International Journal of Population Data Science

Journal Website: www.ijpds.org

Machine learning: how much does it improve the prediction of unplanned hospital admissions?

Weaver, C¹, McBrien, K¹, Sajobi, T¹, Ronksley, PE¹, Lethebe, B¹, and Williamson, T¹

¹University of Calgary

Introduction

Risk prediction models can be used to inform decision-making in clinical settings. With large and detailed electronic medical record data, machine learning may improve predictions. The objective of this work is to determine the feasibility and accuracy of machine learning versus logistic regression to predict unplanned hospital admissions.

try multiple methods, compare their predictive accuracy and interpretability, and then choose a final method.

Objectives and Approach

Data from primary care electronic medical records for community-dwelling adults in Alberta, Canada available from the Canadian Primary Care Sentinel Surveillance Network will be linked to acute care administrative health data held by Alberta Health Services. Two regression methods (forward stepwise logistic, LASSO logistic) will be compared with three machine learning methods (classification tree, random forest, gradient boosted trees). Prior primary and acute care use will be used to predict three outcomes: ≥ 1 unplanned admission within 1 year, ≥ 1 unplanned admission within 90 days, and ≥ 1 unplanned admission within 1 year due to an ambulatory care sensitive condition.

Results

The results of this work in progress will be presented at the conference. 41,142 patients will have their primary and acute care data linked. We anticipate that the machine learning methods will improve predictive performance but will be more challenging for clinicians and patients to understand, including why a given patient is predicted to be at higher risk. The primary comparison of machine learning and regression methods will be based on positive predictive values corresponding to the top 5% predicted risk threshold, and estimated via 10-fold cross-validation.

Conclusion/Implications

This project aims to help researchers decide which statistical methods to use for risk prediction models. When considering machine learning methods the best approach may be to