Disparities in Hospital Readmissions and Lack of Fairness

Huan-ju Shih, MHA[1], Janusz Wojtusiak, PhD, Naren Durbha, MSHI[1], George Mason University[1], Fairfax, VA, USA

Introduction

Hospital readmission reduction programs have been in place for over ten years, yet contradictory evidence exists on gender, racial, and age disparities. Similarly, while a large number of published works describe the construction of machine learning-based models to predict hospital readmissions, very few address the issue of disparities and, consequently, fairness of the constructed models[1]. Unfairness in predicting hospital readmissions may lead to deepening discrimination for underserved groups of patients who would benefit from preventative care aimed at reducing readmissions. The goals of the presented research were to (a) check for disparities in 30-day all-cause hospital readmissions (b) apply machine learning methods to construct models for predicting readmissions for adults, and (c) test if these models propagate biases and may contribute to creating more disparities.

Methods

Data: Data used for this experiment came from AHRQ's Healthcare Cost and Utilization Project (HCUP), New York State 2016 inpatient database (SID). That year was chosen as it included ICD-10-CM, as opposed to ICD-9 in previous years. It contains more than 95% of the state's inpatient discharge records from short-term, acute-care, nonfederal, general, and other specialty hospitals, which can be used to study readmission. A total of 2,293,306 hospitalization records for 1,659,237 was extracted from the data. After excluding patients <18 years old and records with missing key information, the cohort consisted of 1,954,696 hospitalizations for 1,352,577 patients. The outcome was defined as readmission (any hospitalization) within 30 days of index hospital discharge, except for scheduled hospitalizations as defined by CMS[2]. Model inputs included age, gender, race, ethnicity, diagnoses - coded using AHRQ's Clinical Classification Software Refined (CCSR), and procedures – coded using AHRQ's CCS codes. From the total 724 diagnosis and procedure codes, 186 were selected that appear in at least 1% of patients.

Experimental setup: Data were split into training (70%) and testing (30%) sets. Training data was used to build and tune models (10-fold cross-validation and final model construction). Final models were tested on the test set. Random forest (RF), gradient boost (GB), logistic regression (L1, L2, elastic net, plain), and multilayered perceptron (MLP) were used to construct models. Models were evaluated using statistical measures (auc, precision, recall), calibration, sensitivity, and SHAP. Training and test data were stratified by age, gender, race, and ethnicity to evaluate model performance within populations.

Test for bias and potential disparities was done using demographic parity test (societal and algorithmic bias), model evaluation in sub-popualtions (algorithmic bias) as well as regression modeling controlled for other patient characteristics (sociatal and algorithmic bias).

Results

The results indicate that there are disparities (societal bias) in hospital readmissions ($p<0.01$) for age, gender, Hispanic origin, and race. Female patients are less likely to be readmitted (OR .85; CI 0.84-0.86), while Black patients are more likely to get readmitted (OR 1.3; CI 1.3-1.3). These disparities are propagated to the models that predict readmissions. All tested models performed similarly in terms of AUC (cross-validated and on the test set), ranging from 0.68 to 0.7, with MLP being the best. All models were poorly calibrated, with the exception of GB and MLP. However, MLP did not produce any probabilities > 0.4 on the test data, and GB did not produce any probabilities > 0.7. The AUC within subpopulations differs. It is higher for patients who are female (0.71), Hispanic (0.72), Asian (0.74), or Native American (0.73), and lower for male (0.67) and White (0.68) patients. Similarly, the probabilities in sub-populations are significantly differently distributed.

Discussion and Conclusions

Existing disparities are propagated through the prediction of hospital readmission, but no additional bias is added. Accuracy of obtained models is alighed with results available in literature and reasonable, yet bad calibration and biases make them practically unusable for technical and ethical reasons. Further work is needed to remove biases from the data and improve model performance for high-risk cases for whom an intervention may be focused. Generalizability of models to other states and potential concept drift need to be addressed.

References

1. Radovanović S, Petrović A, Delibašić B, Suknović M. Making hospital readmission classifier fair–What is the cost?. InCentral European Conference on Information and Intelligent Systems 2019 (pp. 325-331). Faculty of Organization and Informatics Varazdin.
2. Readmission Measures Hospital-Specific Reports [Internet]. [cited 2023 Mar 20]. Available from: https://qualitynet.cms.gov/inpatient/measures/readmission/reports.