Mortality prediction in patient with atrial fibrillation and 0 or 1 nongender-related CHA2DS2-Vasc risk factor using machine learning: a French nationwide study

A. Bisson¹, Y. Lemrini¹, G.F. Romiti², M. Proietti³, D. Angoulvant¹, S. Bentounes¹, W. El-Bouri⁴, G.Y.H. Lip⁴, L. Fauchier¹

¹CHU de Tours, Tours, France
²University Rome Tre, Department of Translational and Precision Medicine, Sapienza, Rome, Italy
³University of Milan, Milan, Italy
⁴Institute of Ageing and Chronic Disease, Liverpool, United Kingdom of Great Britain & Northern Ireland

Funding Acknowledgements: None.

Background: Atrial fibrillation is associated with important mortality included in patients with low or intermediate thromboembolic risk. Usual clinical risk scores are challenged in this specific population and predict only modestly mortality. Machine learning approach helps to identify new risk factors and outperform existing tools for the prediction of outcomes in atrial fibrillation patients.

Purpose: This study aimed to developpe a machine learning model for the prediction death occurrence within the year following atrial fibrillation diagnosis and to assess performances in patients with 0 or 1 nongender-related CHA2DS2-Vasc risk factor compared to previously described clinical scores.

Methods: We used for this purpose a nationwide cohort of 2,435,541 newly diagnosed atrial fibrillation patients seen in all the French hospitals from 2011 to 2019. Among 110 variables, the 18 most predictive variables were identified and selected using a Random Forest algorithm as adding any other variable did not improve prediction (Figure 1). Three machine learning models (logistic regression, random forests, deep neural networks) were trained to predict mortality within the first year on a train set (70% of the cohort). The best model was selected to be evaluate on the validation set (30% of the cohort). We finally compared our best model with previously published scores in patients with 0 or 1 nongender-related CHA2DS2-Vasc risk factor using the C index.

Results: Among atrial fibrillation patients, 373,155 (15.32%) had 0 or 1 nongender-related risk factor. Within the first year following AF diagnosis, 23,857 patients (6.39%) died after a mean time of 86 (SD 26) days. After training, the best machine learning model selected was a deep neural network with a C index of 0.915 (95% CI, 0.905-0.925) on the validation set. Compared to traditional clinical risk scores, the deep neural network model was significantly superior to the CHA2DS2-VASc and HAS-BLED scores, and dedicated scores such as Charlson Comorbidity Index and Hospital Frailty Risk Score to predict death within the year following AF diagnosis (C indexes: 0.521; 0.633; 0.849; 0.800 respectively. P<0.0001) (Figure 2, Panel A). The ability to predict AF was improved as shown by the net reclassification index and integrated discriminatory improvement increase (P<0.0001, respectively) and decision curve analysis (Figure 2, Panel B).

Conclusion: Patients with atrial fibrillation and 0 or 1 nongender-related CHA2DS2-Vasc risk factor remain at risk of mortality within the first year of the diagnosis. A ML approach may help clinicians to better risk stratify AF patients with low or intermediate thromboembolic risk at high risk of mortality.

Figure 1. Feature importance
Figure 2. ROC and decision curves