Editorial

Mortality Models for Heart Failure: Should they be De Novo or Recalibrated?

JUAREZ R. BRAGA, MD,[1] AND DOUGLAS S. LEE, MD, PhD[1,2,3,4]

Toronto, Canada

Prognostic modelling is the "use of statistical models to determine important predictors for outcomes of interest and to estimate the risk of future outcomes for patients according to different combinations of predictors".[1] The most important use of prognostic models is to guide clinical decision-making in daily practice and select the best individualized risk-based treatment strategy. Prognostic models have received considerable attention in heart failure especially for patients with an acute decompensation. More than 100 prognostic models have been developed for use in this context.[2]

Prognostic models should be validated before clinical use. The first level of validation is internal validation using, for example, split sample or cross-validation approach, where the model is tested on a different subset of the sample from which it was derived. Overoptimism of the model can be assessed using internal validation. The next level of validation can be achieved by testing the prognostic model in an external patient sample. Broader validation entails prospective evaluation and comparison with physician-estimated risk. The broadest level of validation requires an implementation study design or a randomized trial examining the impact of the model on outcomes. In the setting of acute heart failure, the ACUTE study is an example of prospective comparison against physician-estimated risk and the ongoing COACH trial is an example of a randomized implementation trial.[3,4] These studies are evaluating the EHMRG risk model for risk prediction in acute heart failure.[5]

The ideal prognostic model should be generalizable to other patient groups. However, when a prognostic model is transported to individuals who were not represented in the derivation study, the performance may be reduced. There are many reasons why a prognostic model might underperform in another setting including differences across healthcare systems, different definitions for the predictor variables and outcome measures, and changes in any of these aforementioned factors over time.[6] Another important reason for differential performance of prognostic models is differences in patient case mix between study cohorts, including differences in ethnicity.

Race and ethnicity have been of considerable interest in cardiology. Most of the literature examining cardiovascular disease and heart failure has been conducted among white individuals in Western countries. However, ethnicity and race contribute importantly to variations in cardiovascular health. Recent evidence has suggested that there is substantial variability in cardiovascular risk factors and incidence of cardiovascular disease in Asians compared with white populations.[7] Accordingly, the incidence of heart failure also differs across ethnic groups of Asian descent;[8] Asians with a diagnosis of heart failure receive less evidence-based therapies compared with European populations,[9] and prognostic models have been found to have different accuracy when applied in Asian populations.[10] Although the term "Asian" refers to a diversity of ethnic subgroups, there is ground to hypothesize that in general Asians and white populations differ in important characteristics that may lead to differential model performance.[7]

As emphasized by the American guidelines, a heart failure prognostic model can be a useful tool to estimate risk of mortality in patients with heart failure as long as they are validated in an external population.[11] A model that underperforms in a new population can have potentially harmful consequences. For instance, model miscalibration can lead to predicted risks that are systematically different from observed risks, leading to errors in the decision-making process.

From the [1]Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, Ontario, Canada; [2]ICES, Toronto, Ontario, Canada; [3]Peter Munk Cardiac Centre, University Health Network, Toronto, Ontario, Canada and [4]Ted Rogers Centre for Heart Research, Toronto, Canada.

Manuscript received May 10, 2019; revised manuscript accepted May 10, 2019.

Reprint requests: Douglas S. Lee, MD, PhD, ICES and University Health Network, Division of Cardiology, 2075 Bayview Ave, G-wing, Toronto, Ontario, Canada, M4N 3M5. E-mail: dlee@ices.on.ca

Funding: This editorial was supported by a Foundation Grant from the Canadian Institutes of Health Research (FDN 148446) and the Ted Rogers Centre for Heart Research. Dr. Lee is supported by a mid-career investigator award from the Heart and Stroke Foundation and the Ted Rogers Chair in Heart Function Outcomes, a joint Hospital-University Chair of the University Health Network and the University of Toronto.

1071-9164/$ - see front matter

https://doi.org/10.1016/j.cardfail.2019.05.006

Despite their importance, validation studies are scarce. Almost 60% of all prognostic models developed for patients with heart failure have not been externally validated.[2] Based on the notion that prognostic models will underperform in a new population, when confronted with the need to estimate prognosis, researchers may aim to develop a new model and reject previous ones. However, this approach has disadvantages: it wastes prior information, risks overfitting leading to overoptimistic measures of model performance, and ultimately leads to a plethora of prognostic models, which makes it difficult for health professionals to decide which one to use.[12]

In this issue of *The Journal of Cardiac Failure*, Yap et al[13] report findings from a study that externally validated the OPTIMIZE-HF (Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure) score in Singapore. Instead of developing a new prognostic model, the authors validated an existing model to predict mortality in individuals with heart failure. The OPTIMIZE-HF was a registry and performance improvement initiative for patients hospitalized with heart failure in 259 hospitals in the United States. Based on an analysis of more than 48,000 patients between 2003 and 2004, a prognostic model was designed using the 7 most important predictors of in-hospital mortality selected from a multivariable logistic regression model.[14]

Although the OPTIMIZE-HF had been externally validated in two different cohorts,[14] there were marked differences between the ethnic population included in the OPTIMIZE-HF registry and the population in Singapore. Singapore has a high population life expectancy, one of the highest per-capita gross domestic products in the world, and a multiethnic composition, comprised of Chinese, Malay, and Indians. Considering the contrast between the populations included in the original study and in Singapore, it was unclear whether the OPTIMIZE-HF could be used without loss in prognostic performance.

In their analysis of 15,219 participants included in the Singapore Cardiac Databank Heart Failure hospitalized for heart failure between 2008 and 2013, Yap et al[13] found that the OPTIMIZE-HF had a good performance in the new population. In addition to externally validating the model, Yap et al[13] went a step further and recalibrated the original model "tuning in" the model to account for local and contemporary circumstances.[15] This recalibration by the authors entailed refitting the original model and re-estimating the regression coefficients for the same variables that were included in the OPTIMIZE-HF model.[12] Surprisingly, Yap et al[13] found that the recalibrated OPTIMIZE-HF model did not show improved discrimination, which was numerically similar to the original model.

It is not clear why the recalibrated model was not associated with significant gains in performance in relation to the original model, although a key problem of model revision is exactly overoptimistic prediction for the recalibrated model.[15] There are some potential factors that might explain these findings. First, the OPTIMIZE-HF registry included individuals admitted to hospital due to heart failure, but also those who had an underlying diagnosis of heart failure but who were hospitalized for other reasons. In contrast, the population included in the Singapore registry included only patients who had heart failure as the primary reason for hospitalization. Second, the cohort included in the Singapore Cardiac Databank Heart Failure Registry experienced an in-hospital mortality rate of 1.9%, which was approximately half of the mortality rate observed in the OPTIMIZE-HF registry. It is possible that the low number of events may have affected the recalibration of the model. Third, although the predictors included in the original and the recalibrated models could be the same, slightly different definitions or ranges for the same covariate could alter performance. For example, in the OPTIMIZE-HF model, variables such as serum creatinine and systolic blood pressure had explicit ranges defined, and the back-calculated regression coefficient for sodium concentration differed for values less than or above 140 mEq/L. If the recalibrated model does not account for these nuances, performance could be affected.

Despite these questions, the study by Yap et al[13] has two important messages. First, the OPTIMIZE-HF prognostic model can be used in an Asian population without loss of prognostic performance. Second, the use of previously developed models should be encouraged. If derivation of a new prognostic model for the same endpoint in the same context is planned, it may not be necessary. Alternatively, the new model should be compared with existing models to demonstrate the incremental value of the new model. In summary, we agree with the statement by Ahmed et al[16] that "what is needed is a consensus among health professionals towards a single well developed and validated prognostic model, rather than a number of competing non-validated models for the same clinical question".

References

1. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009;338:b375.
2. Ouwerkerk W, Voors AA, Zwinderman AH. Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. JACC Heart Fail 2014;2:429–36.
3. Lee DS, Lee JS, Schull MJ, et al. Prospective validation of the emergency heart failure mortality risk grade for acute heart failure. Circulation 2019;139:1146–56.
4. Dunbar-Yaffe R, Stitt A, Lee JJ, Mohamed S, Lee DS. Assessing risk and preventing 30-Day readmissions in decompensated heart failure: opportunity to intervene? Curr Heart Fail Rep 2015;12:309–17.
5. Lee DS, Stitt A, Austin PC, et al. Prediction of heart failure mortality in emergent care: a cohort study. Ann Intern Med 2012;156:767–75.
6. Riley RD, Ensor J, Snell KIE, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2016;353:i3140.

7. de Souza RJ, Anand SS. Cardiovascular disease in Asian Americans: unmasking heterogeneity. J Am Coll Cardiol 2014;64:2495–7.
8. Giuseppe GD, Chu A, Tu JV, Shanmugasegaram S, Liu P, Lee DS. Incidence of heart failure among immigrants to Ontario, Canada: a CANHEART Immigrant Study [e-pub ahead of print]. J Card Fail doi:10.1016/j.cardfail.2019.03.006. Accessed April 14, 2019.
9. Teng T-HK, Tromp J, Tay WT, et al. Prescribing patterns of evidence-based heart failure pharmacotherapy and outcomes in the ASIAN-HF registry: a cohort study. Lancet Glob Health 2018;6:e1008–18.
10. Barzi F, Patel A, Gu D, et al. Cardiovascular risk prediction tools for populations in Asia. J Epidemiol Community Health 2007;61:115–21.
11. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;62:e147–239.
12. Su TL, Jaki T, Hickey GL, Buchan I, Sperrin M. A review of statistical updating methods for clinical prediction models. Stat Methods Med Res 2018;27:185–97.
13. Yap J, Lim FY, Chia SY, et al. Prediction of survival in Asian patients hospitalized with heart failure: validation of the OPTIMIZE-HF risk score [e-pub ahead of print]. J Card Fail doi:10.1016/j.cardfail.2019.02.016. Accessed March 27, 2019.
14. Abraham WT, Fonarow GC, Albert NM, et al. Predictors of in-hospital mortality in patients hospitalized for heart failure: insights from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). J Am Coll Cardiol 2008;52:347–56.
15. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med 2004;23:2567–86.
16. Ahmed I, Debray TP, Moons KG, Riley RD. Developing and validating risk prediction models in an individual participant data meta-analysis. BMC Med Res Methodol 2014;14:3.