AutoScore: An Automatic Clinical Score Generator with Logistic Regression and Machine Learning

Feng Xie, BS¹, Bibhas Chakraborty, PhD¹, Lian Leng Low, MBBS³, Marcus Eng Hock Ong, MBBS, MPH¹,³, Nan Liu, PhD¹,²
¹Duke-NUS Medical School, National University of Singapore, Singapore; ²Singapore Health Services, Singapore; ³Singapore General Hospital, Singapore

Introduction
Risk score can be useful in clinical risk stratification and accurate allocations of medical resources, helping health providers succeed with value-based care. Point-based scores are more understandable and explainable for clinicians than other complex models and now widely used in clinical decision making. However, the development of risk scores is non-trivial. With the popularity of risk score in practice, there is a need to develop disease-specific or population-specific score conveniently or automatically. To achieve this, we proposed AutoScore, an automatic clinical score generator, based on random forest feature selection, regression model, and categorization of continuous clinical variables.

Methods
We performed variable selection based on random forest algorithm, converted numeric clinical variables to categorical variables, and then calculated scores based on a regression model. The total score is a sum of several breakdown points that are automatically generated. To evaluate the performance of AutoScore, we used data of 1,000 inpatients obtained from Singapore General Hospital. We used our AutoScore algorithm to build a score for 30-day readmission risk prediction and then compared it with the logistic regression model with receiver operating characteristic (ROC) analysis. We also developed an R package for AutoScore implementation.

Results and Evaluation

![Figure 1. Comparison of our readmission risk score with a logistic regression model in ROC.](image)

**Table 1.** The score of readmission prediction generated with our AutoScore algorithm.

Our score was better than logistic regression model in terms of ROC analysis. The automatically generated, point-based score is easy-to-use in clinical practice and able to capture non-linear association than logistic regression model for many continuous clinical features. In this study, our AutoScore algorithm selected less than 10 best predictors from dozens of available variables based on random forest, making it clinically applicable. Given a new dataset, our AutoScore tool can be conveniently implemented to generate a point-based clinical score.

Conclusion
We developed an easy-to-use, automatic clinical score generator, AutoScore, to conveniently build risk scores for big data or population-specific cohort. AutoScore algorithm integrates both the advantage of machine learning in the strong predictive power and the merit of regression analysis in its excellent accessibility and interpretability.