BJR

Received:
30 July 2016

Accepted:
6 September 2016

http://dx.doi.org/10.1259/bjr.20160655

Cite this article as:
Collins GS. Statistical concerns in the development of a prediction model. *Br J Radiol* 2016; **89**: 20160655.

LETTER TO THE EDITOR

Statistical concerns in the development of a prediction model

GARY S COLLINS, PhD

Centre for Statistics in Medicine, Botnar Research Centre, University of Oxford, Oxford, UK

Address correspondence to: Professor Gary S Collins
E-mail: *gary.collins@csm.ox.ac.uk*

To the Editor,

I read with great interest the recent article by Wee et al[1] describing the development of a nomogram to predict the 6-month survival probability for patients with hepatocellular carcinoma treated with radiotherapy for lymph node metastasis. However, in addition to the small sample size, there are numerous aspects surrounding the development, evaluation and reporting of the prediction model that are of concern.

The sample size in the study by Wee et al was small, with only 105 patients used to develop the prediction model. However, it is not the total number of patients in the study but the number of events (*i.e.* deaths) that is the effective sample size in prediction model studies. The number of deaths by 6 months were not reported but clearly fewer than 105. The sample size rule-of-thumb for developing a new prediction model is that a minimum of 10 events per variable are required to minimize the overfitting. Wee et al examined 15 predictors, implying that 150 (15×10) events (deaths) are required, clearly more than the authors have in their data. Reducing the number of predictors examined or increasing the sample size (number of events) would be required to meet this target of 10 events per variable.

Sometimes, collecting more data is problematic, and investigators are faced with using what they have available to them. In these instances, as in the study by Wee et al, statistical methods, such as bootstrapping, are available to quantify overfitting (bias) by adjusting the performance to obtain a bias-corrected estimate of model performance and to adjust the model. Unfortunately, whilst Wee et al used bootstrapping, it is unclear as the Methods section is uninformatively too brief, as to how the authors implemented the bootstrapping. To obtain a bias-corrected estimate of model performance (such as the *c*-index), all variable selection procedures (including univariate analyses, dichotomizing continuous predictors)[2] need to be included in the bootstrapping, which I am not convinced was performed in this study. Omitting the model-building steps and including only the final model in the bootstrapping will bias (overinflate) the performance of the model and thus give a false picture of the predictive accuracy of the model. Correctly carrying out the internal validation component of the model development is crucial, particularly in the absence of a separate data set to evaluate the model.

The next component of model performance examined was model calibration—an important aspect of predictive accuracy and recommended in the recent Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) reporting guideline for prediction model studies.[3,4] However, the calibration plot, as presented in the Wee et al study, is somewhat suboptimal. The authors presented the traditional and commonly seen calibration plot, of predictions against observed outcomes by thirds of predicted risk (*i.e.* 3 equal-sized groups), although usually this is performed by tenths of predicted risk (*i.e.* 10 equal-sized groups)—presumably the small sample size precluded them from doing this by larger groups. However, it is widely recommended (TRIPOD Statement)[3,4] that to improve the calibration plot, the authors could have overlaid the plot with a smoothed regression line using flexible adaptive hazard.[5] This enables readers to judge agreement across the spectrum of predictions (*i.e.* for every 100 patients given a prediction of x%, the observed number of patients with the outcome is close to x).

My final and arguably most important point relates to the presentation of the model—the authors decided to present the model in the form of a nomogram. A nomogram is a graphical representation of the underlying Cox regression model. Unfortunately, for other investigators to evaluate their nomogram, the prediction model needs to be reported in full, namely all the regression coefficients as well as the baseline survival at 6 months; the authors failed to report the baseline survival. Without the full model,

validation of the model by other investigators in their own data is not possible. Thus, without this crucial information, the study has limited value.

Adhering to the TRIPOD statement would improve the transparency and completeness of the published article describing the prediction model.[3,4]

REFERENCES

1. Wee CW, Kim K, Chie EK, Yu SJ, Kim YJ, Yoon JH. Prognostic stratification and nomogram for survival prediction in hepatocellular carcinoma patients treated with radiotherapy for lymph node metastasis. *Br J Radiol* 2016; **89**: 20160383. doi: http://dx.doi.org/10.1259/bjr.20160383
2. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med* 2016; **35**: 4124–35. doi: http://dx.doi.org/10.1002/sim.6986
3. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; **162**: 55–63. doi: http://dx.doi.org/10.7326/M14-0697
4. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, 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**: W1–73. doi: http://dx.doi.org/10.7326/M14-0698
5. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016; **35**: 214–26. doi: http://dx.doi.org/10.1002/sim.6787