON, Canada
e-mail: cmartin1@uwo.ca

London Health Sciences Centre - Victoria Hospital, 800 Commissioners Rd E, London, ON N6A 5W9, Canada

F. Priestap, MSc
Trauma Program, Department of Surgery, London Health Sciences Centre, London, ON, Canada

The authors appropriately applied the TRIPOD guidelines to report their new model.¹⁴ Nevertheless, investigators generally have a mandate or obligation to include knowledge translation in their research programs. For clinical prediction rules, this process begins with model development, continues with external validation, and finishes by evaluating the impact of the new tool.¹⁵ Increasingly, access to big data (including open access data sets) presents opportunities to externally validate prediction models and rules.¹⁶ Evaluating impact requires a more purposeful approach.¹⁵ Notably, a systematic review of selection criteria for triage and transfer of critically ill patients did not find any study that prospectively evaluated the implementation of its cited criteria.¹⁷

We should also consider that limiting aggressive care may not always worsen patient outcomes. Promoting intensive care unit admission for elderly patients, an approach opposite to triage that presumably occurs in standard practice, was tested in a prospective cluster-randomized trial.¹⁸ Mortality was higher in the intervention group. This phenomenon has also been reported in the cancer literature,¹⁹ and mandates that appropriate outcomes should be selected, ideally through consultation with patient partners.²⁰

Despite the noted challenges to use clinical prediction rules for triage, their application may still be appropriate when the intention is of lesser consequence, such as reducing resource utilization and radiation exposure from unnecessary imaging. Two such examples are the *Ottawa Ankle Rules*, which identify ankle injury patients unlikely to have a clinically significant fracture²¹ and the *Canadian CT Head Rule*, which can rule out intracranial injuries requiring neurosurgical intervention in patients with minor head injury.²² The ability of these tools to reduce unnecessary radiation exposure and healthcare costs has been extensively validated and assessed,^{23–26} contributing to the successful adoption and use of these tools in practice.

We challenge researchers who propose algorithms or clinical prediction tools for the purpose of triage or resource allocation (i.e., the decision to withhold or withdraw life-sustaining interventions) to prospectively evaluate the impact of these tools on clinical practice and patient-important outcomes.

Hippocrate et prophéties : la promesse non tenue des règles de prédiction

Vers 2500 av. J.-C. (avant les Jours de COVID-19), Hippocrate déclarait qu'il n'était pas sage de prophétiser la

mort ou la guérison en cas de maladie aiguë.¹ Malgré les quantités de données et l'impressionnante puissance de calcul disponibles pour générer des prédictions, cela semble toujours d'actualité, comme en témoigne l'immense nombre de publications faisant état de nouveaux modèles de prédiction. En réalisant une recherche somme toute grossière sur PubMed des termes « *prognosis* OU *prediction* » (soit pronostic ou prédiction), on trouve plus de 3,3 millions d'articles, et plus de 20 000 lorsque ces termes sont combinés à « COVID-19 OU SARS-CoV-2 ». Une revue systématique publiée en mars 2020, juste au moment où le virus commençait à se propager à l'échelle mondiale, a examiné 2696 titres pour finir par identifier 27 études qui avaient développé ou validé un modèle de prédiction multivariable pour la COVID-19.² Ces auteurs concluent que les modèles proposés sont « mal rapportés, à haut risque de biais, et que leur performance rapportée est probablement optimiste ».

Les modèles statistiques qui relient les caractéristiques des patients aux résultats et issues peuvent servir trois objectifs généraux. Premièrement, l'objectif le moins litigieux est de rapporter de façon cohérente et standardisée les résultats des études cliniques ainsi que les ajustements des risques, ce qui semblait constituer la principale raison de la définition de *Sepsis-3*.³ Le deuxième objectif est de contrôler l'hétérogénéité lorsque l'on rapporte des mesures de qualité, bien qu'il faille faire preuve de prudence lors de l'application et de l'interprétation de ces résultats, qui sont généralement agrégés au niveau institutionnel.⁴ Le troisième objectif est d'utiliser ces résultats pour le pronostic individuel. Souvent mis en œuvre en tant que règles de prédiction clinique, ces résultats se révèlent fréquemment inadéquats.^{5,6} En soins intensifs, une approche précoce pour déterminer la futilité se fondait sur trois défaillances d'organes ou plus pendant trois jours ou plus.⁷ Pourtant, pour des raisons indéterminées, l'amélioration des issues a rapidement dépassé cette règle de prédiction clinique.⁸

« Des prédictions peu fiables pourraient causer plus de torts que de bienfaits pour guider les décisions cliniques » :² voilà une déclaration que nous endossons fermement. Nous sommes particulièrement préoccupés par les propositions visant à utiliser des scores pour le triage. Par exemple, un cadre visant à guider l'allocation des ressources pour les patients gravement malades atteints de COVID-19 comprenait la prédiction de la survie *sans préciser comment y parvenir*.⁹ Néanmoins, une évaluation rétrospective de deux lignes directrices sur le triage en fonction de scores pour l'attribution des ventilateurs mécaniques n'a identifié qu'un nombre restreint de patients comme étant peu prioritaires et la concordance était faible entre les deux lignes directrices concernant les scores pour le triage.¹⁰ En fait, deux modèles de prédiction

de la mortalité bien établis ont été comparés au niveau du patient individuel en utilisant un seuil de prédiction de la mortalité à 50 %, tel que proposé dans les modèles de triage, et le graphique en résultant était un nuage.¹¹

La prédiction pourrait être améliorée grâce à des observations longitudinales et à la modélisation.^{8,12} Dans ce numéro du *Journal*, Bartoszko *et coll.* rendent compte d'une telle approche en utilisant une modélisation dynamique basée sur des intervalles de trois jours.¹³ Leur population a été sélectionnée avec soin dans un seul centre quaternaire dans lequel près d'un patient sur trois de la cohorte a reçu une oxygénation extracorporelle (ECMO). Les auteurs ont conclu à juste titre qu'une validation externe était nécessaire, *mais ont également suggéré que leur outil pouvait être utilisé pour éclairer la prise de décision et l'allocation des ressources, ainsi que pour permettre des comparaisons au niveau de la population entre les institutions.*

Les auteurs ont bien suivi les directives TRIPOD pour présenter leur nouveau modèle.¹⁴ Néanmoins, les chercheurs ont généralement le mandat ou l'obligation d'inclure l'application des connaissances dans leurs programmes de recherche. En ce qui a trait aux règles de prédiction clinique, ce processus commence par le développement d'un modèle, se poursuit par une validation externe, et se termine par l'évaluation de l'impact du nouvel outil.¹⁵ De plus en plus, l'accès aux mégadonnées (y compris les ensembles de données en libre accès) offre des possibilités de valider en externe les modèles et les règles de prédiction.¹⁶ L'évaluation de l'impact nécessite une approche plus ciblée.¹⁵ Soulignons qu'une revue systématique des critères de sélection pour le triage et le transfert des patients en état critique n'a permis de trouver aucune étude évaluant de manière prospective la mise en place de ses critères cités.¹⁷

Nous devrions également tenir compte du fait que les issues des patients n'empirent pas toujours lorsqu'on limite les soins agressifs. La promotion de l'admission en unité de soins intensifs pour les patients âgés, une approche opposée au triage qui se fait probablement dans la pratique courante, a été testée dans une étude prospective randomisée en grappes.¹⁸ La mortalité était plus élevée dans le groupe d'intervention. Ce phénomène a également été rapporté dans la littérature oncologique,¹⁹ c'est pourquoi les résultats appropriés devraient être retenus, idéalement par le biais de consultations avec les patients partenaires.²⁰

Malgré les difficultés relevées dans l'utilisation de règles de prédiction clinique pour le triage, leur application peut tout de même être adaptée lorsque l'intention est de moindre conséquence, comme par exemple pour réduire l'utilisation des ressources et l'exposition aux radiations due à des radiographies inutiles. Voici deux exemples de

telles applications : les *Règles d'Ottawa pour la cheville* permettent d'identifier les patients blessés à la cheville peu susceptibles d'avoir une fracture cliniquement significative,²¹ et les critères de tomodensitométrie crânienne établis dans le *Canadian CT Head Rule* permettent d'exclure les lésions intracrâniennes nécessitant une intervention neurochirurgicale chez les patients présentant un traumatisme crânien mineur.²² La capacité de ces outils à réduire l'exposition inutile aux radiations et les coûts des soins de santé a été largement validée et évaluée,^{23–26} contribuant à l'adoption et à l'utilisation réussies de ces outils dans la pratique.

Nous mettons au défi les chercheurs qui proposent des algorithmes ou des outils de prédiction clinique à des fins de triage ou d'allocation des ressources (c.-à-d. la décision de suspendre ou de retirer des interventions de maintien de la vie) d'évaluer prospectivement l'impact de ces outils sur la pratique clinique et les issues importantes pour le patient.

Disclosures None.

Funding statement None.

Editorial responsibility This submission was handled by Dr. Philip M. Jones Deputy Editor-in-Chief, *Canadian Journal of Anesthesia*.

Déclaration Aucune.

Déclaration de financement Aucune.

Responsabilité éditoriale Cet article a été traité par Dr Philip M. Jones, rédacteur en chef adjoint, *Journal canadien d'anesthésie*.

References

1. Lloyd GE. Hippocratic Writings. Cambridge, UK: Penguin Books; 1983.
2. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020; DOI: <https://doi.org/10.1136/bmj.m1328>.
3. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016; DOI: <https://doi.org/10.1001/jama.2016.0288>.
4. Priestap F, Kao R, Martin CM. External validation of a prognostic model for intensive care unit mortality: a retrospective study using the Ontario Critical Care Information System. *Can J Anesth* 2020; 67: 981-91.
5. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997; 277: 488-94.
6. Ban JW, Empanaza JI, Urreta I, Burls A. Design characteristics influence performance of clinical prediction rules in validation: a meta-epidemiological study. *PLoS One* 2016; DOI: <https://doi.org/10.1371/journal.pone.0145779>.

7. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. Prognosis in acute organ-system failure. *Ann Surg* 1985; 202: 685-93.
8. Zimmerman JE, Knaus WA, Wagner DP, Sun X, Hakim RB, Nystrom PO. A comparison of risks and outcomes for patients with organ system failure: 1982-1990. *Crit Care Med* 1996; 24: 1633-41.
9. Sprung CL, Joynt GM, Christian MD, Truog RD, Rello J, Nates JL. Adult ICU triage during the coronavirus disease 2019 pandemic: who will live and who will die? Recommendations to improve survival. *Crit Care Med* 2020; 48: 1196-202.
10. Wunsch H, Hill AD, Bosch N, et al. Comparison of 2 triage scoring guidelines for allocation of mechanical ventilators. *JAMA Netw Open* 2020; DOI: <https://doi.org/10.1001/jamanetworkopen.2020.29250>.
11. Lemeshow S, Klar J, Teres D. Outcome prediction for individual intensive care patients: useful, misused, or abused? *Intensive Care Med* 1995; 21: 770-6.
12. Beil M, Sviri S, Flaatten H, et al. On predictions in critical care: the individual prognostication fallacy in elderly patients. *J Crit Care* 2021; 61: 34-8.
13. Bartoszko J, Dranitsaris G, Wilcox ME, et al. Development of a repeated-measures predictive model and clinical risk score for mortality in ventilated COVID-19 patients. *Can J Anesth* 2022; this issue.
14. Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; 162: W1-73.
15. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012; 98: 691-8.
16. Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016; DOI: <https://doi.org/10.1136/bmj.i3140>.
17. Dahine J, Hébert PC, Ziegler D, Chenail N, Ferrari N, Hébert R. Practices in triage and transfer of critically ill patients: a qualitative systematic review of selection criteria. *Crit Care Med* 2020; 48: e1147-57.
18. Guidet B, Leblanc G, Simon T, et al. Effect of systematic intensive care unit triage on long-term mortality among critically ill elderly patients in France: a randomized clinical trial. *JAMA* 2017; 318: 1450-9.
19. Temel JS, Greer JA, Muzikansky A, et al. Early palliative care for patients with metastatic non-small-cell lung cancer. *N Engl J Med* 2010; 363: 733-42.
20. Manafò E, Petermann L, Vandall-Walker V, Mason-Lai P. Patient and public engagement in priority setting: a systematic rapid review of the literature. *PLoS One* 2018; DOI: <https://doi.org/10.1371/journal.pone.0193579>.
21. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med* 1992; 21: 384-90.
22. Stiell IG, Lesiuk H, Wells GA, et al. The Canadian CT Head Rule Study for patients with minor head injury: rationale, objectives, and methodology for phase I (derivation). *Ann Emerg Med* 2001; 38: 160-9.
23. Barelds I, Krijnen WP, van de Leur JP, van der Schans CP, Goddard RJ. Diagnostic accuracy of clinical decision rules to exclude fractures in acute ankle injuries: systematic review and meta-analysis. *J Emerg Med* 2017; 53: 353-68.
24. Stiell I, Wells G, Laupacis A, et al. Multicentre trial to introduce the Ottawa ankle rules for use of radiography in acute ankle injuries. Multicentre Ankle Rule Study Group. *BMJ* 1995; 311: 594-7.
25. Smits M, Dippel DW, de Haan GG, et al. External validation of the Canadian CT head rule and the New Orleans criteria for CT scanning in patients with minor head injury. *JAMA* 2005; 294: 1519-25.
26. Rivara FP, Kuppermann N, Ellenbogen RG. Use of clinical prediction rules for guiding use of computed tomography in adults with head trauma. *JAMA* 2015; 314: 2629-31.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.