Correspondence

Explicit causal reasoning is needed to prevent prognostic models being victims of their own success

Matthew Sperrin, David Jenkins, Glen P. Martin, and Niels Peek

Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom

Corresponding Author: Matthew Sperrin, PhD, Vaughan House, Portsmouth Street, University of Manchester, Manchester M13 9PL, UK; matthew.sperrin@manchester.ac.uk

The recent perspective by Lenert et al[1] provides an accessible and informative overview of the full life cycle of prognostic models, comprising development, deployment, maintenance, and surveillance. The perspective focuses particularly on the fundamental issue that deployment of a prognostic model into clinical practice will lead to changes in decision making or interventions, and hence, changes in clinical outcomes. This has received little attention in the prognostic modeling literature but is important because this changes predictor-outcome associations, meaning that the performance of the model degrades over time; therefore, prognostic models become "victims of their own success." More seriously, a prediction from such a model is challenging to interpret, as it implicitly reflects both the risk factors and the interventions that similar patients received, in the historical data used to develop the prognostic model. The authors rightly point out that "holistically modeling the outcome and interventions(s)" and "incorporat[ing] the intervention space" are required to overcome this concern.[1] However, the proposed solution of directly modeling interventions, or their surrogates, is not sufficient. An explicit causal inference framework is required.

When the intended use of a prognostic model is to support decisions concerning intervention(s), the counterfactual causal framework provides a natural and powerful way to ensure that predictions issued by the prognostic model are useful, interpretable, and less vulnerable to degradation over time. The framework allows predictions to be used to answer "what if" questions; for an introduction, see Hernan and Robbins.[2] However, appropriate modeling of these counterfactual scenarios is far more challenging than pure prediction, particularly in the presence of time-dependent confounding. Here, standard regression modeling becomes inadequate and specialist techniques are required.[2] In the scenarios carefully articulated by Lenert et al, in which risk models are used to alert to a high-risk situation and thereby inform inter-

vention, one should primarily be interested in the counterfactual "treatment-naïve" prediction: in other words, "what is the risk of outcome for this individual if we do not intervene?" Failure to explicitly model this treatment-naïve prediction will lead to high-risk patients being classified inappropriately as low risk, as their prediction is reflective of interventions made to lower the risk of similar patients in the past.[3] This situation becomes more pronounced when a successful model is updated, as interventions made based on the predictions from the model are hoped to change the risk. Recently, we illustrated how to calculate treatment-naïve risk in the presence of "treatment drop-in," a scenario in which patients begin taking treatments after the time a prediction is made but before the outcome.[4]

With treatment-naïve risk as a baseline, one can move to evaluating predictions under a range of different interventions; the counterfactual causal framework allows a model to be interrogated with a series of "what if" questions. Comparison of the outcome predictions or distributions under different scenarios can then naturally provide information to support intervention decisions.

Alongside this counterfactual framework, we agree with Lenert et al that "robust performance surveillance of models in clinical use" is required postdeployment as part of prognostic model maintenance and model surveillance. However, doing this through so-called static updating, in which previous iterations of a risk model are refined according to new datasets observed in batches, still requires timely identification of performance drift. This often leads to an identification-action latency period, in which noticing and acting on a deterioration in a model's performance occurs much later in time than should be acceptable in clinical practice. This is amplified by a lower frequency of updating but could be mitigated through continuous surveillance and maintenance of the prognostic models. So-

called dynamic modeling is an emerging area of research[5] that enables the continuous incorporation of surveillance and refinement directly into the modeling processes and could prevent prognostic models being "victims of their own success" if combined appropriately with counterfactual frameworks.

While counterfactual prediction is only beginning to be applied in prognostic model development, it is a technique that will allow many of the issues eloquently described by Lenert and colleagues to be mitigated. Moreover, it provides predictions that are arguably closer to what a decision maker needs, and likely to be more robust over time.

CONFLICT OF INTEREST STATEMENT

None declared.

1. Lenert MC, Matheny ME, Walsh CG. Prognostic models will be victims of their own success, unless. *J Am Med Informatics Assoc* 2019; 26 (12): 1645–1650.
2. Hernan MA, Robins JM. *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC; 2019.
3. Cheong-See F, Allotey J, Marlin N. Prediction models in obstetrics: understanding the treatment paradox and potential solutions to the threat it poses. *BJOG* 2016; 123 (7): 1060–4. doi: 10.1111/1471-0528.13859/pdf.
4. Sperrin M, Martin GP, Pate A, *et al*. Using marginal structural models to adjust for treatment drop-in when developing clinical prediction models. *Stat Med* 2018; 37 (28): 4142–54.
5. Jenkins DA, Sperrin M, Martin GP, *et al*. Dynamic models to predict health outcomes: current status and methodological challenges. *Diagn Progn Res* 2018; 2: 23.