Impact of Area Socioeconomic Deprivation and Demographic Variables on Machine Learning Models for OSA Treatment

Montana Greider, MS*, Chris Fernandez, MS*, Sam Rusk, BS*, Yoav Nygate, MS*, Dana Richardson, BSN, MAB**, Brian Hutchinson, MS, MBA**, Tim Bartholow, MD**, Jessica Arguelles***, Matthew Klimper***, Dennis Hwang, MD***, Nathaniel F. Watson, MD, MSc***, Emerson M. Wickwire, PhD****

* EnsoData Research, Ensodata, Madison, WI, USA
** Wisconsin Health Information Organization
*** University of Washington
**** University of Maryland School of Medicine

KAISER PERMANENTE ensodata research

UW Medicine DEPARTMENT OF NEUROLOGY

WISCONSIN HEALTH INFORMATION ORGANIZATION

UNIVERSITY of MARYLAND SCHOOL OF MEDICINE

Introduction

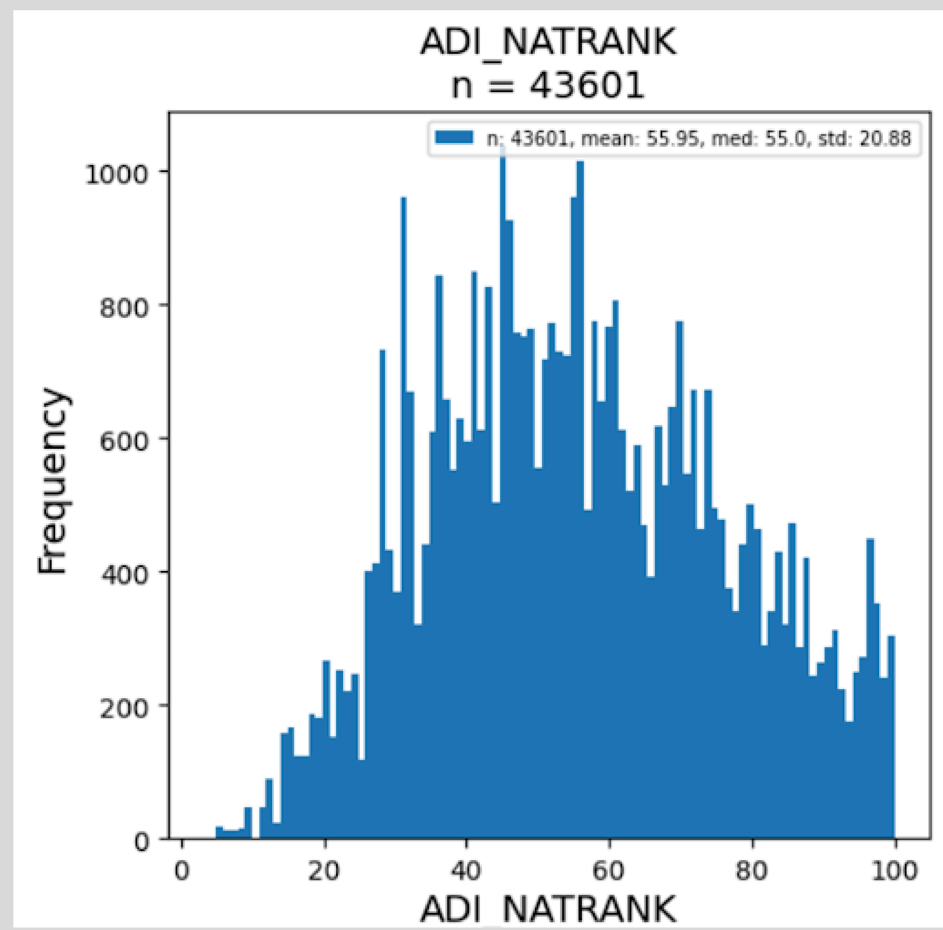
Ensuring equitable performance of sleep-related machine learning (ML) models is vital for public health and health equity. This research sought to determine the impact of sociodemographic and health disparities factors on the performance and characteristics of ML models designed to predict OSA treatment initiation.

What is Obstructive Sleep Apnea? The most common sleep-related breathing disorder, OSA can occur when the throat muscles relax and block the airway while sleeping and it causes people to repeatedly stop breathing throughout the night. OSA is linked to multiple comorbid conditions such as obesity, cardiovascular disease, and diabetes.

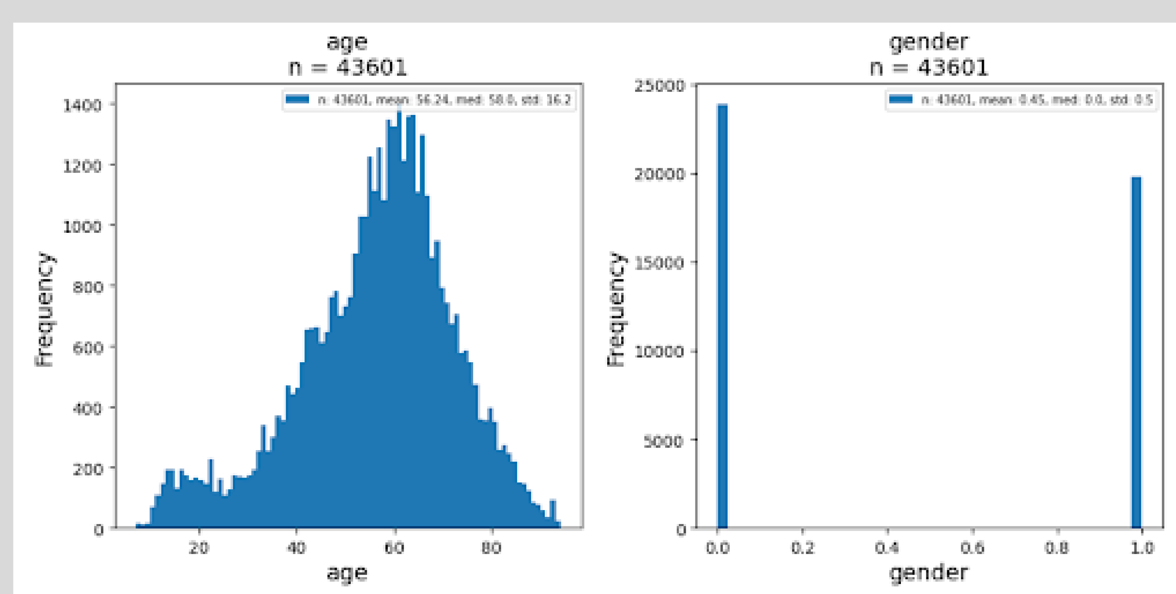
What is Area Socioeconomic Deprivation? Area socioeconomic deprivation, as measured by the Area Deprivation Index (ADI), is associated with numerous adverse health and economic outcomes such as cardiovascular risk, hospital readmissions, and Alzheimer's disease. The composite ADI score is usually ranked on a scale of 1-100 and is based on 17 health disparities indicators including income, education, employment, and housing. It is used to rank relative disadvantage across communities and is a widely utilized key social determinant of health and a validated marker of health risk. The purpose of this study was to determine the association between the ADI and OSA testing and diagnosis. Socioeconomic deprivation as measured by the Area Deprivation Index (ADI) has been associated with numerous adverse health and economic outcomes. This ADI is based on 17 health disparities factors and used to rank relative disadvantage across communities as a key social determinant of health and validated marker of health risk. The purpose of this pedagogical study is to comparatively analyze the performance and characteristics of Machine Learning (ML) models that predict whether a patient will initiate Treatment for OSA following diagnosis, when ML models are trained with versus without access to Socioeconomic and Demographic variables including ADI, Race, Gender, and Age.

The Dataset

Our data source was the All-Payer Claims Database (APCD) for the Wisconsin Health Information Organization from 2017-2022 and linked to the publicly available ADI at the census block level. The APCD includes claims data (e.g., healthcare visits, procedures, pharmacy information) from health insurers, employers, and Medicaid.



Of the N=6,026,463 participants in the overall sample, n=1,310,286 of which met initial inclusion criteria, n=154,821 underwent OSA diagnostic testing and n=43,601 were subsequently diagnosed with OSA.

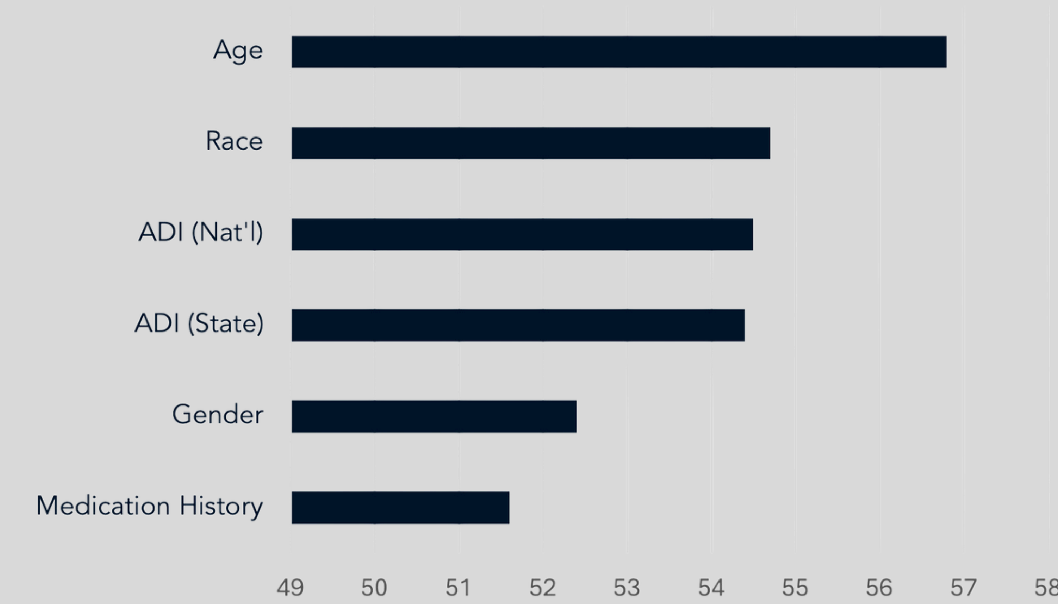


The pie graph on the left shows the number of individuals in the overall sample who underwent diagnostic testing (2.6%) while the right shows those diagnosed with sleep apnea (0.7%).

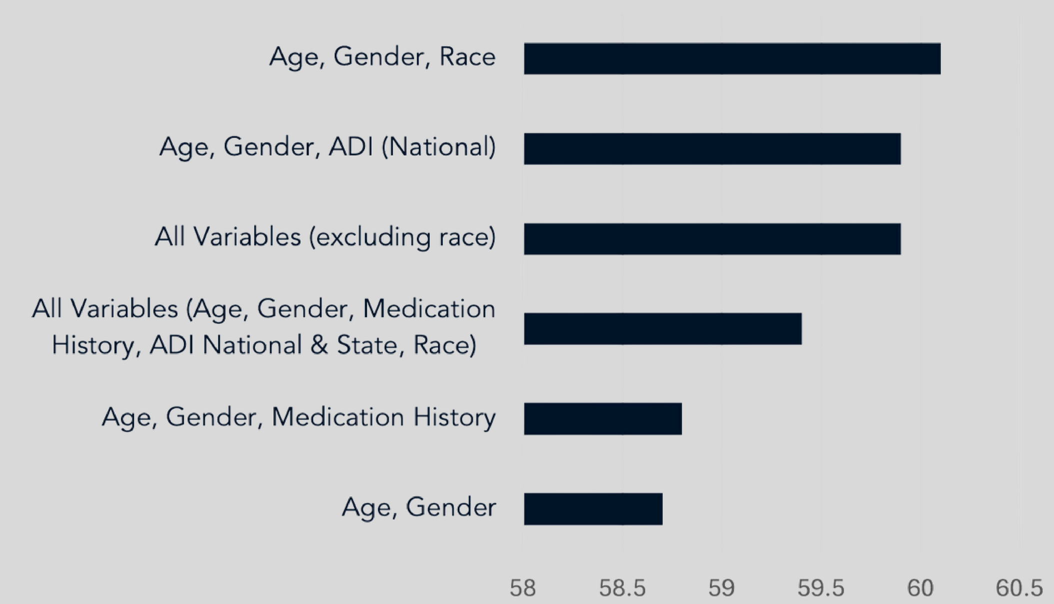
Methodology

OSA diagnosis was established by ICD coding (G47.33), and continuous enrollment for minimum 12-months prior. OSA treatment initiation was established for CPAP therapy by HCSPCS coding, with a definition for treatment initiators of 30-months following diagnosis. Random Forest models were trained to predict whether subjects initiate OSA treatment using variables among National-ADI, State-ADI, Race, Gender, Age, and a Medication History variable counting 39,712 unique drug codes across 94 medication groups.

Performance Results



This graph shows percentages that demonstrate the ability for individual data points to predict OSA.



This graph shows percentages that demonstrate the ability of combining data points to predict OSA.

Results

Of the N=6,026,463 subjects in the ACPD, n=154,821 underwent OSA diagnostic testing, and n=43,601 were diagnosed with OSA. 10-Fold Cross-Validation training-testing was applied to estimate sensitivity-specificity of ML models for predicting treatment initiation.

Receiver operating characteristic curve area-under-the-curve (ROC-AUC) analyses were used to compare relative differences in predictive power of each variable.

In ROC-AUC analysis of individual variables, the power for predicting OSA treatment were observed in relative rank order by age (0.568), race (0.547), ADI-national (0.545), ADI-state (0.544), gender (0.524), and medication history (0.516). In ROC-AUC analysis of combination variables, the highest ML model performance was observed in the combination of only three-variables (age-gender-race, 0.600), while the combination of all-variables showed an ROC-AUC of (0.594), and the ROC-AUC difference was not statistically significant.

Conclusions

Social Determinants of Health variables may play an important role in future development of predictive AI/ML models. Population sleep health data represents an important resource to identify and bridge care gaps, reduce sleep health disparities, and achieve health equity.

Future Work

Future research should seek to increase access to OSA care in areas of socioeconomic disadvantage to reduce sleep health disparities and achieve health equity.