Paper ###-2020

Alternate solution to opioid crisis: Identifying risky providers

Team Cartesian

ABSTRACT

Since the 1990s, the United States has been embraced by the infamous opioid abuse which has caused enormous losses both in terms of lives and money involved. As opposed to identifying patients at risk of becoming opioid-addicts, this paper proposes a supplementary action aimed towards curbing the opioid epidemic: a top-down approach to inspect factors that lead to a provider over-prescribing opioid. To do so, a 3-fold data collection strategy was adopted in which state-level prescription drug related policies, county-level socioeconomic factors and provider-level demographic information were used to create various predictive models. Regression models were run on state and county-level datasets to identify factors that drive the average opioid prescription rate and the factors that turned out to be significant were further utilized at a provider-level for classifying over-prescribers. Among all the models created, the random forest model gave the best output with the highest accuracy (81%), sensitivity (87%) and AUC (87%). The results from the logistic regression model were examined to understand the effect of each variable individually.

INTRODUCTION

In the United States, opioids are used as an effective treatment for chronic and severe pain in cases of surgery or pain management. However, 'The Opioid Crisis', that began in the 1990s, started with over-prescription of opioids eventually leading to them becoming the most prescribed medicines (Wikipedia). This situation has been identified as a uniquely American problem and the current healthcare system in the United States requires people that are not covered by government healthcare insurance policies to obtain private insurance that favor prescribing opioid drugs over more expensive therapies. The prescription rate of pain-killing opioids has been on a rise and the rate of prescription for non-opioid drugs has dramatically dropped in the last decade.

PROBLEM STATEMENT

According to recent CDC statistics, approximately 130 people die every day of opioid overdose and about a third of them die due to an overdose of prescribed opioids such as Morphine. Organizations such as Centers for Disease Control and Prevention are dealing with this epidemic predominantly in two ways: (1) By providing addiction treatments to people which imposes a large financial burden on the economy of the nation; and (2) By identifying risky patients. This paper aims to introduce another strategy as an effective means of curbing the opioid crisis. This paper's objective is to provide a top-down analytical approach to determine factors that lead to a provider over-prescribing opioid drugs by considering:

1. The state policies guiding the supply and prescription of Schedule II and Schedule III pain-relieving drugs.
2. The socio-economic factors that impact the lifestyles of people in a county.
3. Demographic details of the provider as well as the patients/beneficiaries.

This strategy will help in identifying providers at the risk of over-prescribing and educating them appropriately, thereby adding to the formulation of efficient preventive strategies.

METHODS

1. DATA COLLECTION AND PREPARATION

Since the objective of this research is to assess the impact of state policies as well as county-level socioeconomic parameters on over-prescription of opioids, a 3-step data collection process was adopted. The first step had state-level data, the second step had county-level and the last step had provider-level data. The provider-level data was collected from CMS[1]. This dataset contained information for providers who prescribed drugs and were paid under the Medicare Part D Prescription Drug Program for the year 2017 across the nation. The dataset also contained details on the opioid prescription rate for each provider, claims filed by the provider, demographic details of the provider etc. It had about 1.1 million rows and 21 variables. County-level information such as unemployment rate, education, median household income etc. was obtained from AmericanFactFinder[2] for about 3,000 counties. From National Alliance for Model State Drug Laws (NAMSDL)[3], information regarding state statutes, regulations, and guidelines related to the treatment of chronic pain and prescribing practices such as practitioner education requirements, pain treatment requirements or guidelines, and limitations on the prescription of Schedule II and Schedule III prescription drugs was obtained for 50 states. There were approximately 20 such laws in the document which were used to manually create the state-level dataset with states as rows and laws as columns. The columns were binary variables indicating the presence of law in a state with 1, and 0 otherwise. The provider-level dataset was merged with the county as well as the state level dataset on the county and state.

2. DEFINING TARGET VARIABLE

The opioid prescription rate for some of the providers was missing from the dataset. There were about 250,000 such providers out of the 1.1 million that were removed as the opioid prescription rate was used to design the target variable. The next step was creating a binary label for each provider to indicate if they had overprescribed. Since prescription rates vary across specialties of providers as well as states, the combination of the two was used for the process. For instance, the average opioid prescription rate for family practice doctors is 5.6% whereas, for pain medicine doctors, it is 48.6%. Similarly, the average prescription rate per 100 people in Texas is 53.1% whereas, in Missouri, it is 71.8%. To account for this variation across states and specialties, the 150 odd specialties in the dataset were reduced to 9 primary levels based on research. For each joint combination of state and specialty, the 90th percentile value of prescription rate was considered as the cutoff value. Any provider whose prescription rate was greater than the cutoff value was identified as an over-prescriber and labeled as '1'.

3. DATA CLEANING

After removing the providers for whom the opioid prescription rate was missing, the second step was the treatment of missing values for other variables. Variables that had more than 50% of values missing were rejected from the analysis. The rest of the variables that had less than 50% values missing were imputed using tree-based imputation in SAS® Enterprise Miner™. The variables that were imputed are: average age of beneficiaries, average risk score of beneficiaries, etc. Finally, some of the variables that were heavily

[1] https://data.cms.gov/Medicare-Part-D/Medicare-Provider-Utilization-and-Payment-Data-Par/psut-35i4

[2] https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t

[3] https://namsdl.org/wp-content/uploads/Overview-of-State-Pain-Management-and-Prescribing-Policies-1.pdf

skewed were transformed using logarithmic transformation. Figures 3 and 4 in the appendix show the distribution of the variable 'total_drug_cost' before and after transformation.

ANALYSIS

1. Determining the effects of state policies on average opioid prescription rate in a state

To analyze the impact of state policies related to Schedule II and Schedule III drugs, a stepwise polynomial regression model was run in SAS® Enterprise Miner™ with the average opioid prescription rate as the target variable. The level of significance was set at 5%. The results of the regression model are shown in figure 5 in appendix.

2. Determining the effects of socio-economic indicators on county-level on the average opioid prescription rate in a county

The analysis was then narrowed to county-level to comprehend the effects of socio-economic parameters such as education, unemployment, poverty, etc. on the average opioid prescription rate in a county. To do so, a stepwise regression model was run on the county data and the results were used to identify the variables that impact the prescription rate in a county. The results of this model are shown in figure 6 in appendix.

3. Classification model to analyze factors that lead to a provider over-prescribing opioid drugs based on state and county factors.

The significant variables from the aforementioned models were incorporated in the provider-level data. Using the existing variables from the provider-level data as well as the new variables from state policies as well as county-level data, further analysis was conducted. Since the event rate was found to be 4%, under-sampling as well as over-sampling was performed using PROC SURVEY to create a sample data containing equal distribution of providers that over-prescribed and that did not. On both the sampled datasets, imputation was performed using 'Imputation' node in SAS® Enterprise Miner™ followed by data partition of 70% training and 30% validation using the 'Data Partition' node. Since the original event rate in the population is 4%, a 'Decisions' node was used to adjust the prior probabilities. Variable transformation was performed using 'Transformation' node for logistic regression model. Maximum KS statistic was used as the criteria in the cutoff node to change the default cutoff probability value from 0.5 to 0.04. All the models were compared based on their validation misclassification rate. The results of the comparison are shown below.

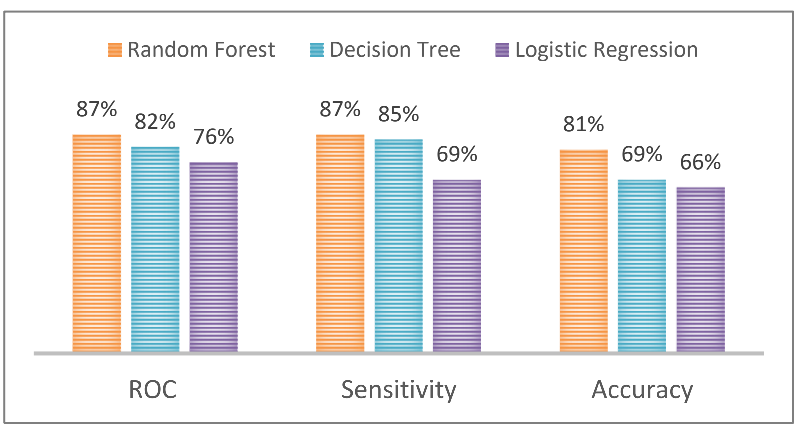

Figure 1 – Comparison of models based on ROC, sensitivity and accuracy for over-sampled dataset

RESULTS

RESULTS FOR REGRESSION MODEL FOR STATE POLICIES

It was found that the interaction of certain policies was significant. The interaction plots for these policies are shown in figures 8 and 9 in appendix.

1. If a state did not have the policy that restricts opioids prescribed by pain management facilities as well as the policy that restricts the no. of days for which opioids can be prescribed, prescription rates were lower; this is most likely because these policies were implemented in states with high prescription rates.
2. Similarly, the average opioid prescription rate goes down in states that authorize private entities to regulate opioids in conjunction with the policy that demands providers provide a written treatment plan in case of prescription of opioids.

These 4 policies were used in the classification model to identify over-prescribers.

RESULTS FOR REGRESSION MODEL FOR COUNTIES

The model helped to identify socio-economic parameters that could help to classify providers. It was found that counties with a higher median household income had lower prescription rates. Similarly, counties with a more educated population (people having at least high school degree) and lower Medicare and Medicaid coverage had lower prescription rates. These factors were used along with the state policies and the provider/beneficiary details to classify providers that over-prescribe.

RESULTS FOR CLASSIFICATION MODEL FOR PROVIDERS

Although the random forest model out-performed all other models in classifying over-prescribers, for the purpose of breaking down the results to examine the effects of individual factors, results of logistic regression were used, and the odds were computed.

The effects of some significant factors from state policies are shown in the figure below:

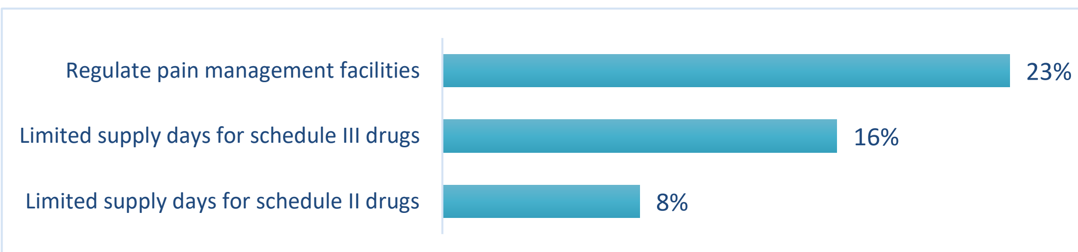

Figure 2 – Increase in the likelihood of over-prescription in states that do not have drug related policies as shown in the figure

The tables below show the effect of county and provider-level information on over-prescription

Variables	Change in the odds of over-prescribing
Population with at least a high school degree	11%
Population in county	15%

Table 1 – Effect of 100 unit increase in the significant variables from the county-level model on identifying over-prescribers

Variables	Change in the odds of over-prescribing
Average age of beneficiaries	-9%
Average risk score of beneficiaries	-22%

Table 2 – Effect of 1 unit increase in the significant variables from the provider-level model on identifying over-prescribers

GENERALIZATIONS

For the purpose of this study, patients (aged 65 and above) and providers enrolled under Medicare Part D were used to do the analysis. However, the study can also be expanded to cover patients from various age groups and analyze the leading causes of over-prescription among people belonging to other age brackets.

FUTURE STUDIES

The scope of this study includes analyzing state policies and county-level socioeconomic data along with provider/beneficiary details. To get deeper insights, spatio-temporal analysis can be performed using the same dataset to find out hot-spots for over-prescription of opioids. Additionally, articles from the media related to the downsides of prescribed opioids as well as data containing details on the manufacture and sales of opioids can be merged with the dataset created for this study and their impact can be assessed on prescription rates of opioids in a state/county.

CONCLUSION

This research follows a top-down approach to identify major factors that help to identify over-prescribers. Apart from the details provided by CMS, this paper broadened its scope by considering the effects of state policies pertaining to the prescription of Schedule II and Schedule III drugs as well as certain socio-economic parameters such as unemployment rate, population educated etc. in a county. Using all the information listed above, various models were built to classify over-prescribers and the random-forest model out-performed decision tree as well as logistic regression. Certain state policies and socio-economic parameters such as median household income, unemployment rate, etc. were significant in classifying over-prescribers. Factors such as the beneficiary's age as well as his/her risk score also helped to determine if a provider will over-prescribe.

REFERENCES

athenainsight, october 2019, "infographic: state-by-state breakdown of opioid regulations" https://www.athenahealth.com/knowledge-hub/clinical-trends/infographic-opioid-regulations-state-by-state

ghertner, robin and groves, lincoln, september 2018. "the opioid crisis and economic opportunity: geographic and economic trends". https://aspe.hhs.gov/system/files/pdf/259261/aspeeconomicopportunityopioidcrisis.pdf

NAMSDL, january 2016. "overview of state pain management and prescribing policies". https://namsdl.org/wp-content/uploads/overview-of-state-pain-management-and-prescribing-policies-1.pdf

APPENDIX

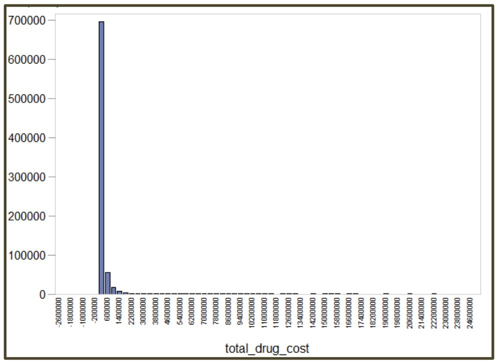

Figure 3 – Distribution of original variable – Total_Drug_Cost

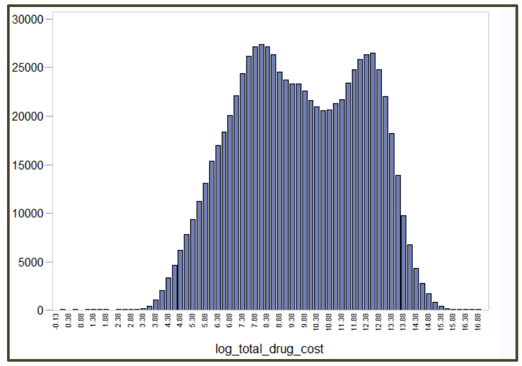

Figure 4 – Distribution of transformed variable – Total_Drug_Cost

Type 3 Analysis of Effects

Effect	DF	Sum of Squares	F Value	Pr > F
Authorization_to_other*Written_Treatment_Plan	1	1591.4468	8.30	0.0060
Regulate_Pain_Facilities*Stautory_limit_14_days	1	2546.1521	13.27	0.0007

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept		1	66.7731	2.4753	26.98	<.0001
Authorization_to_other*Written_Treatment_Plan	0 0	1	-5.9262	2.0574	-2.88	0.0060
Regulate_Pain_Facilities*Stautory_limit_14_days	0 0	1	-8.8648	2.4331	-3.64	0.0007

Figure 5 – Output of regression model for determining the effects of state policies on average opioid prescription rate in a state

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	124661	17809	46.39	<.0001
Error	555	213065	383.90161		
Corrected Total	562	337727			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	158.81696	14.42463	46538	121.22	<.0001
Male_Population_18_to_24_years	-0.00092198	0.00015021	14464	37.68	<.0001
Total_Population_25_years_a_0004	0.00043400	0.00006820	15548	40.50	<.0001
Total_Population_25_years_a_0005	-0.00041462	0.00013218	3777.32099	9.84	0.0018
Percent_Medicare_Coverage	0.61108	0.18696	4101.24153	10.68	0.0011
Percent_Medicaid_Coverage	-0.44548	0.14066	3850.43088	10.03	0.0016
Households_35000_to_49999	-1.70236	0.58744	3223.97653	8.40	0.0039
Households_Median_income	-0.00127	0.00011517	46628	121.46	<.0001

Bounds on condition number: 50.987, 819.98

All variables left in the model are significant at the 0.0500 level.

No other variable met the 0.0500 significance level for entry into the model.

Figure 6 – Output of regression model for determining the effects of socio-economic parameters on average opioid prescription rate in a county

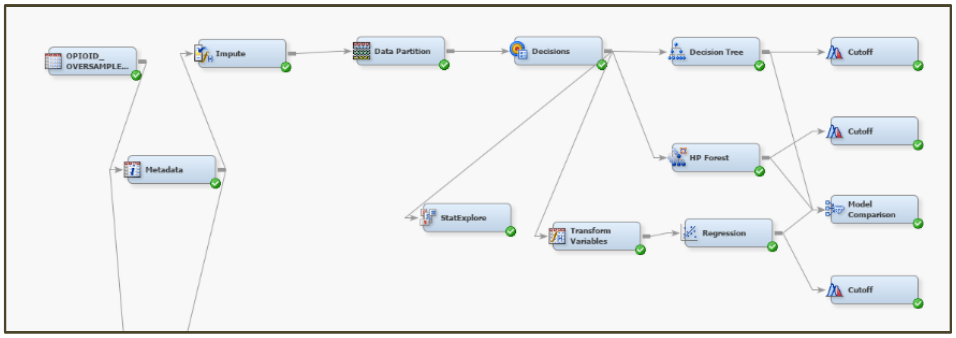

Figure 7 – Process flow for the classification model from SAS® Enterprise Miner™.

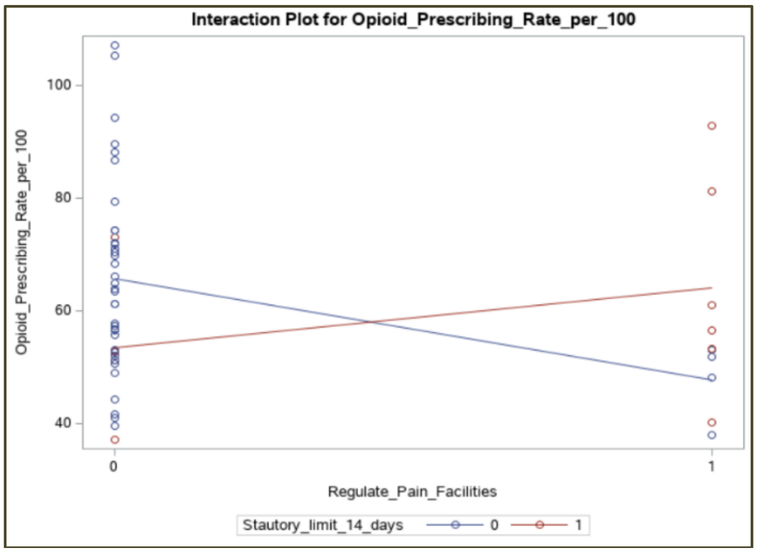

Figure 8 – Interaction plot explaining the interaction between Regulate_Pain_Facilities and Statutory_Limit_14

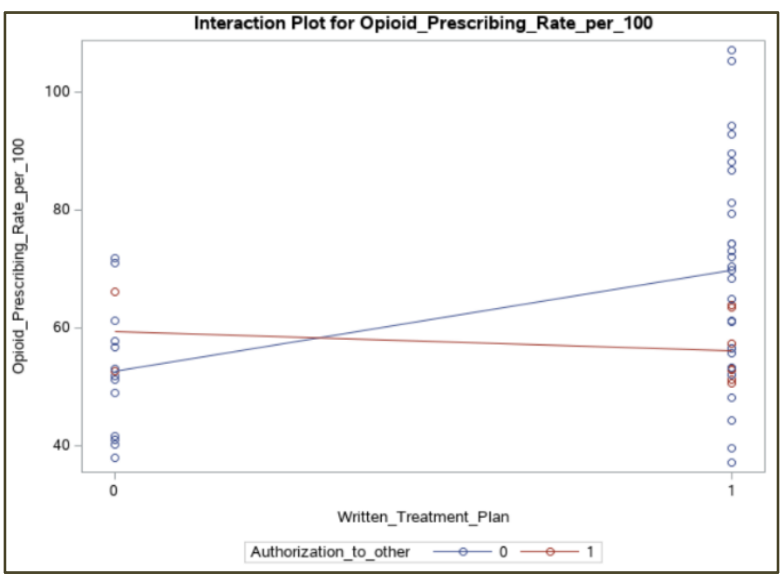

Figure 9 – Interaction plot explaining the interaction between Written_Treatment_Plan and Authorize_To_Others