Clinical prediction models: a fashion or a necessity in medicine?

Alain Bernard

Department of Thoracic surgery, University Hospital of Dijon, Dijon, France

Correspondence to: Alain Bernard, MD, PhD. Department of Thoracic Surgery, University Hospital of Dijon, Bocage central 14 rue Gaffarel 21079 Dijon, France. Email: alain.bernard@chu-dijon.fr.

Provenance: This is an invited Editorial commissioned by Section Editor Dr. Min Zhang (Department of Thoracic Oncology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China).

Comment on: Brunelli A, Salati M, Rocco G, *et al*. European risk models for morbidity (EuroLung1) and mortality (EuroLung2) to predict outcome following anatomic lung resections: an analysis from the European Society of Thoracic Surgeons database. Eur J Cardiothorac Surg 2017;51:490-7.

Surgeons need tools to help them make decisions when discussing therapeutic management. According to PubMed, the number of publications of prognostic studies has increased exponentially in recent years (1). Risk models have been developed for many surgical interventions. One of the most famous is the EuroSCORE for cardiac surgery (2). The first risk model for thoracic surgery was the Thoracoscore, created by Falcoz *et al.* (3). This score was developed using data from the French national thoracic surgery database Epithor®.

The need to create these scores is part of a broader and more complex movement than predictive medicine. The score concept, widely adopted by the media, reflects the need for each patient and doctor to be able to predict what will happen to him or her in cases of a serious illness or a complication after surgery. I must underline here that a risk model cannot answer this question. It is dangerous to make patients believe that this type of score will predict postoperative complications. The risk models found in the literature are not meant to predict the future, but to aid in decision-making. This reminder seems important to clarify the role and limits of risk models.

The latest publication on the European Thoracic Surgery Database (4) developed two new risk models for mortality and morbidity. The study population was large since 47,960 patients with pulmonary resection for cancer were included. The methodology used and the scientific approach were of high enough quality to meet the objectives of the study. In the discussion section, the authors perfectly focused on the limitations of this type of study. I would like to draw attention to some points that can significantly modify the results of a predictive model. The first concerns potential selection bias, which would result in retaining in the database only patients whose characteristics are different from those of the general population. The prognosis for mortality or the occurrence of postoperative complications in these selected patients is different from that in the population of all patients operated on for lung cancer. The other limit, which is rather similar, relates to the centers participating in the database. Indeed, less than 20% of European centers participate actively in this database. The participating teams are motivated and not necessarily representative of all European teams. These motivated teams may have a different medical organization to treat patients with bronchial cancer. These potential biases can significantly influence the selection of variables in the model and modify their coefficients. In 2007, a team from the Columbia Hospital tested the Thoracoscore in patients from their database (5). For some variables of the Thoracoscore, the coefficients were very different from those of the model in the princeps study (3). Take as an example the American Society of Anesthesiologists (ASA) score ≥3, it had a coefficient of 0.6057 in the princeps study (3) and 1.909 in the test sample (5). The same observation is true for the Zubrod score, for which the two coefficients were very different: 0.6890 for the princeps study (3) and 2.655 for the test sample (5). These coefficients how that the patients recruited in the two databases are likely to have different characteristics. Another limitation concerns missing data; variables can be excluded when data are

missing for too many patients or can be retained if patients with the missing data for the variable are excluded, as was the case in the European study (4). Not including a variable in the model may constitute a lack of essential information. The dyspnea score was not collected in the Columbia team's database (5). This variable is an important prognostic factor in the Thoracoscore (3), its absence may explain the differences observed in the model's coefficients.

The foregoing shortcomings are inherent in many prognostic studies, meaning that predictive scores should be used with caution. Theoretically, a predictive score must be validated externally using an independent and representative sample. This validation process should be systematic before a risk model can be used routinely. As with therapeutic trials, reading grids have been developed to determine the validity of a prognostic study.

The work of Steyerberg (1) perfectly described the processes essential to develop a predictive model in clinical practice and the medical research leading up to the score can be used routinely. The first part concerns the different applications of predictive models in medicine. They can be used in public health to set up preventive measures, in clinical practice as a decision-making aid, and finally in research for inclusion in a randomized controlled trial or to make adjustments. A sophisticated analysis cannot salvage a poorly designed study or poor data collection procedures. Then, the steps towards developing a valid predictive model begin with an attentive analysis of the literature to search for validated prognostic factors. A representative sample is then constituted and the methodology described in the paper of Brunelli *et al.* for example, can be used to develop the model (4). The model must then be validated externally in an independent sample so that it can be used routinely.

In conclusion, risk models are increasingly used in medicine to help practitioners improve the quality of care. However, as for randomized controlled trials, prognostic studies must respect certain quality criteria, in particular the representativeness of the patients included and the quality of the data used in the model (6). Finally, a predictive score should only be used routinely if it has been validated in an independent sample.

Footnote

Conflicts of Interest: The author has no conflicts of interest to declare.

References

1. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating (Statistics for Biology and Health). New York: Springer, 2009.
2. Nashef SA, Roques F, Michel P, et al. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardiothorac Surg 1999;16:9-13.
3. Falcoz PE, Conti M, Brouchet L, et al. The Thoracic Surgery Scoring System (Thoracoscore): risk model for in-hospital death in 15,183 patients requiring thoracic surgery. J Thorac Cardiovasc Surg 2007;133:325-32.
4. Brunelli A, Salati M, Rocco G, et al. European risk models for morbidity (EuroLung1) and mortality (EuroLung2) to predict outcome following anatomic lung resections: an analysis from the European Society of Thoracic Surgeons database. Eur J Cardiothorac Surg 2017;51:490-7.
5. Chamogeorgakis TP, Connery CP, Bhora F, et al. Thoracoscore predicts midterm mortality in patients undergoing thoracic surgery. J Thorac Cardiovasc Surg 2007;134:883-7.
6. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ 2015;350:h870.