TITLE: Small datasets to develop and validate prognostic models are problematic

Gary S. Collins, *associate professor*
Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Windmill Road, Oxford OX3 7LD, United Kingdom.
Email: gary.collins@csm.ox.ac.uk

Yannick Le Manach, *assistant professor*
Departments of Anesthesia & Clinical Epidemiology and Biostatistics, Michael DeGroote School of Medicine, Faculty of Health Sciences, McMaster University and the Perioperative Research Group, Population Health Research Institute, Hamilton, Canada

Word count: 819 (max 1000 words)

The authors report no conflict of interest.

Letter Re: Personalised medicine: Development and external validation of a prognostic model for metastatic melanoma patients treated with ipilimumab

We read with great interest, the recent paper by Valpione and colleagues describing the development of a nomogram to predict the 6 month, 12 month and 24 monthsurvival probability for patients with metastatic melanoma treated with ipilimumab(1). Whilst the authors are to be applauded by using advanced statistical methods to develop their prognostic model, there are some aspects surrounding the design and reporting of the study that are of concern.

Sample size considerations for developing a prognostic are typicallycenteredon the concept of events per variable, or more precisely events per estimated parameter; with a value of 10 often cited to reduce the problem of overfitting. The development cohort of the Valpione study comprised 113 patients of which 31 were alive at 11.5 months. The number of predictors examined is not entirely clear but between 30 and 40 predictors (including biomarkers) appear to have been examined from the Tables presented in the paper.To examine 30 predictors requires at least 300 outcome events; a value clearly not achieved in this study. Whilst the authors carried out bootstrapping to quantify overfitting by calculating a shrinkage factor, it is not a replacement for having an adequate sample size in the first place.The bootstrapping procedure should also include all steps of the model building, including variable selection, otherwise, the bootstrapping process is invalid(2).A stronger test of the model is carried out by an external validation on separate data(3). However, in the Valpione study, the external validation cohorts are also small (n=69 and n=34), that fall considerably short of recommended sample sizes for validation studies (4). Whilst we appreciate that collecting adequately sized data for external validation can be prohibitive, shortcomings in external validation data can be offset by conducting high-quality internal validation, leavingexternal validation to be carried out later (5).External validation studies are important to characterize the performance of a prognostic model on a given data set with a particular case-mix (distribution of patient characteristics). Such studies deserve as much attention (if not more) as studies

developing new prognostic models, using sufficiently large data and appropriate statistical methods(6).

The authors used multiple imputation to replace missing data and therefore retained all patients in the cohort but it is unclear what data was missing and how much was missing.Describing what data is missing and how much is missing can provide important information on the data quality. A full description of the missing data, and the multiple imputation model, and whether data were also imputed in the validation data sets is widely recommended(7, 8). A description of missing data for each variable can be reported in a table of patient characteristics for each of the 3 cohorts separately. Describing patient characteristics for each data set (which is missing in this study) can provide information on the case-mix, and therefore whether the validation is assessing reproducibility (similar case-mix) or transportability (different case-mix) of the prognostic model.

Assessing model performance is key, with discrimination and calibration the two main aspects to evaluated(9, 10). Whilst there is nothing wrong with reporting Somers's Dxy for assessing discrimination, it is more customary to calculate the c-index that will be more familiar to readers. The other important aspect of model performance, calibration was assessed graphically in using "Receiver Operating Characteristic curves of the calibration"[sic].Graphically assessing calibration is the preferred approach, but it is important the both the *x*-axis and *y*-axis are on the same scale which should result in a 'square' plot, not a rectangular shaped plot as presented by Valpione.Squashing one of the axes (the *x*-axis), distorts the plot and gives the false impression that model has good calibration.

Finally, when developing a new prognostic model, it is important that the model is presented in full so thatindependent investigators can validate the model in their own data. Presenting a nomogram, as done by Valpione may increase model uptake, but it is not a replacement for fully describing the prognostic model; namely all the regression coefficients and the baseline survival at key time points, as recommended in the recent TRIPOD Statement for reporting prediction models(10). If this information is not presented, then other investigators are unable to accurately calculate individual survival probabilities. Regression coefficients for baseline neutrophil count and lactic dehydrogenase in addition to the baseline survival at 6, 12 and 24 months are unfortunately missing from the paper.

We recommend the authors and indeed other investigators developing prognostic modelsto consult the TRIPOD Statement (www.tripod-statement.org) for key information to report when describing its development and validation, so that readers have the minimal information required to judge the quality of the study and therefore whether to use the model(10). The accompanying TRIPOD Explanation and Elaboration paper highlights the rationale of the importance of transparent reporting, but also discusses various methodological considerations that investigators should consider when developing and validating a prognostic model.

ACKNOWLEDGMENTS
No funding was received.

REFERENCES

1. Valpione S, Martinoli C, Fava P, Mocellin S, Campana LG, Quaglino P, et al. Personalised medicine: Development and external validation of a prognostic model for metastatic melanoma patients treated with ipilimumab. Eur J Cancer 2015;51:2086-2094.

2. Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic regression and survival analysis. New York: Springer; 2001.

3. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19(4):453-73.

4. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 2005;58(5):475-483.

5. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. Journal of Clinical Epidemiology 2015;Apr 18 [Epub ahead of print].

6. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Meth 2014;14:40.

7. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393.

8. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med 2015;162(1):W1-W73.

9. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21(1):128-138.

10. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Ann Intern Med 2015;162(1):55-63.