Name: Roseann E. Peterson, PhD
Email: peterson.roseann@gmail.com
PI Name: Bradley T. Webb, PhD
PI email: bwebb@rti.org

Machine Learning to Predict Alcohol Consumption From Blood-Based Biomarkers

Amanda E. Gentry[1], Mohammed F. Hassan[1], Roseann E Peterson[2,1], Bradley T. Webb[3]

[1]Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia, USA; [2]Department of Psychiatry and Behavioral Sciences, Institute for Genomics in Health, SUNY Downstate Health Sciences University, Brooklyn, New York, USA; [3]Omics, Epidemiology, and Analytics Program, RTI International, Research Triangle Park, North Carolina, USA

Structured diagnostic interviews, screening instruments, and self-report are important tools for identifying individuals with or at increased risk of substance use disorders (SUD). The development of biological based indices may serve as complementary tools. Towards this goal, we evaluated if Elastic Nets (EN), a machine learning (ML) approach, could predict alcohol consumption (AC) using blood-based biomarkers. Predictive performance for AC was compared to other complex outcomes including height, body mass index (BMI), body fat percentage (BF%), and major depression symptoms (MDDsx).

This research has been conducted using the UK Biobank Resource (ID30782). The blood-based measures examined included 89 that are routinely collected and 249 NMR-spectroscopy measures including metabolites and lipids. LASSO and Ridge regression, two implementations of EN, were tested and predictive performance compared.

Both EN implementations performed well, with LASSO slightly outperforming Ridge regression. Baseline LASSO models using solely age and sex as features showed r-squared of 0.53, 0.48, 0.01, 0.07, and 0.07 for height, BF%, BMI, AC, and MDDsx, respectively, within the hold-out set. Inclusion of blood-based measures improved prediction for BF%, BMI, and AC to 0.71, 0.46, and 0.33, respectively, but did not significantly change for height (0.55) or MDDsx (0.002). These results demonstrate blood-based biomarkers and ML based models can be useful in predicting AC in persons where it is not assessed and may be a useful screener to identify individuals for follow up clinical assessments. Future directions include assessing additional ML approaches including Gradient Boosting Machines, which may offer additional predictive performance above EN.