Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Surupa Sarkar Date

Estimating County-Level Drug Overdose Mortality in Georgia with

Mixed-Effects Poisson Regression Modeling

By

Surupa Sarkar
Master of Public Health

Department of Biostatistics

_____ Thesis Advisor's signature

Lance A. Waller, PhD
Thesis Advisor

_____ Reader's signature

Xiangqin Cui, PhD
Reader

Estimating County-Level Drug Overdose Mortality in Georgia with Mixed-Effects Poisson Regression Modeling

By

Surupa Sarkar

B.A.
University of Washington
2017

Thesis Committee Chair: Lance A. Waller, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics
2020

Abstract

Estimating County-Level Drug Overdose Mortality in Georgia with Mixed-Effects Poisson Regression Modeling

By Surupa Sarkar

Background: With the recent increase in stimulant and cocaine drug overdose mortality in Georgia and the continuing impact of the opioid epidemic, it is critical for public health researchers to have reliable methods to estimate and evaluate drug overdose mortality rates and risk factors. This study used statistical modeling to predict the drug overdose public health burden in Georgia counties. A multivariable Poisson mixed-effects model was developed to estimate county-level drug overdose mortality rates using 2017 opioid epidemic data.

Methods: An empirical review was performed on 73 county-level indicators to assess each indicator for potential association with drug overdose mortality. Principal component analysis was implemented for dimension reduction and was followed by a multicollinearity assessment and univariate analysis. Stepwise selection methods were performed, and potential effect modification was explored to finalize the predictive model. Final model fit was assessed by comparing estimated drug overdose mortality cases and spatially smoothed rates to the observed values.

Results: The predictors significantly associated with drug overdose mortality were Race (β=0.022; $p<0.001$), Opioid and Benzodiazepine Prescription Overlap (β=0.062; $p<0.001$), STD Rate (β=0.001; $p<0.001$), Opioid Prescription Length (β=-0.090; $p=0.010$), and Vehicle Inaccessibility (β=-0.036; $p=0.032$). A significant interaction was found between STD Rate and Race (β=-1.136e-05; $p=0.018$). 83% of predicted drug overdose mortality estimates were within 1 case of the observed 2017 death cases, 91% were within 2 cases, and 95% were within 3 cases. After performing spatial smoothing on the estimated cases for 2017, Bacon, Gilmer, Pickens, Fannin, and Haralson were identified as counties with the largest estimated smoothed mortality rates.

Conclusion: The final model provides researchers with a tool to identify which Georgia counties may demonstrate high drug overdose mortality rates and counts based on race, opioid and benzodiazepine prescription overlap, vehicle inaccessibility, opioid prescription length, and STD rate. The model's accuracy in estimating the mortality cases for 2016 and 2017 indicates it is a helpful tool to understand the spatial spread of drug overdose mortality throughout Georgia. It is important for public health researchers to explore the identified risk factors further to understand how preventative measures can be implemented for high-risk counties in Georgia.

Estimating County-Level Drug Overdose Mortality in Georgia with Mixed-Effects Poisson Regression Modeling

By

Surupa Sarkar

B.A.
University of Washington
2017

Thesis Committee Chair: Lance A. Waller, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics
2020

Acknowledgements

I would like to express my sincere gratitude to my thesis advisor, Dr. Lance Waller. I am greatly appreciative of your guidance and support throughout the thesis process and throughout my education as a biostatistics student. I have learned an incredible amount from you during my time at Emory University. Thank you for always sharing your wisdom and insights.

I would also like to extend my appreciation to my mentors at the Georgia Department of Public Health, Lee Hundley and Michael Thomas. Thank you for offering your help and suggestions with all the projects I have been lucky to collaborate with you on. Your excellent mentorship provided me with a greatly rewarding experience at the Georgia Department of Public Health. Your dedication towards your research inspired me to explore different exciting areas of biostatistics and public health.

Table of Contents

1. Introduction

1.1. *Drug Overdose Mortality in the United States*

Drug overdose mortality in the United States persists as a major public health concern. In 2017, more than 70,200 Americans died from drug-involved events (Hedegaard et al., 2018). The number of drug overdose deaths in 2017 surpassed the number of deaths caused by homicides, vehicle accidents, gun violence, or HIV/AIDS in any year in American history (*Drug Overdose | Drug Policy Alliance*, n.d.). According to the Centers for Disease Control and Prevention, the age-adjusted rate of American drug overdose mortality in 2017 was 21.7 deaths per 100,000 population, which was a 9.6% increase from the rate in 2016. Between 1999 and 2017, the age-adjusted rate of drug overdose mortality increased for both sexes and all age groups (Hedegaard et al., 2018). Approximately 68% of drug-involved deaths in 2017 involved prescription or illicit opioids (*Understanding the Epidemic | Drug Overdose | CDC Injury Center*, n.d.). More than 28,000 Americans died from involvement with synthetic opioids other than methadone, such as fentanyl, fentanyl analogs, and tramadol in 2017, which led to a 45.2% increase in the rate of drug overdose mortality from 2016 to 2017 (*Synthetic Opioid Overdose Data | Drug Overdose | CDC Injury Center*, n.d.). Methadone, heroin, and natural and semisynthetic opioids such as morphine, codeine, hydrocodone, and oxycodone contributed to nearly the same rate of drug overdose mortality in both 2016 and 2017 (Hedegaard et al., 2018). Resultingly, opioid misuse remains the most pressing concern when addressing the burden of drug overdose mortality in the United States. Non-opioid drug misuse also continues to pose a major public health threat. Between 2016 and 2017, the drug overdose mortality rate for deaths involving cocaine increased by 34%, and the rate for deaths involving

psychostimulants increased by 37% (*Other Drugs | Drug Overdose | CDC Injury Center*, n.d.). As a result of these trends, in October 2017, the opioid crisis was declared a national public health emergency (*Opioid Epidemic | Georgia Department of Public Health*, n.d.).

1.2. *Drug Overdose Mortality and Surveillance in Georgia*

The state of Georgia has not been spared from the national opioid epidemic. Georgia saw a 245% increase (426 to 1,043 deaths) in opioid-involved mortality between 2010 to 2017. The year 2013 marked the beginning of a sharp increase in the number of deaths involving illicit opioids which includes heroin and fentanyl (Georgia Department of Public Health, 2017). This trend increased until 2018 when Georgia saw a 14% yearly decrease (996 to 873 deaths) in opioid-involved overdose mortality which was the first decrease seen since 2013 (Georgia Department of Public Health, 2018). However, the number of overdose deaths involving stimulant and cocaine misuse continued to rise with a 11% increase (631 to 703 deaths) and an 18% increase (289 to 340 deaths), respectively, from 2017 to 2018 in Georgia (Georgia Department of Public Health, 2018).

The Georgia Department of Public Health's Drug Surveillance Unit routinely monitors drug overdose morbidity and mortality in Georgia with the goal of providing accurate data to public health partners to combat the opioid epidemic. The Drug Surveillance Unit also takes on the responsibility of identifying and responding to sudden increases or trends of drug overdose across the state (*Drug Surveillance Unit | Georgia Department of Public Health*, n.d.). To efficiently anticipate counties or regions of Georgia that may require immediate public health assistance, it is crucial that the Georgia Department of Public Health be able to accurately predict drug overdose mortality and morbidity rates and identify Georgia counties most vulnerable to drug overdose.

1.3. *Public Health Responses to the Opioid Epidemic*

Over the last decade, policymakers, medical physicians, and public health researchers collaborated to develop and propose numerous initiatives in hopes of ending the nation's opioid epidemic. The American Medical Association created an Opioid Task Force with the goal of providing important suggestions to physicians to eliminate the opioid crisis (*Reversing the opioid epidemic | American Medical Association*, n.d.). The Opioid Task Force encourages physicians to use state prescription drug monitoring programs (PDMPs), support comprehensive care for patients with substance use disorders, end substance use disorder stigma, increase patient access to naloxone which is a drug that reverses opioid overdose, and promote proper storage and disposal of opioids (*Reversing the opioid epidemic | American Medical Association*, n.d.). The presidential administration also allocated $6 billion to fund measures to end opioid abuse. This national initiative aims to decrease the demand for opioids through education and preventative measures, prevent illicit drugs from entering United States borders, and increase options for evidence-based treatment for opioid addiction (*Ending America's Opioid Crisis | The White House*, n.d.). In September 2018, the presidential administration supplied over $1 billion to organizations including the Substance Abuse and Mental Health Services Administration, Health Resources and Services Administration, and Centers for Disease Control and Prevention to battle the ongoing opioid crisis (*President Donald J. Trump's Initiative to Stop Opioid Abuse and Reduce Drug Supply and Demand*, n.d.).

The Georgia Department of Public Health developed three programs to combat the statewide opioid epidemic. The first program, the Opioid and Substance Misuse Response Unit, leads various strategic planning efforts to respond to opioid misuse in Georgia. The

unit develops response plans to address sudden increases in overdose numbers and creates campaigns and partnerships with external organizations to increase funding available to address the epidemic in Georgia (*Georgia's Opioid Response | Georgia Department of Public Health*, n.d.). The second state program, the Prescription Drug Monitoring Program, is an electronic system that tracks information on controlled substance prescriptions and dispensations. The purpose of this database is to end overprescribing and provide physicians with patient prescription history. This system ensures that physicians are aware of patient behaviors that may contribute to drug misuse (*Prescription Drug Monitoring Program | Georgia Department of Public Health*, n.d.). The third state program is the Drug Surveillance Unit which monitors and collects drug overdose morbidity and mortality data (*Drug Surveillance Unit | Georgia Department of Public Health*, n.d.). Collectively, these state programs work to ensure the Georgia Department of Public Health has the financial, personnel, and technological resources required to address and tackle the opioid crisis.

1.4. *Literature Review*

Because the opioid epidemic is the most dangerous drug crisis in American history, numerous public health researchers have explored potential risk factors for opioid-involved morbidity and mortality by developing statistical models to identify significant associations for different patient populations. Glanz and others (2018) developed a prediction model for two-year risk of opioid overdose among patients prescribed chronic opioid therapy. In their study, they used Cox proportional hazards regression to understand the relationships between baseline predictors and opioid overdose incidence. They found age, mental health diagnosis, substance abuse diagnosis, tobacco use, and Hepatitis C diagnosis to be significantly associated with opioid overdose trends (Glanz et al., 2018). Similarly, Liang

and others (2016) developed sex-specific prediction models for drug overdose among opioid users with non-cancer pain. Using logistic regression, they found pain conditions, mental illness, and alcohol or substance abuse to be significantly associated with drug overdose for both sexes (Liang et al., 2016). Several other public health researchers used a variety of statistical and machine learning methods to perform similar analyses for other patient populations including patients with chronic pain and Medicare beneficiaries with opioid prescriptions (Dunn et al., 2010; Lo-Ciganic et al., 2019). These studies frequently demonstrated that, not surprisingly, mental illness and substance abuse diagnoses are commonly associated with drug overdose events.

Similar research focused on populations from specific geographic regions, instead of for patient populations, was limited and challenging to find. The most comparable studies performed for a specific geographic location were vulnerability assessments which identified state counties at greatest risk for HIV/HCV infection. Van Handel and others (2016) from the Centers for Disease Control and Prevention performed a vulnerability assessment for county-level HIV/HCV outbreak. Opioid epidemic data were used because of the increasing link between HIV outbreak caused by sharing of needles among injection drug users. Van Handel and others (2016) identified 15 predictors within a final multilevel Poisson model including several socioeconomic factors such as education, income, race, and unemployment. This model was used to identify United States counties falling above the 95^{th} percentile for vulnerability to HIV/HCV. Forty-one of the 220 most vulnerable counties in the U.S. were found to be in Tennessee (Van Handel et al., 2016). As a result, Rickles and others (2018) from the Tennessee Department of Public Health performed a similar vulnerability assessment for Tennessee county-level HIV/HCV outbreak using

opioid epidemic data, again using a multivariable Poisson regression model (Rickles et al., 2018). These vulnerability assessments were successful in identifying the most vulnerable counties based on significant epidemiological risk factors. However, the statistical models developed from this research were not created to predict or improve estimation of the county-level mortality rate estimates themselves.

The Georgia Department of Public Health has not yet developed a statistical model to evaluate potential drug overdose risk factors or predict drug overdose mortality rate estimates in Georgia counties. Several previous studies have developed statistical models to predict opioid overdose mortality for particular patient populations, but not predictions or estimates of drug overdose mortality for a geographic location. Because the drug overdose mortality rates involving stimulants and cocaine continue to increase in Georgia while the opioid overdose mortality rate slightly decreased in 2018, it is critical to consider the burden of non-opioid drugs along with opioids when predicting drug overdose mortality rates in Georgia.

1.5. *Purpose*

The objective of this thesis is to address a gap in statistical modeling used to predict the drug overdose public health burden in Georgia by developing a multivariable Poisson model to better estimate county-level drug overdose mortality rates. To achieve this, we used 2017 opioid epidemic data from the Georgia Department of Public Health that were originally used to perform a county-level HIV/HCV vulnerability assessment in Georgia. This thesis will discuss the statistical methods implemented for dimension reduction and model selection to identify a prognostic statistical model that can be used by public health researchers in Georgia to estimate county-level drug overdose mortality rates and identify

counties that are at greatest risk for drug overdose mortality. Understanding the epidemiological risk factors that play the greatest roles in the opioid epidemic in Georgia will allow public health professionals to more efficiently plan efforts to respond to opioid misuse in Georgia and to monitor geographic areas of highest risk.

2. Methods

2.1. *Study Design*

An ecologic study design was used with the 159 counties in the state of Georgia serving as the study sample. In 2018, the Drug Surveillance Unit at the Georgia Department of Public Health (GDPH) collected 2016 and 2017 county-level measures relevant to HIV/HCV transmission which were used for this thesis. These data were used by epidemiologists at GDPH to perform a county-level vulnerability assessment for both HIV/HCV infection and opioid overdose. Epidemiologists at GDPH collected the data from a variety of sources including the Centers for Disease Control and Prevention, US Census Bureau, US Centers for Medicare & Medicaid Services, US Department of Health and Human Services, Georgia Prescription Drug Monitoring Program, Georgia Department of Community Health, Georgia Department of Transportation, and GDPH's HIV/HCV and Drug Surveillance teams. Eighty-nine county-level variables were collected out of which 16 were specific to 2016. These 16 indicators were removed and 73 indicators that were specific to 2017 remained.

2.2. *Outcome Variable*

To evaluate county-level incidence to drug overdose mortality, 2017 mortality counts from GDPH's Online Analytical Statistical Information System were obtained. Drug overdose mortality was defined by the Centers for Disease Control and Prevention's specified ICD-10 (International Classification of Diseases) codes for all drug poisoning mortality (Table 1) (CDC NCIPC DUIP PDO Team, n.d.). All drug overdose mortality was chosen as the outcome of interest due to the increasing health burden of non-opioid drug misuse in Georgia. County-level drug overdose mortality counts were modeled with a multivariate

Poisson regression model that included an offset to adjust for differing county population sizes. A log link function was used with the generalized linear regression model so the conditional mean of the outcome, expected drug overdose mortality rate, was modeled. A random intercept for county was also incorporated into the model to account for spatial heterogeneity. The model was given by:

$y_i \sim Poisson(\lambda_i)$

$log(\lambda_i) = log(County\ Population) + \beta_0 + \boldsymbol{x\beta} + \theta_i$

$\theta_i \sim N(0, \tau^2)$

where,

$log(County\ Population) = \text{offset}$

$\beta_0 = \text{overall intercept}$

$\theta_i = \text{difference between mean of county i and overall intercept}$

$\beta_0 + \theta_i = \text{county i's intercept}$

$\boldsymbol{\beta} = \text{vector of coefficients}$

$\tau^2 = \text{heterogeneity variance}$

The model assumed that the random coefficients θ_i were independent with a mean of 0 and normally distributed for all *i.* The model also assumed that the drug overdose mortality counts in each county were conditionally independent of one another, given the county-level random effects.

2.3. *Potential Model Predictors*

Empirical Review

73 county-level indicators were investigated for potential association with drug overdose mortality. Individual distributions of each indicator were reviewed, and only 2 indicators,

household income and per-capita income, were log-transformed. The first step in identifying potential predictors was conducting an empirical review. Each of the 73 indicators was assessed for potential association with drug overdose mortality based on findings from prior research and an understanding of the risk factors for drug overdose mortality. After this review, 23 indicators were removed. The remaining 50 indicators were examined for similarity. This assessment identified 5 categories that each of the 50 indicators fell into: 1) Employment/Income, 2) Drug-related activity, 3) HIV/HCV infection status, 4) Household measures, and 5) Community measures. The 50 indicators were grouped into 1 of the 5 categories (Table 2). If multiple indicators in a category collected similar measurements, the indicator that had the strongest association with drug overdose mortality was retained. 17 variables were eliminated during empirical review leaving 33 remaining indicators for further analysis.

Principal Component Analysis

Principal component analysis was used for dimension reduction. Principal component analysis is a widely used technique that reduces dimensionality of data while maintaining the variability of the original data. Principal component analysis uses orthogonal transformation of the data to alter a set of potentially correlated variables into a new set of uncorrelated variables. The new uncorrelated variables are called principal components and are created with linear functions of variables from the original data (Jolliffe & Cadima, n.d.). A covariance matrix is generated with the eigenvalues and eigenvectors from the data. Eigenvectors with the largest eigenvalues represent the primary principal components. The first principal component accounts for the greatest proportion of total variability in the data, and the succeeding principal component accounts for the second

largest proportion of total variability. This decreasing trend continues with the last principal component explaining the minimum proportion of total variability in the data (*Principal Component Analysis 4 Dummies: Eigenvectors, Eigenvalues and Dimension Reduction – George Dallas*, n.d.). This procedure is depicted below.

Let the data be represented by

$$\boldsymbol{\alpha'_k x} = \sum_{j=1}^{n} \alpha_{kj} x_j$$

where $\boldsymbol{x}$ is a vector of n random variables and $\boldsymbol{\alpha_k}$ is a vector of n constants.

During principal component analysis, a linear function of $\boldsymbol{x}$ and $\boldsymbol{\alpha'_1 x}$ with maximum variance is found and identified as the first principal component. Next, a linear function of $\boldsymbol{x}$ and $\boldsymbol{\alpha'_2 x}$ that is uncorrelated with $\boldsymbol{\alpha'_1 x}$ and has maximum variance is identified. This is the second principal component. This process is continued for all $\boldsymbol{\alpha'_k x}$. The goal of principal component analysis is to identify m principal components that explain almost all the variability in $\boldsymbol{x}$ where $m < n$ (Wood, 2009).

Using R version 3.6.1, the predictors were standardized to have a mean of 0 and a standard deviation of 1 with the scale function (R Core Team, 2019). The binary indicator for highway presence was not included in this step since principal component analysis is intended for continuous data. Twenty-six of the 32 principal components accounted for 98.3% of the total variability in the data (Table 3). A Scree plot was created to display the proportion of variance explained by each principal component (Figure 1). Six principal components accounted for less than 2% of the data's variability, indicating not all 32 variables were required for analysis.

Principal component analysis can also be conducted using a correlation matrix as opposed to a covariance matrix. This method uses the correlation matrix to calculate principal

component loadings. Principal component loadings are the correlation coefficients between each variable and each principal component (*What is Variable Loadings in PCA? - Articles - STHDA*, n.d.). Using SAS version 9.4 software, standardized variables with a principal component loading less than 0.40 were removed (SAS Institute Inc, 2018). A loading less than 0.40 indicates the variable has low correlation with the principal component and does not contribute much to the overall variance of the data. There was only 1 indicator with a principal component loading less than 0.40 that was removed. Remaining variables with a principal component loading less than 0.50 were then identified. These 5 variables were removed, and 26 continuous indicators remained.

Collinearity Assessment

To assess any remaining collinearity or multicollinearity in the data, a correlation matrix for the remaining 26 continuous indicators was constructed. Any combination of indicators with a correlation coefficient greater than 0.80 or less than -0.80 was identified. Using this standard, multicollinearity was found among 4 predictors (Figure 2). 1 of the 4 predictors also showed collinearity with a 5^{th} predictor (Figure 3). To select which of these 5 predictors to retain, univariate Poisson regression models with and without a random intercept were constructed. The predictor with the most significant relationship with drug overdose mortality based on the lowest p-value was selected. This predictor showed the strongest association with the outcome for both the random intercept models and the marginal models. The other 4 predictors were removed, and 22 continuous predictors remained for additional analysis.

2.4. Regression Modeling

Univariate Modeling

The remaining 23 variables (22 continuous and 1 binary) were examined for significant association with drug overdose mortality by fitting univariate Poisson regression models. A random-intercept model and marginal model was constructed for each of the 23 predictors. Only 8 predictors demonstrated a significant association with drug overdose mortality based on a resulting p-value less than 0.10 in both the random-intercept model and the marginal model.

Multivariate Modeling

The remaining 8 predictors were fit in a multivariable Poisson regression model. At a significance level of 0.10, 3 predictors in the model were not significantly associated with drug overdose mortality, although they were significant in the univariate models. Stepwise selection methods were used on the multivariate model. First, backwards selection based on the Akaike information criterion (AIC) was performed which resulted in the elimination of 2 of the 3 insignificant predictors. Forward selection based on AIC did not result in any changes to the original model. To ensure a significant relationship between all predictors and the outcome, the 3 insignificant variables were removed resulting in 5 remaining predictors.

The next step in developing the prognostic model was considering potential effect modification. Each possible two-way interaction was fit in a multivariate random-intercept Poisson regression model. Five significant interaction terms were found based on a significance level of 0.10. However, only 2 of the 5 interaction terms resulted in models where all other predictors remained significant. Of these 2 interaction terms, the more

significant interaction term based on a lower p-value was selected. Including more than 1 interaction term also resulted in other model predictors losing significance. Therefore, the final model consisted of 6 predictors, one of which was an interaction term.

2.5. *Model Fit Assessment*

While fitting the final model in R version 3.6.1, warning messages suggested rescaling the predictor variables because some were on very different scales (R Core Team, 2019). To account for this warning, a second model was fit with standardized predictors. Just as the predictors were standardized for principal component analysis, the variables were standardized to have a mean of 0 and a standard deviation of 1 using the scale function in R version 3.6.1 (R Core Team, 2019). The final Poisson regression model was refit with the scaled predictors to create a second final model. The predictions of the unscaled and scaled Poisson regression models were compared alongside the observed 2017 drug overdose mortality counts for each county.

Estimating county-level drug overdose death rates in Georgia is difficult due to several counties with small populations and low overdose death counts, resulting in the small number problem. The small number problem occurs when rates calculated from regions with small populations and low death counts misleadingly display elevated rates due to limited data instead of true increased risk. To stabilize these county-level rate statistics, spatial smoothing methods can be utilized through which data from surrounding counties are incorporated to calculate a county-level rate estimate (Waller & Gotway, 2004). Geographic maps of adjacency-based smoothed drug overdose mortality rates were constructed for the observed and predicted values. County-level adjacency-based smoothed rates were calculated by taking the mean of the raw mortality rates from the county and all

neighboring counties. If $r_1, r_2, \ldots, r_n$ are the observed raw mortality rates, then the adjacency-based smoothed rate was calculated as (Waller & Gotway, 2004)

$$\tilde{r}_i = \frac{\sum_{j=1}^{N} w_{ij} r_j}{\sum_{j=1}^{N} w_{ij}}$$

$$N = number\ of\ counties\ in\ Georgia$$

where the weights are given by

$$w_{ij} = \begin{cases} 1, & if\ regions\ i\ and\ j\ share\ a\ boundary \\ 0, & otherwise \end{cases}$$

To account for any uncertainty in the model estimates, a simulation was run on the final two Poisson regression models while assuming the regression coefficients from the models were normally distributed. Using the regression coefficients and standard deviations of the significant predictors, 10,000 sets of regression coefficients were simulated for the scaled and unscaled model. The distributions of the regression coefficients were reviewed to assess the variability of the model prediction. In addition, the final unscaled Poisson regression model was used to predict drug overdose mortality using 2016 data in order to evaluate the model's predictive ability for non-2017 data.

3. Results

3.1. *Risk Factors Associated with Drug Overdose Mortality*

In the final Poisson regression models, all predictors were significant with p-values less than 0.05, although a significance level of 0.10 was established throughout the analysis. The unscaled predictors significantly associated with drug overdose mortality from most to least significant were Race (β=0.022; p<0.001), Opioid and Benzodiazepine Prescription Overlap (β=0.062; p<0.001), STD Rate (β=0.001; p<0.001), Opioid Prescription Length (β=-0.090; p=0.010), and Vehicle Inaccessibility (β=-0.036; p=0.032) (Table 4). A significant interaction was found between STD Rate and Race (β=-1.136e-05; p=0.018). The final unscaled Poisson regression model was given by:

$$y_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \log(\text{County Population}) + \beta_0 + \theta_i + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

$$\theta_i \sim N(0, \tau^2)$$

where,

$\beta_0 + \theta_i = \text{county i's intercept}$

$X_1 = \text{Opioid and Benzodiazepine Prescription Overlap}$

(Percent of days that patients had an opioid and benzodiazepine prescription on the same day)

X_2

$= \text{Vehicle Inaccessibility (Percent of total housing units with no vehicle available)}$

X_3

$= \text{Opioid Prescription Length (Average days of opioid analgesics supplied per prescription)}$

$X_4 = \text{STD Rate}$

(Age

$-$ adjusted STD rate per 100,000 population (all STD except Congenital Syphilis))

$X_5 = \text{Race (Percent of non} - \text{Hispanic or Latino that is white race alone)}$

$X_6 = \text{Interaction between } X_4 \text{ and } X_5$

Figure 5 displays the distribution of each regression coefficient after simulating 10,000 sets of regression coefficients while assuming the coefficients were normally distributed. Based on the simulation, the distribution of two variables, Vehicle Inaccessibility and Opioid Prescription Length, range from -0.09945 to 0.02749 and from -0.20579 to 0.03753, respectively. These ranges indicate that Opioid Prescription Length and Vehicle Inaccessibility can have either a positive or negative association with drug overdose mortality based on simulated values from the final model.

There existed minimal heterogeneity in the baseline counts for drug overdose mortality with a between-county standard deviation of 0.145. Based on the model coefficients, a one percent increase in non-Hispanic or non-Latino white residents, one day increase in days that patients had an opioid and benzodiazepine prescription on the same day, and unit increase in the age-adjusted STD rate increases the average county-level drug overdose mortality rate by 2.2%, 6.4%, and 0.1% respectively. On the other hand, a one day increase in the average days of opioid analgesics supplied per prescription, one percent increase in total housing units with no vehicle available, and unit increase in the interaction between percent of non-Hispanic or non-Latino white residents and age-adjusted STD rate decreases the average county-level drug overdose mortality rate by 8.6%, 3.5%, and 1.1e-05% respectively.

In the final model with scaled predictors, the predictors from most to least significantly associated with drug overdose mortality were Race (β=0.239; p<0.001), Opioid and Benzodiazepine Prescription Overlap (β=0.229; p<0.001), Opioid Prescription Length (β=-0.147; p=0.011), Vehicle Inaccessibility (β=-0.128; p=0.032), and STD Rate (β=0.131; p=0.037). This is the same order of significance that was seen from the model covariates in the unscaled model. The interaction term between STD Rate and Race was also significant (β=-0.063; p=0.019).

3.2. *County-Level Drug Overdose Mortality Estimates*

To predict drug overdose mortality cases, the final Poisson regression models were used to estimate the drug overdose mortality rate and calculate the estimated mortality count for each county. The drug overdose death count estimates from the unscaled and scaled Poisson regression models were identical for all counties, indicating the model with unscaled predictors can be used exclusively to model drug overdose mortality estimates in Georgia. Table 5 displays the predicted county-level drug overdose mortality cases compared to the observed cases in 2017. 83% of predicted drug overdose mortality estimates were within 1 case of the observed 2017 death cases, 91% were within 2 cases, and 95% were within 3 cases. Figure 4 shows that the northern and southeastern regions of Georgia contain the highest burden of drug overdose mortality in the state. Both the observed and estimated drug overdose mortality cases identified Fulton, Cobb, Gwinnett, DeKalb, and Richmond counties to have the greatest 2017 drug overdose death cases in Georgia. The counties with the highest observed adjacency-based smoothed drug overdose mortality rates were Camden, Charlton, Dade, Fannin, and Gilmer. After performing spatial smoothing on the estimated cases, Bacon, Gilmer, Pickens, Fannin, and Haralson

were identified as counties with the largest estimated smoothed rates. Both the observed and estimated smoothed mortality rates identified Fannin and Gilmer to have one of the 5 highest county-level drug overdose mortality rates in the state. However, 3 of the 5 highest predicted county-level smoothed mortality rates did not match the highest observed mortality rates.

To confirm the model can be used to estimate drug overdose mortality in Georgia for years other than 2017, the final model was also used to predict county-level drug overdose death counts using 2016 data (Table 6). 82% of predicted drug overdose mortality estimates were within 1 count of the observed 2016 death counts, 91% were within 2 counts, and 96% were within 3 counts. Both the observed and estimated drug overdose mortality counts identified Fulton, Cobb, Gwinnett, DeKalb, and Cherokee counties to have the greatest 2016 drug overdose death counts in Georgia. When modeling drug overdose mortality using 2016 data, Opioid Prescription Length (β=-0.069; p=0.141) and Vehicle Inaccessibility (β=-0.031; p=0.140) were no longer significant in the final Poisson regression model. However, a one percent increase in non-Hispanic or non-Latino white residents, one day increase in days that patients had an opioid and benzodiazepine prescription on the same day, and unit increase in the age-adjusted STD rate still increases the average county-level drug overdose mortality rate by 1.9%, 5.5%, and 0.1% respectively.

4. Discussion

4.1. *Understanding the Final Model*

The final model obtained from the analysis provides researchers with a tool to identify which Georgia counties may demonstrate high drug overdose mortality rates and counts based on race, opioid and benzodiazepine prescription overlap, vehicle inaccessibility, opioid prescription length, and STD rate. The northern and southeastern regions of Georgia carry the highest burden of drug overdose mortality in the state. Both regions also include counties with the highest percentages of white residents. The southeastern region of Georgia contains counties with the highest percentages of opioid and benzodiazepine prescription overlap and the highest STD rates. It is important for public health researchers to explore these risk factors further from an epidemiological perspective to understand how preventative measures can be implemented for these high-risk regions. Two of the model covariates, Opioid Prescription Length and Vehicle Inaccessibility, were negatively associated with drug overdose mortality. However, when the final model was fit with 2016 data, these variables became statistically insignificant indicating their inverse relationship with drug overdose mortality should be assessed with caution. The simulation of the regression coefficients also showed that the coefficient distributions for Opioid Prescription Length and Vehicle Inaccessibility ranged from negative values to positive values, further confirming the need to assess these variables with caution. While the final model should not be used exclusively to predict drug overdose mortality, its accuracy in estimating the death counts for 2016 and 2017 indicates it is a strong and helpful tool to gain an understanding of the spatial spread of drug overdose mortality throughout Georgia.

4.2. *Comparison with Prior Research*

Epidemiologists at GDPH performed a county-level vulnerability assessment for both HIV/HCV infection and opioid overdose using the same data that was used for this thesis analysis. While the goal of these vulnerability assessments was not to develop a predictive model, a final Poisson regression model was obtained to evaluate the vulnerability of each county to HIV/HCV infection and opioid overdose. The risk factors found to be significantly associated with opioid overdose in this assessment were opioid analgesic prescription rate, multiple opioid prescription overlap, opioid and benzodiazepine prescription overlap, sex, and percent of households with food stamp benefits. Out of these risk factors, only opioid and benzodiazepine prescription overlap was found to be significantly associated with drug overdose mortality in this thesis analysis. The difference in results could be attributable to the different outcome variables (opioid overdose versus drug overdose mortality). In addition, the vulnerability assessment analyzed data for both 2016 and 2017 while our model was developed from 2017 data only. It was surprising to see that no risk factors measuring income or employment remained in our final model. In future analysis, a variable measuring financial status could be forced into the final model to evaluate how the model estimates change when considering the effect of income or employment.

4.3. *Limitations and Challenges*

The greatest challenge in this thesis was utilizing variable selection methods for a generalized linear model with a Poisson distributed response variable and a random intercept. Variable selection techniques for simple linear regression are widely known and used. Such techniques include forward, backward, and stepwise selection. However, the

translation of these methods to the generalized linear model framework requires some adjustments and there is limited literature that discusses the effectiveness of these techniques for Poisson regression. Famoye and Rothe (2003) identified the AIC as a goodness-of-fit measure that can be used to compare Poisson regression models where a smaller AIC indicates stronger fit. The AIC was assessed in our analysis with the step function in R version 3.6.1 which performs variable selection based on the AIC (R Core Team, 2019). However, the AIC of each potential model was very similar which led us to eventually remove 2 covariates that were insignificant based on a high p-value. Famoye and Rothe (2003) also discussed a modified R^2 statistic that assesses goodness-of-fit and can be used to perform variable selection for Poisson regression. The modified R^2 statistic was not considered in our analysis but should be incorporated into the variable selection process for future analysis.

Limitations in our analysis arose when considering interaction terms for our final model. Including more than one two-way interaction term in our model led to other covariates losing significance based on their associated p-values. In addition, three-way interaction terms also resulted in this same issue. Due to this limitation, three-way interactions and multiple two-way interactions were not included in our final model which likely restricts our model's predictive capability. Variable selections methods for mixed-effect models are challenging to implement, and the random intercept term often leads to complications when considering the inclusion of interaction terms.

4.4. *Next Steps*

Further investigation should repeat similar model and variable selection methodology using data from multiple years, possibly from 2015 to 2020, to obtain a more

comprehensive model that reduces bias that may arise from outliers observed within any specific year. In addition, because the northern and southeastern regions of Georgia are impacted the most by drug overdose mortality, it would be beneficial to also include data from Georgia's surrounding states when developing a predictive model. There is a possibility that drug misuse activity in the border regions of Georgia's surrounding states also contribute to drug overdose mortality in Georgia's northern and southeastern regions. With the recent increase in stimulant and cocaine drug overdose mortality in Georgia and the continuing impact of the opioid epidemic, it is critical for public health researchers to have reliable methods to estimate and evaluate drug overdose mortality rates and risk factors.

5. References

1. CDC NCIPC DUIP PDO Team, H. (n.d.). *PRESCRIPTION DRUG OVERDOSE DATA & STATISTICS GUIDE TO ICD-9-CM AND ICD-10 CODES RELATED TO POISONING AND PAIN Introduction to ICD-9-CM and ICD-10 Codes Related to Poisoning and Pain*. Retrieved March 17, 2020, from http://www.cdc.gov/nchs/icd.htm
2. *Drug Overdose | Drug Policy Alliance*. (n.d.). Retrieved January 30, 2020, from http://www.drugpolicy.org/issues/drug-overdose
3. *Drug Surveillance Unit | Georgia Department of Public Health*. (n.d.). Retrieved January 30, 2020, from https://dph.georgia.gov/epidemiology/drug-surveillance-unit
4. Dunn, K. M., Saunders, K. W., Rutter, C. M., Banta-Green, C. J., Merrill, J. O., Sullivan, M. D., Weisner, C. M., Silverberg, M. J., Campbell, C. I., Psaty, B. M., & Von Korff, M. (2010). Opioid prescriptions for chronic pain and overdose: A cohort study. *Annals of Internal Medicine*, *152*(2), 85–92. https://doi.org/10.7326/0003-4819-152-2-201001190-00006
5. *Ending America's Opioid Crisis | The White House*. (n.d.). Retrieved January 30, 2020, from https://www.whitehouse.gov/opioids/
6. Famoye, F., & Rothe, D. E. (2003). Issue 2 Article 11 11-1-2003 Recommended Citation Famoye. Journal of Modern Applied Statistical Methods, 2(2), 11. https://doi.org/10.22237/jmasm/1067645460
7. *Georgia's Opioid Response | Georgia Department of Public Health*. (n.d.). Retrieved January 30, 2020, from

https://dph.georgia.gov/stopopioidaddiction/georgias-opioid-response

8. Georgia Department of Public Health. (2017). *Opioid Overdose Surveillance Preliminary Report*. 1–27. https://dph.georgia.gov/sites/dph.georgia.gov/files/2017 Preliminary Georgia Opioid Overdose Report.pdf
9. Georgia Department of Public Health. (2018). *2018 Georgia Opioid Overdose Surveillance Preliminary Report*. 1–27. https://dph.georgia.gov/document/document/opioid-overdose-surveillance-2018-preliminary-report/download
10. Glanz, J. M., Narwaney, K. J., Mueller, S. R., Gardner, E. M., Calcaterra, S. L., Xu, S., Breslin, K., & Binswanger, I. A. (2018). Prediction Model for Two-Year Risk of Opioid Overdose Among Patients Prescribed Chronic Opioid Therapy. *Journal of General Internal Medicine*, 1–8. https://doi.org/10.1007/s11606-017-4288-3
11. Hedegaard, H., Miniño, A. M., & Warner, M. (2018). NCHS Data Brief. Drug Overdose Deaths in the United States, 1999–2017. *NCHS Data Brief, No 239*, *329*, 8. https://www.cdc.gov/nchs/data/databriefs/db329_tables-508.pdf#3.
12. Jolliffe, I. T., & Cadima, J. (n.d.). *Principal component analysis: a review and recent developments*. https://doi.org/10.1098/rsta.2015.0202
13. Liang, Y., Goros, M. W., & Turner, B. J. (2016). Drug overdose: Differing risk models for women and men among opioid users with non-cancer pain. *Pain Medicine (United States)*, *17*(12), 2268–2279. https://doi.org/10.1093/pm/pnw071
14. Lo-Ciganic, W. H., Huang, J. L., Zhang, H. H., Weiss, J. C., Wu, Y., Kwoh, C.

K., Donohue, J. M., Cochran, G., Gordon, A. J., Malone, D. C., Kuza, C. C., & Gellad, W. F. (2019). Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions. *JAMA Network Open*, *2*(3), e190968. https://doi.org/10.1001/jamanetworkopen.2019.0968

15. *Opioid Epidemic | Georgia Department of Public Health*. (n.d.). Retrieved January 30, 2020, from https://dph.georgia.gov/stopopioidaddiction
16. *Other Drugs | Drug Overdose | CDC Injury Center*. (n.d.). Retrieved January 30, 2020, from https://www.cdc.gov/drugoverdose/data/otherdrugs.html
17. *Prescription Drug Monitoring Program | Georgia Department of Public Health*. (n.d.). Retrieved January 30, 2020, from https://dph.georgia.gov/pdmp
18. *President Donald J. Trump's Initiative to Stop Opioid Abuse and Reduce Drug Supply and Demand*. (n.d.). Retrieved January 30, 2020, from https://www.whitehouse.gov/briefings-statements/president-donald-j-trumps-initiative-stop-opioid-abuse-reduce-drug-supply-demand-2/
19. *Principal Component Analysis 4 Dummies: Eigenvectors, Eigenvalues and Dimension Reduction – George Dallas*. (n.d.). Retrieved March 18, 2020, from https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/
20. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
21. *Reversing the opioid epidemic | American Medical Association*. (n.d.). Retrieved

January 30, 2020, from https://www.ama-assn.org/delivering-care/opioids/reversing-opioid-epidemic

22. Rickles, M., Rebeiro, P. F., Sizemore, L., Juarez, P., Mutter, M., Wester, C., & McPheeters, M. (2018). Tennessee's In-state Vulnerability Assessment for a "rapid Dissemination of Human Immunodeficiency Virus or Hepatitis C Virus Infection" Event Utilizing Data about the Opioid Epidemic. *Clinical Infectious Diseases*, *66*(11), 1722–1732. https://doi.org/10.1093/cid/cix1079
23. SAS Institute Inc (2018). SAS Software Version 9.4M6. Cary, North Carolina.
24. *Synthetic Opioid Overdose Data | Drug Overdose | CDC Injury Center*. (n.d.). Retrieved January 30, 2020, from https://www.cdc.gov/drugoverdose/data/fentanyl.html
25. *Understanding the Epidemic | Drug Overdose | CDC Injury Center*. (n.d.). Retrieved January 30, 2020, from https://www.cdc.gov/drugoverdose/epidemic/index.html
26. Van Handel, M. M., Hallisey, E. J., Kolling, J. L., Zibbell, J. E., Lewis, B., Bohm, M. K., & Jones, C. M. (2016). County-level Vulnerability Assessment for Rapid Dissemination of HIV or HCV Infections among Persons who Inject Drugs, United States. *Journal of Acquired Immune Deficiency Syndromes*, *73*(3), 323–331. https://doi.org/10.1097/QAI
27. Waller, Lance A. & Gotway, Carol A. (2004). Applied Spatial Statistics for Public Health Data. Hoboken, New Jersey: John Wiley & Sons, Inc.
28. *What is Variable Loadings in PCA? - Articles - STHDA*. (n.d.). Retrieved March 18, 2020, from http://www.sthda.com/english/articles/17-tips-tricks/68-what-is-

variable-loadings-in-pcae/

29. Wood, F. (2009). *Principal Component Analysis*.

6. Tables and Figures

Table 1.

ICD-10 (International Classification of Diseases) codes for all drug poisoning mortality as specified by the Centers for Disease Control and Prevention.

ICD-10 Code	Definition
Accidental poisoning by drugs as underlying cause of death	
X40	Accidental poisoning by and exposure to nonopioid analgesics, antipyretics and antirheumatics
X41	Accidental poisoning by and exposure to antiepileptic, sedative-hypnotic, antiparkinsonism and psychotropic drugs, not elsewhere classified
X42	Accidental poisoning by and exposure to narcotics and psychodysleptics (hallucinogens), not elsewhere classified
X43	Accidental poisoning by and exposure to other drugs acting on the autonomic nervous system
X44	Accidental poisoning by and exposure to other and unspecified drugs, medicaments and biological substances
Intentional self-poisoning by drugs as underlying cause of death	
X60	Intentional self-poisoning by and exposure to nonopioid analgesics, antipyretics and antirheumatics
X61	Intentional self-poisoning by and exposure to antiepileptic, sedative-hypnotic, antiparkinsonism and psychotropic drugs, not elsewhere classified
X62	Intentional self-poisoning by and exposure to narcotics and psychodysleptics (hallucinogens), not elsewhere classified
X63	Intentional self-poisoning by and exposure to other drugs acting on the autonomic nervous system
X64	Intentional self-poisoning by and exposure to other and unspecified drugs, medicaments and biological substances
Assault by drug poisoning as underlying cause of death	
X85	Assault by drugs, medicaments and biological substances
Drug poisoning of undetermined intent as underlying cause of death	
Y10	Poisoning by and exposure to nonopioid analgesics, antipyretics and antirheumatics, undetermined intent
Y11	Poisoning by and exposure to antiepileptic, sedative-hypnotic, antiparkinsonism and psychotropic drugs, not elsewhere classified, undetermined intent
Y12	Poisoning by and exposure to narcotics and psychodysleptics [hallucinogens], not elsewhere classified, undetermined intent
Y13	Poisoning by and exposure to other drugs acting on the autonomic nervous system, undetermined intent
Y14	Poisoning by and exposure to other and unspecified drugs, medicaments and biological substances, undetermined intent

Poisoning by narcotics and psychodysleptics (hallucinogens) as other cause of death	
T40.0	Poisoning by opium
T40.1	Poisoning by heroin
T40.2	Poisoning by natural and semisynthetic opioids
T40.3	Poisoning by methadone
T40.4	Poisoning by synthetic opioids, other than methadone
T40.6	Poisoning by other and unspecified narcotics

Table 2.

Potential model variables, variable descriptions, data sources, and analysis steps at which variables were removed.

Variable		Description	Source	Removal Step
Community Measures				
1	Age	Median age (years)	US Census Bureau	Univariate Model
2	Disability	Percent of civilian non-insitutionalized population age 18-64 with a disability	US Census Bureau	Univariate Model
3	Highway	At least 1 major US highway in county or within 5 miles of county border (1=yes)	Georgia Department of Transportation	Univariate Model
4	Rurality	Rurality classification based on population size, population density, remoteness, and urban land area	GDPH	Univariate Model
5	Sex	Percent of total population that is male	US Census Bureau	Univariate Model
6	Mental Health Provider	Rate of registered mental health providers per 100,000 pop., 2017 (most recent year)	Centers for Medicare & Medicaid Services	Univariate Model
7	No Insurance	Percent of civilian non-insitutionalized population with no health insurance coverage	US Census Bureau	Univariate Model
8	Primary Care Provider	Rate of registered primary care providers per 100,000 pop., 2016 (most recent year)	US Department of Health and Human Services	Principal Component Analysis
9	Social Vulnerability Rank	Overall percentile social vulnerability rank relative to other counties in GA, 2016 (most recent year available)	Centers for Disease Control and Prevention	Correlation Matrix
10	Housing Unit Vacancy	Percent of total housing units that are vacant	US Census Bureau	Univariate Model
11	Veteran Status	Percent of civilians age 18+ that are veterans	US Census Bureau	Principal Component Analysis
12	Homeowner Vacancy	Estimate of homeowner vacancy rate	US Census Bureau	Univariate Model
13	Rental Vacancy	Estimate of rental vacancy rate	US Census Bureau	Univariate Model

14	Race	Percent of non-hispanic or latino that is white race alone	US Census Bureau	In Final Model
15	Years of Potential Life Lost	Years of potential life lost before age 75 that occur per 100,000 population less than 75 years of age, 2017	GDPH	Univariate Model
16	High School Graduate	Percent of population age 25+ high school grad or higher	US Census Bureau	Empirical Review
17	Non-High School Graduate	Percent of population age 25+ without a high school diploma or higher	US Census Bureau	Univariate Model
Drug-Related Activity				
18	Opioid and Benzodiazepine Prescription Overlap	Percent of days that patients had an opioid and benzodiazepine Rx on the same day, 2017	Georgia Prescription Drug Monitoring Program	In Final Model
19	Drug Abuse Treatment Facility	Drug abuse treatment facilities per 100,000 population	GDPH	Univariate Model
20	Narcotics Treatment Program	Narcotics treatment programs per 100,000 population	GDPH	Principal Component Analysis
21	Opioid Overdose	Opioid overdoses, combined ER/hospital discharge	GDPH	Empirical Review
22	Opioid Overdose Rate	Two year average rate of opioid-involved overdoses, per 100,000	GDPH	Empirical Review
23	Opioid Overdose Mortality	Count of overdose deaths, opioids only (including heroin)	GDPH	Empirical Review
24	Opioid Overdose Mortality Rate	Two year average rate of overdose deaths, per 100,00 (opioids only)	GDPH	Empirical Review
25	Opioid Prescription Overlap	Percent of days that patients had >1 prescribed opioid on the same day, 2017	Georgia Prescription Drug Monitoring Program	Principal Component Analysis
26	Opioid Prescription	Opioid analgesic prescriptions dispensed and reported to PDMP per 1,000 population	Georgia Prescription Drug Monitoring Program	Stepwise Selection

27	Opioid Prescription Length	Avg. days of opioid analgesics supplied per prescription	Georgia Prescription Drug Monitoring Program	In Final Model
Employment/Education				
28	Natural Resources, Construction, or Maintenance	Percent of civilian employed population age 16+ with occupation classified as natural resources, construction, or maintenance	US Census Bureau	Univariate Model
29	Construction	Percent of civilian employed population age 16+ in construction industry	US Census Bureau	Empirical Review
30	Mean household Income	Estimate of mean household income in 2017 inflation-adjusted dollars	US Census Bureau	Empirical Review
31	Median Household Income	Estimate of median household income in 2017 inflation-adjusted dollars	US Census Bureau	Empirical Review
32	Log-transformed Median household Income	Estimate of median household income in 2017 inflation-adjusted dollars, log10 transformed	US Census Bureau	Correlation Matrix
33	Log-Transformed Per-Capita Income	Log-transformed per capita income in past 12 months	GDPH	Empirical Review
34	Manufacturing Employment	Percent of civilian employed population age 16+ in manufacturing industry	ACS 2017 5-year estimates	Empirical Review
35	Labor Force	Percent of population 16+ not in labor force	ACS 2017 5-year estimates	Empirical Review
36	Per-Capita Income	Per capita income in past 12 months	ACS 2017 5-year estimates	Empirical Review
37	Family Poverty	Percent of all families whose income in past 12 months is below poverty level	ACS 2017 5-year estimates	Empirical Review
38	Poverty	Percent of all people whose income in past 12 months is below poverty level	ACS 2017 5-year estimates	Stepwise Selection
39	Unemployment Rate	Civilian labor force unemployment rate	ACS 2017 5-year estimates	Principal Component Analysis

HIV/HCV Infection Status				
40	Acute HCV Rate	Two year average rate of acute HCV, per 100,000	DPH	Empirical Review
41	All HCV Rate	Two year average rate of acute HCV + chronic HCV age <40, per 100,000	DPH	Principal Component Analysis
42	Chronic HCV Rate	Two year average rate of chronic HCV age <40, per 100,000	DPH	Empirical Review
43	HIV Rate	Two year average rate of persons living with HIV, per 100,000	DPH	Stepwise Selection
44	STD Rate	Age-Adjusted STD Rate per 100,000 population, All STD except Congenital Syphilis, 2017	GA Online Analytical Statistical Information System	In Final Model
Household Measures				
45	Alone	Percent of total households with householder living alone	ACS 2017 5-year estimates	Univariate Model
46	Family	Percent of total households that are family households	ACS 2017 5-year estimates	Empirical Review
47	Food Stamp	Percent of total households with food stamp/SNAP benefits past 12 months	ACS 2017 5-year estimates	Correlation Matrix
48	Internet Accessibility	Percent of total households with broadband internet connection	ACS 2017 5-year estimates	Empirical Review
49	Internet Inaccessibility	Percent of total households without broadband internet connection	ACS 2017 5-year estimates	Correlation Matrix
50	Vehicle Inaccessibility	Percent of total housing units with no vehicle available	ACS 2017 5-year estimates	In Final Model

Table 3.

Variances and standard deviations explained by components from principal component analysis.

Principal Component	Proportion of Variance Explained	Cumulative Variance	Standard Deviation
1	29.8%	29.8%	3.089
2	15.5%	45.3%	2.224
3	6.4%	51.6%	1.426
4	5.6%	57.2%	1.342
5	5.3%	62.5%	1.299
6	4.6%	67.1%	1.214
7	3.6%	70.7%	1.068
8	3.4%	74.1%	1.036
9	3.2%	77.2%	1.009
10	2.5%	79.8%	0.903
11	2.4%	82.1%	0.871
12	2.1%	84.3%	0.826
13	1.8%	86.1%	0.757
14	1.5%	87.6%	0.688
15	1.4%	89.0%	0.673
16	1.4%	90.3%	0.662
17	1.2%	91.5%	0.608
18	1.0%	92.5%	0.579
19	1.0%	93.5%	0.558
20	0.9%	94.4%	0.541
21	0.8%	95.2%	0.516
22	0.8%	96.1%	0.505
23	0.7%	96.7%	0.466
24	0.6%	97.4%	0.450
25	0.5%	97.8%	0.389
26	0.4%	98.3%	0.378
27	0.4%	98.7%	0.360
28	0.4%	99.1%	0.353
29	0.3%	99.4%	0.318
30	0.3%	99.7%	0.292
31	0.2%	99.9%	0.266
32	0.1%	100.0%	0.186

Figure 1.

Scree plot displaying the proportion of variance explained by each principal component.

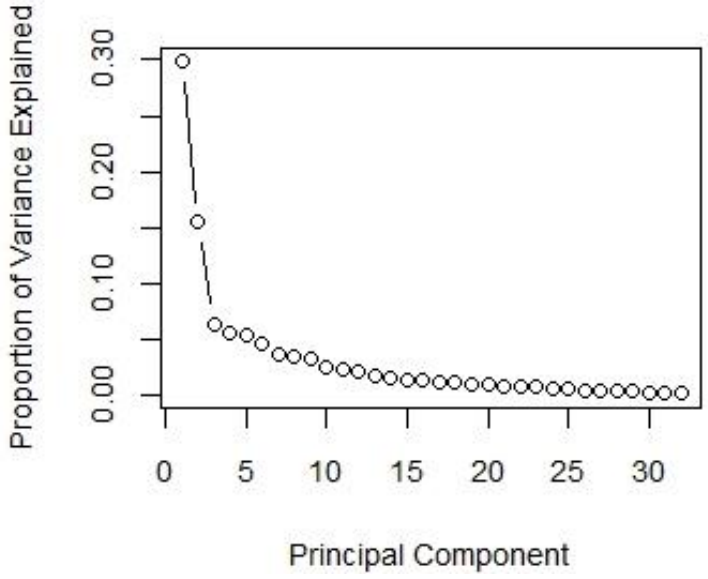


Figure 2.

Log-transformed median household income (hhincome_medlog), poverty (pov_people), food stamp (foodstamp), and social vulnerability rank (SVI_16) exhibited multicollinearity. The poverty variable had the most significant relationship with the outcome based on lowest p-value so it was the only variable retained for further analysis.

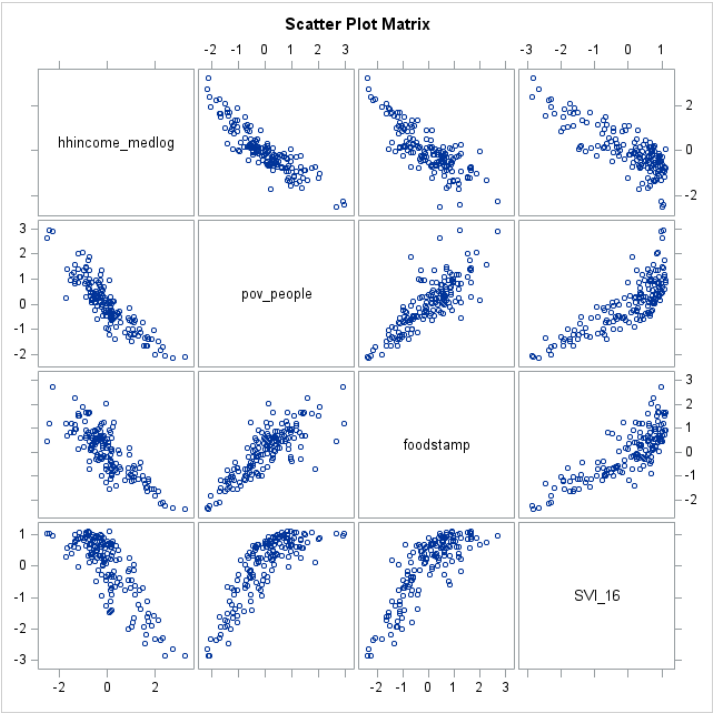


Figure 3.

Log-transformed median household income (hhincome_medlog) and no internet (no_internet) exhibited collinearity. Both variables were removed from further analysis. Household income also showed multicollinearity with poverty, food stamp, and social vulnerability rank. The poverty variable had the most significant relationship with the outcome based on lowest p-value so it was the only variable retained for further analysis.

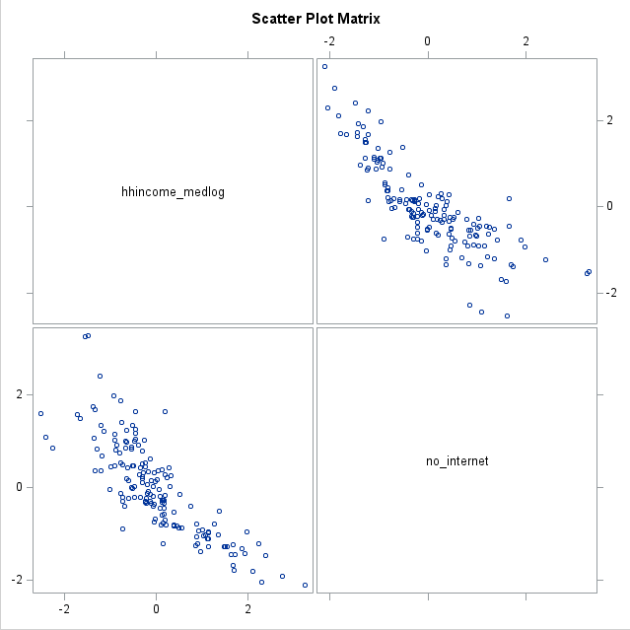


Table 4.

Final unscaled Poisson regression model coefficient estimates predicting 2017 log-transformed drug overdose mortality rates.

Variable	Coefficient	Standard Error	Z-Statistic	P-Value	95% CI
Intercept	-9.852	0.533	-17.979	<0.001	(-10.626,-8.538)
Opioid and Benzodiazepine Prescription Overlap	0.062	0.017	3.774	<0.001	(0.030,0.095)
Vehicle Inaccessibility	-0.036	0.017	-2.141	0.032	(-0.003,-0.068)
Opioid Prescription Length	-0.090	0.035	-2.561	0.010	(-0.021,-0.159)
STD Rate	0.001	3.044e-04	3.665	<0.001	(5.184e-04,0.002)
Race	0.022	0.005	4.778	<0.001	(0.032,0.013)
STD Rate*Race	1.136e-05	4.790e-06	-2.372	0.018	(-1.972e-06, -2.075e-05)

Figure 4.

Geographic maps of adjacency-based smoothed drug overdose mortality rates for the observed and estimated 2017 drug overdose death cases.

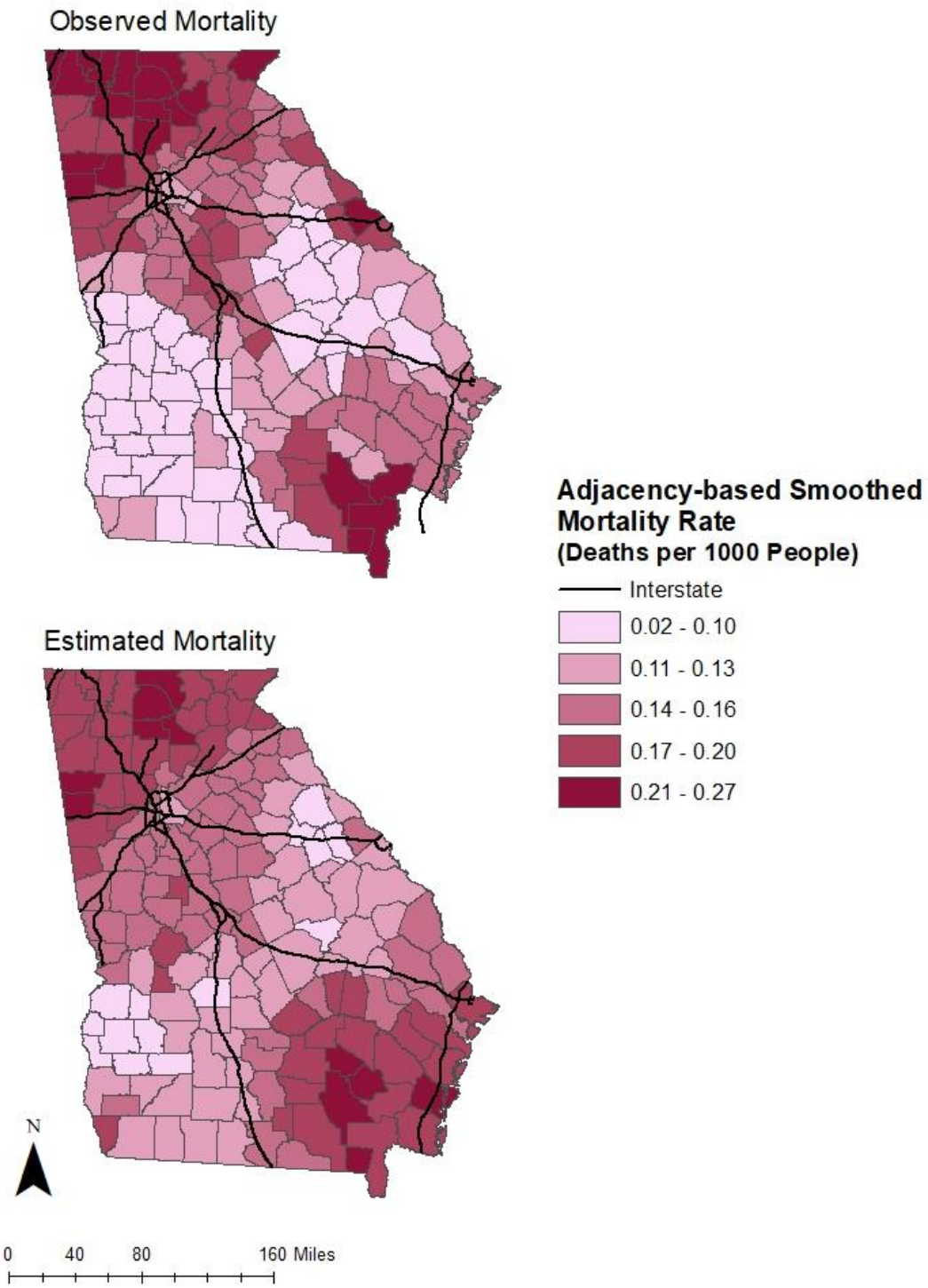
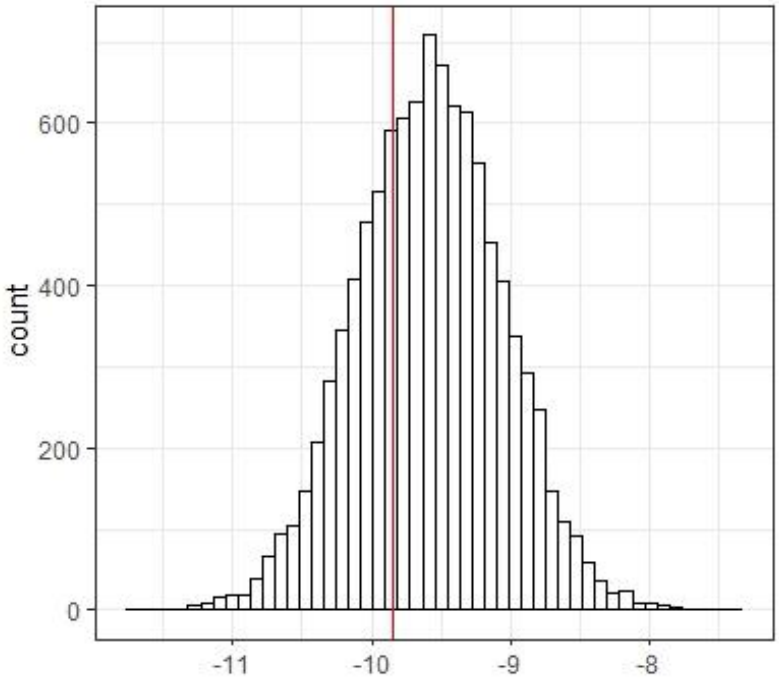


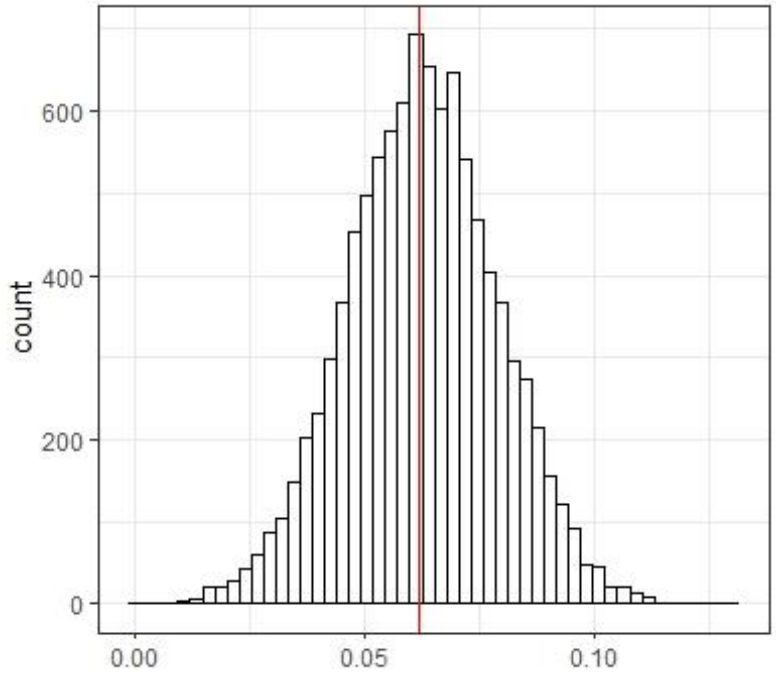
Figure 5.

Histograms of 10,000 simulations of random regression coefficient values from the final unscaled Poisson regression model based on the Normal distribution. Red line indicates the true value of the coefficient.

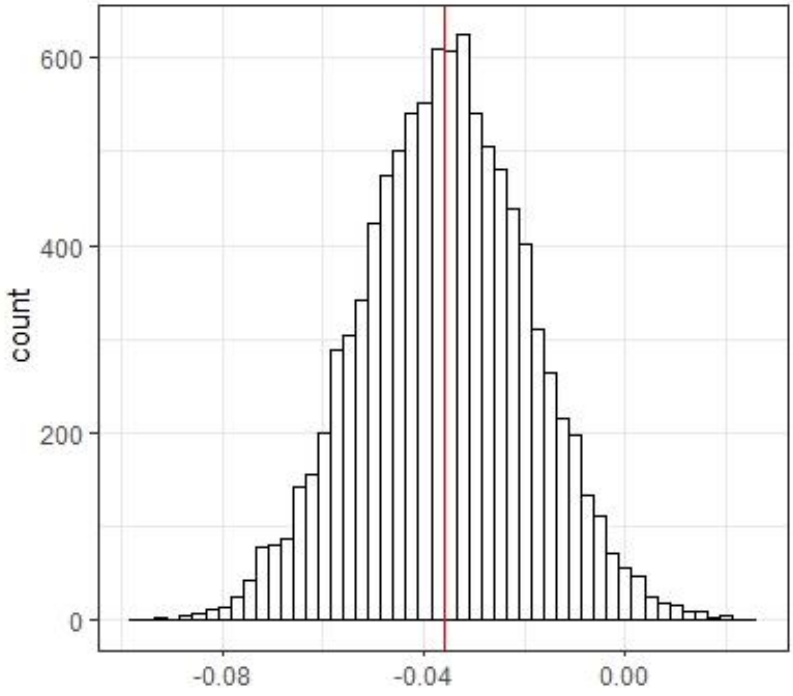
A. Intercept (Coefficient=-9.852).



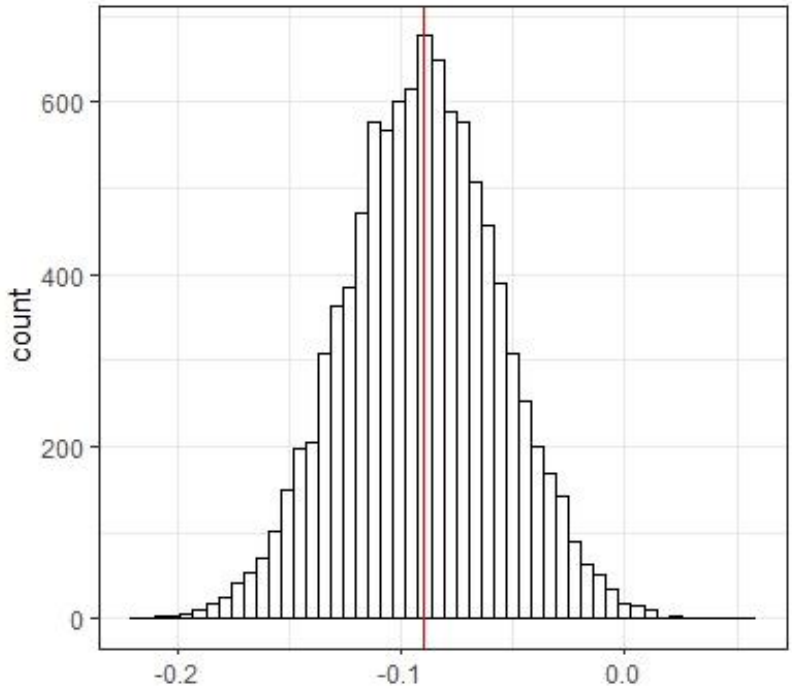
B. Opioid and Benzodiazepine Prescription Overlap (Coefficient=0.062)



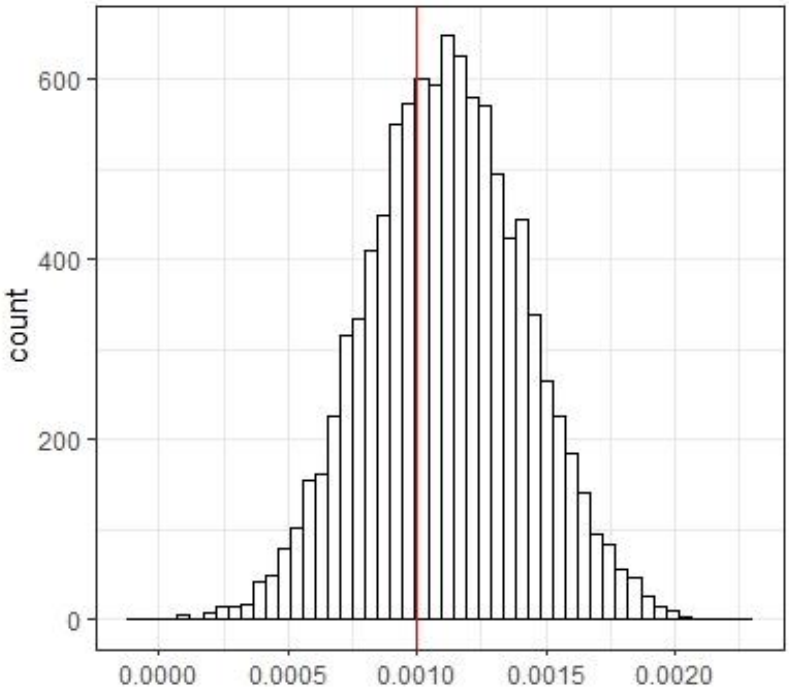
C. Vehicle Inaccessibility (Coefficient=-0.036)



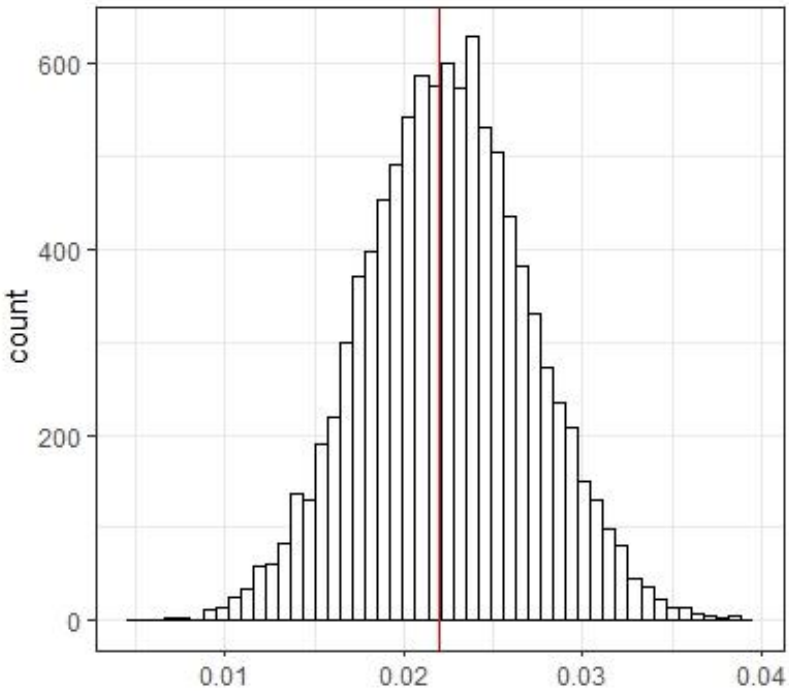
D. Opioid Prescription Length (Coefficient=-0.090)



E. STD Rate (Coefficient=0.001)



F. Race (Coefficient=0.022)



G. Interaction between STD Rate and Race (Coefficient=-1.136e-05)

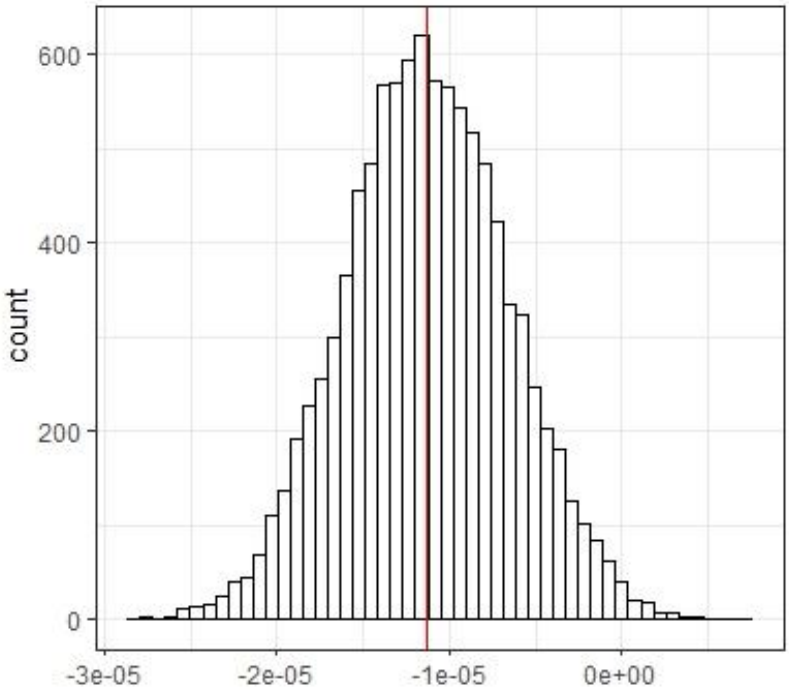


Table 5.

Observed and estimated county-level mortality cases for Georgia in 2017.

County	Observed Mortality	Estimated Mortality
Appling	1	3
Atkinson	4	2
Bacon	1	3
Baker	0	0
Baldwin	4	4
Banks	5	4
Barrow	17	14
Bartow	25	23
Ben Hill	2	2
Berrien	1	3
Bibb	20	18
Bleckley	2	2
Brantley	5	5
Brooks	0	2
Bryan	5	6
Bulloch	7	10
Burke	4	3
Butts	5	4
Calhoun	0	0
Camden	5	8
Candler	1	2
Carroll	32	25
Catoosa	24	16
Charlton	7	2
Chatham	37	40
Chattahoochee	1	1
Chattooga	6	4
Cherokee	40	43
Clarke	12	11
Clay	0	0
Clayton	34	37
Clinch	0	1
Cobb	142	131
Coffee	10	9
Colquitt	3	4
Columbia	25	24
Cook	2	2

Coweta	23	24
Crawford	3	2
Crisp	3	2
Dade	5	3
Dawson	6	6
Decatur	1	4
DeKalb	79	82
Dodge	3	3
Dooly	0	1
Dougherty	6	9
Douglas	24	20
Early	1	2
Echols	1	1
Effingham	12	12
Elbert	3	2
Emanuel	2	2
Evans	1	1
Fannin	7	5
Fayette	14	15
Floyd	12	15
Forsyth	31	37
Franklin	4	4
Fulton	163	156
Gilmer	10	7
Glascock	0	1
Glynn	16	13
Gordon	9	10
Grady	3	3
Greene	2	2
Gwinnett	94	93
Habersham	4	6
Hall	34	28
Hancock	1	1
Haralson	8	8
Harris	1	5
Hart	0	4
Heard	2	2
Henry	28	30
Houston	27	24
Irwin	1	1
Jackson	12	13
Jasper	4	2

Jeff Davis	3	5
Jefferson	2	1
Jenkins	0	1
Johnson	1	1
Jones	5	5
Lamar	4	3
Lanier	1	1
Laurens	6	5
Lee	5	4
Liberty	7	9
Lincoln	2	1
Long	4	2
Lowndes	5	12
Lumpkin	11	8
Macon	0	1
Madison	7	5
Marion	0	1
McDuffie	3	3
McIntosh	1	3
Meriwether	2	2
Miller	0	1
Mitchell	3	2
Monroe	4	5
Montgomery	1	1
Morgan	3	3
Murray	8	8
Muscogee	25	27
Newton	16	15
Oconee	3	7
Oglethorpe	2	2
Paulding	25	28
Peach	1	3
Pickens	9	8
Pierce	1	4
Pike	2	4
Polk	9	8
Pulaski	3	1
Putnam	2	3
Quitman	0	0
Rabun	6	3
Randolph	1	0
Richmond	52	41

Rockdale	12	12
Schley	0	1
Screven	1	1
Seminole	3	2
Spalding	10	9
Stephens	2	3
Stewart	0	0
Sumter	2	3
Talbot	0	1
Taliaferro	0	0
Tattnall	6	3
Taylor	2	1
Telfair	3	2
Terrell	0	1
Thomas	4	6
Tift	3	4
Toombs	5	4
Towns	2	2
Treutlen	0	1
Troup	12	10
Turner	1	1
Twiggs	1	1
Union	3	4
Upson	3	4
Walker	14	14
Walton	15	16
Ware	4	7
Warren	0	0
Washington	0	2
Wayne	5	4
Webster	0	0
Wheeler	0	1
White	2	5
Whitfield	16	14
Wilcox	0	1
Wilkes	2	1
Wilkinson	1	1
Worth	1	3

Table 6.

Final unscaled Poisson regression model coefficient estimates predicting 2016 log-transformed drug overdose mortality rates.

Variable	Coefficient	Standard Error	Z-Statistic	P-Value	95% CI
Intercept	-9.735	0.723	-13.466	<0.001	(-11.152,-8.318)
Opioid and Benzodiazepine Prescription Overlap	0.053	0.020	2.635	0.008	(0.014,0.092)
Vehicle Inaccessibility	-0.031	0.021	-1.476	0.140	(-0.072,0.010)
Opioid Prescription Length	-0.069	0.047	-1.472	0.141	(-0.161,0.023)
STD Rate	0.001	4.526e-04	2.359	0.018	(1.113e04, 0.002)
Race	0.019	0.006	2.962	0.003	(0.007,0.031)
STD Rate*Race	-1.390e-05	6.786e-06	-2.048	0.041	(-5.994e-07, -2.720e-05)