CORR Insights

CORR Insights®: Can A Multivariate Model for Survival Estimation in Skeletal Metastases (PATHFx) Be Externally Validated Using Japanese Patients?

Raphaël Porcher PhD

Where Are We Now?

Prediction models play an important role in informing patients about the future course of their disease and in aiding medical decision-making. For example, in a primary care setting, the Framingham

cardiovascular risk score is used to identify patients who meet indications to receive cholesterol-lowering and antihypertensive drugs [6, 7]. In a surgical-oncology setting, patients undergoing resection for a soft-tissue sarcoma who have a low risk of local recurrence could be spared the short- and long-term complications of adjuvant radiotherapy, whereas those considered at high risk are commonly offered this treatment [10].

In their study, Ogura and colleagues performed external validation in a Japanese population of the PATHFx model (www.pathfx.org) that predicts 1, 3, 6 and 12-month postsurgical survival of patients with skeletal metastases. The PATHFx model was developed and validated for Western countries patients [2–4, 9], and the validation in an Asian population therefore represents a step forward in the evaluation of the applicability of model worldwide.

In this Japanese population, the PATHFx model showed a fair-to-good discrimination based on receiver operating characteristic curves, with the c index (the area under the curve, [AUC]) ranging between 0.75 and 0.83, comparable to previous development and validation studies (AUCs between 0.76 and 0.85). Interestingly, this occurred despite substantial differences in the characteristics of patients as compared to the original development sample, including a markedly better survival from 3 months and beyond.

This CORR Insights® is a commentary on the article “Can A Multivariate Model for Survival Estimation in Skeletal Metastases (PATHFx) Be Externally Validated Using Japanese Patients?” by Ogura and colleagues available at: DOI: [10.1007/s11999-017-5389-3](https://doi.org/10.1007/s11999-017-5389-3).

The author certifies that neither he, nor any members of his immediate family, have any commercial associations (such as consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article. All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request. The opinions expressed are those of the writers, and do not reflect the opinion or policy of CORR® or The Association of Bone and Joint Surgeons®.

This CORR Insights® comment refers to the article available at DOI: [10.1007/s11999-017-5389-3](https://doi.org/10.1007/s11999-017-5389-3).

R. Porcher PhD (✉)
Centre d'Epidémiologie Clinique,
Hôpital Hôtel-Dieu, 1 Parvis Notre-
Dame - Place Jean-Paul II, 75004 Paris,
France
e-mail: raphael.porcher@aphp.fr

R. Porcher PhD
Team METHODS, Centre de Recherche
en Epidémiologie et Statistiques
(CRESS), INSERM U1153, Paris, France

R. Porcher PhD
Paris Descartes University, Paris, France

CORR Insights

Where Do We Need To Go?

Risk-prediction models should be validated, and their usefulness in improving medical decision making, and ultimately patient outcome [7], should be evaluated.

Sample sizes in validation studies have scarcely been studied. Recent work suggests that substantial sample sizes (at least 100 events [11]) are needed, which may explain the lack of precision of the estimates of the c index at earlier timepoints in the current study. Additional research should be dedicated to missing data in validation studies. Risk prediction models will likely yield poor results if the models are overfitted or missing important predictors, or if there are differences between the original study settings and the new data, such as patient characteristics, variables assessment, standard of care, or healthcare system.

The PATHFx model was developed using Bayesian belief network modeling, which has the advantage of allowing model predictions to be obtained even when one of the predictors are missing, such as the surgeon's estimate of survival. Classical regression models are only applicable if all predictors are available, except if additional work was

undertaken to propose risk estimates in the presence of missing data. Whatever the approach, the risk predictions would be expected to be more accurate when all predictors are available.

Since prediction models should be transferable to other populations, the assessment of their performance in other settings—external validation—is crucial, yet too-rarely undertaken [5]. Validation of a prediction model should examine both calibration (how predicted probabilities of event match the observed probabilities of event) and discrimination (how well the model separates between patients who will experience the event of interest and the others) [1].

The authors of the current study did not present calibration, thus it remains unknown whether the model predictions were close to the observed survival probabilities or if the model was miscalibrated to some extent for this population. Indeed, a prediction model can retain good discrimination even when being miscalibrated. If the model was well-calibrated in the Japanese population, then it is plausible that the difference in overall survival would be due to the differences in case mix. Otherwise, updating the model for the Japanese population through recalibration or a more-complete model revision [8] should be considered.

How Do We Get There?

Determining whether a prediction model really improves clinical care is better shown in an impact study [8]. Impact studies quantify the effects of prognostic models on physician behavior, clinical results, or cost effectiveness when compared to usual care (care delivered without guidance from the model). They also require different experimental approaches than do studies that validate prediction models. In this setting, an impact study might compare the decision to perform surgery and possibly patient survival in a control group where the model is not used with care delivered to an experimental group where the surgeons' decisions would be informed by the PATHFx model (an assistive approach) or the PATHFx prediction with a suggested decision (a decisive approach). Individual patients' randomization is usually not the preferred approach for impact studies because of learning effects, but cluster randomized studies (surgeons would be randomized to either group) or before-and-after studies could be used. In practice, impact studies are rarely performed, but they still remain the best way to demonstrate the usefulness of a prediction model, and should therefore be considered.

CORR Insights

References

1. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
2. Forsberg JA, Eberhardt J, Boland PJ, Wedin R, Healey JH. Estimating survival in patients with operable skeletal metastases: An application of a Bayesian belief network. *PLoS One*. 2011;6:e19956.
3. Forsberg JA, Wedin R, Bauer HC, Hansen BH, Laitinen M, Trovik CS, Keller JØ, Boland PJ, Healey JH. External validation of the Bayesian Estimated Tools for Survival (BETS) models in patients with surgically treated skeletal metastases. *BMC Cancer*. 2012;12:493.
4. Forsberg JA, Wedin R, Boland PJ, Healey JH. Can we estimate short- and intermediate-term survival in patients undergoing surgery for metastatic bone disease? *Clin Orthop Relat Res*. 2017;475:1252–1261.
5. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, Tibshirani R. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014;383:166–175.
6. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham study. *Am J Cardiol*. 1976;38:46–51.
7. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: What, why, and how? *BMJ*. 2009;338:b375.
8. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.
9. Piccioli A, Spinelli MS, Forsberg JA, Wedin R, Healey JH, Ippolito V, Daolio PA, Ruggieri P, Maccauro G, Gasbarrini A, Biagini R, Piana R, Fazioli F, Luzzati A, Di Martino A, Nicolosi F, Camnasio F, Rosa MA, Campanacci DA, Denaro V, Capanna R. How do we estimate survival? External validation of a tool for survival estimation in patients with metastatic bone disease-decision analysis and comparison of three international patient populations. *BMC Cancer*. 2015;15:424.
10. Sherman KL, Wayne JD, Chung J, Agulnik M, Attar S, Hayes JP, Laskin WB, Peabody TD, Bontrem DJ, Pollock RE, Bilimoria KY. Assessment of multimodality therapy use for extremity sarcoma in the United States. *J Surg Oncol*. 2014;109:395–404.
11. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol*. 2005;58:475–483.