

Individual risk prediction using data beyond the medical clinic

Robert M. Califf MD, Frank E. Harrell Jr. PhD

■ Cite as: *CMAJ* 2018 August 13;190:E947-8. doi: 10.1503/cmaj.180967

See related article at www.cmaj.ca/lookup/doi/10.1503/cmaj.170914

At their cores, the disciplines of public health and clinical medicine are both based on decisions. A rational approach to decision-making relies on weighing the predicted benefits and risks of one alternative versus those of others. Accordingly, the quest for accurate predictions has consumed considerable effort for generations of researchers. Although the general framework for medical decision-making is decades old and many robust prediction models exist, we are entering an era in which the amount, quality and latency of data have changed dramatically. And as more data become available faster, many more analysts are crafting predictions pertinent to individuals and populations.

In linked research, Manuel and colleagues advance the field by showing that information from population surveys can produce estimates of cardiovascular disease risk of comparable accuracy to estimates created from data collected within the health care system using more traditional medical risk measures.¹ This approach offers much less expensive data collection and is more easily repeatable than detailed measures that require a physical visit to a clinic and often an evaluation by a specialist. Although in-depth symptoms, physical examinations and specialized medical tests would likely refine such predictions, the survey-based model is efficient and could be used to better rationalize the ordering of medical tests. Manuel and colleagues' model particularly used smoking history in a more powerful way than most previous models have done. One downside of the approach is its oversimplification of measures that require physical assessment and are continuous, such as information about blood pressure; however, the authors should be congratulated on their exceptionally thorough and transparent analysis using state-of-the-art statistical modelling.

An important element of the linked research is an in-depth analysis of the operating characteristics of the predictive model. Although appropriate methods for characterizing predictions are well-described,² many published models fail to consider essential elements. At a basic level, those developing a predictive model should define its ability to discriminate between people with and without the outcome of interest, and its calibration with the actual event proportions in the relevant population. A major strength of the present model is that the researchers were able to assess the calibration within 205 subpopulations that would be of interest for

KEY POINTS

- Amid growth in the quantity and availability of data, patients and their clinicians should expect ready access to high-quality information that predicts likely health outcomes over time, given alternative decisions about health and health care.
- Recent research shows that information from population surveys can produce estimates of cardiovascular disease risk of comparable accuracy to estimates created from traditionally collected health care data.
- Risk predictions using population survey data will be increasingly subject to scrutiny because, whereas the consequences of high-quality predictions will be better outcomes, erroneous predictions that lead to wrong decisions have capacity for harm.
- The rate at which actionable predictions result in better outcomes will depend on commitment to investing in high-quality data collection and curation; ensuring transparency about data provenance, analytical methods and operating characteristics; fostering methodologic appropriateness and quality; and monitoring the impact of algorithms when applied in practice.

decision-making. In fields with an advanced understanding of decision thresholds, further information about whether the information is useful in moving beyond a decision point is also advantageous. Knowing about a change in risk when nothing would be done differently may not be useful to patients or clinicians.

The only notable fault we find with the authors' analysis is that they, like many others, used percentiles of risk estimates in some characterizations of model performance. Grouping participants into percentiles would be appropriate in settings in which groups influence individual outcomes, but, in general, "percentiling" and the variable heterogeneity of absolute risk within intervals it entails obscures the fact that an individual's risk is a function of her own characteristics and not of some group to which she may belong (i.e., interval boundaries are functions of how many participants have similar predicted risks). Continuous calibration curves showing absolute accuracy over the entire risk spectrum without subgrouping provide a better way of understanding the characteristics of a model.

As we look forward, we must recognize the changes now occurring in the health information environment. Electronic health records

are universally available in North America, and transactions are recorded in real time. As technology improves, curated information will be available with shorter latency, making “just-in-time” decision support possible. At the same time, the costs of obtaining biomolecular data, including genome sequences and details of molecular and immune system function, are rapidly diminishing, and “digital phenotyping”³ with data from social media, sensors and cell phones provides deep information about behaviour, values and social interactions. In addition, fixed aspects of the built environment and variable environmental elements are available in a geospatial context.⁴ The same qualities of the digital era that make so much information available also enable bidirectional exchange, which in turn makes outputting of predictions in the form of decision support directly to the decision-maker increasingly possible.

These advances raise questions of how to ingest and analyze these newly accessible data sources efficiently and effectively to enable better understanding of health and disease, and to inform the many health and health care decisions being made every day. To this end, there is considerable interest in applications from the fields of machine learning and other kinds of artificial intelligence that can automate analysis to speed access to actionable results. In fact, multiple evolving approaches share a characteristic, in that the algorithm used changes in real time to optimize prediction. While such automation is attractive to those anxious for more effective decision support for health decisions, there is a robust discussion ongoing about which methods are best suited for which predictions. The spectrum of possibilities includes traditional biostatistical approaches that take into account explicit definitions of key analytical factors and assessment of biological and clinical plausibility; more flexible adaptive Bayesian approaches; and “black box” deep learning models that are difficult to explain in terms of discrete causal pathways.

Distinguishing predictions from causal inference is essential.⁵ There is a tendency to believe that predictions about likely outcomes can substitute for randomization in determining whether an outcome is causally related to an intervention or risk factor. While causal inference from observation and prediction is reasonable in many situations, for therapeutic clinical interventions, there is currently no substitute for randomization when expected effects are modest.

Converting probabilistic predictions to effective decision-making will require a major educational effort directed at both the health care workforce and the public, given that empirical studies reveal low lev-

els of numeracy.⁶ Our health care and public health systems must also invest in information curation and storage, and substantial funding is needed to develop an understanding of which analytical methods are best suited to which types of problems. In addition, regulatory agencies are developing approaches to oversee predictive instruments, which almost certainly will require evaluation of the quality of systems rather than oversight of individual algorithms. But the power of predictions to influence decisions requires regulatory oversight, as substantial harm could result from a widely propagated error.

After decades of preparation, we are at a point where people and their clinicians should expect to have ready access to high-quality information that predicts likely outcomes over time, given a set of alternatives to consider when making decisions about health and health care. The basis for these predictions will no longer be limited to traditional medical evaluation data, as shown by Manuel and colleagues in their linked paper.¹ It should be focused on actionable information that informs decisions or focused efforts to understand the mechanisms that generate health outcomes. These predictions will be increasingly subject to scrutiny because, whereas the consequences of high-quality predictions will be better outcomes, errors or inappropriately promoted predictions that lead to wrong decisions have an obvious capacity for harm. The rate at which actionable predictions lead to better outcomes will depend on a broad commitment to the following: investing in high-quality data through better collection and curation; fostering methodologic appropriateness and quality; ensuring transparency about data provenance, analytical methods and operating characteristics; and monitoring the impact of algorithms when applied in practice.

References

1. Manuel DG, Tuna M, Bennett C, et al. Development and validation of a cardiovascular disease risk-prediction model using population health surveys: the Cardiovascular Disease Population Risk Tool (CVDPORT). *CMAJ* 2018;190:E871-82.
2. Harrell FE. *Regression modeling strategies*. 2nd ed. New York: Springer; 2015.
3. Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA* 2017;318:1215-6.
4. Miranda ML, Ferranti J, Strauss B, et al. Geographic health information systems: a platform to support the ‘triple aim’. *Health Aff (Millwood)* 2013;32:1608-15.
5. Hernán MA, Hsu J, Healy B. Data science is science’s second chance to get causal inference right: a classification of data science tasks. Ithaca (NY): Cornell University Library; (preprint; v. 4, 2018 July 12). Available: <https://arxiv.org/abs/1804.10846> (accessed 2018 July 19).
6. Understanding literacy & numeracy. Atlanta: US Centers for Disease Control and Prevention; (updated 2016 Dec. 19). Available: www.cdc.gov/healthliteracy/learn/UnderstandingLiteracy.html (accessed 2018 July 19).

Competing interests: Robert Califf was the Commissioner of Food and Drugs for the US Food and Drug Administration from February 2016 to January 2017, and Deputy Commissioner for Medical Products and Tobacco for the US Food and Drug Administration from February 2015 to January 2016. Robert Califf serves on the corporate board for Cytokinetics. He also reports receiving consulting fees from Merck and Boehringer Ingelheim, and is employed as a scientific advisor by Verily Life Sciences (Alphabet). Frank Harrell has no conflicts of interest to disclose.

This article was solicited and has not been peer reviewed.

Affiliations: Duke Forge (Califf), Duke University School of Medicine, Durham, NC; Verily Life Sciences (Alphabet) (Califf), South San Francisco, Calif.; Department of Biostatistics (Harrell), Vanderbilt University Medical Center, Nashville, Ten.

Contributors: Both authors drafted the manuscript, gave final approval of the version to be published and agreed to be accountable for all aspects of the work.

Correspondence to: Robert Califf, robert.califf@duke.edu