Leading Predictors of Economic Burden Among Postmenopausal Women with Heart Failure: An Application of Machine Learning with XGBoost and SHapley Additive exPlanations

Arshama Dehghan, BS[1]; Chanhyun Park, PhD[2]; Nethra Sambamoorthi, PhD[3]; Chan Shen, PhD[4]; Nawar Shara, PhD[5]; Usha Sambamoorthi, PhD[1]

[1]University of North Texas Health Science Center, Fort Worth, TX; [2]The University of Texas at Austin, Austin, TX; [3]Northwestern University, Evanston, WV; [4]Penn State College of Medicine, Hershey, PA; [5]Medstar Health Institute, Hyattsville, MD

THE UNIVERSITY *of* NORTH TEXAS HEALTH SCIENCE CENTER *at* FORT WORTH

BACKGROUND

- **Heart Failure (HF)** is a long-term and costly health condition that impacts numerous individuals, particularly postmenopausal women. There are gaps in our understanding of HF research within this specific population.

OBJECTIVE

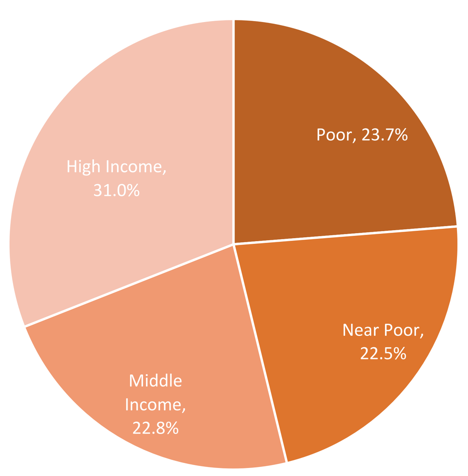
- Identify key predictors and their associations with economic burden, specifically third-party expenditures (3rd Party) and out-of-pocket spending (OOP), among postmenopausal women with HF using nationally representative data and interpretable machine learning methods.

METHODS

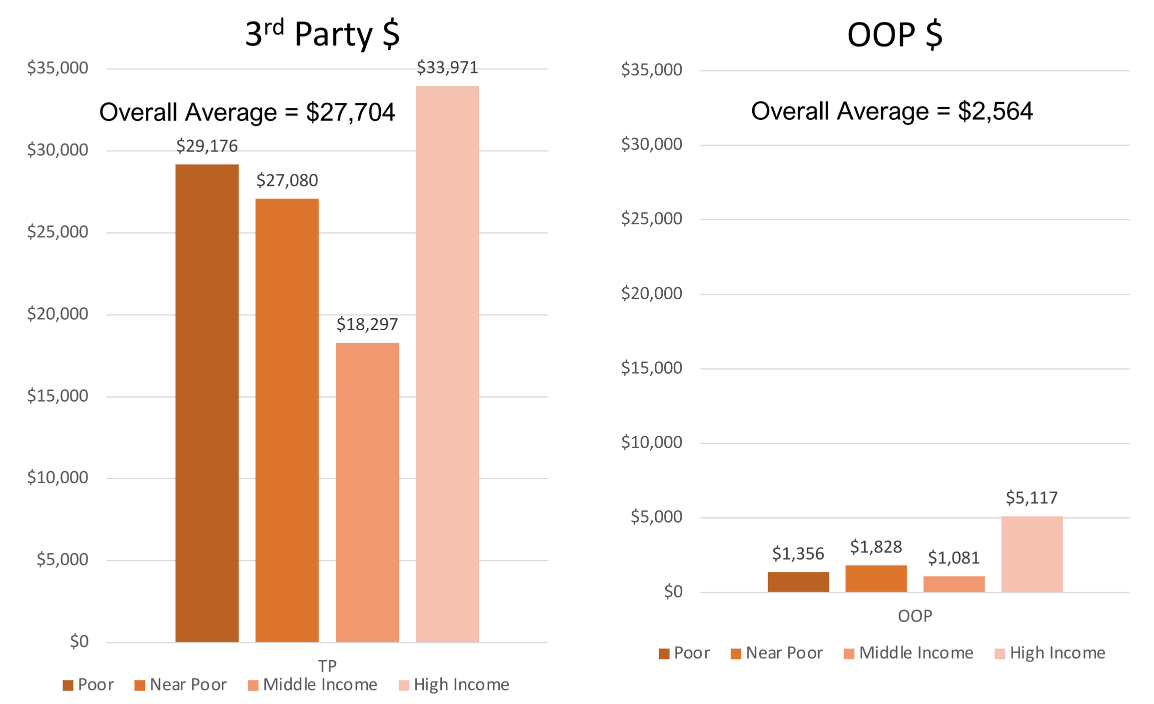
- **Study Design:** Cross-sectional
- **Data Source:** 2020 Medical Expenditure Panel Survey (MEPS), a nationally representative survey of households in the United States.
- ***Analytical Sample***: Postmenopausal women (i.e., age $\geq$ 50 years) with HF (weighted N=600,742).
- ***Target***: Economic burden was assessed with 3rd Party expenditures paid by the insurers, and total OOP spending on healthcare paid by the respondents or their families.
- ***Features***: 21 features were used: age, health status including comorbidities (anxiety, arthritis, asthma, cancer, COPD, depression, diabetes, high cholesterol, hypertension, and thyroid disease), perceived physical and mental health status, polypharmacy, social determinants of health (SDoH) (marital status, health insurance coverage, prescription drug coverage, education, poverty status, region).
- ***Methods***: XGBoost regressions and SHapley Additive exPlanations for key predictors identification and interpretations
- ***Model building and Performance***:
- 70% training and 30% testing split of the data,
- 10-fold cross-validations, and
- Six rounds of optimization using Python 3.9.12.
- Model performance metrics included absolute mean squared errors, root mean squared error and coefficient of determination.
- ***Results***: Model performance metrics were excellent with: mean absolute errors (0.442, 0.310), root mean square errors (0.452, 0.342), and coefficients of determination (0.935,0.987) for 3rd Party and OOP spending, respectively.

RESULTS

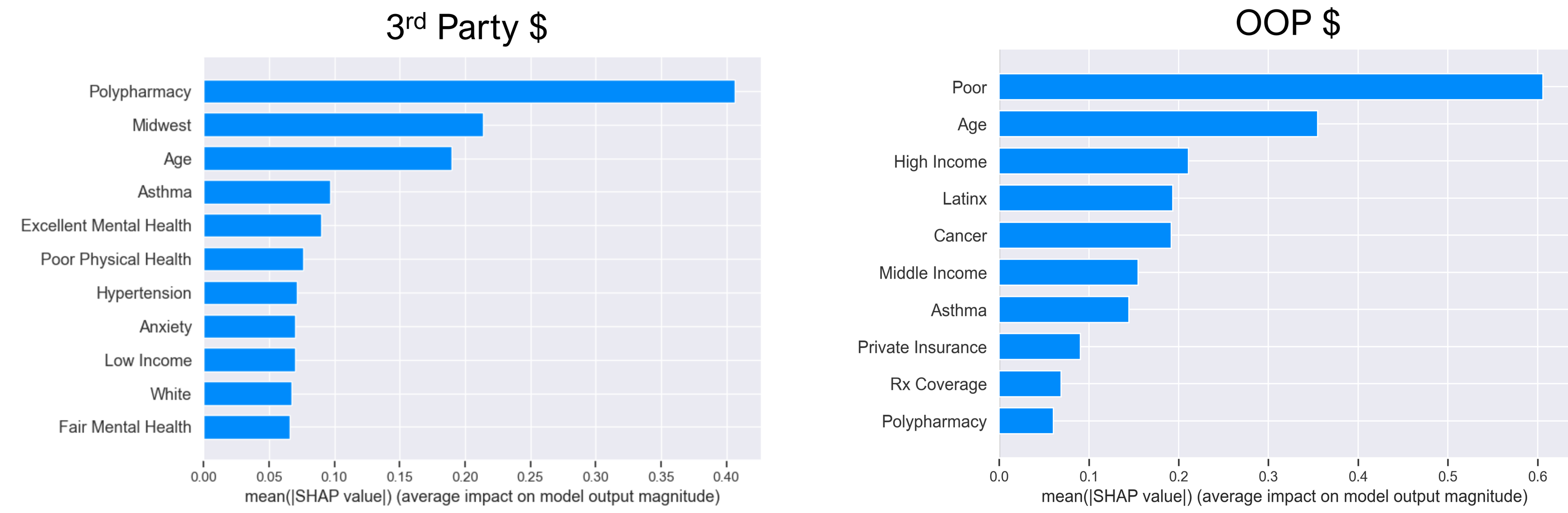
Weighted %, Poverty Status, MEPS 2020



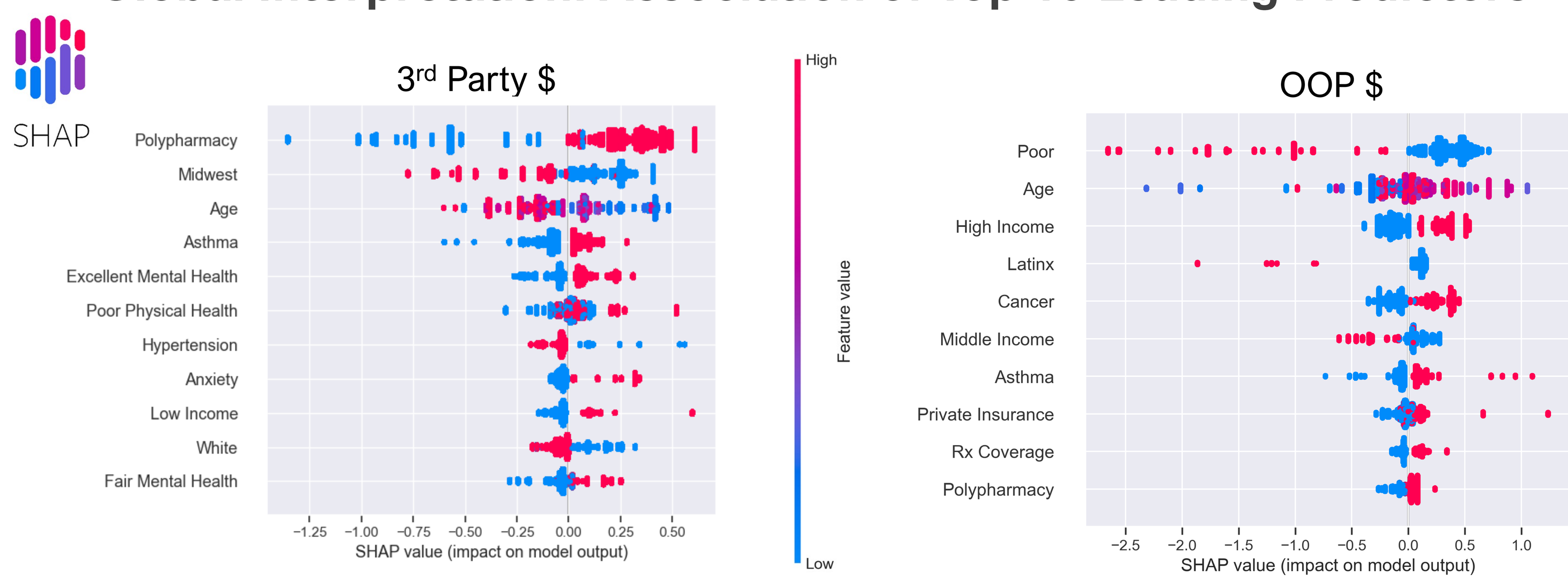
Average $ by Poverty Status



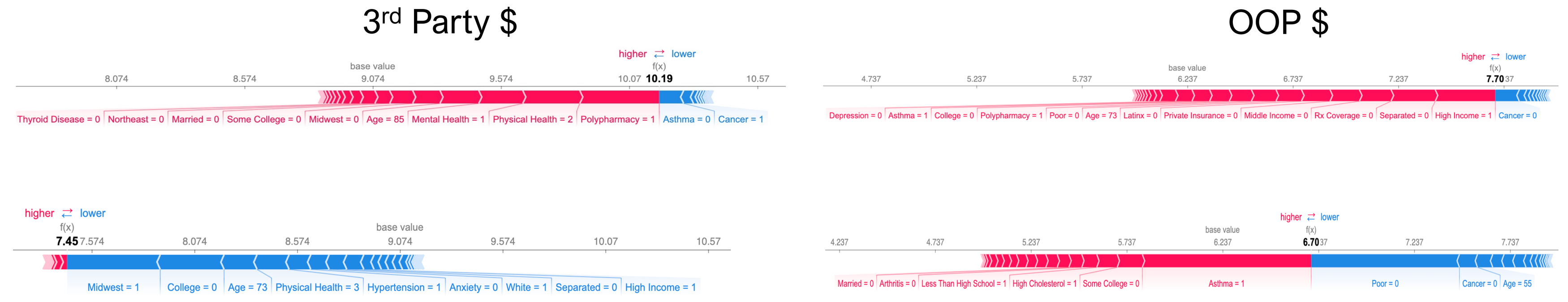
Leading Predictors: Feature Importance



Global Interpretation: Association of Top 10 Leading Predictors

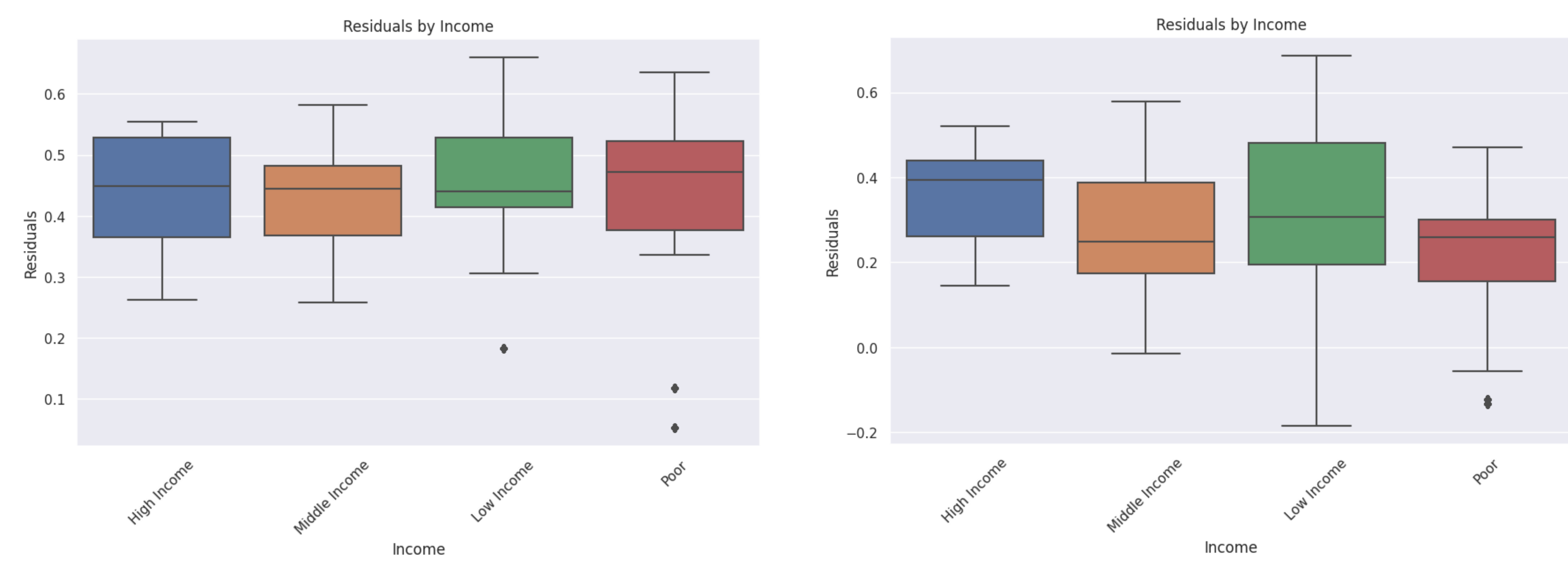


Local Interpretation: Drivers of Healthcare Expenditures



RESULTS CONT'D.

Predicted Residuals by Poverty Status



- Our global interpretations suggest that there were common drivers of 3rd party and OOP expenditures. These were biological factors (age) and clinical factors such as polypharmacy and comorbid conditions.
- We also observed different drivers of 3rd party and OOP expenditures.
- For example, SDoH (low income) additionally predicted higher 3rd party expenditures, but not OOP spending.
- Predicting 3rd party and OOP spending drivers for a specific person may help customize strategies to reduce economic burden.
- The Prediction Errors varied by subpopulations, for example the poor and low-income groups.

STRENGHTS & LIMITATIONS

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

CONCLUSIONS

- Drivers affecting 3rd party and OOP spending differed.
- Unsurprisingly, comorbid conditions were associated with higher expenditures for both payors and patients.
- Cost-containment policies, programs, and interventions at payor and patient levels must include effective comorbidity management strategies.
- Responsible AI/ML modeling strategies are needed to minimize underestimation of expenditures for some subpopulations.