

TITLE: Statistical inefficiencies in the development of a prediction model

Gary S. Collins, *professor of medical statistics*

Centre for Statistics in Medicine, Botnar Research Centre,
University of Oxford, Windmill Road, Oxford OX3 7LD, United Kingdom
Email: gary.collins@csm.ox.ac.uk. Tel: +44 (0)1865 223460

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
Email: yannick.lemanach@phri.ca

476 words (max 500 word); 3 references (max 3 references)

The authors report no conflict of interest.

Letter Re: A Clinical Prediction Model for Postcardiac Surgery Atrial Fibrillation in an Asian Population By Zhang et al, Anesth Analg 2016; 123: 283-289.

Our attention was drawn to the recent paper by Zhang and colleagues which exhibits a number statistical inefficiencies and flaws during their development a clinical prediction model, of which due to word limits, we highlight just a few so that future investigators do not fall into the same trap.

The premise of their study was existing models are focused on Western populations; then before deciding to build another prediction model, why not evaluate these existing models to see if they showed any promise, and possibly re-calibrate or update these existing models? Far too many prediction models are being developed when existing models can be used.

Our next point relates to study design. The sample size calculations the authors have presented bear no resemblance to recommended approaches; model development should be constrained to the rule-of-thumb of 10 events-per-variable, whilst model validation should comprise a minimum of 100 outcome events (1).

The authors proceed to select variables based on univariate analyses, which is a flawed approach as nuances in the data may prevent important predictors be selected. The authors also carried out a practice long regarded as being unnecessary and a waste, by throwing important prognostic information away by dichotomizing continuous variables (2). Even if an easy-to-use model is required, it is preferable to retain all continuous predictors and simplify the model afterwards – retaining as much predictive ability as possible (3).

Model validation, an important part of model development, was carried out on a separate cohort collected from the same institutions during the same time period; this is a weak test of model performance is nothing but random split-validation; again a practice that has been shown to be unnecessary and inefficient (2).

The authors evaluated calibration by calculating the Hosmer-Lemeshow test, widely known as being affected by sample size but more importantly as being uninformative in that it provides no indication of magnitude or direction of (mis)calibration. In accordance with recent recommendations on the reporting of prediction model studies, calibration should be assessed graphically by plotting predicted outcome probabilities (x-axis) against observed outcomes (y-axis) using a high resolution smoothed (loess) line (1). It is important to also highlight that the performance characteristics reported are of the logistic regression (which is never published in full, the intercept is missing) and not the simplified scoring system which the authors are promoting to be used. Therefore, model performance will be substantially worse than reported in the paper.

Our final point relates to the presentation of the simplified model. Whilst a simplified model may increase model uptake, it is important to indicate what a score of 1, 2, ..., 9 relates to in terms of predicted risk. Creating arbitrary risk groups, low, moderate and high without quantifying what this actually means is uninformative. For example, it would be useful to indicate that low risk relates to predicted risk between x% and y%.

REFERENCES

1. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015;162:W1-W73.
2. Collins GS, Ogundimu EO, Cook JA, Le Manach Y, Altman D, G. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med* 2016.
3. Sullivan LM, Massaro JM, D'Agostino Sr RB. Presentation of multivariate data for clinical use: the Framingham study risk score functions. *Stat Med* 2004;23:1631-60.