When Predictive Models Collide

Peter D. Stetson, MD, MA¹; Michael N. Cantor, MD, MA²; and Mithat Gonen, PhD³

PREDICTIVE MODELS RISING

No trend is more visible in medicine than integrating machine learning and artificial intelligence (AI) into clinical care. Terms like “random forests” and “convolutional neural networks” that were previously the domain of computer science PhDs are now part of the vocabulary of both C-suite administrators and front-line clinicians. As the high-performance storage and computing power that underlie these models become a commodity, both the supply and demand for predictive models based on machine learning will only increase. Advances in genomics¹ and patient-generated health data will enrich predictive models, but will also increase model complexity even as they increase precision. How can clinicians and modelers prevent the messy accidents that could happen when these models collide?

GENERALIZING PREDICTIVE MODELS TO COMPLEX PATIENTS

Predictive models in oncology have long been used to improve early diagnosis and better identify ideal treatment and monitoring based on responses predicted by these models,²⁻⁵ including recent work taking advantage of novel markers from emerging technologies.⁶⁻⁸ Yet, although models may perform well for a single disease or clinical situation, the circumstances of the laboratory where they are developed are rarely replicated in real-world clinical settings. Advanced cognitive computing solutions have been shown to have variable performance in the wild⁹⁻¹² and may require additional patient characteristics (especially genetic data) to perform optimally in certain diseases.¹³

Development of predictive models often involves dealing with competing risks. Being discharged is not an eventuality for an inpatient; dying in the hospital is an outcome, too. Yet all length-of-stay (LOS) models censor patients who die in the hospital, essentially implying that death is an event that happens independent of being discharged. Statistical methodology to deal with such competing risks is well developed and understood, but largely ignored by the field of medical machine learning. Incorporating these methods into predictive model building will improve transparency and allow the providers to anticipate some of the collisions in the field.

The current state of predictive modeling parallels that of clinical practice guidelines, particularly in the case of patients with multiple comorbidities (aka the

“multimorbid”), with the potential for contradictory or even detrimental recommendations when each model is applied individually, rather than taking into account the overall clinical context.¹⁴ Recent, high-profile examples of conflicting guidelines include the differences among the Veterans Administration, American Diabetes Association, and American Society of Clinical Endocrinologists for managing diabetes and the US Preventive Services Task Force, American Congress of Obstetricians, and American College of Surgeons recommendations for mammography.¹⁵

One approach to managing these conflicts is identifying and targeting likely subgroups at the beginning of the guideline development process, knowing that guideline recommendations will never be evidence-based at the individual patient level.¹⁵ Where guidelines and predictive models differ, however, is that predictive models can be based on local data, tailored for a particular patient population, and therefore approach the ideal of individualized, evidence-based recommendations. Although models may only be as good as their training set, the increased availability of clinical data on diverse patient populations means that models may be precisely tailored for those high-risk subgroups during their development.

PRIORITIZATION CONSIDERATIONS: PATIENTS, PROVIDERS, AND SYSTEMS

Effective implementation of predictive models requires a three-part approach: (1) prioritization or context-sensitive weighting, (2) integration into workflows, and (3) a cognitive support framework for clinicians. Prioritization and decision modeling is a common challenge in medicine and has been successfully applied in prioritizing preventive care recommendations in the outpatient setting.¹⁶ With this approach in place, colliding models could then provide opportunities for shared decision making, giving patients a line of sight into their complex conditions and the multiple interventions for which they are eligible. The flexibility and dynamism of many models may help patients envision their futures based on decisions made in the present, while also helping them make their own decisions about prioritizing issues like length versus quality of life.

Models and their user interface displays must be dynamically prioritized. For example, the expected LOS model (with popup reminders to prepare for timely discharge by doing your medication reconciliation)

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on April 30, 2020 and published at ascopubs.org/journal/cci on June 16, 2020; DOI <https://doi.org/10.1200/CCI.20.00024>

should be suppressed for an admitted patient with acute myeloid leukemia post–bone marrow transplantation who has neutropenic fever and is on the verge of sepsis according to its model (a care-level model conflict). In the outpatient setting, a patient with congestive heart failure, major depression, and obesity, and in active treatment for recurrent breast cancer, may not need to know that she is at high risk for developing diabetes in 6 months when she is more likely to develop chemotherapy-related toxicity requiring an emergency room visit or admission in the next 30 days (a temporality model conflict). That same patient with breast cancer, congestive heart failure, depression, and obesity, when ultimately admitted to the hospital, could have a range of models applied, including readmission risk, in-hospital mortality risk, long-term mortality risk, risk of progression to renal failure, risk of breast cancer progression, and so on (a relative-risk model conflict). A dynamic prioritization system is essential to not only highlight interventions that are in the best interests of patients and align with their current priorities, but also to protect clinicians from alert fatigue and burnout. The last thing we want to do is add predictive model fatigue to the already omnipresent alert fatigue, documentation fatigue, and inbox fatigue.

INTEGRATING COMPLEX PREDICTIVE MODELS INTO CLINICAL WORKFLOW

A major challenge for any clinical decision support system, from guidelines to alerts to predictive models, is the actionable display of complex information so that clinicians can act effectively. The relatively basic tools available in most electronic health records (EHRs) would be quickly overwhelmed in the case of a multimorbid patient at an institution with a sophisticated predictive modeling infrastructure. Would a series of interruptive alerts fire? Would there be a line of red-yellow-green traffic signals across the patient's chart header, each representing a particular risk? Would providers get multiple, conflicting secure text messages for the same patient from a loquacious, AI-driven chat bot? To effectively apply the power of machine learning to improve health, we must address implementation challenges that, because they involve human rather than

computer-based decision making, may be more difficult than the development of effective algorithms. Ongoing work in the domain called “Explainable Artificial Intelligence”¹⁷⁻²⁰ and “augmented intelligence”²¹ may offer methods for the transparency clinicians will require to make sense of recommendations from single and multiple competing predictive models. Clinicians can't realistically be expected to do shared decision making with patients if we do not understand the processes behind the algorithms.

RECOMMENDATIONS: A COGNITIVE SUPPORT FRAMEWORK

Predictive models have great potential to improve patient outcomes and improve the efficiency and effectiveness of medical interventions. To be effective, simultaneously active predictive models must be implemented with an accompanying cognitive support framework that permits their most efficient use. In Table 1, we propose the features of a multimodel cognitive support framework that would facilitate understanding of models and resolution of model collisions.

For models to reach their potential, developers and implementers must learn from experience with clinical guidelines to determine the optimal approach to integrate models into clinical care. Expertise from the fields of informatics and biostatistics should be applied in all three aspects of the proposed approach. EHR vendors must design human-computer user interfaces that help clinicians manage complex recommendations for complex patients and open up application programming interfaces for developers to extend EHRs with advanced platforms for predictive model delivery at the point of decision. Regulations must be developed that allow for the interactions of various priorities and that incentivize addressing complexity rather than checking the boxes with various isolated process measures.

Significant research opportunities exist regarding handling competing models. Given that implementation into real-world workflows remains a challenge for AI and machine learning models,²³ research regarding explainable AI or augmenting traditional clinical decision support with AI²⁴ has the potential to identify better ways to improve user

TABLE 1. Proposed Features of a Multimodel Cognitive Support Framework

Feature	Description
Transparent	Methods behind a predictive model are well described and explainable. Black Box models (eg, neural networks) have well-described inputs, outputs, and performance characteristics. ²²
Nonredundant	Results of models with overlapping suggestions or predictions are only displayed once.
Disambiguated	Recommendations are given in context. Directly conflicting suggestions are eliminated.
Weighted	Recommendations are prioritized based on an established set of context-sensitive criteria (which may be personalized at the patient level).
Embedded	Models are integrated into the EHR and into clinical workflows (to reduce cognitive burden for providers and avoid burnout).
Dynamic	Models are consistently updated (to keep up with shifting evidence).

Abbreviation: EHR, electronic health record.

experience. Research regarding the effectiveness of multimodel predictions on oncology care outcomes should be prioritized. Finally, capturing audit log data during trade-off decisions made by clinicians when considering multiple models at play could assist the machine learning community to further refine their models, potentially resulting in better weighting and dynamism as proposed in our framework.

Developing complex algorithms is difficult, but ensuring their successful implementation to benefit patients is even harder. Applying the lessons learned from dealing with colliding guidelines and taking advantage of the dynamism of predictive models can help overcome this challenge and lead to more engaged patients and providers, better outcomes, and improved health overall.

AFFILIATIONS

¹Department of Medicine, Digital Informatics and Technology Solutions, Memorial Sloan Kettering Cancer Center, New York, NY

²Regeneron Genetics Center, Tarrytown, NY

³Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY

CORRESPONDING AUTHOR

Peter D. Stetson, MD, MA, Digital Informatics and Technology Solutions, Memorial Sloan Kettering Cancer Center, 633 3rd Ave, #1201, New York, NY 10017; Twitter: @inforMD, @sloan_kettering, @MSKCC_Library; e-mail: stetsonp@mskcc.org.

SUPPORT

Supported by Memorial Sloan Kettering Cancer Center Core Grant, No. P30CA008748 (P.D.S. and M.G.).

AUTHOR CONTRIBUTIONS

Conception and design: All authors

Collection and assembly of data: Peter D. Stetson

Data analysis and interpretation: Peter D. Stetson, Mithat Gonen

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](http://OpenPayments)).

Michael N. Cantor

Employment: Regeneron Genetics Center

Stock and Other Ownership Interests: Regeneron

No other potential conflicts of interest were reported.

REFERENCES

- Schram AM, Reales D, Galle J, et al: Oncologist use and perception of large panel next-generation tumor sequencing. *Ann Oncol* 28:2298-2304, 2017
- Gönen M, Schrag D, Weiser MR: Nodal staging score: A tool to assess adequate staging of node-negative colon cancer. *J Clin Oncol* 27:6166-6171, 2009
- Kattan MW, Eastham JA, Stapleton AM, et al: A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 90:766-771, 1998
- Lloyd-Jones DM, Wang TJ, Leip EP, et al: Lifetime risk for development of atrial fibrillation: The Framingham Heart Study. *Circulation* 110:1042-1046, 2004
- Partin AW, Kattan MW, Subong EN, et al: Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update. *JAMA* 277:1445-1451, 1997
- Paudyal R, Lu Y, Hatzoglou V, et al: Dynamic contrast-enhanced MRI model selection for predicting tumor aggressiveness in papillary thyroid cancers. *NMR Biomed* 33:e4166, 2019
- Pulvirenti A, Javed AA, Landoni L, et al: Multi-institutional development and external validation of a nomogram to predict recurrence after curative resection of pancreatic neuroendocrine tumors. *Ann Surg* 10.1097/SLA.0000000000003579 (epub ahead of print on September 24, 2019)
- Golia Pernicka JS, Gagniere J, Chakraborty J, et al: Radiomics-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation. *Abdom Radiol (NY)* 44:3755-3763, 2019
- Choi YI, Chung JW, Kim KO, et al: Concordance rate between clinicians and Watson for oncology among patients with advanced gastric cancer: Early, real-world experience in Korea. *Can J Gastroenterol Hepatol* 2019:8072928, 2019
- Lee WS, Ahn SM, Chung JW, et al: Assessing concordance with Watson for oncology, a cognitive computing decision support system for colon cancer treatment in Korea. *JCO Clin Cancer Inform* 2:1-8, 2018
- Somashekhar SP, Sepúlveda MJ, Puglielli S, et al: Watson for oncology and breast cancer treatment recommendations: Agreement with an expert multi-disciplinary tumor board. *Ann Oncol* 29:418-423, 2018
- Zhou N, Zhang CT, Lv HY, et al: Concordance study between IBM Watson for oncology and clinical practice for patients with cancer in China. *Oncologist* 24:812-819, 2019
- Kim YY, Oh SJ, Chun YS, et al: Gene expression assay and Watson for oncology for optimization of treatment in ER-positive, HER2-negative breast cancer. *PLoS One* 13:e0200100, 2018
- Boyd CM, Darer J, Boulton C, et al: Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: Implications for pay for performance. *JAMA* 294:716-724, 2005
- Greenfield S, Kaplan SH: When clinical practice guidelines collide: Finding a way forward. *Ann Intern Med* 167:677-678, 2017
- Taksler GB, Keshner M, Fagerlin A, et al: Personalized estimates of benefit from preventive care guidelines: A proof of concept. *Ann Intern Med* 159:161-168, 2013
- Gordon L, Grantcharov T, Rudzicz F: Explainable artificial intelligence for safe intraoperative decision support. *JAMA Surg* 154:1064, 2019

18. Lundberg SM, Nair B, Vavilala MS, et al: Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2:749-760, 2018
19. Spinner T, Schlegel U, Schafer H, et al: explAIner: A visual analytics framework for interactive and explainable machine learning. *IEEE Trans Vis Comput Graph* 1 26:1064-1074, 2019
20. Zhang J, Wang Y, Molino P, et al: Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models. *IEEE Trans Vis Comput Graph* 25:364-373, 2019
21. Barbour AB, Frush JM, Gatta LA, et al: Artificial intelligence in health care: Insights from an educational forum. *J Med Educ Curric Dev* 6:2382120519889348, 2020
22. Steyerberg EW, Vickers AJ, Cook NR, et al: Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* 21:128-138, 2010
23. Romero-Brufau S, Wyatt KD, Boyum P, et al: A lesson in implementation: A pre-post study of providers' experience with artificial intelligence-based clinical decision support. *Int J Med Inform* 137:104072, 2020
24. Bizzo BC, Almeida RR, Michalski MH, et al: Artificial intelligence and clinical decision support for radiologists and referring providers. *J Am Coll Radiol* 16:1351-1356, 2019