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 performed 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

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

Descriptive statistics:

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

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

Figure 3: Cross-tabulation of active use or not and gender
Figure 4: Cross-tabulation of race and gender – AU=No

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

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](image1)

![Figure 12: Graph of Move_Age vs. Current_Age, including AU status](image2)

![Figure 13: Box plot for move age vs. AU](image3)

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.

URL: http://dx.doi.org/10.14738/assrj.65.6580.
Table 2: List of variables additionally compiled (second dataset)

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Role</th>
<th>Type of data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>Variable with the observation ID label</td>
<td>Qualitative</td>
<td>51 US states (including DC but not including Puerto Rico)</td>
</tr>
<tr>
<td>GDP</td>
<td>Test variable</td>
<td>Quantitative</td>
<td>Gross domestic product (GDP) by state (millions of current dollars)- 2016</td>
</tr>
<tr>
<td>PopDensity</td>
<td>Independent variable used to determine clusters</td>
<td>Quantitative</td>
<td>Population Density (pop/sq miles)- 2016</td>
</tr>
<tr>
<td>MedianIncome</td>
<td>Test variable</td>
<td>Quantitative</td>
<td>Median household income in 2016</td>
</tr>
<tr>
<td>AppAlcohol Comp</td>
<td>Independent variable used to determine clusters</td>
<td>Quantitative</td>
<td>Apparent Alcohol Consumption- 2016</td>
</tr>
<tr>
<td>AlcoholCompRate</td>
<td>Independent variable used to determine clusters</td>
<td>Quantitative</td>
<td>Alcohol consumption rate- 2015</td>
</tr>
<tr>
<td>AlcoholDeath</td>
<td>Independent variable used to determine clusters</td>
<td>Quantitative</td>
<td>Age-adjusted death rate due to alcohol poisoning- 2015</td>
</tr>
<tr>
<td>OpPrecRate</td>
<td>Independent variable used to determine clusters</td>
<td>Quantitative</td>
<td>Opioid prescription rate- 2016</td>
</tr>
<tr>
<td>OpDeathRate</td>
<td>Independent variable used to determine clusters</td>
<td>Quantitative</td>
<td>Opioid-Related Overdose Deaths/100,000(2016)</td>
</tr>
<tr>
<td>DrugDeathRate</td>
<td>Independent variable used to determine clusters</td>
<td>Quantitative</td>
<td>Drug overdose death rate- 2016</td>
</tr>
<tr>
<td>SuicideDeathRate</td>
<td>Independent variable used to determine clusters</td>
<td>Quantitative</td>
<td>Suicide death rate- 2016</td>
</tr>
<tr>
<td>Education Level</td>
<td>Independent variable used to determine clusters</td>
<td>Quantitative</td>
<td>Educational attainment + Quality of education &amp; attainment gap; measured on a scale of 100 points- 2016</td>
</tr>
</tbody>
</table>

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 thought 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.
Figure 14: Basic statistics for GDP

Figure 15: Histogram of GDP distribution

Figure 16: Basic statistics for opioid prescription rate

Figure 17: Histogram of opioid prescription rate distribution

Figure 18: Basic statistics for opioid death rate

Figure 19: Histogram of opioid prescription rate distribution

URL: http://dx.doi.org/10.14738/assrj.65.6580.
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
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 overdose 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:

\[ Y_i \sim \text{Binomial}(n = 1, p = f(E_i)) \]

where

\[ E_i = \beta_0 + \beta_1 \text{ Gender}_i + \beta_{21} \text{ Race}_1i + \beta_{22} \text{ Race}_2i + \beta_{31} \text{ Education}_1i + \beta_{32} \text{ Education}_2i + \beta_{33} \text{ Education}_3i + \beta_4 \text{ MoveAge}_i + \beta_5 \text{ CurrentAge}_i \]
\( f \) is the logistic link function; i.e., \( f(E) = \frac{e^{(E)}}{1 + e^{(E)}} \)

- \( Y_i \) is whether or not the \( i \)th respondent is an active user (in the last six months) of drugs (methamphetamine, cocaine, crack, or heroin).

- **Gender** is a (0, 1) dummy variable for the gender of the \( i \)th respondent recorded and where “Male” is the reference/base/dummy level

- **Race\(_{1i}^{(1)}\) and Race\(_{2i}^{(2)}\)** are (0, 1) dummy variables for the race of the \( i \)th respondent recorded and where “White” is the reference/base/dummy level, Race\(_{1i}^{(1)}\) refers to Black and Race\(_{2i}^{(2)}\) refers to “other.”

- **Education\(_{1i}^{(1)}\), Education\(_{2i}^{(2)}\) and Education\(_{3i}^{(3)}\)** 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, Education\(_{1i}^{(1)}\) refers to College degree, Education\(_{2i}^{(2)}\) refers to High School Diploma/GED, and Education\(_{3i}^{(3)}\) refers to less than high school.

- **CurrentAge\(_{i}^{(4)}\)** is the age of \( i \)th respondent when he/she recorded the interview (2009/2010)

- **MoveAge\(_{i}^{(5)}\)** 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.
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%.

**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:

URL: http://dx.doi.org/10.14738/assrj.65.6580.
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.

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.

![Figure 29: Full output for logistic regression](image)

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.
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.
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.

**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 overuse. 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


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


**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.