Alcohol Lapse Risk Prediction with Machine Learning: Evaluating Algorithmic Fairness

Jiachen Yu, Kendra Wyant, Sarah Sant'Ana, Gaylen Fronk, John Curtin

(Department of Psychology)

Addiction Research Center
UNIVERSITY OF WISCONSIN – MADISON

Overview and Goals

Alcohol Lapse Prediction

- Developed an XGBoost machine learning model to predict alcohol lapses in the next 24 hours using ecological momentary assessments (EMA; Wyant et al., 2023)
- Model has exceptional performance when predicting lapses for new individuals (mean auROC = .90)
- Locally important features can identify the factors that contribute to lapse risk for any specific person and moment in time
- A "smart" digital therapeutic (smart DTx) could use algorithms based on this model to monitor lapse risk and recommend personalized, optimal interventions and supports for momentary risks

Algorithmic Fairness

- Less privileged, marginalized groups display mental health treatment disparities due to barriers related to affordability, accessibility, availability, and acceptability of mental healthcare
- Smart DTx can partially address these barriers by providing 24/7/365, affordable, personalized support
- However, if embedded algorithms perform relatively worse for less privileged groups, their use may exacerbate rather than reduce treatment disparities

GOAL: Evaluate algorithmic fairness across historically privileged and unprivileged groups

Conclusions

Summary

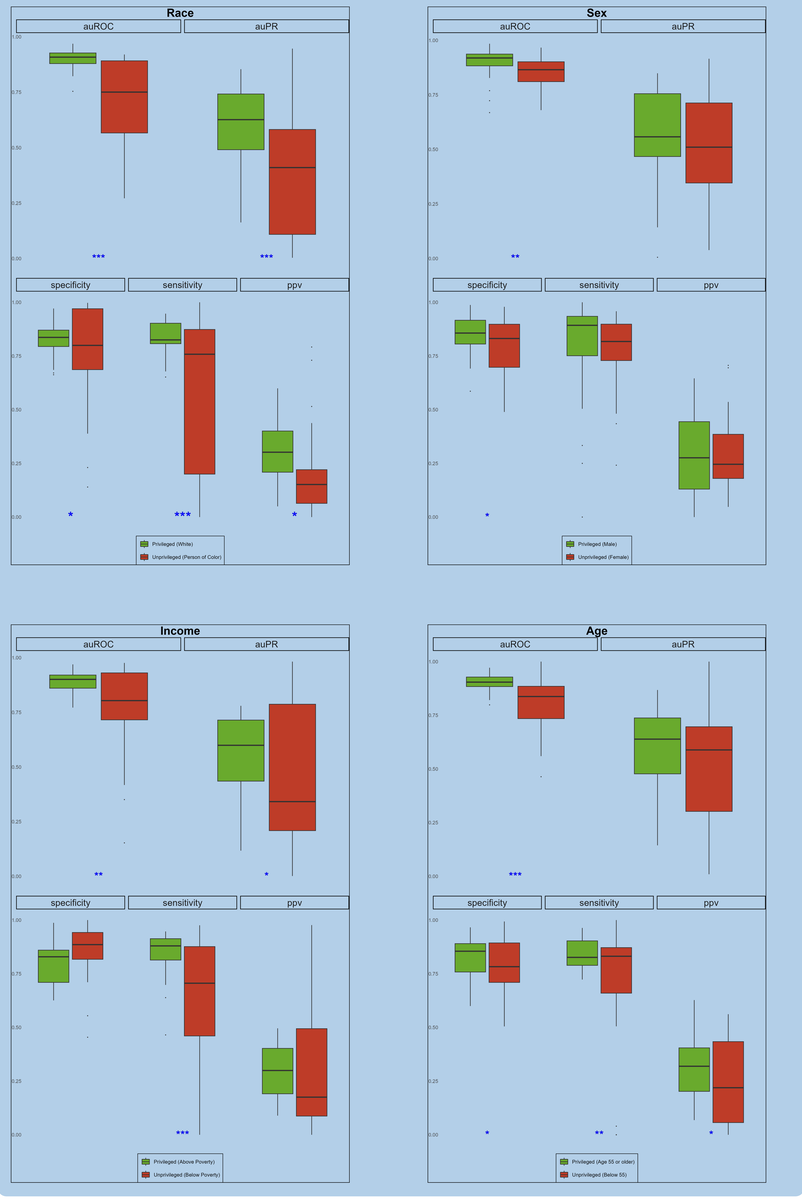
- Substantially poorer model performance for people of color
- Modestly poorer model performance for groups defined by sex, income, and age
- Representation in training data is clearly important
- Poorer performance for some unprivileged groups may also result from selection of features using domain expertise based on decades of research focused on predominately white men

Next Steps

- Evaluate methods to reduce algorithmic bias
  - More representative training data (NIDA project)
  - Resampling to increase representation
  - Modified cost functions to differential penalize errors based on privilege
  - Control for privilege when predicting
  - More representative features
- Expand evaluation of privilege
  - Rural vs urban/suburban; Education level
  - AUD severity
  - Intersectional analyses
  - Fairness in treatment/support recommendations

Research supported by NIH under NIDA award R01DA047315 and NIAAA award R01AA024391.

Results



Methods

Participants

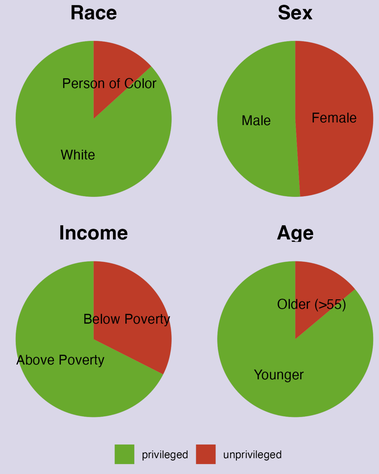
- N = 151
- Early remission from Alcohol Use Disorder
- Endorsed abstinence goal

Procedure

- Personal sensing via smartphone for up to 3 months
- 4x daily (craving, affect, efficacy, risky situations, stressful events, pleasant events)
- Self-reported lapses back to alcohol use

Sensitive Attributes

- **Race/Ethnicity**: Non-hispanic White vs. People of Color
- **Sex**: Male vs. Female
- **Income**: above 50% of median personal income in Madison ($15k) vs. below
- **Age**: 55 or younger vs. above 55



Machine Learning Model

- XGBoost classification model
- Features based on previous EMA
- Provides hour-by-hour probability for future (next 24 hour) lapse

Statistical Analysis

- Compared groups (priviledged vs. unpriviledged) on two aggregate (auROC; auPR) and three specific (specificity, sensitivity, postive preditive value) peformance metrics
- 30 held-out performance estimates (from nested k-fold cross validation) for each metric
- Posterior probabilities of group differences estimated using Bayesian generalized mixed effect models