
Editorial

The science of informatics and predictive analytics

Leslie Lenert

Institution of Medical University of South Carolina, Charleston, SC

As an interdisciplinary technologically driven field, the science of informatics is rapidly evolving. In this issue of *Journal of the American Medical Informatics Association*, we bring together a series of articles and commentaries that describe various aspects of the science of predictive modeling. These articles describe work to ensure that models are useful and valid on release and, perhaps more importantly, continue to be so as clinical processes and patient populations evolve over time. The upshot of the collection is to point out a new direction for informatics research and policy advocacy in the development of models for predictive analytics. Rather than focus on the mechanics of model building and validation, scientists should now be focused on how to document the model,¹ when it is likely to yield benefits,² what the model life cycle is,³ how to maintain models in a sustainable way,⁴ and even which types of health care offer the optimal predictive performance.⁵

What accounts for this change in context? In the past, bringing the resources, data, and analytical methods together to develop a predictive model was viewed as an innovative and valuable contribution to the science of informatics. However, times have changed. The presence of ubiquitous electronic health record (EHR) systems makes data for modeling commonplace. Standardized clinical data models have been developed, such as the Observational Health Data Sciences and Informatics model, to support low-effort replication of methodologies across studies. Data warehousing methods also have evolved, from the mere storage of data in applications such as Informatics for Integrating Biology and the Bedside (i2b2), to the linkage of data to analytic tools to the Health Insurance Portability and Accountability Act-compliant storage in the cloud (eg, Google Health, Azure, Amazon), lowering most barriers to model development.

In addition, methods for unsupervised machine learning (ML) have also evolved and become more user-friendly, bringing together algorithms for data compression, bootstrap dataset regeneration, and analytics into standardized packages. There is widespread agreement on basic statistical measures of performance such as the C-statistic⁶ and growing agreement on the importance of measures of calibration such as the Brier score—which is the primary metric in Davis et al's⁴ article on model maintenance—as a supplement to measures of diagnostic accuracy. EHRs and clinical

data warehouses ensure that there are sufficient data available in most circumstances for split-sample validation methods further ruggedized by the bootstrap resampling when necessary. As a result, unsupervised ML methods can often produce models with acceptable clinical accuracy (receiver-operating characteristic curves >0.7 or 0.8) in many circumstances; though, as Liu et al² suggest, threshold performance for clinical use depends on a wide range of factors. Propensity score methods are widely recognized as important in predictions that can compensate for confounding variables and there is growing confidence in the ability of neural networks to deal with the complex problems caused by missing not-at-random data. In sum, developers have a full toolbox of data systems and methods.

So, if model development for predictive analytics using existing methods of ML is no longer “informatics science,” what is the science now? This issue offers a view. First and foremost, in van Calster et al's¹ commentary, “Predictive analytics in health care: How can we know what works?” calls for transparency in models as the foundation for the new science of clinical usefulness. There is no place for black-box algorithms in our new endeavor. Research must look at the relative performance of any given method, particularly innovations, and characterize the context for the model's use. Liu et al² propose a metric for assessing the usefulness of a model in a given clinical context, called number needed to benefit. This approach borrows from the literature on evaluation of diagnostic testing to create a metric for the number of patients that need to be screened with a model to capture benefit. This metric sums up in a single number, the decision analytically derived number, much of the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework⁷ for informatics program evaluation. Lenert et al (yes, there is some relation)³ proposes the concept of a life cycle of predictive models. Namely, in addition to development, there is a maintenance phase, in which a model may need to be recalibrated, and eventual obsolescence. These authors argue that if widely applied, models might become victims of their own success, changing the rates of observed events and negating correlations producing the model. Davis et al⁴ go on to propose criteria for the assessment of the sensitivity of a model to changes in key attributes of the clinical context of application: changes in event rate, case mix,

and correlation among variables. Their data suggest that recalibration of the model, rather than redevelopment with a “new population,” may be the optimal approach to maintain many models over time. Last, in a world in which more data is always viewed as better, Simon et al⁵ looked at the types and scope of data available for prediction of suicide risk in clinical settings, comparing the scope of availability (when the model could be applied) and predictive performance. Their findings challenge the “more data” paradigm, showing minimal improvements in predictive performance with the addition of EHR-based patient-level data. Although person-level information on symptoms of depression during the visit did improve performance for prediction of future suicide attempts, the logistical requirements for the collection of this information are much greater. Programs based on administrative data and claims data may have a greater impact.

Taken together, these articles show how the science of informatics has evolved within the context of infrastructure, data, and algorithms available to it, based on the maturity of tools and methods for prediction. The availability of commercial and open source environments that can largely automate most of the aspects of predictive model development does not necessarily make modeling development more “engineering” than science, but it will challenge investigators to move beyond mere model development to find the informatics science in their endeavors. It is not enough just to build tools that predict and describe them; authors who want to publish in

Journal of the American Medical Informatics Association need to write about the science that ensures that they are predicting something that matters.

REFERENCES

1. Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works? *J Am Med Inform Assoc* 2019; 26 (12): 1651–1654.
2. Liu VX, Bates DW, Wiens J, Shah NH. The number needed to benefit: estimating the value of predictive analytics in healthcare. *J Am Med Inform Assoc* 2019; 26 (12): 1655–1659.
3. Lenert MC, Matheny ME, Walsh CG. Prognostic models will be victims of their own success, unless... *J Am Med Inform Assoc* 2019; 26 (12): 1645–1650.
4. Davis SE, Greevy RA Jr, Fonnesbeck C, Lasko TA, Walsh CG, Matheny ME. A nonparametric updating method to correct clinical prediction model drift. *J Am Med Inform Assoc* 2019; 26 (12): 1448–1457.
5. Simon GE, Shortreed SM, Johnson E, et al. What health records data are required for accurate prediction of suicidal behavior? *J Am Med Inform Assoc* 2019; 26 (12): 1458–1465.
6. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc* 2017; 24 (1): 198–208.
7. Bakken S, Ruland CM. Translating clinical informatics interventions into routine clinical care: how can the RE-AIM framework help? *J Am Med Inform Assoc* 2009; 16 (6): 889–97.