Praise to robust prediction modeling in large datasets

Steyerberg, E.W.; Bonneville, E.F.

JACC: CARDIOONCOLOGY VOL. 2, NO. 3, 2020

EDITORIAL COMMENT

Praise to Robust Prediction Modeling in Large Datasets*

Ewout W. Steyerberg, PhD, Edouard F. Bonneville, MSc

It is known that patients with cardiovascular disease have a higher risk of cancer compared to the general population; this is likely attributable to the presence of several common risk factors such as smoking. Adequate estimation of absolute individual cancer risk requires the combination of such risk factors in clinical prediction models. In this issue of *JACC: CardioOncology*, a team of international researchers should be applauded for presenting a promising risk calculator for cancer in individual patients with cardiovascular disease (1).

As risk prediction is receiving increasing attention in the current Big Data era, a fundamental question is: Are the predictions valid? Validity of a prediction model may be limited by poor reproducibility (internal validity) and poor transportability (external validity) (2). Prediction models are often developed in relatively small datasets with data hungry methods, that is, methods that only work in very large datasets (3). We see this unfortunate practice in the many prediction models that appear for coronavirus disease 2019 diagnosis and prognosis (4). This current study stands out positively with sensible modeling approaches in carefully collected datasets and large sample sizes with complete follow-up. Impressive numbers are noted in the development dataset (UCC-SMART [Utrecht Cardiovascular Cohort Second Manifestation of Arterial Diseases]: 1,143 incident cancers among 7,280 patients with median follow-up of 8.1 years) and the validation dataset (CANTOS [Canakinumab Anti-inflammatory Thrombosis Outcomes Study]: 509 incident cancers among 9,322 patients with median follow-up of 3.8 years). For specific cancers, numbers are smaller (CANTOS: 123 lung cancers, 72 colorectal cancers). As such, detailed modeling can be difficult, even in overall large datasets. Indeed, some shrinkage of model coefficients was needed, especially for the colorectal cancer prediction model (1). Internal validity of the calculator may be improved with larger numbers of patients, particularly for the lung cancer and colorectal cancer prediction models. Longer follow-up is also important to corroborate the current 10-year and lifetime risk estimates, which are to some extent extrapolations.

An additional threat to validity is the limited transportability due to various sources of heterogeneity between populations. The current models were developed within a broader range of clinically manifest cardiovascular disease, as compared to the validation setting (UCC-SMART: coronary artery disease, cerebrovascular disease, or peripheral artery disease; CANTOS: myocardial infarction and elevated C-reactive protein) (1). The geography also differed (the Netherlands vs. the United States). Nevertheless, adequate performance was found in this rather strong test of validity. Given this, we might be tempted to claim that the predictions from the calculator apply to all patients in the Western world. However, such a strong claim would require further validation. To further support generalizability, additional clinical settings need to be studied, as some form of local updating is often required for prediction models (5). Preferably, extensions are sought to non-Western hospitals, with large numbers of patients, high-quality data, and long-term follow-up. Such extensions may reveal heterogeneity in performance,

*Editorials published in *JACC: CardioOncology* reflect the views of the authors and do not necessarily represent the views of *JACC: CardioOncology* or the American College of Cardiology.

From the Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands. Both authors have reported that they have no relationships relevant to the contents of this paper to disclose.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the *JACC: CardioOncology* author instructions page.

ISSN 2666-0873 https://doi.org/10.1016/j.jaccao.2020.08.004

motivating locally adapted versions of the calculator, that is, model updating (6).

Many methodological issues in the presented study are dealt with in an exemplary way. The study may serve well for teaching purposes to those interested in predictive analytics. The high-quality data are analyzed with regression techniques that take into account competing risks. Absolute risks of cancer are thus adequately estimated (with cumulative incidence functions) in a context where many die of cardiovascular disease and other causes. Missing data are minimal and statistically imputed. Predictors were pre-selected based on previous literature with easy clinical applicability. Age was used as the timescale, which is natural and allows for 10-year and lifetime prediction horizons. A simple model only considered sex and smoking as additional predictors, whereas a full model added height and weight, C-reactive protein, diabetes mellitus, alcohol use, and antiplatelet use. This robust approach in selecting predictors is in contrast to modern machine-learning approaches, which may explore a wider range of potential predictors with highly flexible functions. Interestingly, with external validation, the more complex model ("full model") versus the simple model did not demonstrate improved discrimination (1). Two lessons might be drawn. First, the key set of predictors in prediction of cancer may be age, sex, and smoking status. This is in line with other prediction studies where a simple model may generalize better than a more complex model (7). Second, complex models based on machine learning would not be expected to provide improved performance or better prognostication in the current study, consistent with other large-scale external validation studies (8,9).

From a clinical perspective, a key concern should be whether the predictions from the risk calculator are well calibrated (5). For example, the 70-year-old male in the example calculation sheet has a predicted 1.1% risk of lung cancer, while the risk may in fact be 0.5%, or 2%. This type of miscalibration has been common in earlier evaluations of lung cancer prediction models (10). Indeed, the investigators state that "calibration is a more clinically relevant for risk prediction accuracy than the C-statistic." We should be specific on what "clinically relevant" refers to in this risk prediction context. Predictions may serve to inform patients about their individual risk. The C-statistic is a commonly used measure to indicate discrimination, or how well we can separate low-risk patients from high-risk patients. As the investigators state, this may be less of a concern to individual patients. Beyond informing patients, clinical relevance may refer to decision support, such as selecting patients for lung cancer screening. Prioritizing calibration over discrimination may be an oversimplification because both calibration and discrimination properties of the calculator are relevant to quantify its clinical usefulness. Better discrimination is needed for better decision support than possible with the current models, where the lung cancer prediction model was best with a C-statistic of only 0.74 (1). An increasingly popular summary measure for clinical usefulness is "net benefit." Net benefit is a classic measure, proposed in 1884 to quantify the quality of predictions (11). It was recently rediscovered and presented through a "decision curve" (12,13), which is unfortunately missing from the current report. Future work should consider the clinical decision-making perspective more fully, with net benefit in a decision curve as a step towards a more comprehensive cost-effectiveness analysis.

In sum, the presented risk calculator is very promising given its high-quality data sources, large numbers, and sensible methodology. Some caution is needed with regard to clinical application, and we await further model validation and the potential need for local updating. Further research is also needed to identify more robust predictors beyond age, sex, and smoking status, which may provide a solid knowledge base when counseling individual patients on their cancer risks and emphasizing healthy lifestyle changes.

ADDRESS FOR CORRESPONDENCE: Dr. Ewout W. Steyerberg, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, the Netherlands. E-mail: e.w.steyerberg@lumc.nl. Twitter: @ESteyerberg.

REFERENCES

1. Klooster van 't. Prediction of Lifetime and 10-Year Risk of Cancer in Individual Patients with Established Cardiovascular Disease. J Am Coll Cardiol CardioOnc 2020;2:400–10.

2. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York, NY: Springer, 2019.

3. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014;14:137.

4. Wynants L, Van Calster B, Collins GS, et al. prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328.

5. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:230.

6. Steyerberg EW, Nieboer D, Debray TPA, van Houwelingen HC. Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: an overview and illustration. Stat Med 2019;38:4290–309.

7. Dijkland SA, Foks KA, Polinder S, et al. Prognosis in moderate and severe traumatic brain injury: a systematic review of contemporary models and validation studies. J Neurotrauma 2020;37:1–13.

8. van der Ploeg T, Nieboer D, Steyerberg EW. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. J Clin Epidemiol 2016;78:83–9.

9. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12–22.

10. Ten Haaf K, Jeon J, Tammemagi MC, et al. Risk prediction models for selection of lung cancer screening candidates: a retrospective validation study. PLoS Med 2017;14:e1002277.

11. Peirce CS. The numerical measure of the success of predictions. Science 1884;4:453–4.

12. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.

13. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ 2016;352:i6.

KEY WORDS big data, prediction, regression analysis, validation