

EDITORIAL OPEN



Beyond performance metrics: modeling outcomes and cost for clinical machine learning

Advances in medical machine learning are expected to help personalize care, improve outcomes, and reduce wasteful spending. In quantifying potential benefits, it is important to account for constraints arising from clinical workflows. Practice variation is known to influence the accuracy and generalizability of predictive models, but its effects on cost-effectiveness and utilization are less well-described. A simulation-based approach by Mišić and colleagues goes beyond simple performance metrics to evaluate how process variables may influence the impact and financial feasibility of clinical prediction algorithms.

npj Digital Medicine (2021)4:119; <https://doi.org/10.1038/s41746-021-00495-4>

Advances in medical machine learning are expected to help personalize care, improve outcomes, and reduce wasteful spending. In quantifying potential benefits, it is important to account for constraints arising from clinical workflows. Practice variation is known to influence the accuracy and generalizability of predictive models^{1,2}, but its effects on cost-effectiveness and utilization are less well-described. A simulation-based approach by Mišić and colleagues³ goes beyond simple performance metrics to evaluate how process variables may influence the impact and financial feasibility of clinical prediction algorithms.

Mišić et al.'s study builds on previous work that developed equations for predicting unplanned readmissions⁴. Readmission rate is a publicly reported metric that commands significant attention in quality improvement and cost management initiatives. As part of these efforts, prediction equations are used to stratify readmission risk and allocate limited interventions to the patients who need it most. Still, many important questions—how many interventions are applied, how many readmissions are prevented, and how much spending is averted—are often unclear.

Mišić and colleagues provide answers by simulating patient flow for a hypothetical clinical workflow. Under their model, each patient is assigned a risk score based on four separate algorithms (LACE, HOSPITAL, and two locally-designed equations) for each day a prediction is available. On each day that a provider is available, the eight highest-risk patients are “treated” with a hypothetical intervention that has a 10% chance of preventing readmission. The authors applied this model to a dataset of 19,331 post-operative surgical patients from the UCLA Ronald Reagan Medical Center, including 969 (5.0%) who were later readmitted. Because the described workflow is speculative, it cannot be validated with data. Instead, the design and parameters were chosen to reflect typical staff and time constraints.

Using these simulation conditions, Mišić and colleagues compute utilization and volume metrics for the Ronald Reagan Medical Center, including interventions conducted, readmissions anticipated, and expected readmissions prevented. To compute net savings, the authors added the highest-cost ICD-10 codes for each “prevented” readmission and subtracted expected labor costs. By toggling simulation parameters, the authors also show how differences in accuracy, prediction timing, and provider

availability translate into differences in outcomes. For example, algorithms that rely on length of stay are unable to assign risk scores before the day of discharge, potentially constraining opportunities to intervene. The authors also find several parameter settings where costs outweigh savings, consistent with earlier studies showing that interventions for preventing readmission are not always cost-effective⁵.

The simulation approach relies on a broad set of simplifying assumptions and therefore has several limitations. In particular, the assumption of fixed, limited availability (e.g., one nurse practitioner providing readmission interventions for a 520-bed hospital) may be overly stringent, or may overlook the need for additional funding to support effective programs. Assumptions for intervention timing may also be inexact, as strategies for preventing hospital readmission increasingly comprise multiple components administered before, during, and after discharge⁶. Last, the evaluated algorithms do not account for many important drivers of readmissions, such as language and cultural barriers, mental illness, and poverty. Allocating interventions based on clinical risk alone may not represent the most common or effective strategy for reducing rehospitalization. Together, these considerations indicate the need to validate simulation results against real-world data and recalibrate assumptions where necessary. Beyond validation, future work should extend the model to provide estimates of uncertainty and evaluate health and equity-based outcomes.

Ultimately, the proposed simulation model provides estimates for utilization and financial feasibility in the setting of a specific clinical workflow. Preventing rehospitalization is only one application; others include prevention of sepsis⁷ or acute kidney injury⁸. While not a substitute for randomized trials⁹, simulation modeling can provide initial answers and insights for all stakeholders involved, including researchers developing prediction algorithms, administrators optimizing clinical workflows, executives evaluating business models, and regulators seeking to understand performance in context.

For decades, medical practice has proved impervious to algorithmic reinvention¹⁰. One contributor is imperfect communication of a clear value proposition centered on outcomes, costs, and metrics that matter. Performance metrics like sensitivity and specificity are only part of the puzzle. A simulated modeling approach may help contextualize and complement traditional accuracy metrics to strengthen the case for new prediction models.

Received: 11 June 2021; Accepted: 27 July 2021;
Published online: 10 August 2021

James A. Diao¹✉, Leia Wedlund¹ and Joseph Kvedar^{1,2}
¹Harvard Medical School, Boston, MA, USA. ²Mass General Brigham,
Boston, MA, USA. ✉email: james_diao@hms.harvard.edu

REFERENCES

1. Agniel, D., Kohane, I. S. & Weber, G. M. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ* **361**, k1479 (2018).
2. Beaulieu-Jones, B. K. et al. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? *NPJ Digit. Med.* **4**, 62 (2021).
3. Mišić, V. V., Rajaram, K. & Gabel, E. A simulation-based evaluation of machine learning models for clinical decision support: application and analysis using hospital readmission. *NPJ Digit. Med.* **4**, 98 (2021).
4. Mišić, V. V., Gabel, E., Hofer, I., Rajaram, K. & Mahajan, A. Machine learning prediction of postoperative emergency department hospital readmission. *Anesthesiology* **132**, 968–980 (2020).
5. Nuckols, T. K. et al. Economic evaluation of quality improvement interventions designed to prevent hospital readmission: a systematic review and meta-analysis. *JAMA Intern. Med.* **177**, 975–985 (2017).
6. Kripalani, S., Theobald, C. N., Anctil, B. & Vasilevskis, E. E. Reducing hospital readmission rates: current strategies and future directions. *Annu. Rev. Med.* **65**, 471–485 (2014).
7. Reyna, M. et al. Early Prediction of Sepsis from Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. *2019 Computing in Cardiology Conference (CinC)* (2019) <https://doi.org/10.22489/cinc.2019.412>.
8. Flechet, M. et al. Machine learning versus physicians' prediction of acute kidney injury in critically ill adults: a prospective evaluation of the AKIpredictor. *Crit. Care* **23**, 1–10 (2019).
9. Nikolova-Simons, M. et al. A randomized trial examining the effect of predictive analytics and tailored interventions on the cost of care. *NPJ Digit. Med.* **4**, 92 (2021).
10. Schwartz, W. B., Patil, R. S. & Szolovits, P. Artificial intelligence in medicine. Where do we stand? *N. Engl. J. Med.* **316**, 685–688 (1987).

AUTHOR CONTRIBUTIONS

Initial draft by J.A.D. Critical revisions by L.W. and J.K. All authors approved the final draft.

COMPETING INTERESTS

J.K. is the Editor-in-Chief of *npj Digital Medicine*. J.A.D. and L.W. declare no competing interests.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021