Association of Depression and Political Climate in Patients Undergoing Treatment at Mental Health Facilities

Authors: Trent Yu and Matthew Naftilan

INTRODUCTION:

Depression is a widespread diagnosis with a reported 22.5 million adults experiencing a depressive episode in 2022.[1] The impact of depression on quality of life has led prior researchers to explore various factors that may contribute to and/or predict its development. One aspect that has been explored in limited fashion is the contribution of political affiliation on mental health. Previous findings have suggested that Democrats and Independents had worse mental health outcomes than Republicans at the onset of the COVID19 pandemic in 2020.[2] This disparity has led to some questions about the influence of the wider political environment on an individual. Previous studies have also shown that positive screening for depression did not differ across political affiliation; however, Republicans were shown to be less likely to receive treatment.[3]

While partisan differences were apparent in mental health outcomes early in the pandemic there are questions as to whether this difference would remain as the pandemic receded and a new administration was in place. Therefore, we sought to examine the frequency of and the covariate factors associated with a depression diagnosis amongst those who sought care for mental health disorders in 2022. We hypothesize that there will be increased association of depression in red states compared to blue states while living under the Democratic administration.

METHODS/RESULTS:

Data from the MH-CLD 2022 data source(https://www.samhsa.gov/data/data-we-collect/mh-cld-mental-health-client-level-data) served as the primary source of information. This data is published by the Substance Abuse and Mental Health Services Administration (SAMHSA) The report includes information on all clients who received services from a State Mental Health Agency(SMHA) during the 12 month period during 2022. The data includes demographics, regionality information, outcome measures, and substance use characteristics.

Initial cleaning of the data included creating dummy variables to parse categorical variables into binary outcome variables for age, ethnicity, employment status, education level, and marriage status.
However, in order to try to answer our question about partisan differences in depression we needed to introduce new features into the data set. We decided that given that regionality and state location data was included in the original set then we would split the states into Republican

(red) and Democrat (blue) states. We defined the political climate of each area based on the outcome of the 2020 presidential election at the state level. The breakdown of states is listed in table 1 below as well as the total number of cases for the red and blue states

Table 1:

Republican (Red) States (n=2,561,412)	Democrat (Blue) States(n=4,393,001)_
Alabama	Arizona
Alaska	California
Arkansas	Colorado
Florida	Connecticut
Idaho	Delaware
Indiana	District of Columbia
Iowa	Georgia
Kansas	Hawaii
Kentucky	Illinois
Louisiana	Maryland
Mississippi	Massachusetts
Missouri	Michigan
Montana	Minnesota
Nebraska	Nevada
North Carolina	New Hampshire
North Dakota	New Jersey
Ohio	New Mexico
Oklahoma	New York
South Carolina	Oregon
South Dakota	Pennsylvania
Tennessee	Vermont

Texas	Virginia
Utah	Washington
West Virginia	Wisconsin
Wyoming	Vermont

Given the large differences in care received across blue vs red states we wanted to ensure that the analysis would not simply be a reflection of access to care and therefore ran a simple cross tabulation of the state vs depression diagnosis. This revealed that there was approximately a 29% depression diagnosis rate amongst those who lived in red states and a 25.2% diagnosis rate amongst those in blue states. The similar rates of diagnosis gave us some confidence that it was not only an access to care issue that would be affecting the analysis.

We then chose to initially examine a logistic regression model in order to better understand the influence of living in a certain state on the diagnosis of depression. The results are listed in Figure 1 below.

Figure 1: Logistic Regression Summary

```
Coefficients:
                             Estimate Std. Error  z value Pr(>|z|)    
(Intercept)                 -3.737958   0.068285  -54.740  < 2e-16 ***
Political_Stance_Republican  0.098776   0.001895   52.132  < 2e-16 ***
GENDER_FEMALE                0.612988   0.024289   25.237  < 2e-16 ***
GENDER_Male                  0.084634   0.024303    3.482 0.000497 ***
MARSTAT_Never_married       -0.302867   0.002185 -138.621  < 2e-16 ***
MARSTAT_Now_married         -0.091055   0.003785  -24.054  < 2e-16 ***
MARSTAT_Divorced_widowed    -0.124551   0.003750  -33.213  < 2e-16 ***
MARSTAT_Separated           -0.112376   0.005736  -19.591  < 2e-16 ***
EMPLOY_Part_time             0.430863   0.004879   88.313  < 2e-16 ***
EMPLOY_Unemployed            0.348718   0.003055  114.162  < 2e-16 ***
EMPLOY_Full_time             0.467477   0.003990  117.150  < 2e-16 ***
EMPLOY_Not_in_labor_force    0.311180   0.002699  115.299  < 2e-16 ***
EDUC_Special_education      -0.519700   0.014681  -35.399  < 2e-16 ***
EDUC_0to8                    0.118533   0.003403   34.831  < 2e-16 ***
EDUC_9to11                   0.450416   0.003449  130.600  < 2e-16 ***
EDUC_12_or_GED               0.300190   0.002924  102.679  < 2e-16 ***
EDUC_More_than_12            0.313683   0.003499   89.661  < 2e-16 ***
Children                     1.724187   0.068228   25.271  < 2e-16 ***
Young_Adult                  2.281324   0.068240   33.431  < 2e-16 ***
Adult                        2.088542   0.068213   30.618  < 2e-16 ***
Middle_Aged_Adults           2.252993   0.068220   33.026  < 2e-16 ***
Senior                       2.232333   0.068279   32.694  < 2e-16 ***
Hispanic                     0.145292   0.003243   44.805  < 2e-16 ***
Non_Hispanic                 0.080043   0.002827   28.314  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8058502  on 6954412  degrees of freedom
Residual deviance: 7698181  on 6954389  degrees of freedom
  (3506 observations deleted due to missingness)
AIC: 7698229

Number of Fisher Scoring iterations: 5

> 
> # Exponentiate the coefficient for Political_Stance_Republican to interpret odds ratio
> exp(coef(model)["Political_Stance_Republican"])
Political_Stance_Republican 
                   1.103819 
```

Based on the analysis, there is a statistically significant relationship between depression and the political stance of the state. Patients from red states are more likely to have depression as one of their primary diagnoses. As you can see from the output above the coefficient of living in a red state is 0.098776 ($p \sim 0$) which would suggest that living in a republican state at this time increases the log-odds of depression. We also demonstrate that living in a republican state has an odds ratio of 1.103819 indicating an approximately 10% increase in having a depression diagnosis as compared to those patients who are not in a republican state with other characteristics held constant.

Given that there seems to be an increase in association between political state and depression diagnosis, we sought to further understand this relationship by looking at several machine learning models to discover significant features outside of political state that were likely to be associated with depression in individuals living in a republican state.

In order to do this, we first needed to take a sample of the total dataset as the original dataset was too large to run on the machines that we had available. We chose to select 200,000 cases at random as this would maintain a large enough sample without needing increased computing power. Clients from blue states were filtered out as we were only interested in examining the factors for those in the red states to better understand what the factors outside of political climate were affecting the diagnosis.

Three machine learning algorithms were selected to compare to the logistic regression that can natively handle categorical variables without requiring one-hot encoding to compare their performance. These methods were random forest, gradient boosting, and naïve bayes.

Initial runs demonstrated that the AUC score for the models was very high (nearly 1) indicating there were some variables which were highly correlated to the target variable. We inspected if there was any data leakage or overfitting happening in the dataset by closely examining the codebook and the definitions for each variable**.** After checking the feature importance and codebook, there were several variables(NUMMHS, ANXIETYFLG, MH1, MH2, MH3) that overlapped with the target variable which we removed from the dataset. After removing those variables, we proceeded to run the algorithms again and got the results below.

Figure 2:

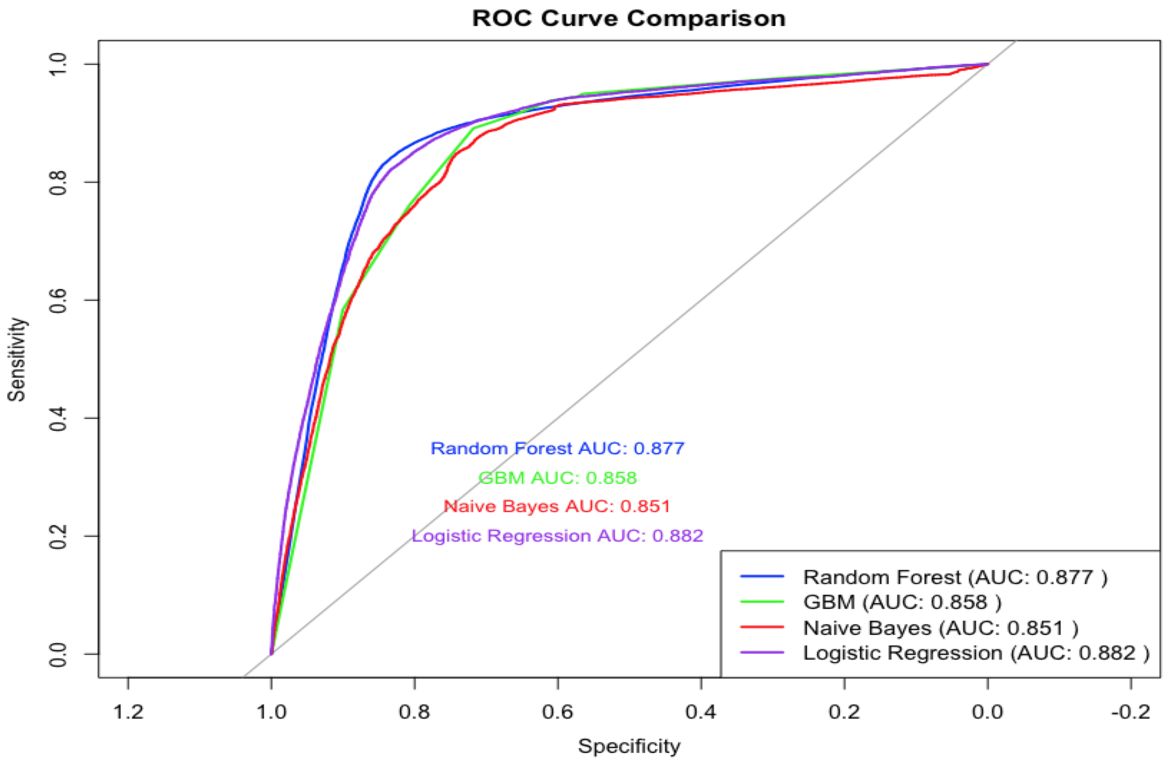
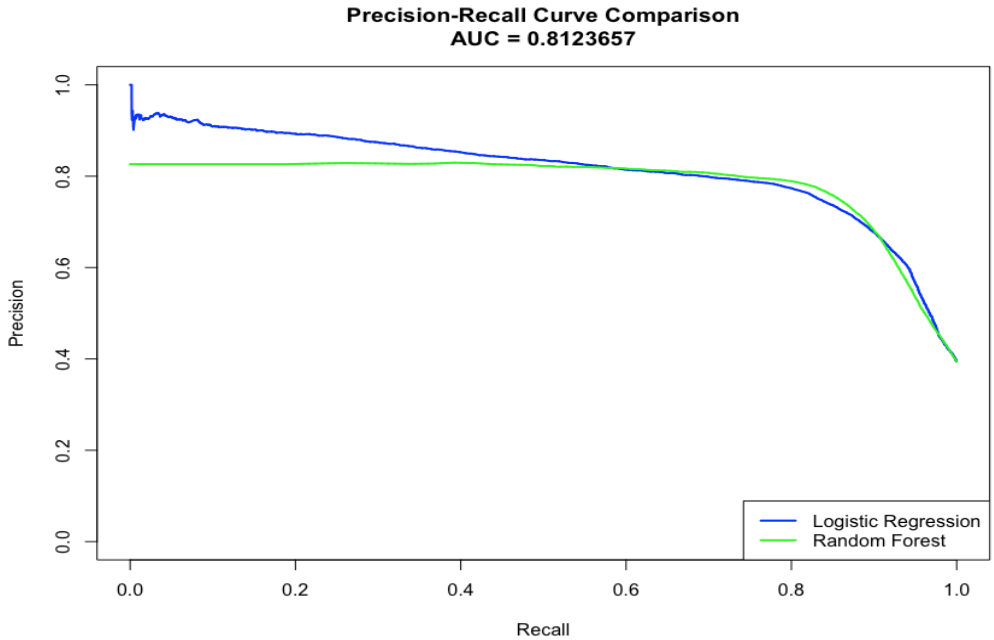


Figure 2 plots the AUC for each of the models. Each model demonstrated high AUC in the 0.85-0.88 range making selection based purely on AUC difficult. We then compared the top two models (Random Forest and Logistic Regression). using precision-recall curve comparison (Figure 3) to further analyze the two models' performance.

Figure 3:



The precision-recall curve demonstrated similar performance of the two models overall. However, since our question aims to identify all factors associated with depression for those living in red states we decided that recall was a more critical metric.

To prioritize recall, we lowered the cutoff from 0.5(default) to 0.3 and examined the confusion matrix for both models(Table 2).

Table 2:

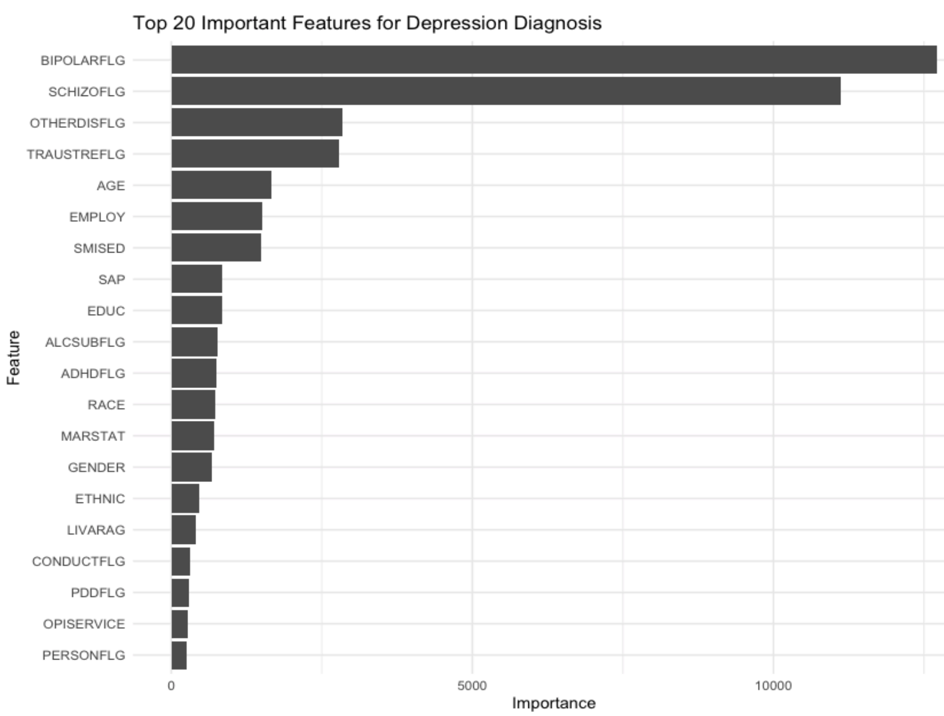
Logistic Regression. Accuracy = 0.79		**Actual**	
Predicted		0	1
	0	17358	1550
	1	6855	14236

Random Forest Accuracy = 0.834		**Actual**	
Predicted		0	1
	0	19919	2360
	1	4294	13426

The random forest model demonstrated a higher accuracy(0.834 vs 0.79) and therefore we decided to use the random forest model for the next piece of the analysis.

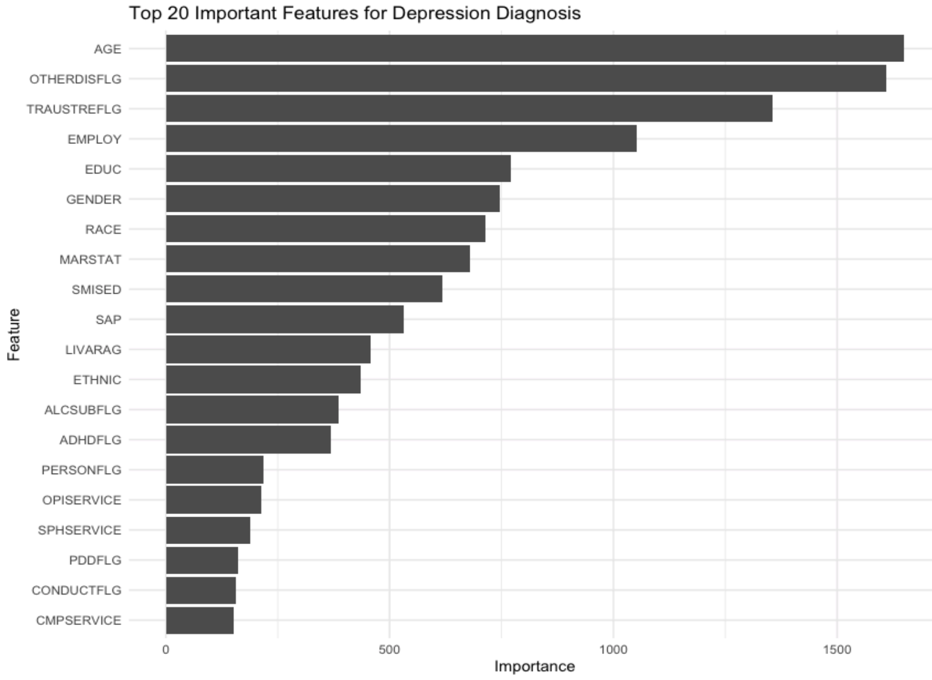
In order to examine more closely the factors associated with the depression diagnosis, the importance of the random forest model was calculated(figure 4). The output demonstrates that a coexisting diagnosis of Bipolar disorder (BIPOLARFLG) as well as schizophrenia disorder(SCHIZOFLG) are the most important features by multiple factors.

Figure 4:



Because these two were so dominant in the model we wanted to test how the model would perform when removing those two variables from the analysis(Figure 5)

Figure 5:



While removing those factors from the analysis demonstrates a smoother importance gradient the predictive ability of the model is greatly decreased. Table 3 shows the comparison of AUC and accuracies for both the logistic regression and random forest models.

Table 3:

		With Bipolar and Schizophrenia Diagnosis	Without Bipolar and Schizophrenia Diagnosis
Logistic Regression	AUC	0.882	0.65
	Accuracy	0.79	0.531
Random Forest	AUC	0.877	0.667
	Accuracy	0.834	0.64

The Final Piece of analysis we did was to explore whether adding the political state to the model would create a better prediction. In order to do this, we resampled from the large data set but did not restrict to only republican states. All other cleaning pieces were the same as what was outlined for the analysis of factors affecting those in republican states. The random forest model including the political stance demonstrated an AUC of 0.8536 and accuracy of 0.8005 while eliminating the state data resulted in an AUC of 0.8439 and accuracy of 0.7917.

Discussion/Limitations/Future Directions:

Our initial question of whether there were associations between living in a red or blue state and the diagnosis of depression was shown to be true. Given the prior research showing a higher incidence of depression amongst democrats at the beginning of covid our project lends some support to the notion that depression is at least moderately linked to the political environment. Further analysis including measures of access to care may help to control for some confounders. However, published access to care rankings indicate that in general blue states have much better access to mental health services including 13 of the top 15 in rankings of access to care (Figure 6).[4] Therefore, at face value we might expect the association between political belief and depression to be even stronger than what we demonstrated.

Figure 6:

Access to Care Ranking 2022 Rank	State	Political Stance
1	Vermont	Blue State
2	Massachusetts	Blue State
3	Maine	Blue State
4	Wisconsin	Blue State

5	Minnesota	Blue State
6	New Hampshire	Blue State
7	Rhode Island	Blue State
8	Pennsylvania	Blue State
9	Connecticut	Blue State
10	District of Columbia	Blue State
11	Washington	Blue State
12	Montana	Red State
13	Illinois	Red State
14	Maryland	Blue State
15	New York	Blue State

However, there are several limitations in our project that do not allow us to make larger conclusions. Primarily, the dataset used does not allow us to generalize the findings to a larger general population vs the population of people that have received care at a state mental health agency. It is likely different patterns may emerge if we were able to obtain data on the general population who then go on to receive a diagnosis of depression. Another major limitation of our project is the use of state residency as a surrogate for political belief given that we did not have the individual data for each client.

Time Statement:
In the preparation and completion of this project our group met together for well over 40 minutes.

References:

1. Abuse S, Administration MHS. Key substance use and mental health indicators in the United States: Results from the 2022 National Survey on Drug Use and Health. 2023.
2. Kwon S. The interplay between partisanship, risk perception, and mental distress during the early stages of the COVID-19 pandemic in the United States. *Psychology, Health & Medicine.* 2023;28(1):69-85.
3. Ettman CK, Hatton CR, Castrucci BC, Galea S. Mental Health and Mental Health Care Utilization Across Political Affiliation in U.S. Adults. *J Public Health Manag Pract.* 2024.
4. Reinert M, Fritze D, Nguyen T. The State of Mental Health in America 2022. *Mental Health America.* 2022.