The HKU Scholars Hub   The University of Hong Kong   香港大學學術庫

Title	Reply to Nomograms need to be presented in full
Author(s)	Choi, CW; Xu, W; Lee, WMA
Citation	Cancer, 2017, v. 123 n. 1, p. 178-179
Issued Date	2017
URL	http://hdl.handle.net/10722/244542
Rights	

Response to "Nomograms need to be presented in full"

Horace C.W. Choi, PhD [1]; Wei Xu, PhD [2]; Anne W.M. Lee, MD * [1,3]

[1] Department of Clinical Oncology, the University of Hong Kong, Hong Kong, China

[2] Department of Biostatistics, Princess Margaret Cancer Centre, Canada

[3] Department of Clinical Oncology, the University of Hong Kong-Shenzhen Hospital, Hong Kong, China

* Corresponding author: Anne Lee, MD, Department of Clinical Oncology, the University of Hong Kong, 1/F, Professorial Block, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong. Email: awmlee@hku.hk; Office telephone: (852) 2255 5050

Running title:

Response: Presenting nomograms in full

Dear Sir,

We deeply appreciate the interest in our recent publication[1] and the comments on the statistical methodology by Collins and Le Manach. We would like to take this opportunity to respond to their comments.

Firstly, we agree that bootstrapping could be another technique in model selection. However, there are also published reports that showed the potential drawbacks of bootstrapping method (for instance, retrieving overly complex models),[2,3] and a simulation study by Austin comparing bootstrap method and conventional backward elimination method showed similar performance by both methods in identifying variables.[4] Furthermore, there are studies showing that alternative method by subsampling may have merits over bootstrapping and are worth considering in future investigations.[2,5] Nevertheless, we did repeat our analyses using the bootstrap resampling approach as suggested by the captioned Letter, the final model returned is the same as that by the backward elimination method used in our original paper.[1] Specifically, the four contributing factors (overall stage, age, GTV-p and LDH) were selected in >80% of 200 bootstrap replications while the two unselected variables (sex and performance status, which were chosen via univariable analysis) were excluded in >65% of the replications. We hereby confirm that our selection of prognostic factors for the nomogram are appropriate.

Secondly, in the calibration plots we compared observed vs. predicted survival probability for the 5-year overall survival endpoint. We presented the intercept and slope of the joined lines based on the calibration plots, rather than the "calibration slope" as the captioned Letter suggested. We agree that the number of groups may affect the estimation of the intercepts and slopes. We have explored different number of groups for calibration, the estimated values remained similar regardless of using 4 (reported in the paper),[1] 5 or 10 groups. Take the example of the calibration plots based on the training cohort, the corresponding intercept (slope) for 4, 5 and 10 groups were –0.05 (1.06), –0.08 (1.10) and –0.02 (1.03), respectively. The three sets of results were insignificantly different from intercept=0 and slope=1.

We have performed additional analysis using the methods proposed in the captioned Letter, the results show similar findings as published in our original paper.[1] Given baseline hazard $h_0(t)$ as the intercept and lp(X) as the linear predictor, "calibration slope" refers to the slope b in the linear fit of $\log(h(t|X)) = \log(h_0(t)) + b\text{lp}(X)$.[6] For the training cohort, the "calibration slopes" were 1.000, 0.997 and 0.978 for 4, 5 and 10 groups, respectively; all had no significant difference from 1. The smoothed regression line from flexible adaptive hazard regression (the blue line in Figure, https://goo.gl/FlA8Xs) also showed good calibration over the training cohort, consistently supporting our conclusion.

Lastly, baseline hazard simply refers to the hazard for the standard set of conditions that continuous variables equal 0 and categorical variables equal corresponding references.[6] With this approach, investigators could directly obtain the baseline hazard from the nomogram and assess prognostication of individual patients with their data.

In conclusion, we thank Collins and Le Manach for their comments and agree that their suggested approaches are useful alternatives. However, repeating the analyses with the suggested methods and performing sensitivity analysis to have comprehensive evaluation of the predictive model showed comparable results as those obtained by the statistical methods used in our original paper.[1] We confirm that the clinical conclusions on the selection of prognostic factors and the nomogram calibration are robust and consistent, irrespective of the statistical approaches used. Our original paper provides a valid predictive model for NPC patients;[1] the developed nomogram based on the newly proposed AJCC/UICC staging (8th edition) together with additional independent prognostic factors, provide a practical supplementary tool for refining the prediction of overall survival and tailoring of treatment strategies for individual patients.

Yours sincerely,

References

1. Pan JJ, Ng WT, Zong JF, et al. Prognostic nomogram for refining the prognostication of the proposed 8th edition of the AJCC/UICC staging system for nasopharyngeal cancer in the era of intensity-modulated radiotherapy. *Cancer*.

2. Rospleszcz S, Janitza S, Boulesteix A-L. Categorical variables with many categories are preferentially selected in bootstrap-based model selection procedures for multivariable regression models. *Biom J* 2016; **58**(3):652-673.

3. Janitza S, Binder H, Boulesteix A-L. Pitfalls of hypothesis tests and model selection on bootstrap samples: causes and consequences in biometrical applications. *Biom J* 2016; **58**(3):447-473.

4. Austin PC. Bootstrap model selection had similar performance for selecting authentic and noise variables compared to backward variable elimination: a simulation study. *J Clin Epidemiol* 2008; **61**(10):1009-1017.e1.

5. De Bin R, Janitza S, Sauerbrei W, at al. Subsampling versus bootstrapping in resampling-based model selection for multivariable regression. *Biometrics* 2016; **72**(1):272-280.

6. Cox DR. Regression Models and Life-Tables. *J R Stat Soc Ser B Methodol* 1972; **34**(2):187-220.