|  |
| --- |
| **DoD Manuscripts – Prospectus/Abstract Form**  Proposing Author: Lauren Gaydosh  Sponsoring Investigator (if necessary): N/A  Proposed co-authors: Harris, Shanahan, Copeland, Goodwin, Hill, others can opt in?  Provisional Paper Title: Predicting Diseases of Despair  Potential Journals: *Demography*  Intended Submission Date: Abstract for PAA 2020 submitted September 29, 2019 |
| Objective:  Can we use supervised machine learning strategies to predict the diseases of despair (compared to traditional statistical methods)? Are “despair” measures relevant in predicting the diseases of despair? Are different measures of the diseases of despair predicted similarly, and with similar important of despair measures? |
| Measures used:  Dependent variables: diseases of despair, including suicidal ideation, heavy/problem drinking, drug use, DSM-IV criteria for alcohol, marijuana, and other drug abuse and dependence   * Suicidal ideation at Wave IV   + Thoughts of suicide in the last 12 months * Alcohol use at Wave IV   + Days drank alcohol in last 12 months   + Drinks each time in last 12 months   + Days drank alcohol in last 3 months   + Drinks each time in last 3 months   + Occasions of more than 4/5 drinks in a row in last 12 months   + DSM4 lifetime diagnosis of alcohol abuse or dependence * Marijuana use at Wave IV   + Times used in last 30 days   + DSM4 lifetime diagnosis of cannabis abuse or dependence * Illegal drug use at Wave IV   + DSM4 lifetime diagnosis of other drug use/dependence   + Lifetime use of cocaine, crystal meth, heroin, other illegal drugs (separately) * Prescription drug abuse at Wave IV   + Lifetime use of sedatives, tranquilizers, stimulants, painkillers (separately)   Key independent variables: measures of cognitive and emotional despair   * Cognitive despair   + Wave III personality measures of self-consciousness, persistence * Emotional despair   + Wave I – III CES-D items   Other variables:   * Demographic controls for age at Wave IV, sex, race/ethnicity * Wave I – III predictors   Preliminary descriptives:   |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | CES-D4 | CES-D9 | Suicidal ideation | Any marijuana 30d | Binge last year | Risky drink 30 | DSM4 alcohol | DSM4 marijuana | DSM4 drug | | CES-D4 | 1.00 | 0.87 | 0.09 | 0.03 | -0.03 | 0.00 | 0.05 | 0.06 | 0.04 | | CES-D9 | 0.87 | 1.00 | 0.10 | 0.02 | -0.04 | 0.00 | 0.07 | 0.09 | 0.06 | | Suicidal ideation | 0.09 | 0.10 | 1.00 | 0.07 | 0.03 | 0.02 | 0.12 | 0.10 | 0.07 | | Any marijuana 30d | 0.03 | 0.02 | 0.07 | 1.00 | 0.21 | 0.14 | 0.12 | 0.06 | -0.02 | | Binge last year | -0.03 | -0.04 | 0.03 | 0.21 | 1.00 | 0.38 | 0.19 | 0.00 | -0.06 | | Risky drink 30 | 0.00 | 0.00 | 0.02 | 0.14 | 0.38 | 1.00 | 0.17 | -0.01 | 0.00 | | DSM4 alcohol | 0.05 | 0.07 | 0.12 | 0.12 | 0.19 | 0.17 | 1.00 | 0.26 | 0.20 | | DSM4 marijuana | 0.06 | 0.09 | 0.10 | 0.06 | 0.00 | -0.01 | 0.26 | 1.00 | 0.27 | | DSM4 drug | 0.04 | 0.06 | 0.07 | -0.02 | -0.06 | 0.00 | 0.20 | 0.27 | 1.00 |   Low correlation between Wave I depression and Wave IV diseases of despair. |
| Proposed Statistical Analysis:   * The Wave IV analytic sample will be split into three – training, validation, and hold out. * Machine learning algorithms will be used in the training set to predict each of the diseases of despair.   + Decision trees and random forest   + Lasso regression * We will evaluate the importance of cognitive and emotional despair measures, as well as relative importance of other features in the model. * Model performance will be evaluated with and without the sets of cognitive and emotional despair measures. * Machine learning results will be compared to traditional statistical methods. |
| Significance of work:  The rise in midlife mortality in the United States has been attributed to causes of death including suicide, poisoning/overdose, and alcohol-related liver disease. This cluster of causes, referred to as deaths of despair, are postulated to reflect an underlying feeling of hopelessness and resulting self-destructive behaviors. However, there is limited research testing whether individual feelings of cognitive and emotional despair are indeed predictive of such behaviors.  This project will provide a data-driven test of the despair hypothesis, predicting disease of despair behaviors using supervised machine learning approaches.  Brand et al. 2019 demonstrate the utility of machine learning approaches for identifying subgroups with heterogeneous treatment effects. |
| Primary References cited: |

**Confidentiality and Data Security Agreement**

|  |  |
| --- | --- |
| Provisional Paper Title | Predicting diseases of despair |
| Proposing Author | Lauren Gaydosh |
| Today’s Date | September 16, 2019 |

***Please keep one copy for your records***

(Please initial your agreement)

\_\_X\_\_ I am current on Human Subjects Training (CITI (www.citiprogram.org) or training in human subject protection through my post or courses.

\_PENDING\_\_\_ My project has IRB approval from my home institution.

\_X\_\_\_ I will treat all data as “restricted” and store in a secure fashion.

\_X\_\_\_ I will not share the data with anyone, including students or other collaborators not specifically listed on this prospectus.

\_X\_\_\_ Before submitting my paper to a journal, I will submit my draft manuscript and scripts for data checking, and my draft manuscript for co-author mock review, allowing three weeks.

**Signature:** 