Defining County level Social Determinants of Health Based Phenotypes Related to Depression in the United States Using Machine Learning Approach

Md Fitrat Hossain, PhD; Taraneh Mousavi, PharmD; Fadia T. Shaya, PhD, MPH, FAMIA

University of Maryland School of Pharmacy

Introduction

- Depression is a significant public health concern in the United States, with varying prevalence rates [1].
- It is associated with other increased cardiovascular disease incidence and mortality [1].
- During August 2021–August 2023, the prevalence of depression in adolescents and adults aged 12 and older was 13.1% which was 8.2% in 2013–2014 [2].

Aims

SDoH significantly impact the prevalence and severity of depression. This research emphasizes on identifying areas where prevalence of depression is high based on SDoH based phenotypes using machine learning algorithms.

Study Design and Data Collection

- Data on 19 SDoH variables were collected for 2347 counties across the US, from a geography information system platform PolicyMap®.
- Counties were divided into two clusters using k-means clustering based on these predictors.
- 70% and 30% of the counties were taken from each cluster to form training and the test set, respectively.
- Classification and regression tree (CART) model was used on the training set to develop phenotypes from SDoH. Random forest analysis was also conducted to identify additional risk factors.
- One way ANOVA test was conducted to check whether these phenotypes were significantly different.

Results and Principal Findings

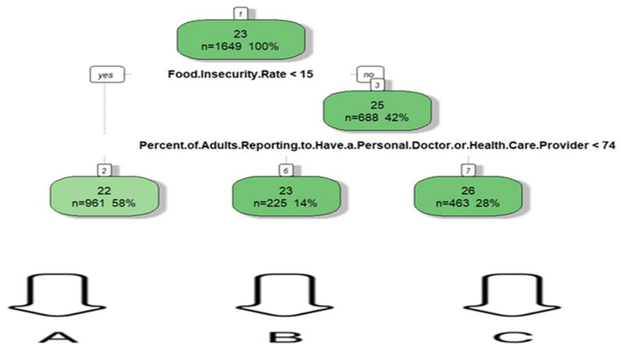

Figure 1: Three Different Phenotypes Based on SDoH Using CART

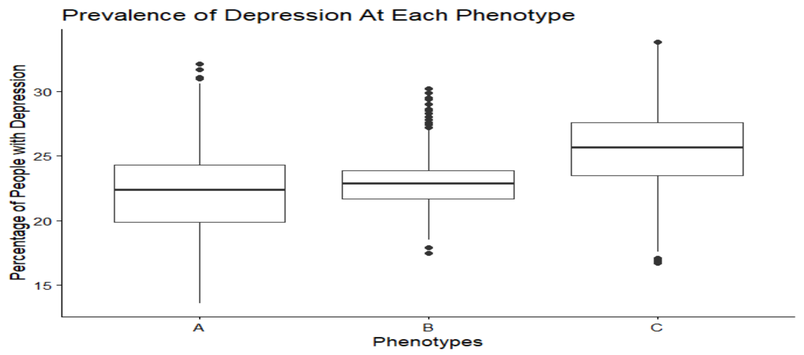

Figure 2: Percentage of People with Depression at Each Phenotype

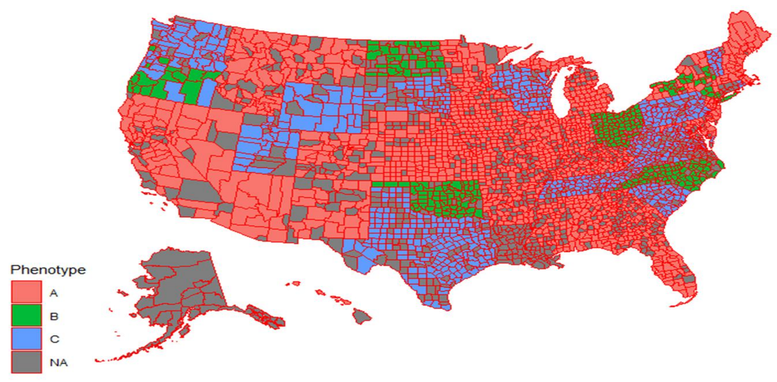

Figure 3: Phenotypes Based on SDoH Related To Depression Across United States

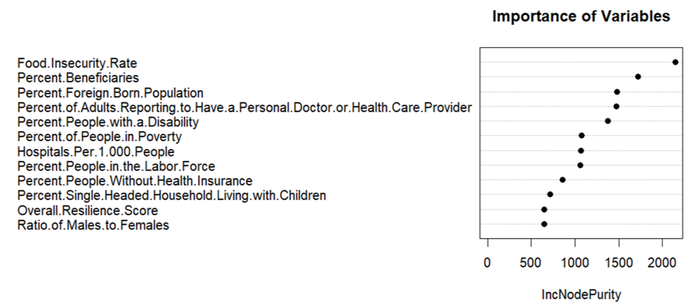

Figure 4: Relative Importance of SDoH Based on Node Purity from Random Forest

Conclusion

Three phenotypes for depression characterized by food insecurity and access to healthcare provider. These divided the counties into three distinct groups with significantly different mean depression rates (F-statistics: 280.7, p-value < 0.001). Counties with food insecurity rate ≥ 15 % and access to healthcare ≥ 74% have the highest depression rates (mean 25.57, 95% CI [25.3, 25.8]). There was little difference in model performance between training and test sets (training RMSE: 2.76, test RMSE: 2.86). Random forest models identified additional risk factors: beneficiaries, nativity, and disability, based on node purity.

Policy/Practice Implications

This research will help policy makers to identify geographic areas with high risks of depression based SDoH and make targeted interventions on those population.

References

[1] Zhang Z, Jackson SL, Gillespie C, Merritt R, Yang Q. Depressive Symptoms and Mortality Among US Adults. JAMA Netw Open. 2023;6(10):e2337011. doi:10.1001/jamanetworkopen.2023.37011

[2] Brody DJ, M.P.H., Hughes JP, M.P.H. Products - Data Briefs - Number 527 - April 2025. April 16, 2025. doi:10.15620/cdc/174579

Acknowledgement

We acknowledge the support of the University of Maryland, Baltimore, Institute for Clinical & Translational Research (ICTR) and the National Center for Advancing Translational Sciences (NCATS) Clinical Translational Science Award (CTSA), UM1TR004926, Maryland Department of Health, and SAMHSA.

Contact

1. Dr. Md Fitrat Hossain, Department of Practice, Sciences, and Health Outcomes Research, University of Maryland School of Pharmacy, 220 Arch Street, 12th Floor, Room 01-205, Baltimore, MD 21201 (Email: mdfhossain@rx.umaryland.edu)

1. Dr. Fadia T. Shaya, Department of Practice, Sciences, and Health Outcomes Research, University of Maryland School of Pharmacy, 220 Arch Street, 12th Floor, Room 01-204, Baltimore, MD 21201 (Email: fshaya@rx.umaryland.edu)