 doi:10.20944/preprints201905.0165.v1

Article

Geospatial-Temporal and Demand Models for Opioid Admissions, Implications for Policy

Lawrence Fulton [1*], Zhijie Dong [2], Benjamin Zhan [3], Clemens Scott Kruse[4], and Paula Stigler Granados [5*]

1 School of Health Administration, Texas State University; larry.fulton@txstate.edu, 210-837-9977
2 School of Engineering, Texas State University; sasha.dong@txstate.edu
3 Department of Geography, Texas State University; zhan@txstate.edu
4 School of Health Administration, Texas State University; scottkruse@txstate.edu
5 School of Health Administration, Texas State University; psgranados@txstate.edu

Abstract: *Background:* As the opioid epidemic continues, understanding the geospatial, temporal and demand patterns is important for policymakers to assign resources and interdict individual, organization, and country-level bad actors. *Methods:* GIS geospatial-temporal analysis, k-means cluster analysis, and extreme-gradient boosted random forests are used to evaluated ICD-10 F11 opioid-related admissions. The period of analysis was January 2016 through September 2018. *Results:* The analysis shows existing high-intensity areas in Chicago and New Jersey with emerging areas in Atlanta, Salt Lake City, Phoenix, and Las Vegas. Further, cluster analysis supports the current inflow from China through Mexico and Canada with another cluster in the Northeast likely associated with the Dominican flow. Explanatory models suggest that hospital overall workload and financial variables may be used for allocating opioid-related funds effectively, as the gradient-boosted random forest models accounted for 88% of the variability on a blinded test data set. *Conclusions:* Based on GIS analysis, the opioid epidemic is likely to spread or diffuse through the country, and interdiction efforts require demand-analysis such as that provided in this study to allocate scarce resources for supply-side and demand-side interdiction: prevention, treatment, and enforcement.

Keywords: opioids, GIS, random forests

1. Introduction

In the 1990's, pharmaceutical companies began marketing to the medical community that opioids were non-addictive, and medical providers began prescribing them at a higher rate. [1] This marketing opened the door to the U.S. opioid epidemic. Federal funding alone to fight this epidemic was estimated at $7.4 billion in 2018. [2]

The U.S. Department of Health and Human Services estimates that 91,548 people died from opioids, synthetic opioids, and heroine in 2016. [1] The National Survey on Drug Use and Health estimated that 4.2% of the U.S. population misused opioids in 2017. [3] The toll in morbidity and mortality is facilitated by over-prescription and bad actors, both individuals and countries.

In April of 2019, 31 physicians, 7 pharmacists, 8 nurse practitioners, and 7 other licensed medical professionals in 7 different states were charged as part of a law enforcement investigation of providing opioid prescriptions for cash or sex. These individuals prescribed more than 32 million pills. [4] In May 2019, a podiatrist was convicted of operating an opioid pill mill. [5] In another example from May 2019, a Virginia doctor was convicted on 861 counts of drug distribution. The oxycodone and oxymorphone that the physician prescribed to a West Virginian patient resulted in her death. [6] As a final example, 162 individuals including doctors were charged for prescribing and distributing opioids in June of 2018. [7] The problem with bad actors is real and widespread, yet the distribution of opioid and opioid-like products outside of the medical system may be an even larger problem.

China is the largest US source of illicit fentanyl and fentanyl-like substances, and it distributes that product through Canada, Mexico, and directly to the US. [8] The reason for China's involvement in our markets is that its pharmaceutical system is poorly regulated [9] and that its manufacturers create new and uncontrolled substances to stay ahead of regulators. [10] One estimate is that Chinese fentanyl and derivatives supply 90% of the illicit product in the United States. [11] Even so, Mexico's two largest criminal organizations traffic the product largely through San Diego. Dominican traffickers supply the heavily stricken Northeast. [12]

The net result of over-prescription, illicit actors, and illicit suppliers is an increase in morbidity and mortality. Policy considerations for addressing these three problems and providing funding for prevention, treatment, and enforcement require an understanding of the geospatial, temporal spread of the epidemic as well as models for demand of services. This research describes the geospatial, temporal spread of opioid inpatient demand, an analysis of the clusters associated with inpatient admissions, and explanatory models for opioid admissions that might be used to estimate state, zip-code level, hospital-level demand as well as resource requirements. The significance of this research is that it provides decision support for policymakers by identifying areas which require additional enforcement as well as funding.

2. Experimental Section

2.1 Data

Data for this research derive from Definitive Healthcare, through the hospital "inpatient diagnosis analytics" query. Only principal diagnoses ICD-10 codes beginning with F11 were used. F11 codes are opioid related disorders. Complete annual data were available for 2016 and 2017; 2018 data were only available through September. The Definitive Healthcare data largely derive from the Centers for Medicare and Medicaid Services [CMS] Standard Analytical Files [SAF], although the organization estimates all-payor claims through parochial algorithms. [13]

2.2 Geospatial Analysis

Heat maps are used to plot zip-code unit of analysis claims data by year. These maps illustrate the intensity of opioid admissions by geographic region and city by color-coding areas of intensity. When used properly, they can highlight geographic variation. [14] The use of heat maps in healthcare is ubiquitous, as they have been used for improving minority health surveillance [15], examining birth outcomes [16], and many other applications. The value in geospatial-temporal analysis is the graphical depiction of change in demand over time.

As part of the geographic assessment, this research leverages k-means cluster analysis (unsupervised learning) to evaluate inpatient admissions for F11 ICD-10 codes (opioids). Individual observations are zip codes for each claim. For example, a hospital with latitude X, longitude Y, and claims Z would have Z repeated observations. Use of cluster analysis in public health GIS provides advantages in that visualization of complex, multidimensional data is readily possible. [17]

2.3 Explanatory Analysis

Stepwise linear regression, lasso regression, robust regression, elastic net regression, and extreme gradient-boosted random forests estimate the ICD-10 F11 opioid admissions. Models are built on a random 80% training set and evaluated on a 20% test set. These models are exploratory to see which workload, financial, technical, and geospatial-temporal features might be explanatory and thus useful for allocation of resources by policymakers.

 doi:10.20944/preprints201905.0165.v1

Stepwise regression models add and subtract variables based on criteria to produce reasonable multiple regression models. In this research, the Akaike Information Criterion (AIC) is used to select the stepwise model using a forward and backwards method. In this method, variables are added in sequence and removed if they no longer contribute significantly to the model's performance. [18]

Lasso regression is a constrained regression that penalizes any model with too many variables using an L1-norm penalty function (absolute value). Ridge regression is similar to Lasso regression but penalizes using (squared coefficient estimates, L2-norm). Elastic combines both L1 and L2 penalty functions. Equations 1 through 4 are the parameter estimation models for linear, lasso, ridge, and elastic net regression. [18] The parameter λ in all models is a Lagrangian multiplier, while the parameter α in Equation 4 mixes the squared penalty with the absolute value penalty.

Linear regression (OLS):
$$\hat{\beta} = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\,\beta_j)^2 \tag{1}$$

Lasso regression (L1-norm):
$$\hat{\beta} = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\,\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{2}$$

Ridge regression (L2-norm):
$$\hat{\beta} = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\,\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{3}$$

Elastic Net (L1 & L2 Norm):
$$\hat{\beta} = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\,\beta_j)^2 + \lambda \sum_{j=1}^{p} (\alpha\beta_j^2 + (1-\alpha)|\beta_j|) \tag{4}$$

Random Forests, are a machine learning technique that use an ensemble of de-correlated tree models. The tree forecasts are averaged (ensembled) to produce an estimate. A tree model classifies counts of observations by splitting variables based on some decision criteria. Trees must be pruned or truncated, so that they do not overfit. [18] Figure 1 is an example of a tree built from the Definitive Healthcare F11 dataset with depth of two branches. The tree splits observations by surgeries less than / greater than or equal to 2600.5 then again by region = Middle Atlantic and state equal to New York. The "<0 .5" indicates that the region is not the Middle Atlantic and that the state is not New York, as those are dichotomous variables.

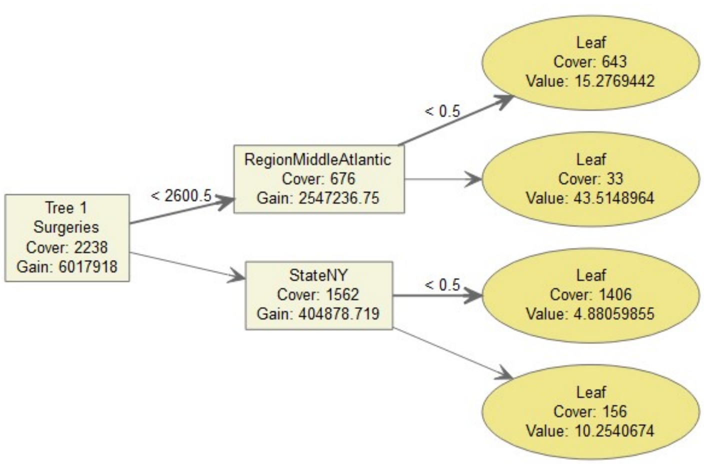

Figure 1. An example of a tree model to classify opioid admissions

Gradient boosted random forests use nonlinear optimization to optimize a cost function based on the (pseudo)-residuals of a function. The residuals of each tree are re-fitted with the possible independent variables in other tree models to estimate a better fit. A more complete discussion of gradient boosting is provided in *The Elements of Statistical Learning*. [18]

2.4 Variables.

The primary variable of interest is the inpatient admissions for ICD-10 code F11 ("Opioid-Related Disorders") which are measured by hospital claims associated with opioids. There are 55 total, opioid-related F11 codes used in this study (Appendix 1). This variable is measured at the hospital level and aggregated by zip code / year for geospatial mapping. The inpatient admissions provides a measure of the met demand for services and is suggestive of which areas may need additional funding and resources from health policy decisionmakers.

Independent variable groups evaluated in the explanatory models included financial variables, workload variables, technical variables, and geo-spatial temporal variables. The financial variables included net patient revenue, net income, cash on hand, assets, and liabilities. Workload variables included discharges, emergency room visits, surgeries, and acute beds. Technical variables included staffed beds, affiliated physicians, employees, percent Medicare or Medicaid patients, ownership, medical school affiliation, and hospital type. Geographic / temporal variables included the Census Bureau region, the urban / rural status, the state, and the year. These models will identify characteristics of the facilities providing inpatient care to opioid abusers. As the epidemic spreads or diffuses, these features might be used to anticipate which local facilities are likely to experience an increase in care for these patients.

2.5 Software

All analysis was performed in R Statistical Software [19] and Microsoft Excel 2016. [20] Packages used for the primary analysis in R are cited.

3. Results.

3.1 Descriptive Statistics-Missing Data

A missingness map depicts that 1% of the data were missing after both ER visits and surgeries for psychiatric hospitals were imputed with zeros. The assumption for this imputation was simply that these values were true zeros rather than missing data. Even without this imputation, only 3% of the data were missing. Because the percent of missing was so small, means were imputed rather than leveraging more sophisticated techniques like multiple imputation. The total number of valid observations at the hospital unit of analysis was N=2,090.

3.2 Descriptive Statistics-Quantitative Data

Descriptive statistics for the quantitative data are in Table 1. The average number of ICD-10 claims for F11 was 103.47 with a median of 33; however, the standard deviation was 198.19 indicating significant variability. The average reporting hospital had 274 beds, 14K discharges, 50K ER visits, 10K surgeries, 72K acute days, 385 affiliated physicians, 2K employees, and had 42% of the claims paid by Medicare / Medicaid. The average facility had $527K in payments for F11, $1.5M in charges, $403M net patient revenue, $30.1M net income, $37M cash on hand, $524M in assets, and $213M in liabilities. On average, a facility was paid about 35% of charges. (Neither payments nor charges were used in models, as they derive directly from claims.)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2019 doi:10.20944/preprints201905.0165.v1

Table 1. Descriptive statistics for the quantitative variables

N=2090	Mean	SD	Median	10% Trimmed
Claims	103.47	198.19	33.00	57.14
Staffed Beds	273.53	227.59	214.00	240.19
Discharges	14,039.67	13,572.68	10,306.00	11,934.51
ER Visits	50,642.37	49,459.27	43,564.50	44,229.24
Surgeries	10,113.16	12,165.17	7,142.00	7,992.79
Acute Days	72,140.06	72,658.04	49,710.00	59,421.91
Physicians	384.89	391.09	312.00	318.58
Employees	1,968.09	2,458.91	1,222.50	1,475.25
% Medicare/caid	42.00%	16.00%	42.00%	42.00%
Payments ($1M)	$.527	$1.077	$.160	$.281
Charges ($1M)	$1.505	$2.818	$.548	$.858
Net Patient Revenue ($1M)	$403.07	$519.72	$247.83	$298.78
Net Income ($1M)	$30.25	$109.98	$8.12	$18.94
Cash on Hand ($1M)	$36.59	$182.67	$1.35	$10.53
Assets ($1 M)	$524.23	$961.34	$206.53	$317.09
Liabilities ($1 M)	$212.69	$542.10	$70.82	$125.23

3.3 Descriptive Statistics-Categorical Data

The modal Census Bureau region for reporting F11 claims was the South Atlantic (373, 17.8%) followed closely by the East North Central (366, 17.5%). The bar chart of the frequencies by Census Bureau region is Figure 3. Most of the admissions for F11 codes were in urban areas (1732, 82.9%) with the remainder being areas classified as rural. The South Atlantic region extends from Florida to Washington, D.C., where there is a significant intensity of opioid abuse. The East North Central region includes Chicago, which has extremely high intensity of abuse.

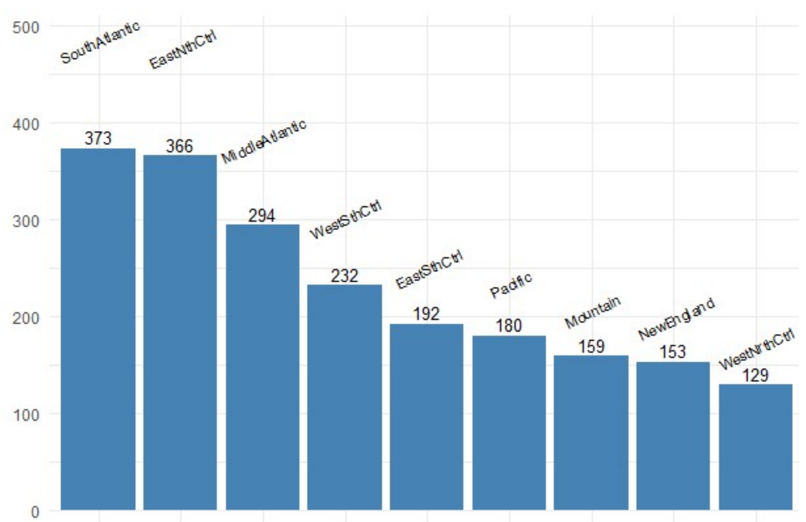

Figure 3. Bar chart of frequencies by Census Bureau region shows raw frequencies, not adjusted for population.

Most of the hospitals reporting these admissions were short-term acute care facilities (1682, 80.5%) with psychiatric hospitals being the next most common type (385, 18.4%). The majority had no affiliation with a medical school (1159, 55.5%), although 21.2% (443) reported a major affiliation. In terms of ownership, 965 (46.2%) were voluntary non-profit non-church owned, 567 (27.1%) were proprietary corporation owned, and 279 (13.3%) were voluntary non-profit church owned. The observations were split nearly evenly between 2016 and 2017 with 1051 and 1039 observations respectively. Most interestingly to policymakers is that the non-profit community of hospitals appears to provide the majority of inpatient care for opioid patients.

3.4 Descriptive Statistics-Correlational Analysis

Graphical hierarchical clustered correlational analysis [21] revealed strong relationships among most of the workload and financial variables. The strongest correlation is between discharges and acute days (r=.97), while the next strongest correlation is between employees and net patient revenue (r=.97). All correlations in Figure 4 are statistically significant at the α=.05 level unless an "X" appears in the correlation plot. The number of claims appears to be weakly correlated with other variables indicating that the relationships are either nonlinear or not present. What is also interesting from a policy perspective is that as the facility increases in workload and financial metrics, there is a negative relationship with the number of inpatient admissions for F11. This would seem to indicate that smaller hospitals are bearing the brunt of the opioid epidemic for inpatient services. The effect size is small and requires investigation.

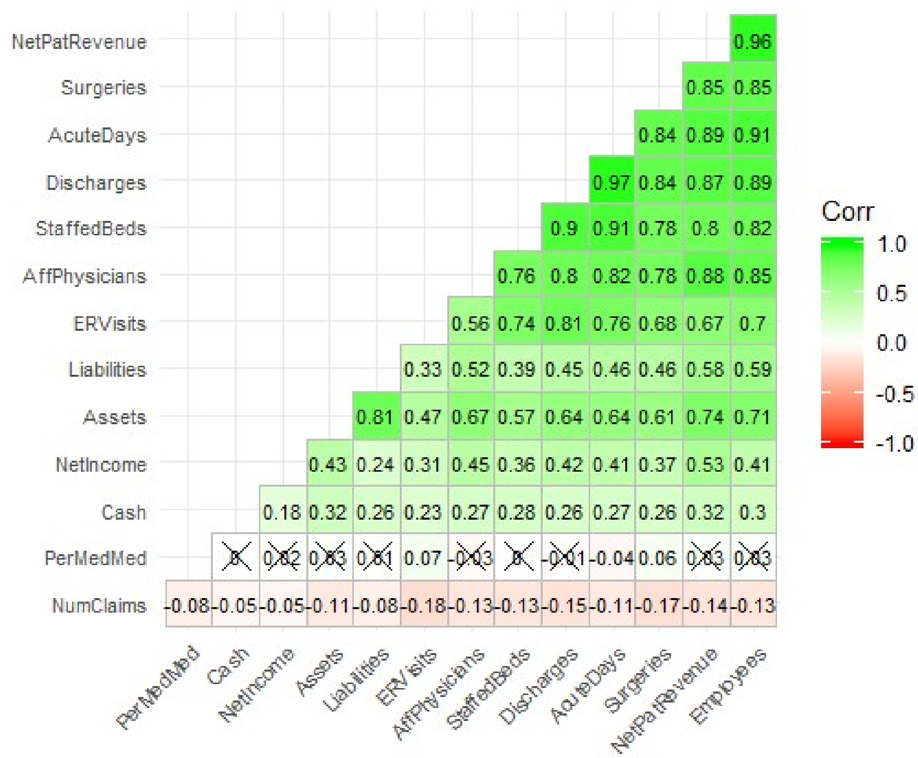

Figure 4. The correlation plot reveals strong relationships among financial and workload variables.

3.5 Exploratory Data Analysis-Feature Engineering and Transformations

Random forest regression models are scale invariant; however, the other methods used in this research are not. [18] The "car" package in R [22] facilitated a multivariate Box-Cox transformation for all modeled quantitative variables simultaneously after adjustment. Box-Cox transformations

 doi:10.20944/preprints201905.0165.v1

require that variables be strictly positive definite. With positive definite variables, the transformation seeks power transformations (powers of λ) that make the data multivariate normal enough for use in traditional linear models. [23] This multivariate transformation is particularly useful for random effects models, models where the independent variables are assumed to be not fully observed or the result of random variable draws. The likelihood ratio test of the null (multivariate normal) vs. the alternative (not multivariate normal) after location transformation to make all variables positive definite resulted in a p-value of 1.0. The actual vector of transformations follows: λ={-.39, .31, .38, .24, .23, .34, .25, .14, .1, .21, .22, .48, .21, .71 }for x={number of claims, number of staffed beds, number of discharges, ER visits, total surgeries, acute days, net patient revenue, net income, cash, assets, liabilities, affiliated physicians, employees, percent Medicare/Medicaid}, respectively. Figure 5 is a correlation plot [24] post-transformation which reveals the strength, direction, and bivariate shape of the bivariate normal between variable pairs. With successful transformation, the forecasting using linear methods is likely to improve.

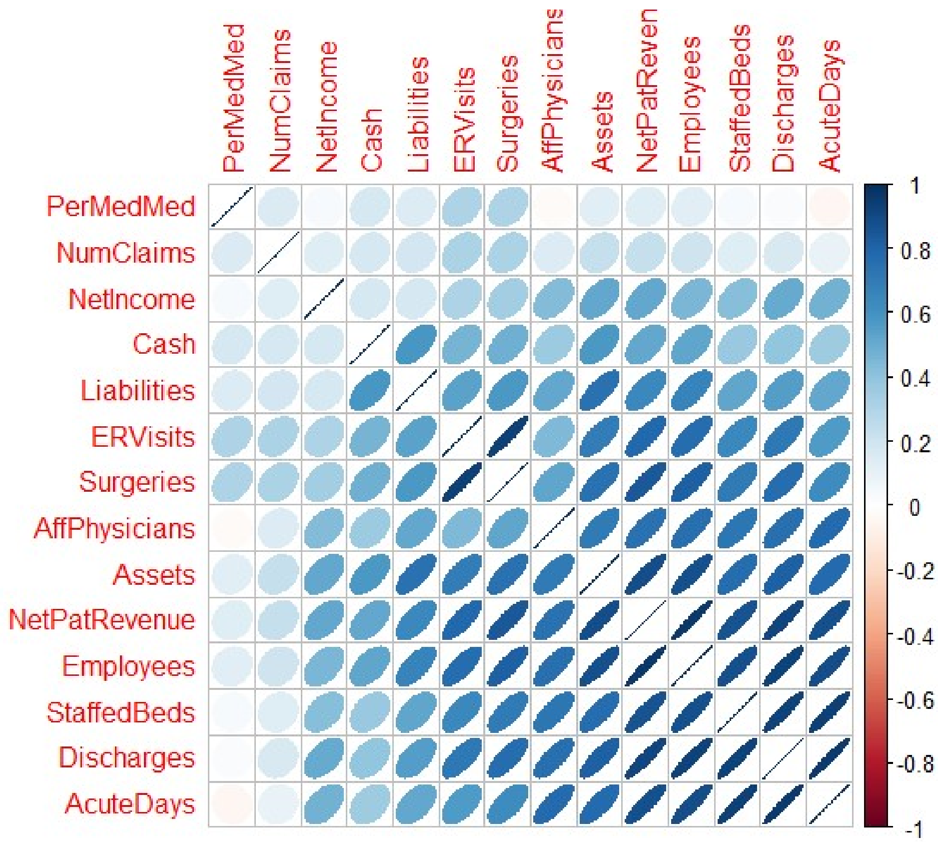

Figure 6. The correlation plot post-transform depicts the bivariate pairs.

3.6. Geospatial Analysis Results-Zip Code Unit of Analysis

Geospatial heat map analysis of F11 claims by year and zip code is shown in the Figure 5 panels. The maximum scale is 3000 claims for the diagrams. The problem areas of the opioid epidemic become clear with simple graphical analysis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2019 doi:10.20944/preprints201905.0165.v1

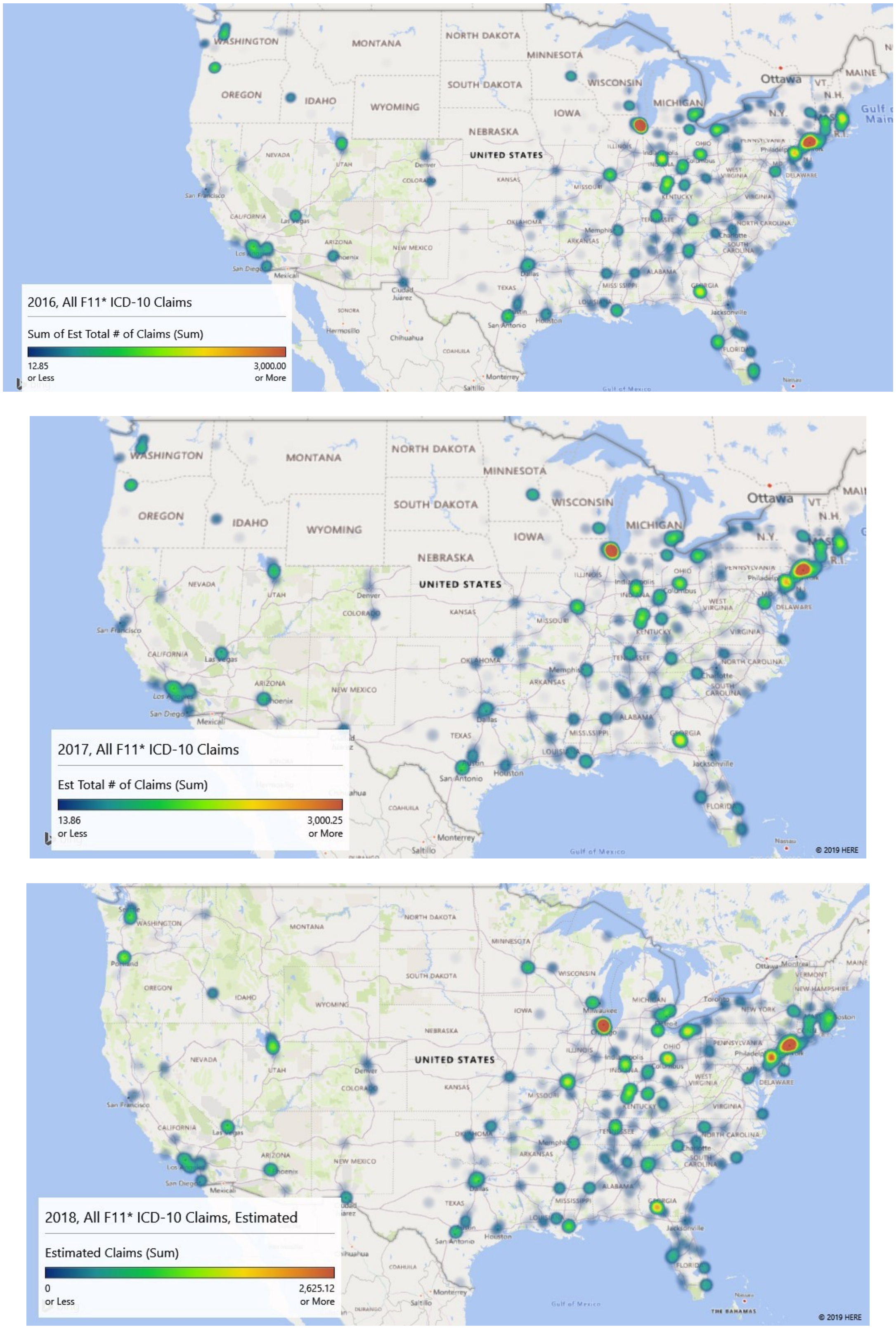

Figure 5. Excel-based [20] heat panel maps of the opioid admissions (ICD-10 F11) for 2016, 2017, and 2018 (extrapolated from September) show emerging areas of interest..

In 2016, the level of intensity for admissions is strongest around Chicago, Illinois and large swaths of New Jersey, where drug overdose is its leading cause of accidental death. [25] The heat map depicts extreme intensity (dark red) in Chicago). Emerging areas appear to be Washington, DC; Atlanta, GA; and areas of Kentucky, Indiana, and Ohio.

By 2017, the area of intensity around New Jersey had grown, Atlanta saw more intensity, and Chicago remained the most intense. The usage in Los Angeles had expanded but remained sub-intense. Areas in Kentucky, Indiana, and Ohio remained problematic.

Data for 2018 were complete only through September, so they are excluded in the explanatory modeling. However, linear extrapolation produced the 2018 chart which indicates significant intensity in Chicago, New Jersey, and Atlanta with emerging "red" intensity levels in Ohio. If this extrapolation held, then Salt Lake City, Phoenix, and Las Vegas will have increased in intensity to a brighter level of "green-yellow." Montana, the Dakotas, Iowa, and Wyoming appear to be inoculated against the epidemic.

Overall, the maps are suggestive of areas where intervention efforts are needed most or are emerging. From a policy perspective, opioid prescriptions in the highest afflicted areas like Illinois and New Jersey should be screened more closely than those (say) in Montana, South Dakota, and North Dakota. Machine learning techniques should be used to identify outliers similar to Ekin et al. [26] Further, interdiction efforts should focus on Chicago as a major transportation hub along with the emerging problem city, Atlanta, for the same reason, and (of course) New Jersey.

It is interesting that while California and Florida have large populations, none of their major population centers reached the same level of high intensity scales of other large cities. The questions then become how these patterns might be explained and possibly forecast, and what are the federal and local policy implications for funding based on the expansion / diffusion associated with the epidemic.

3.7. Cluster Analysis

A k-means cluster analysis was evaluated with k=2 to 8 clusters for the inpatient data, and the within sum of squares was evaluated. The results of the k=4 analysis is overlaid on the inpatient admissions map for 2018 (Figure 6). Interestingly, this map may provide some insights into the flow of the illicit opioids into the country.

The cluster analysis highlights possible inflow through various countries discussed previously. The red on the map may reflect direct in-flow from China (West Coast) as well as some in flow through the Tijuana border from Mexico, while the green might reflect flow through Canada. The blue may be an indicator of trafficking from China through Mexico, and it is possible that the black indicates the Dominican flow. This cluster analysis is at least partially congruent with known flow [8, 12]; however, as an unsupervised machine learning method, it provides insights only. The four cluster centers are located in Nevada, Alabama, Pennsylvania, and Indiana. The geospatial temporal analysis provides insight into the opioid epidemic and decision support for resource allocation. Explanatory models in the next section provide additional decision support.

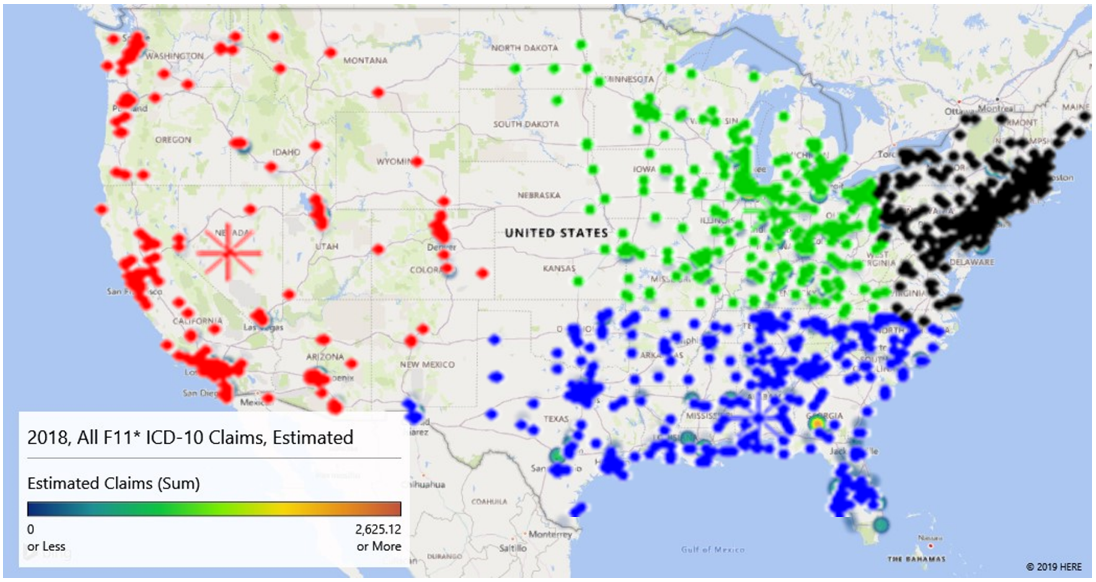

Figure 6. A cluster analysis is super-imposed on the 2018 map of ICD-10 F11 claims.

3.8. Explanatory Modeling Results

The first explanatory model, stepwise regression, investigated the number of inpatient opioid claims as a function of the independent variables. Models were built on an 80% training set and applied to 20% blinded test set for analysis of performance. The final stepwise model, the one with the smallest Akaike Information Criterion, included 1) staffed beds, 2) discharges, 3) emergency room visits, 4) surgeries, 5) assets, 6) affiliated physicians, 7) percent Medicare / Medicaid, 8) medical school affiliation, 9) hospital type, 10) year, and 11) state. Unfortunately, this model was only able to account for 17.73% of the dependent variable's variability. The root mean squared error (RMSE) of the forecast predictions was 1.76. The largest contributions to the model were from the ER Visits (SS=1.49, 1 df) and from the state (SS=1.25, 51 df). All variables in the model were statistically significant largely, due to sample size. The overall effect size, however, is small.

Lasso, ridge, and elastic net regression models built using "glmnet" [27] provided only slightly more variance capture with $R^2 = \{17.82\%, 17.77\%, 17.77\%\}$, respectively. The RMSE's were 1.75 for all three models. The elastic net selected a lasso model by assigning parameter $\alpha=0$. These models produced are essentially equivalent to the stepwise regression analysis.

Gradient-boosted random forests [28] performed well on the unobserved test set and untransformed data, achieving an $R^2=.878$ with hyperparameter tuning (depth of 6 trees, 500 rounds, learning rate of .1). To compare the results more fairly with the regression models, the same random forest configuration was run on the transformed data resulting in $R^2=.550$ and an RMSE=.06. Figure 7 is a plot of the observed claims versus the random forest predicted claims for the training and unobserved test set data. From this plot, it appears reasonable to forecast demand for opioid inpatient services based on factors important to the random forest model. The implication for policymakers at the state and local level is that resource allocation might reasonably to treat opioid abuse might reasonably be based on these models.

 doi:10.20944/preprints201905.0165.v1

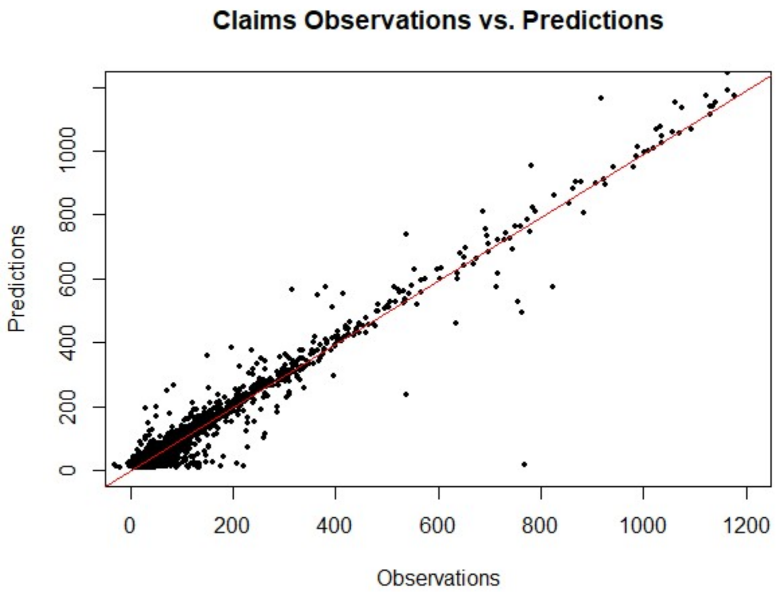

Figure 7. Predictions vs. observations for the claim data based on an extreme gradient boosting random forest provides reasonable predictive accuracy. The R^2 for the fit on the entirety of the data using the model built only on the training set is .965.

Figure 7 is a side-by-side plot of the gain (improvement of an estimate when a feature is used in a tree) and cover (the average proportion of samples affected by splitting using this feature) for the top five items in the importance matrix. The most important features for predicting the F11 opioid claims appear to be the staffed beds (10.1% gain and 5.5% cover), surgeries (9.8% gain and 3.6% cover), and liabilities (7.3% gain and 6.2% cover). Most interesting is that workload and financial variables are the most explanatory. Table 2 shows the top 10 most important features by gain. Because of their predictive accuracy, random forests may be used by policymakers to assign funds and resources to states and localities based on the estimated inpatient demand.

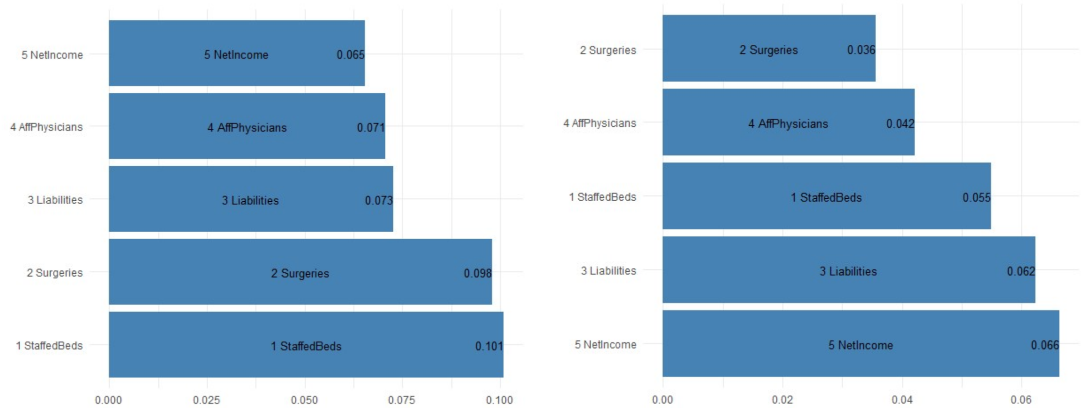

Figure 7. The gain chart (right) and the cover chart (left) show that hospital overall workload and financial variables are explanatory to opioid F11 admissions.

Table 2. This table provides the gain for the top 10 most important features.

Feature	Gain
Staffed Beds	10.09%
Surgeries	9.79%
Liabilities	7.27%
Affiliated Physicians	7.07%
Net Income	6.54%
ER Visits	5.46%
% Medicare/caid	5.32%
Employees	5.20%
Year 2018	5.04%
Illinois	4.35%

4. Discussion

The opioid abuse problem in the United States is non-static. While the US may have seen a decline in prescriptions from 2012 to the present, the average days of supply per prescription has increased [29] and illicit provider activity continues. The contribution of this illicit activity to the problem is likely to intensify the epidemic, which requires no additional assistance. In fact, a March 2018 CDC report showed a 35% increase in ER visits for the 16 states most affected by opioids. [30]Policymakers need to consider additional provider controls to ensure that opioids are distributed in accordance with the law.

The GIS mapping of F11 ICD-10 cases through 2018 suggests that the intensity of the epidemic is not fading and that there are new growth areas emerging including areas like Salt Lake City, Phoenix, and Las Vegas. Further, the unsupervised case cluster analysis is suggestive of the flow of drugs through Mexico, Canada, and the Dominican Republic (Northeast) as indicated by DEA analysis. Policymakers should consider funding prevention, treatment, and interdiction activities according to the GIS trends and demand for inpatient services and should focus analytical techniques to the most highly afflicted cities to target illicit activity by providers.

The gradient-boosted random forest model was effective in estimating the demand for inpatient services associated with ICD-10 F11. This type of model may be used policymakers for the allocation of resources and funding to appropriate states, zip codes, or even hospitals. The model suggests that hospital technical and workload factors are important in determining the demand for inpatient services. Specifically, the most important features for predicting the F11 opioid claims appear to be the staffed beds (10.1% gain and 5.5% cover), surgeries (9.8% gain and 3.6% cover), and liabilities (7.3% gain and 6.2% cover). Further analysis of facilities with high demand might be indicative of illicit actors in the community, either individual or otherwise. Such a finding would help prioritize interdiction efforts (enforcement and prevention) and potentially reduce the requirement for treatment, treatment that cost the Federal Government alone $7.4 billion in 2018.

5. Conclusions

This research is largely descriptive and explanatory in nature, yet it provides some insights about the spread of the opioid epidemic over time and space. This battle is likely to continue for the near future, and with limited assets, policymakers will have to use techniques like those presented here to allocate resources for supply-side and demand-side interventions (prevention and enforcement). While the research only focused on inpatient admission (exceedingly resource intensive), analogous studies for outpatient visits and deaths might be done. This research team will continue describing, explaining, and forecasting opioid-related incidents.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2019 doi:10.20944/preprints201905.0165.v1

Appendix 1. Opioid-Related ICD-10 Codes

F11 Opioid related disorders

F11.1 Opioid abuse

F11.10 …… uncomplicated

F11.11 …… in remission

F11.12 Opioid abuse with intoxication

F11.120 …… uncomplicated

F11.121 …… delirium

F11.122 …… with perceptual disturbance

F11.129 …… unspecified

F11.14 …… with opioid-induced mood disorder

F11.15 Opioid abuse with opioid-induced psychotic disorder

F11.150 …… with delusions

F11.151 …… with hallucinations

F11.159 …… unspecified

F11.18 Opioid abuse with other opioid-induced disorder

F11.181 Opioid abuse with opioid-induced sexual dysfunction

F11.182 Opioid abuse with opioid-induced sleep disorder

F11.188 Opioid abuse with other opioid-induced disorder

F11.19 …… with unspecified opioid-induced disorder

F11.2 Opioid dependence

F11.20 …… uncomplicated

F11.21 …… in remission

F11.22 Opioid dependence with intoxication

F11.220 …… uncomplicated

F11.221 …… delirium

F11.222 …… with perceptual disturbance

F11.229 …… unspecified

F11.23 …… with withdrawal

F11.24 …… with opioid-induced mood disorder

F11.25 Opioid dependence with opioid-induced psychotic disorder

F11.250 …… with delusions

F11.251 …… with hallucinations

F11.259 …… unspecified

F11.28 Opioid dependence with other opioid-induced disorder

F11.281 Opioid dependence with opioid-induced sexual dysfunction

F11.282 Opioid dependence with opioid-induced sleep disorder

F11.288 Opioid dependence with other opioid-induced disorder

F11.29 …… with unspecified opioid-induced disorder

F11.9 Opioid use, unspecified

F11.90 …… uncomplicated

F11.92 Opioid use, unspecified with intoxication

F11.920 …… uncomplicated

F11.921 …… delirium

F11.922 …… with perceptual disturbance

F11.929 …… unspecified

F11.93 …… with withdrawal

F11.94 …… with opioid-induced mood disorder

F11.95 Opioid use, unspecified with opioid-induced psychotic disorder

F11.950 …… with delusions

F11.951 …… with hallucinations

F11.959 …… unspecified

F11.98 Opioid use, unspecified with other specified opioid-induced disorder

F11.981 Opioid use, unspecified with opioid-induced sexual dysfunction

F11.982 Opioid use, unspecified with opioid-induced sleep disorder

F11.988 Opioid use, unspecified with other opioid-induced disorder

F11.99 …… with unspecified opioid-induced disorder

 doi:10.20944/preprints201905.0165.v1

Author Contributions: Lawrence Fulton was responsible for conceptualization, writing, and methodology. Zhijie Dong was responsible for methodology, validation, and writing. Ben Zhan was responsible for methodology and validation. Clemens Scott Kruse was responsible for validation and investigation. Paula Stigler Granados was responsible for validation, writing, review, and editing.

Conflicts of Interest: "The authors declare no conflict of interest."

References

1. *What is the U.S. Opioid Epidemic?* 2017 [cited 2019 May 12]; Available online: http://www.webcitation.org/78Jgiy2wz, (accessed on May 12, 2019).
2. Center, B.P. *Tracking federal funding to combat the opioid crisis*. 2019 [cited 2019 May 12]; Available online: https://bipartisanpolicy.org/library/tracking-federal-funding-to-combat-the-opioid-crisis/, (accessed on May 12, 2019).
3. McCance-Katz, E. *The national survey on drug use and health*. 2017 [cited 2019 May 12]; Available online: https://www.samhsa.gov/data/sites/default/files/nsduh-ppt-09-2018.pdf, (accessed on May 12, 2019).
4. *Nearly 60 Doctors, Other Medical Workers Charged In Federal Opioid Sting*. 2019 [cited 2019 12 May]; Available online: https://www.npr.org/2019/04/17/714014919/nearly-60-docs-other-medical-workers-face-charges-in-federal-opioid-sting, (accessed on 12 May, 2019).
5. *Jury finds podiatrist guilty of operating pill mill*. 2019 [cited 2019 12 May]; Available online: https://www.justice.gov/usao-ndga/pr/jury-finds-podiatrist-guilty-operating-pill-mill, (accessed on 12 May, 2019).
6. *Virginia Doctor Convicted on 861 Federal Counts of Drug Distribution, Including Distribution Resulting in Death: Faces Mandatory Minimum of 20 Years in Federal Prison*. 2019 [cited 2019 12 May]; Available online: https://www.justice.gov/usao-wdva/pr/virginia-doctor-convicted-861-federal-counts-drug-distribution-including-distribution, (accessed on 12 May, 2019).
7. *Southern District of Florida Charges 124 Individuals Responsible for $337 Million in False Billing as Part of National Healthcare Fraud Takedown*. 2018 [cited 2019 12 May]; Available online: https://www.justice.gov/usao-sdfl/pr/southern-district-florida-charges-124-individuals-responsible-337-million-false-billing, (accessed on 12 May, 2019).
8. *Fentanyl flows from China: an update since 2017*. 2018 [cited 2019 12 May]; Available online: https://www.uscc.gov/sites/default/files/Research/Fentanyl%20Flows%20from%20China.pdf, (accessed on 12 May, 2019).
9. *China policies to promote local production of pharmaceutical products and protect health*. Vol. Licence: CC BY-NC-SA 3.0 IGO. 2017, Geneva: World Health Organization.
10. Deprez, E., L. Hui, and K. Wills, *Deadly Chinese fentanyl is creating a new era of drug kingpins*, in *Bloomberg*. 2018.
11. *China's fentanyl production, export fueling U.S. opioid crisis, Schumer says*. 2018 [cited 2019 12 May]; Available online: https://www.amny.com/news/china-fentanyl-schumer-1.18531808, (accessed on 12 May, 2019).
12. Dudley, S., et al. *Mexico's role in the deadly rise of fentanyl*. 2019 [cited 2019 12 May]; Available online: https://www.insightcrime.org/wp-content/uploads/2019/02/Fentanyl-Report-InSight-Crime-19-02-11.pdf, (accessed on 12 May, 2019).
13. *Definitive Healthcare*. 2019 [cited 2019 12 May]; Available online: http://www.webcitation.org/788Bun3Wb, (accessed on 12 May, 2019).
14. Gehlenborg, N. and B. Wong, *Heat maps are useful for visualizing multivariate data but must be applied properly.* Nature Methods, 2012. **9**(213): p. online.
15. Lee, D.C., et al., *Using Geospatial Analysis and Emergency Claims Data to Improve Minority Health Surveillance.* Journal of Racial and Ethnic Health Disparities, 2018. **5**(4): p. 712-720.
16. MacQuillan, E.L.A.-O.h.o.o., et al., *Using GIS Mapping to Target Public Health Interventions: Examining Birth Outcomes Across GIS Techniques.* (1573-3610 (Electronic)).
17. Yasobant, S., K.S. Vora, and A. Upadhyay, *Geographic Information System Applications in Public Health: Advancing Health Research*. 2019: IGI Global. 24.
18. Hastie, T., R. Tibshirani, and J.H. Friedman, *The elements of statistical learning* 2nd ed. Springer series in statistics,. 2017, New York, NY: Springer.
19. *R: A Language and Environment for Statistical Computing*. 2018, R Core Team.
20. *Microsoft Excel*. 2019, Microsoft: Seattle, Washington.

 doi:10.20944/preprints201905.0165.v1

21. Kassambra, A., *ggcorplot: Visualization of a correlation matrix using 'ggplot2'*. 2018.
22. Weisburg, S. and J. Fox, *A {R} companion to applied regression*. 2 ed. 2011, Thousand Oaks, CA: Sage.
23. Box, G.E.P. and D.R. Cox, *An Analysis of Transformations.* Journal of the Royal Statistical Society. Series B (Methodological), 1964. **26**(2): p. 211-252.
24. Wei, T. and V. Simko, *R package "corrplot:" Visualization of a correlation matrix*. 2017.
25. *New Jersey's drug addiction crisis*. 2019 [cited 2019 12 May]; Available online: https://www.njtvonline.org/addiction/, (accessed on 12 May, 2019).
26. Ekin, T., R. Musal, and L. Fulton, *Overpayment models for medical audits.* Journal of Applied Statistics, 2015. **42**(11): p. 2391-2405.
27. Friedman, J., T. Hastie, and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent.* Journal of Statistical Software, 2010. **33**(1): p. 1-22.
28. Chen, T., et al., *xgboost: Extreme gradient boosting*. 2018.
29. *Data trends: Opioid prescribing, overdose deaths, and more*. 2019 [cited 2019 12 May]; Available online: http://core-rems.org/data-trends-opioid-prescribing-overdose-deaths-and-more/, (accessed on 12 May, 2019).
30. *Emergency department data show rapid increases in opioid overdoses*. [cited 2019 12 May]; Available online: https://www.cdc.gov/media/releases/2018/p0306-vs-opioids-overdoses.html (accessed on 12 May, 2019).