Advances in Social Sciences Research Journal – Vol.6, No.5
Publication Date: May. 25, 2019
DoI:10.14738/assrj.65.6580.
Ganguli, D., Zhang, K., Zhu, X., Vaughan, G., & Berger, P. D. (2019). Opioid Usage In The United States. *Advances in Social Sciences Research Journal, 6(5)* 232-249.

Opioid Usage In The United States

Debjani Ganguli
Bentley University
Waltham, MA 02452, U.S.A.

Kevin Zhang
Bentley University
Waltham, MA 02452, U.S.A.

Xuemeng Zhu
Bentley University
Waltham, MA 02452, U.S.A.

Gregory Vaughan
Bentley University
Waltham, MA 02452, U.S.A.

Paul D. Berger
Bentley University
Waltham, MA 02452, U.S.A.

ABSTRACT
Our paper focuses on studying drivers of individual drug user relapse as well as broader systematic drug use patterns across states in the United States of America. For this study, two datasets were used; the first is a subset of the survey results from Dr. Miriam Boeri's co-authored study, "Older Drug Users: A Life Course Study of Turning Points in Drug Use [in a large Southeastern Metropolitan Area], 2009-2010". The dataset included variables such as the gender, race, education level, the age at which the respondents moved away from their guardian's home, and the age at which the respondent filled the survey. With this dataset, a logistic regression analysis is preformed to identify factors that may be associated with drug relapse. We found that the only "somewhat important" factor, as compared to all the factors studied, seemed to be "the current age of the respondent when they interviewed." The second dataset is composed of variables such as the state-wise population density, GDP, median household income, opioid prescription rate, death rate due to overuse of opioids, alcohol consumption rate, death rate due to alcohol overuse, death rate due to drug overuse, suicide death rate, and education attainment level across the United States.

Key Words: Opioid use, general drug use, multiple binary logistic regression analysis, cluster analysis

INTRODUCTION

A drug is any substance (with the exception of food and water) which, when taken into the body, alters the body's function either physically and/or psychologically. Drugs may be legal (e.g., alcohol, caffeine, and tobacco) or illegal (e.g., ecstasy, cocaine, and heroin). Psychoactive drugs affect the central nervous system and alter a person's mood, thinking, and behavior. Psychoactive drugs may be divided into four categories: depressants, stimulants, hallucinogens, narcotics, and "other." Opioids are a class of drugs in the narcotics family that include the illegal drug, heroin, synthetic opioids such as fentanyl, methamphetamine, crack cocaine and pain relievers available legally by prescription, such as oxycodone (OxyContin®), hydrocodone (Vicodin®), codeine, morphine and many others.

All opioids are chemically related and interact with opioid receptors on nerve cells in the body and brain. Opioid pain-relievers are generally safe when taken for a short time and as prescribed by a doctor, but because they produce euphoria in addition to pain relief, they can be misused (taken in a different way or in a larger quantity than prescribed, or taken without a doctor's prescription). Regular use—even as prescribed by a doctor—can lead to dependence and, when misused, opioid pain-relievers can lead to addiction, overdose incidents, and deaths. From 1999-2016, more than 350,000 people in the U.S. died from overdoses involving opioids. This included both prescription opioids and illicit opioids. This rise in opioid overdose deaths can be outlined in three distinct waves.

1. The first wave began with increased prescribing of opioids in the 1990s, with overdose deaths involving prescription opioids (natural and semi-synthetic opioids and methadone) increasing since at least 1999.
2. The second wave began in 2010, with rapid increases in overdose deaths involving heroin.
3. The third wave began in 2013, with significant increases in overdose deaths involving synthetic opioids, particularly those involving illicitly-manufactured fentanyl (IMF). The IMF market continues to change, and IMF can be found in combination with heroin, counterfeit pills, and cocaine.

As per reports released from the International Narcotics Control Board, the cost of drug abuse that is often cited is the loss in productivity that can occur when drug users are under the influence of drugs or are experiencing the consequences of their drug use (e.g., while in treatment, incarceration, or hospital.) Studies have put the cost of lost productivity borne by employers at tens of billions of (U.S.) dollars. The focus in this paper is in the opioids' overdose and dependency situation in the U.S.

Our paper focuses on describing our data and findings. We conducted both multiple logistic regression analysis and a cluster analysis. In the following sections, we have the characteristics of our two data sets, an in depth analysis of all testing and findings, summary & conclusions, which are followed then by the References and Acknowledgment sections.

DATA CHARACTERISTICS

Our primary dataset consisted of cross-sectional data. For the first data set we used, we selected all our variables, both the response/dependent variable and covariates, from Dr. Miriam Boeri's study (2012) published and available online on *Inter-University Consortium for Political and Social Research* called, "Older Drug Users: A Life Course Study of Turning Points in Drug Use [in a large Southeastern Metropolitan Area], 2009-2010". She conducted ninety-two face-to-face interviews with former and active drug (opioid) users from a Southeastern metropolitan area in the U.S. from 2009 and 2010. Her questionnaire had more than forty questions, but we included only those variables which we thought would address our proposed core question, the likelihood of a relapse in adults who have undertaken a rehabilitation treatment. All the respondents in the survey had received at least one kind of treatment for going off drug dependence. The variables are listed in Table 1. We follow Table 1 with a descriptive statistics section containing a variety of Figures and Tables indicating cross-tabulations of relevant variables. It will be seen that the database consists of 92 respondents.

Table 1: List of variables considered from Dr. Boeri's study

Variable	Variable Role	Type of data	Response options
Active User	Response/ Dependent variable	Qualitative	Yes or No (whether or not the respondent is an active (up to last six-months) user of an opioid)
Gender	Predictor / Independent variable	Qualitative	Male or Female
Race	Predictor / Independent variable	Qualitative	African American or White or Other (masked for confidentiality)
Move Age	Predictor / Independent variable	Quantitative	Discrete variable- the age at which the respondent left the home of their guardian
Current Age	Predictor / Independent variable	Quantitative	Discrete variable- the age at which the respondent interviewed (2009/2010)
Education	Predictor / Independent variable	Qualitative	College degree / Advanced degree, High school diploma / GED, Less than high school, Some college

Descriptive statistics:

Table of Active_User by Education

Frequency Percent Row Pct Col Pct	Education				
Active_User	College degree / Advanced degree	High school diploma / GED	Less than high school	Some college	Total
No	3 3.26 8.33 33.33	10 10.87 27.78 43.48	5 5.43 13.89 33.33	18 19.57 50.00 40.00	36 39.13
Yes	6 6.52 10.71 66.67	13 14.13 23.21 56.52	10 10.87 17.86 66.67	27 29.35 48.21 60.00	56 60.87
Total	9 9.78	23 25.00	15 16.30	45 48.91	92 100.00

Figure 1: Cross-tabulation of active use (AU) or not and education

Table of Active_User by GENDER

Frequency Percent Row Pct Col Pct	GENDER		
Active_User	Female	Male	Total
No	15 16.30 41.67 40.54	21 22.83 58.33 38.18	36 39.13
Yes	22 23.91 39.29 59.46	34 36.96 60.71 61.82	56 60.87
Total	37 40.22	55 59.78	92 100.00

Table of Active_User by RACE

Frequency Percent Row Pct Col Pct	RACE			
Active_User	African American	Other	White	Total
No	16 17.39 44.44 34.78	2 2.17 5.56 50.00	18 19.57 50.00 42.86	36 39.13
Yes	30 32.61 53.57 65.22	2 2.17 3.57 50.00	24 26.09 42.86 57.14	56 60.87
Total	46 50.00	4 4.35	42 45.65	92 100.00

Figure 2: Cross-tabulation of active use or not and race

Figure 3: Cross-tabulation of active use or not and gender

Table 1 of RACE by GENDER
Controlling for Active_User=No

Frequency / Percent / Row Pct / Col Pct

RACE	Female	Male	Total
African American	8 22.22 50.00 53.33	8 22.22 50.00 38.10	16 44.44
Other	0 0.00 0.00 0.00	2 5.56 100.00 9.52	2 5.56
White	7 19.44 38.89 46.67	11 30.56 61.11 52.38	18 50.00
Total	15 41.67	21 58.33	36 100.00

Table 2 of RACE by GENDER
Controlling for Active_User=Yes

Frequency / Percent / Row Pct / Col Pct

RACE	Female	Male	Total
African American	11 19.64 36.67 50.00	19 33.93 63.33 55.88	30 53.57
Other	1 1.79 50.00 4.55	1 1.79 50.00 2.94	2 3.57
White	10 17.86 41.67 45.45	14 25.00 58.33 41.18	24 42.86
Total	22 39.29	34 60.71	56 100.00

Figure 4 : Cross-tabulation of race and gender –AU=No

Figure 5 Cross-tabulation of race and gender - AU = Yes

The UNIVARIATE Procedure
Variable: MOVE_AGE

Moments

N	92	Sum Weights	92
Mean	19.1304348	Sum Observations	1760
Std Deviation	3.72168398	Variance	13.8509317
Skewness	1.60528256	Kurtosis	5.52411256
Uncorrected SS	34930	Corrected SS	1260.43478
Coeff Variation	19.4542572	Std Error Mean	0.38801238

Basic Statistical Measures

Location		Variability	
Mean	19.13043	Std Deviation	3.72168
Median	19.00000	Variance	13.85093
Mode	18.00000	Range	25.00000
		Interquartile Range	3.00000

The UNIVARIATE Procedure
Variable: Current_Age

Moments

N	92	Sum Weights	92
Mean	51.4673913	Sum Observations	4735
Std Deviation	4.57525856	Variance	20.9329909
Skewness	0.54006885	Kurtosis	-0.2027766
Uncorrected SS	245603	Corrected SS	1904.90217
Coeff Variation	8.88962593	Std Error Mean	0.47700368

Basic Statistical Measures

Location		Variability	
Mean	51.46739	Std Deviation	4.57526
Median	51.00000	Variance	20.93299
Mode	50.00000	Range	20.00000
		Interquartile Range	6.50000

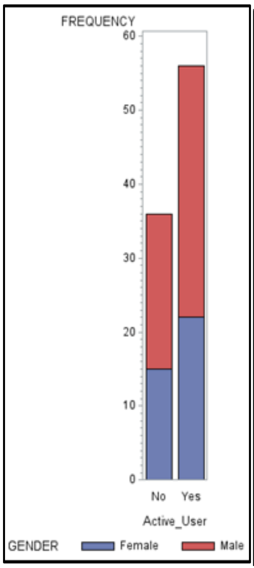
Note: The mode displayed is the smallest of 3 modes with a count of 9.

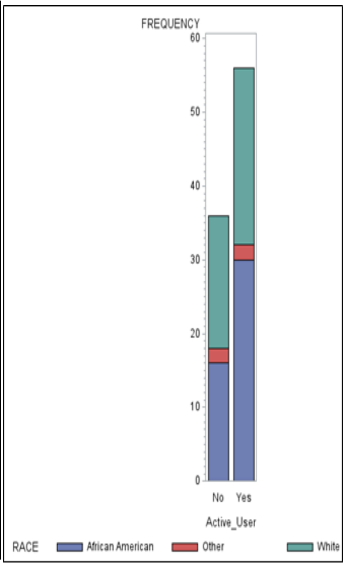
Figure 6: Basic statistics for Move_Age (moved out of care)

Figure 7 Basic statistics for Current_Age

Some Visual Summaries:

Please recall that "AU" stands for Active User or Not.





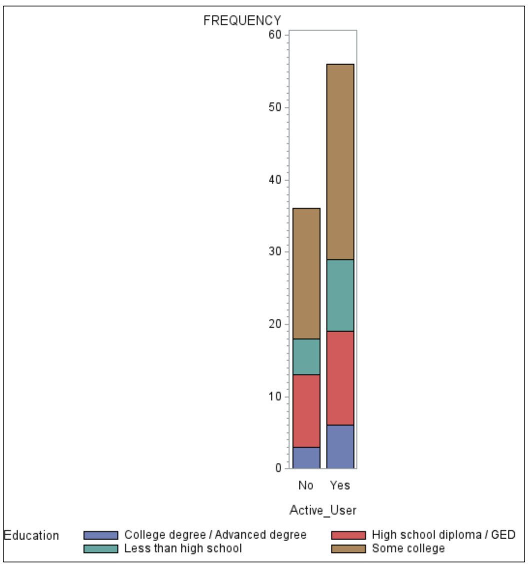


Figure 8: Gender by AU Figure 9: Race by AU Figure 10: Education by AU

From the summaries above we can see that men are more likely to be active users of drugs than women (Figure 8). Also, Figure 9 shows that older African Americans are more likely to be active drug users, while Figure 10 indicates that each education group in our sample is higher in the "Yes" category than the "No" category.

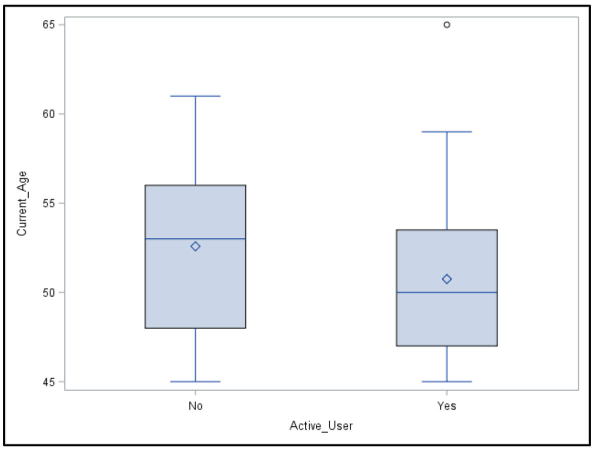


Figure 11: Box plot for current age vs. AU

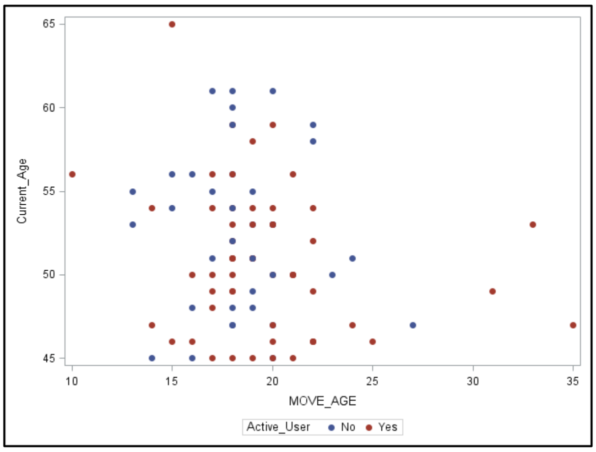


Figure 12: Graph of Move_Age vs. Current_Age, including AU status

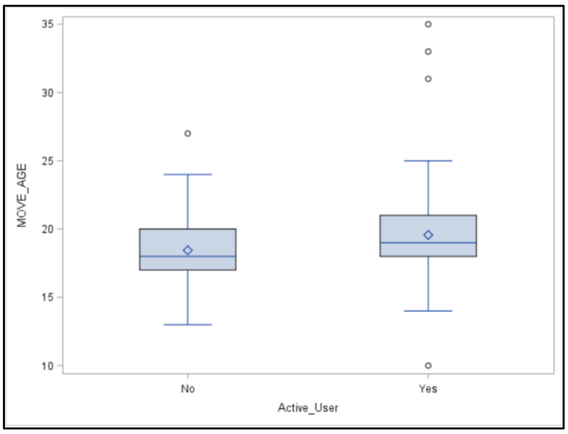


Figure 13: Box plot for move age vs. AU

From the above summaries, it appears that those respondents who moved out of care at an older age are more likely to be active drug users; it could be that the respondents moved out of care because they had been dependent on drugs for a long time and couldn't be independent. However, we can't demonstrate any causation, but, rather, can only identify association.

We composed our own data set to determine whether there is any underlying structure among the states in the U.S. when it comes to drug & opioid use and related substance (drugs, opioids, and alcohol) abuse deaths. Based on our domain knowledge, we determined that economic and socioeconomic factors impact psychological behavior of human beings. Hence, we collected related data available online from sources such as Statista, Centers for Disease Control and Prevention, and National Institute on Drug Abuse, among others, for the year 2015/2016. The detailed sources are available in the Reference section. One limitation is that there are some missing values, since, for a few independent variables, five U.S. states didn't have enough data points to view the specific independent variables to be reported with sufficient accuracy.

Table 2: List of variables additionally compiled (second dataset)

Variable Name	Variable Role	Type of data	Description
State	Variable with the observation ID label	Qualitative	51 US states (including DC but not including Puerto Rico)
GDP	Test variable	Quantitative	Gross domestic product (GDP) by state (millions of current dollars)- 2016
PopDensity	Independent variable used to determine clusters	Quantitative	Population Density (pop/ sq miles)- 2016
MedianIncome	Test variable	Quantitative	Median household income in 2016
AppAlcohol Comp	Independent variable used to determine clusters	Quantitative	Apparent Alcohol Consumption- 2016
AlcoholCompRate	Independent variable used to determine clusters	Quantitative	Alcohol consumption rate- 2015
AlcoholDeath	Independent variable used to determine clusters	Quantitative	Age-adjusted death rate due to alcohol poisoning- 2015
OpPrecRate	Independent variable used to determine clusters	Quantitative	Opioid prescription rate- 2016
OpDeathRate	Independent variable used to determine clusters	Quantitative	Opioid-Related Overdose Deaths/100,0001(2016)
DrugDeathRate	Independent variable used to determine clusters	Quantitative	Drug overdose death rate- 2016
SuicideDeathRate	Independent variable used to determine clusters	Quantitative	Suicide death rate- 2016
Education Level	Independent variable used to determine clusters	Quantitative	Educational attainment + Quality of education & attainment gap; measured on a scale of 100 points- 2016

Descriptive Statistics and Visual Summaries

Since, in Table 2, we have a relatively large number of variables, we decided to focus only on selected variables that we though related to our first data set, i.e., opioid prescription rate, opioid related overuse death, drug overuse death, and state wise GDP (an indicator of the economic performance of a state). Also, as mentioned in our introduction, productivity of a state is impacted by employee’s drug use; thus, we examined the correlation of state-wise GDP with respect to opioid use/death rate.

The UNIVARIATE Procedure
Variable: GDP

Moments			
N	51	Sum Weights	51
Mean	361888.118	Sum Observations	18456294
Std Deviation	459966.022	Variance	2.11569E11
Skewness	3.10300023	Kurtosis	11.7264402
Uncorrected SS	1.72576E13	Corrected SS	1.05784E13
Coeff Variation	127.10172	Std Error Mean	64408.126

Basic Statistical Measures			
Location		Variability	
Mean	361888.1	Std Deviation	459966
Median	209716.0	Variance	2.11569E11
Mode	.	Range	2571580
		Interquartile Range	403322

Figure 14: Basic statistics for GDP

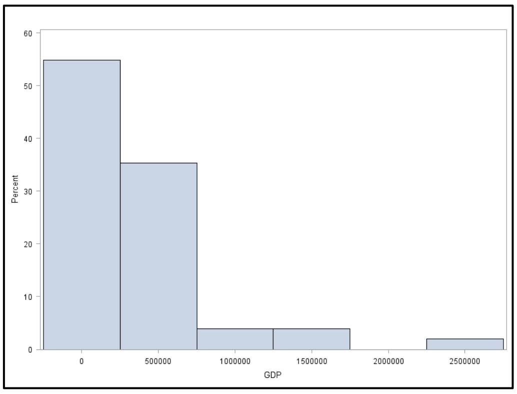


Figure 15: Histogram of GDP distribution

The UNIVARIATE Procedure
Variable: OpPrecRate

Moments			
N	51	Sum Weights	51
Mean	70.8176471	Sum Observations	3611.7
Std Deviation	19.3736492	Variance	375.338282
Skewness	0.58490558	Kurtosis	0.13649366
Uncorrected SS	274539.01	Corrected SS	18766.9141
Coeff Variation	27.3570924	Std Error Mean	2.71285351

Basic Statistical Measures			
Location		Variability	
Mean	70.81765	Std Deviation	19.37365
Median	66.90000	Variance	375.33828
Mode	.	Range	88.50000
		Interquartile Range	22.10000

Figure 16: Basic statistics for opioid prescription rate

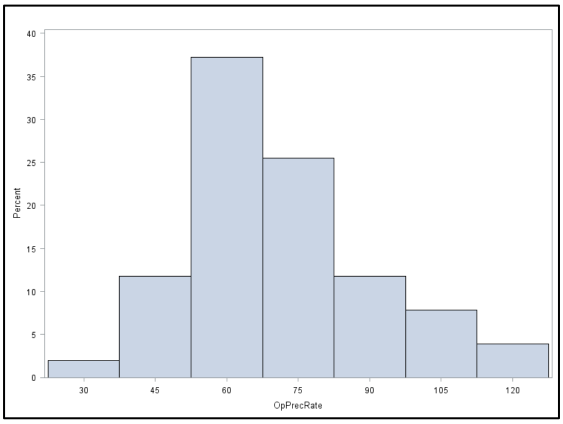


Figure 17: Histogram of opioid prescription rate distribution

The UNIVARIATE Procedure
Variable: OpDeathRate

Moments			
N	51	Sum Weights	51
Mean	14.7823529	Sum Observations	753.9
Std Deviation	9.18310853	Variance	84.3294824
Skewness	1.0987877	Kurtosis	0.92604666
Uncorrected SS	15360.89	Corrected SS	4216.47412
Coeff Variation	62.1221031	Std Error Mean	1.2858924

Basic Statistical Measures			
Location		Variability	
Mean	14.78235	Std Deviation	9.18311
Median	13.30000	Variance	84.32948
Mode	4.90000	Range	41.00000
		Interquartile Range	10.90000

Figure 18: Basic statistics for opioid death rate

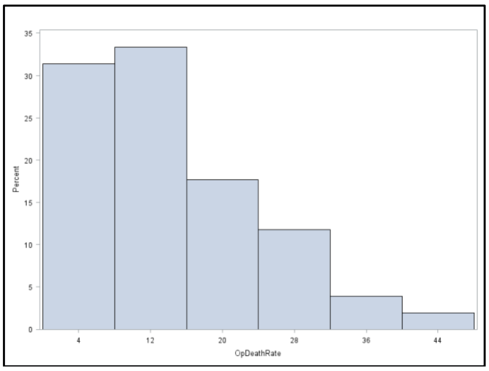


Figure 19: Histogram of opioid prescription rate distribution

The UNIVARIATE Procedure
Variable: DrugDeathRate

Moments			
N	50	Sum Weights	50
Mean	20.964	Sum Observations	1048.2
Std Deviation	9.37207663	Variance	87.8358204
Skewness	1.01891159	Kurtosis	1.24077887
Uncorrected SS	26278.42	Corrected SS	4303.9552
Coeff Variation	44.7055745	Std Error Mean	1.32541179

Basic Statistical Measures			
Location		Variability	
Mean	20.96400	Std Deviation	9.37208
Median	19.50000	Variance	87.83582
Mode	10.60000	Range	45.60000
		Interquartile Range	11.20000

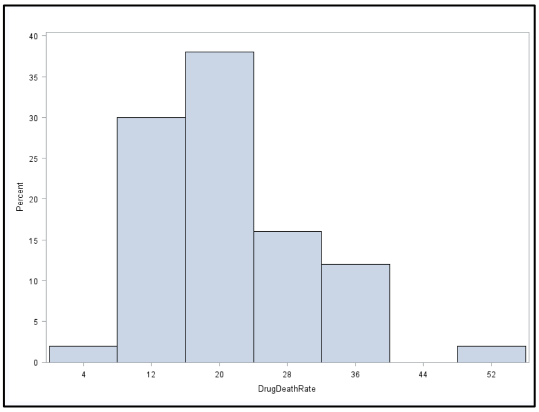


Figure 20: Basic statistics for drug death rate

Figure 21: Histogram of drug related death rate distribution

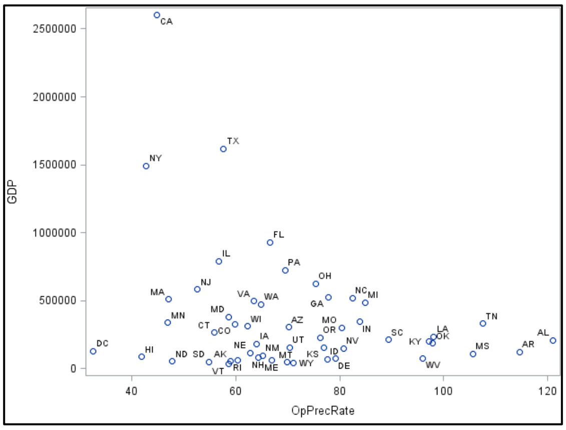


Figure 22: Scatter plot of GDP vs opioid prescription rate identified state wise

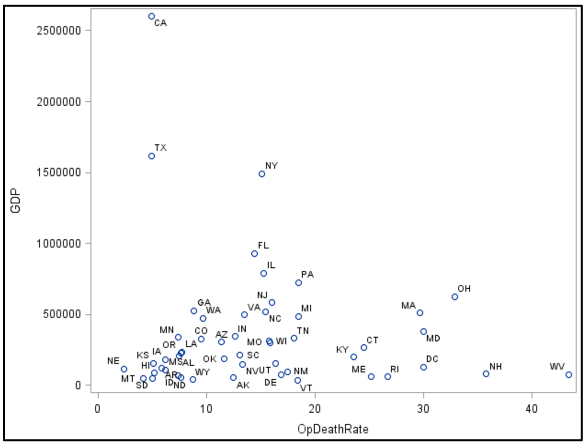


Figure 23: Scatterplot of GDP vs opioid death rate identified state wise

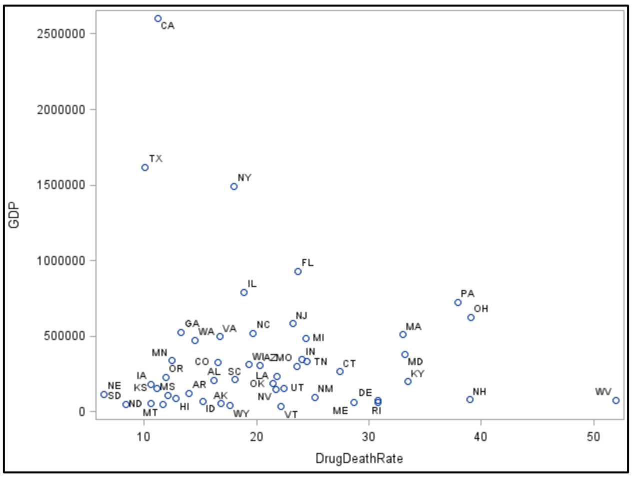


Figure 24: Scatterplot of GDP vs drug death rate identified state wise

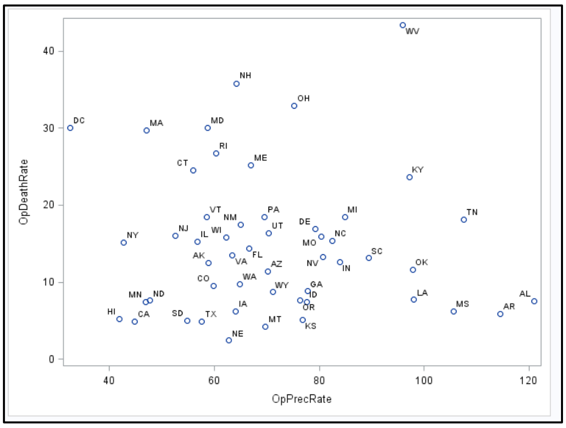


Figure 25: Scatterplot of opioid prescription rate vs opioid death rate identified state wise

From the scatter plots (Figures 22, 23, and 24) above, we can identify a trend that states with lower GDPs, tend to have higher opioid prescription rates, opioid related death rates, and drug overuse related death rates.

We have used this as a precursor to our later-described cluster analysis.

IN DEPTH ANALYSIS

Logistic Regression Model

For the purpose of the paper, our primary focus was to examine whether the variables selected would help build a good regression model that could help researchers understand the relapse pattern of the drugs (former and recently-turned active users) users (in older patients), and eventually, help predict the likelihood of relapse in older active/former drug users/abusers. Since our dependent variable is a qualitative binary response, and our predictors are a mix of both quantitative and qualitative variables, we implemented logistic regression for our analysis.

The model specification and assumptions are summarized as follows:

$$\boxed{Y_i \overset{indp}{\sim} Binomial(n = 1, p = f(E_i))}$$

where

$$E_i = \beta_0 + \beta_1 Gender_i + \beta_{21} Race_{1i} + \beta_{22} Race_{2i} + \beta_{31} Education_{1i} + \beta_{32} Education_{2i} + \beta_{33} Education_{3i} + \beta_4 MoveAge_i + \beta_5 CurrentAge_i$$

- ***f*** is the logistic link function; i.e. $f(E) = e^{(E)} / (1+e^{(E)})$
- $\boldsymbol{Y_i}$ is whether or not the i^{th} respondent is an active user (in the last six months) of drugs (methamphetamine, cocaine, crack, or heroin).
- $\boldsymbol{Gender_i}$ is a (0, 1) dummy variable for the gender of the i^{th} respondent recorded and where "Male" is the reference/base/dummy level
- $\boldsymbol{Race_{1i}}$ and $\boldsymbol{Race_{2i}}$ are (0, 1) dummy variables for the race of the i^{th} respondent recorded and where "White" is the reference/base/dummy level, $\boldsymbol{Race_{1i}}$ refers to Black and $\boldsymbol{Race_{2i}}$ refers to "other."
- $\boldsymbol{Education_{1i}}$, $\boldsymbol{Education_{2i}}$ *and* $\boldsymbol{Education_{3i}}$ are (0, 1) dummy variables for the education level of the i^{th} respondent recorded and where "Some College" is the reference/base/dummy level, $\boldsymbol{Education_{1i}}$ refers to College degree, $\boldsymbol{Education_{2i}}$ refers to High School Diploma/GED, and $\boldsymbol{Education_{3i}}$ refers to less than high school.
- $\boldsymbol{CurrentAge_i}$ is the age of i^{th} respondent when he/she recorded the interview (2009/2010)
- $\boldsymbol{MoveAge_i}$ is the age of i^{th} respondent when he/she left the home of their parents/guardians/caretakers

It should be noted that the choice of reference/base/dummy levels here is arbitrary; the estimates for the dummy values and the intercept may change if we used different reference/base/dummy levels, but the ultimate interpretations (and significance) would, of course, be the same. Also, the event (i.e., the "1") selected in our logistic regression model is "Yes." In this model, we have included the main effects since an initial look at our various visual summaries didn't indicate any evidence of interaction. Additionally, if indeed we are missing any complex predictors, such as quadratic terms or an interaction term, we may be able to identify it when we do a model validity check.

Model Utility

Misclassification Rate:

We see from Figure 26 that the optimal % correct that the model produces is 60.9% (i.e., a misclassification rate of 100 – 60.9 = 39.1%) when a cutoff of 0.12 is used.

Classification Table

Prob Level	Correct		Incorrect		Percentages				
	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.120	56	0	36	0	60.9	100.0	0.0	39.1	.
0.140	55	0	36	1	59.8	98.2	0.0	39.6	100.0
0.160	55	0	36	1	59.8	98.2	0.0	39.6	100.0
0.180	55	0	36	1	59.8	98.2	0.0	39.6	100.0
0.200	55	0	36	1	59.8	98.2	0.0	39.6	100.0
0.220	55	0	36	1	59.8	98.2	0.0	39.6	100.0
0.240	55	0	36	1	59.8	98.2	0.0	39.6	100.0
0.260	55	0	36	1	59.8	98.2	0.0	39.6	100.0
0.280	55	0	36	1	59.8	98.2	0.0	39.6	100.0
0.300	55	0	36	1	59.8	98.2	0.0	39.6	100.0
0.320	52	0	36	4	56.5	92.9	0.0	40.9	100.0
0.340	52	0	36	4	56.5	92.9	0.0	40.9	100.0
0.360	51	1	35	5	56.5	91.1	2.8	40.7	83.3
0.380	51	1	35	5	56.5	91.1	2.8	40.7	83.3
0.400	48	1	35	8	53.3	85.7	2.8	42.2	88.9
0.420	47	3	33	9	54.3	83.9	8.3	41.3	75.0
0.440	46	4	32	10	54.3	82.1	11.1	41.0	71.4
0.460	44	4	32	12	52.2	78.6	11.1	42.1	75.0
0.480	42	6	30	14	52.2	75.0	16.7	41.7	70.0
0.500	41	8	28	15	53.3	73.2	22.2	40.6	65.2
0.520	40	11	25	16	55.4	71.4	30.6	38.5	59.3
0.540	37	11	25	19	52.2	66.1	30.6	40.3	63.3
0.560	36	12	24	20	52.2	64.3	33.3	40.0	62.5
0.580	33	15	21	23	52.2	58.9	41.7	38.9	60.5
0.600	32	18	18	24	54.3	57.1	50.0	36.0	57.1
0.620	31	20	16	25	55.4	55.4	55.6	34.0	55.6
0.640	27	22	14	29	53.3	48.2	61.1	34.1	56.9
0.660	23	23	13	33	50.0	41.1	63.9	36.1	58.9
0.680	20	24	12	36	47.8	35.7	66.7	37.5	60.0
0.700	15	24	12	41	42.4	26.8	66.7	44.4	63.1

From our classification table, on the left, we see that the cutoff level of 0.12 results in a Sensitivity of 100%. So, at this cut off, we see that the model is actually preferring to identify “events" (Active user = “Yes”) correctly over identifying “non-events" (Active user = “No”) correctly because the sensitivity (the % of “events" correctly identified) is 100% whereas the specificity (the % “non-events" correctly identified) is 0%.

0.720	11	25	11	45	39.1	19.6	69.4	50.0	64.3
0.740	8	25	11	48	35.9	14.3	69.4	57.9	65.8
0.760	8	26	10	48	37.0	14.3	72.2	55.6	64.9
0.780	6	26	10	50	34.8	10.7	72.2	62.5	65.8
0.800	3	29	7	53	34.8	5.4	80.6	70.0	64.6
0.820	2	33	3	54	38.0	3.6	91.7	60.0	62.1
0.840	2	36	0	54	41.3	3.6	100.0	0.0	60.0
0.860	2	36	0	54	41.3	3.6	100.0	0.0	60.0
0.880	1	36	0	55	40.2	1.8	100.0	0.0	60.4
0.900	1	36	0	55	40.2	1.8	100.0	0.0	60.4
0.920	0	36	0	56	39.1	0.0	100.0	.	60.9

Figure 26: Classification Table

ROC Curve and AUC:

Next, we consider, in Figure 27, the receiver operating characteristic (ROC) curve and corresponding area under the ROC curve:

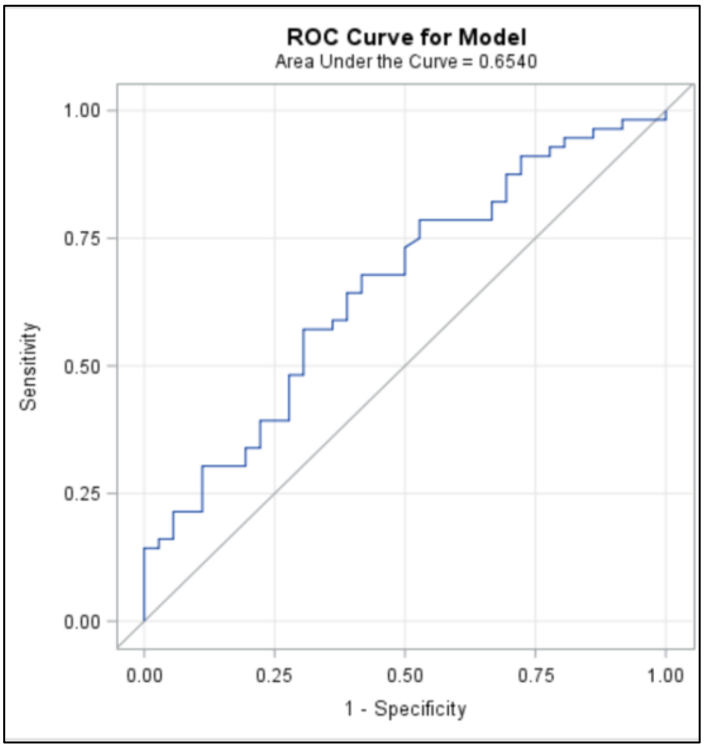


Figure 27: ROC Curve

From the area under the ROC curve of (0.654), we see that the model performs just slightly better than random guessing (half of the AUC= 0.5). While the model outperforms random guessing across most of the specificity values, it is generally just a relatively small improvement. Based on these measures, it seems evident that the model does provide predictive benefit, albeit minimal, indicating either that there is additional information that could be gathered to improve the model (i.e., additional variables) or the task of predicting whether older adults will have a relapse in drug/substance addiction can be too challenging to capture.

Model Validity

To assess the logistic model's validity, we checked “Goodness of Fit” (i.e., “lack of fit"). We didn’t need to check for Independence of observations since we didn’t have any longitudinal data. We looked for evidence of “lack of fit," using the Hosmer-Lemeshow Test.

Ho: The logistic model used accurately describes the data
Ha: The logistic model used DOES NOT accurately describe the data

Our results are shown in Figure 28.

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
1.8623	8	0.9849

Figure 28: Hosmer and Lemeshow test results

From Figure 28, we see that at the significance level of 0.05, the test (overwhelmingly) fails to rejects the null hypothesis, that the model describes the data well. This conclusion, coupled with the model's (only) “slightly better” predictive performance noted earlier, further supports the notion that the task of predicting whether an older adult will have a substance use relapse or not is a challenging problem.

Full logistic regression output

While our model has been shown to not be "all that predictive," we show the complete logistic regression output in Figure 29. While it is possible that a particular variable is significant, while the overall model test is not, that is not the case here; none of our predictors are statistically significant at the 5% significance level.

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	4.6663	3.4869	1.7908	0.1808
GENDER	Female	1	-0.6221	0.5288	1.3841	0.2394
GENDER	Male	0	0	.	.	.
RACE	African American	1	0.3745	0.4761	0.6186	0.4316
RACE	Other	1	-0.4822	1.0994	0.1924	0.6609
RACE	White	0	0	.	.	.
Education	College degree / Advanced degree	1	0.6861	0.8361	0.6734	0.4119
Education	High school diploma / GED	1	-0.3507	0.5526	0.4028	0.5256
Education	Less than high school	1	0.0686	0.6844	0.0100	0.9202
Education	Some college	0	0	.	.	.
MOVE_AGE		1	0.0634	0.0729	0.7568	0.3843
Current_Age		1	-0.1030	0.0553	3.4664	0.0626

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
GENDER Female vs Male	0.537	0.190	1.513
RACE African American vs White	1.454	0.572	3.697
RACE Other vs White	0.617	0.072	5.325
Education College degree / Advanced degree vs Some college	1.986	0.386	10.226
Education High school diploma / GED vs Some college	0.704	0.238	2.080
Education Less than high school vs Some college	1.071	0.280	4.096
MOVE_AGE	1.065	0.924	1.229
Current_Age	0.902	0.809	1.005

Figure 29: Full output for logistic regression

Ultimately, the findings of the analysis of the model's coefficients further support the earlier hypothesis about the difficulty of accurately predicting whether or not an older adult will have a drug relapse or not. At the very least, we can say this data does not provide enough evidence that the variables we studied greatly improve our ability to predict the likelihood of a relapse and that there is likely a need for additional variables which may provide more information and allow the model to make more accurate predictions.

Cluster Analysis

We then conducted a Cluster Analysis. We conducted a hierarchical cluster analysis using an average linkage in SAS with the all variables excluding Median Income and GDP, which we kept aside as test variables.

We got the following Dendrogram, as shown in Figure 30.

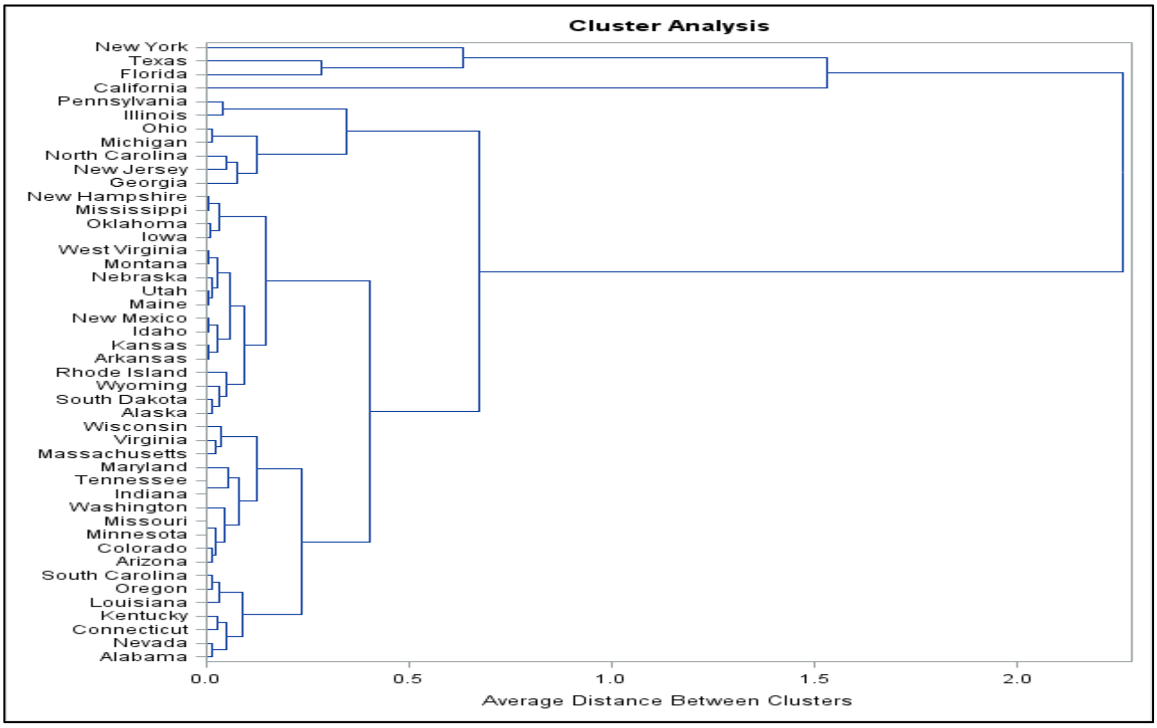


Figure 30: Dendrogram

In examining the plot for Jumps, we saw one clear jump (from three to two clusters). So, based on the jumps, we concluded that there is only one candidate clustering of three clusters. We then looked to assign labels to these clusters of U.S. states based on the typical values of the variables (excluding Median Income and GDP) for each cluster. These labels were attempted to ideally describe a typical state of each cluster. Our conclusions are highlighted in Figure 31.

The MEANS Procedure

CLUSTER	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
1	42	PopDensity	42	163.5058258	219.7955238	1.1144271	1029.33
		AppAlcoholComp	42	9631.52	6302.88	1271.00	25302.00
		AlcoholCompRate	42	2.4333333	0.5539247	1.3400000	4.7600000
		AlcoholoDeath	42	11.2833333	7.7447360	5.3000000	46.5000000
		OpPrecRate	42	74.7619048	18.0782035	46.9000000	121.0000000
		OpDeathRate	42	15.1547619	9.3996939	2.4000000	43.4000000
		DrugDeathRate	42	21.6380952	9.6055104	6.4000000	52.0000000
		SuicideDeathRate	42	16.2976190	4.5166083	7.2000000	25.9000000
		EducationPercentage	42	49.1609524	15.0479779	21.0600000	81.9200000
		GDP	42	272802.52	199940.61	37858.00	791608.00
		MedianIncome	42	57074.02	9894.70	41754.00	78945.00
2	3	PopDensity	3	260.5400000	137.8993731	103.8900000	363.6000000
		AppAlcoholComp	3	44910.67	7733.72	36604.00	51903.00
		AlcoholCompRate	3	2.4033333	0.2218859	2.2200000	2.6500000
		AlcoholoDeath	3	6.9666667	1.7156146	5.4000000	8.8000000
		OpPrecRate	3	55.6333333	12.0707636	42.7000000	66.6000000
		OpDeathRate	3	11.4666667	5.6976603	4.9000000	15.1000000
		DrugDeathRate	3	17.2666667	6.8295925	10.1000000	23.7000000
		SuicideDeathRate	3	11.5666667	3.0827477	8.1000000	14.0000000
		EducationPercentage	3	48.3233333	9.1653823	39.1100000	57.4400000
		GDP	3	1343872.00	366876.97	926817.00	1616801.00
		MedianIncome	3	56778.00	6027.32	50860.00	62909.00
3	1	PopDensity	1	240.0600000	.	240.0600000	240.0600000
		AppAlcoholComp	1	75148.00	.	75148.00	75148.00
		AlcoholCompRate	1	2.3300000	.	2.3300000	2.3300000
		AlcoholoDeath	1	9.9000000	.	9.9000000	9.9000000
		OpPrecRate	1	44.8000000	.	44.8000000	44.8000000
		OpDeathRate	1	4.9000000	.	4.9000000	4.9000000
		DrugDeathRate	1	11.2000000	.	11.2000000	11.2000000
		SuicideDeathRate	1	10.5000000	.	10.5000000	10.5000000
		EducationPercentage	1	50.2800000	.	50.2800000	50.2800000
		GDP	1	2602672.00	.	2602672.00	2602672.00
		MedianIncome	1	67739.00	.	67739.00	67739.00

Figure 31: Means procedure for clusters

We make the following observations:

- The first cluster has the highest percentage of suicide death rate, opioid prescription rate, opioid related overuse death rate, and drug related death rates, with the lowest population density.
- The second cluster has the highest population density, with higher suicide death rate, opioid prescription rate, opioid related overuse death rate, and drug related death as compared to the third cluster.
- The third cluster is an outlier cluster with only one observation. From our dendrogram in Figure 30, we find this includes only California. This cluster has a higher population density than cluster one, and the lowest suicide death rate, opioid prescription rate, opioid related overuse death rate, and drug related death as compared to both other clusters. And, California differs from the rest of the U.S. in many ways!!!!

Based on our observations, we propose the following cluster labels:

- The first cluster represents states where there is a major opioid and drug related issue, along with suicide cases. We can label this cluster as Opioid Priority 1 cluster.
- The second cluster represents states where there is a medium level of opioid and drug related issues, along with suicide cases as compared to the other clusters. We can label this cluster as Opioid Priority 2 cluster.
- The third cluster represents a state (the largest in the U.S) where the level of opioid and drug related issues, along with suicide cases, are much lower than the other clusters. We can label this cluster as Opioid Low-Priority cluster.

Looking back at our Means table, Figure 31, and now examining the validation variables, GDP and Median Income, we can note that the states in Cluster 1 have the lowest amount of mean GDP, whereas the one state in Cluster three has the highest GDP. Also, the mean median income in the three clusters appears to be quite different.

To confirm that the differences in the means are statistically significant, we conducted an ANOVA test at the 5% Significance level for both the test variables, GDP and the Median Income. Results are shown in Figures 32 and 33.

Class Level Information		
Class	Levels	Values
CLUSTER	3	1 2 3

Number of Observations Read	51
Number of Observations Used	46

The SAS System

The ANOVA Procedure

Dependent Variable: GDP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	8.201911E12	4.1009555E12	92.41	<.0001
Error	43	1.9082236E12	44377293791		
Corrected Total	45	1.0110135E13			

R-Square	Coeff Var	Root MSE	GDP Mean
0.811256	53.56139	210659.2	393304.2

Source	DF	Anova SS	Mean Square	F Value	Pr > F
CLUSTER	2	8.201911E12	4.1009555E12	92.41	<.0001

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
CLUSTER	3	1 2 3

Number of Observations Read	51
Number of Observations Used	46

The SAS System

The ANOVA Procedure

Dependent Variable: MedianIncome

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	111926610	55963305	0.59	0.5594
Error	43	4086767867	95041113		
Corrected Total	45	4198694477			

R-Square	Coeff Var	Root MSE	MedianIncome Mean
0.026657	17.01778	9748.903	57286.57

Source	DF	Anova SS	Mean Square	F Value	Pr > F
CLUSTER	2	111926610.3	55963305.2	0.59	0.5594

Figure 32: ANOVA results for GDP **Figure 33: ANOVA results for Median Income**

We see from the ANOVA table in Figure 32, that the ANOVA test rejects at the 5% significance level and thus, we conclude that the means truly are different and the clusters are reasonable based on our validation variable, GDP of each state. However, we also see from the ANOVA table in Figure 33, that the ANOVA test fails to reject at the 5% significance level (p = .5594) and thus, we cannot conclude that the means are different and the clusters may not be reasonable based on the validation variable, Median Income of each state. Overall, the results from our Cluster Analysis confirm the descriptive and visual summaries, where we based our plots on the correlation of opioid prescription rate, opioid related death, and drug overuse death rate with respect to GDP (based on domain knowledge.)

SUMMARY AND CONCLUDING REMARKS

In conclusion, we would like to reiterate our findings.

Our logistic regression model was shown to be of only minimal help in predicting the response variable, and none of our predictors were found to be significant at 5% significance level – although one variable was "close" (p = .0620 for Current_Age.) In essence, to identify older adults who are more likely to have a relapse in their use of opioid drug abuse, predictors such as gender, race, education level, the age at which they moved away from their parent's home and current age don't appear to provide practical material benefit. Other predictors should be looked into, such as access to health care insurance, support from family, job satisfaction, and, likely, others we have not thought about. While we examined the literature, none of the authors have direct expertise in this area.

Interestingly enough, the results of our cluster analysis, which used a dataset that was different than that of the logistic-regression study, show that the states of the U.S. can potentially be usefully divided into three clusters. This may allow us to dive more deeply into how rampant drug and opioid use is in each cluster and the associated variable values.

We note from the results of our cluster analysis that in the states of Cluster 2: Texas, New York, and Florida, the GDP is generally much higher, and opioid prescription rate, deaths related to opioid or drug overuse, and suicide deaths, are much lower than the rest of the forty-two states included in Cluster 1. Cluster one has most of the Southeastern states, barring Florida. Cluster 1 states have lower population densities, lower GDP and median income, but have much larger rates of higher opioid prescription rates and associated deaths, with respect to opioid and drug over use. Indeed, it is also true that suicide rates are higher. And, in Cluster 3, we see that in California, not only is that one state's GDP and median income higher than the corresponding values for the other two clusters, but also, opioid and drug related issues are much lower, as well as a lower death rate due to suicide and opioid/drug overuse. We are assuming that none of our conclusions are affected by difference in how states "measure and report" the variables we used in this paper.

We suspect that economic, socio-economic, and emotional factors play a role in these trends. These are things that could be looked into in the future.

References

Boeri, Miriam, and Whalen, Thor. Older Drug Users: A Life Course Study of Turning Points in Drug Use [in a large Southeastern Metropolitan Area], 2009-2010 (2012). Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2012-07-31. https://doi.org/10.3886/ICPSR34296.v1

CDCP contributor. (2015, January 6). Alcohol Poisoning Deaths infographic. Retrieved on December 07, 2018, from https://www.cdc.gov/vitalsigns/alcohol-poisoning-deaths/infographic.html#infographic

CDCP contributor. (2018, March 28). *Alcohol and Public Health.* Retrieved on December 06, 2018, from https://www.cdc.gov/alcohol/data-stats.htm

CDCP contributors. (2017, July 31). *U.S. State Prescribing Rates, 2016.* Retrieved on December 06, 2018, from https://www.cdc.gov/drugoverdose/maps/rxstate2016.html

CDCP contributors. (2018, January 10). *Drug Overdose Mortality by State.* Retrieved on December 06, 2018, from https://www.cdc.gov/nchs/pressroom/sosmap/drug_poisoning_mortality/drug_poisoning.htm

CDCP contributors. (2018, January 11). *Suicide Mortality by State.* Retrieved on December 06, 2018, from https://www.cdc.gov/nchs/pressroom/sosmap/suicide-mortality/suicide.htm

Drug Aware. (2018). *What is a drug?.* Retrieved on December 13, 2018 from https://drugaware.com.au/getting-the-facts/faqs-ask-a-question/what-are-drugs/#what-is-a-drug

Guzman, Gloria G. "Household Income: 2016". *U.S. Department of Commerce.* (2017). Census.gov. Web. Retrieved on December 07, 2018, from https://www.census.gov/content/dam/Census/library/publications/2017/acs/acsbr16-02.pdf

Haughwout, Sarah P. and Slater, Megan E. (2018). *Apparent Per Capita Alcohol Consumption: National, State, And Regional Trends, 1977–2016.* Retrieved on December 6, 2018, from https://pubs.niaaa.nih.gov/publications/surveillance110/CONS16.pdf

International Narcotics Control Board. (2013). Annual Reports: Chapter 1 (Pg. 3). Retrieved on December 14 from *https://www.incb.org/documents/Publications/AnnualReports/Thematic_chapters/English/AR_2013_E_Chapter_I.pdf*

Missouri DHSS contributor. (). *Age-Adjusted Rate.* Retrieved on December 13, 2018 from https://health.mo.gov/data/mica/CDP_MICA/AARate.html

NIDA. (2018). *Opioids*. Retrieved on 2018, December 13 from https://www.drugabuse.gov/drugs-abuse/opioids

NIDA. (2018, February). *Opioid Summaries by State*. Retrieved on December 06, 2018, from https://www.drugabuse.gov/drugs-abuse/opioids/opioid-summaries-by-state

Statista contributor. (2018, April). *Alcohol consumption per capita from all beverages in the U.S. in 2016, by state (in gallons of ethanol)*. Retrieved on December 07, 2018, from https://www.statista.com/statistics/442848/per-capita-alcohol-consumption-of-all-beverages-in-the-us-by-state/

WalletHub contributor. (2018, January 2018). *2018's Most & Least Educated States in America.* Retrieved on December 13, 2018 from https://wallethub.com/edu/most-educated-states/31075/

Wikipedia contributors. (2018, November 16). List of U.S. states and territories by area. In *Wikipedia, The Free Encyclopedia*. Retrieved 23:29, December 6, 2018, from https://en.wikipedia.org/w/index.php?title=List_of_U.S._states_and_territories_by_area&oldid=869038675

Wikipedia contributors. (2018, October 11). List of U.S. states and territories by GDP. In *Wikipedia, The Free Encyclopedia*. Retrieved 01:20, December 7, 2018, from https://en.wikipedia.org/w/index.php?title=List_of_U.S._states_and_territories_by_GDP&oldid=863472848

ACKNOWLEDGMENT

We want to express our gratitude to Dr. Miriam Boeri, who granted us permission to use her report and survey findings for the purpose of this paper. Additionally, she spent a considerable amount of time explaining her research work and this helped us to better understand this general field of study. Her insights greatly aided us in writing the paper.

