Examining Fairness in Machine Learning Predictions of Healthcare Costs in Older Women with Osteoarthritis: XGBoost Regression

Sahar Elchehabi¹, Arshama Dehghan², Chanhyun Park³, Nethra Sambamoorthi¹, Chan Shen⁴, Usha Sambamoorthi¹

¹The University of North Texas Health Science Center; ²UT Southwestern Medical Center; ³The University of Texas at Austin; ⁴Penn State Health

BACKGROUND

• Osteoarthritis (OA) is the most common joint disorder worldwide with women more prone to develop symptomatic disease than men. This condition has shown to be associated with high direct healthcare costs. However, gaps persist in our understanding of economic burden specific to this demographic.

OBJECTIVE

• Investigate the leading predictors of economic burden among older women with OA utilizing Machine Learning (ML) methods as well as the fairness of the ML algorithm in its predictions for sensitive attributes.

METHODS

- Study Design: Cross-sectional
- Data Source: 2021 Medical Expenditure Panel Survey, a nationally representative survey of the non-institutionalized civilian households in the United States (US)
- Analytical Sample: Older women (age \geq 65 years) with OA (weighted N=1,262,496)
- *Target:* High-cost users were identified as having greater than the 90th percentile (>\$39,388) in total healthcare expenditures • *Features:* Age, health status including comorbidities (anxiety, arthritis, asthma, cancer, COPD, depression, diabetes, high cholesterol, hypertension, and thyroid disease), and social determinants of health (SDoH) (race and ethnicity, marital status, health insurance coverage, education, and poverty status) Methods: XGBoost classifications and SHapley Additive
- exPlanations (SHAP) for leading log-transformed predictor identification and interpretations

Model Building and Performance: 70% training and 30% testing split of the data

Six rounds of optimization using Python 3.10.12 Model performance metrics included root-mean-square error (RMSE) and R-squared

Fairness metrics included mean absolute error (MAE), RMSE, and R-squared across subgroups and counterfactuals using swap test



- The overall mean healthcare expenditures was \$18,619.
- poverty status, and college for education level.

In comparison to NHW, <u>NHB</u> and <u>Other</u> race had much higher expenditures.

Middle income class had the highest incremental expenditure out of all groups when compared to high income.

Less than high school (HS) and HS education levels had **much higher** expenditures versus college education level

- The model fit was exceptional: RMSE 0.20 and R-squared 0.96
- SHAP analysis revealed age, high school (HS), some college, and comorbidities as leading predictors





Income Level, and **Education Level**

when compared to Non-Hispanic White (NHW) race.

RESULTS

• The privileged groups are: NHW for race and ethnicity, high income for

Incremental Expenditures Compared to **Privileged Counterparts**

Race and Ethnicity	Expenditures (\$)
NHB	3,398
Hispanic	1,255
NHA	-1,903
Other	4,005
Income	Expenditures (\$)
Poor	4,866
Near Poor	4,866
Middle	8,717
Education Level	Expenditures (\$)
Less Than HS	4,291
HS	4,544
Some College	1,574



MSR16

RESULTS CONT.

Individual Fairness Counterfactual Predictions

STRENGTHS & LIMITATIONS

• The strengths of the study include a nationally representative survey, a comprehensive list of features, enhanced interpretability, and accurate predictions. However, limitations such as the crosssectional study design and self-reported data must be considered when interpreting our findings.

CONCLUSIONS

• While the ML model performed well in accuracy by its low RMSE and high R-squared, the group and individual fairness metrics suggested bias in predictions.

• Research studies need to routinely incorporate fairness metrics for sensitive attributes in ML models.



THE UNIVERSITY of NORTH TEXAS HEALTH SCIENCE CENTER at FORT WORTH