Spatial Association Between Social Vulnerability and Accidental Poisoning Mortality In Pennsylvania and its Adjacent States

by

Kendra J. Bobby

BS, University of Pittsburgh, 2016

Submitted to the Graduate Faculty of

the Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Kendra J. Bobby

It was defended on

October 14, 2019

and approved by

Jenna C. Carlson, PhD, Assistant Professor, Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Donald S. Burke, MD, Dean Emeritus, Epidemiology,
Graduate School of Public Health, University of Pittsburgh

Thesis Advisor: Jeanine M. Buchanich, PhD, Research Associate Professor, Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Jeanine M. Buchanich, PhD

Spatial Association Between Social Vulnerability and Accidental Poisoning Mortality in Pennsylvania and its Adjacent States

Kendra J. Bobby, MS

University of Pittsburgh, 2019

Abstract

The purpose of the current study is to investigate the spatial relationship between county-level social factors that are a proxy to measure the social vulnerability to adverse events and accidental poisoning mortality, or overdose deaths, in Pennsylvania and surrounding states. This study explores the suggestion that the opioid epidemic is an aggregate of individual sub-epidemics by looking at mortality rates and social determinants of health at the county level. We evaluate this by examining local spatial distribution of overdose mortality rates and social factors, such as poverty, overcrowding, and unemployment. This regional analysis compares spatial and non-spatial models for modeling overdose mortality, and the magnitude of influence of each variable at the county level compared to its global impact. We discuss the public health importance of non-homogeneous effects of social determinants of health on the opioid epidemic, and how they may help us to better understand and diagnose its contributing factors.

Table of Contents

List of Tables

List of Figures

Preface

I would like to thank Drs. Jeanine Buchanich, Jenna Carlson, and Don Burke for their guidance in completing this project.

1.0 Introduction

National mortality from accidental drug overdose reached an all-time high of 18.82 per 100,000 population in 2017, almost doubling the rate (9.72 per 100,000) in 2010,[9] and has been exponentially increasing since 1979.[12] The opioid epidemic is constantly changing: over time, urbanicity, and across demographic groups,[12] so that mortality in one subset of the population looks dramatically different from mortality in another. Each sub-epidemic presents uniquely in its social and economic features, which makes it particularly challenging for researchers and policymakers to quantify the causes or formulate solutions. As rates of mortality from accidental drug poisoning continue to rise, so do calls from public health professionals to address its root causes in ways not previously considered for drug use or addiction.[6; 8] Case and Deaton historically called attention to these deaths of despair when they showed that drug and alcohol poisonings were largely responsible for the increase in midlife mortality for non-Hispanic whites.[7] Woolf et. al. later confirmed that deaths of despair, specifically drug overdose, had caused an increase in midlife mortality for not only non-Hispanic whites, but all other race groups as well.[22] Dasgupta et. al. calls on researchers to look past the problem of overprescribing and explore the, "role of opioids as a refuge from physical and psychological trauma, concentrated disadvantage, isolation, and hopeless."[8]

One strategy to expand the current approach to researching the opioid epidemic is to address spatial variation in mortality rates and the social factors influencing those rates. By addressing spatial heterogeneity, we can better understand the influential factors in regions of concentrated social disadvantage, and its impact on opioid mortality rates in particular. The goal of the present analysis is to identify a group of uncorrelated variables from the CDC's Social

Vulnerability Index (SVI)[1] that best predict overdose death counts from years 2012 to 2016 in a regional analysis that spans the state of Pennsylvania and its adjacent states, Delaware, Maryland, West Virginia, New York, New Jersey and Ohio. The SVI is a group of variables taken from the American Community Survey for years 2012 to 2016. These variables were selected to represent the community's social vulnerability to the impact of external stressors. The given region was chosen for this analysis because it includes the highest rates of accidental poisoning mortality in the United States during this time period, as well as areas with lower mortality rates, and there is variability in the social factors across this region. All states adjacent to Pennsylvania are included so that we may address the question of whether administrative boundaries, such as state lines, and differences in reporting across them, play a significant role in the spatial trends.

This analysis can be used to gain insights into the geographical impact of social determinants of health on accidental poisoning mortality rates.

2.0 Methods

2.1 Data

County-level mortality information, including death and population counts, were obtained from the Mortality Information and Research Analytics (MOIRA) system[9], a detailed mortality database maintained by the University of Pittsburgh. Cases were identified as accidental poisoning deaths using ICD-10 codes X40-X44 that occurred during the period 2012 to 2016.

Social determinants of health were taken from the Center for Disease Control's Social Vulnerability Index (SVI)[1] and are county-level data that measure housing units, per capita income, percentage of population of persons below poverty, civilian (age 16+) unemployed, persons with no high school diploma (25+), persons aged 65 and older, persons aged 17 and younger, civilian non-institutionalized population with a disability, single parent households with children under 18, minority (all persons except white, non-Hispanic), persons (age 5+) who speak English 'less than well', persons in institutionalized group quarters, percentage of housing in structures with 10 or more units, mobile homes, occupied housing units with more people than rooms, and households with no vehicle available. The SVI is based on the results of the 2012 to 2016 American Community Survey (ACS), and is used to quantify the strength of a community to resist public health crises such as natural disasters or disease outbreaks. The SVI also includes aggregate measures of these variables, classified into summary theme percentile ranking variables, socioeconomic, household composition & disability, minority status & language, and housing & transportation. The SVI summary themes percentile rankings were scaled by a factor of 100 so that they would represent percentiles between 0 and 100, similar to the other SVI variables.

2.2 Analysis

2.2.1 Correlation

The relationship between each combination of SVI variables was assessed by calculating the Spearman correlation coefficient using R package *stats*.[19] Only data from counties in Pennsylvania and its adjacent states Delaware, Maryland, West Virginia, New York, New Jersey and Ohio, were used. Spearman correlation coefficient was selected over Pearson correlation coefficient because most of the SVI variables are expressed as percentages of population, and so are bound between 0 and 1 and non-normally distributed. The non-parametric Spearman correlation is more appropriate in this case because it assigns and evaluates correlation based on the ranks of the data values, rather than parametric Pearson correlation which calculates at the linear correlation of the raw data values. The Spearman correlation of the SVI variables was used to inform whether their inclusion in a multivariable model would be appropriate or result in collinearity of predictors. A cut-point of absolute Spearman correlation coefficient 0.50 was used when considering pairs or groups of highly correlated SVI variables for the multivariable model.

2.2.2 Negative Binomial Regression

Discrete data in the form of counts typically follow a Poisson distribution, where mean and variance are equal, and are usually modeled using Poisson regression techniques. Poisson regression was inappropriate for this analysis because the mean death count (158) is much smaller than its variance (81187.2), resulting in overdispersion. A likelihood ratio test was performed to confirm that the negative binomial model was more appropriate for these data than Poisson regression. A likelihood ratio test has the null hypothesis that two models are equally appropriate and is tested by using the ratio of their likelihoods. The test statistic for the likelihood ratio test is $-2 \cdot \ln[\, L_{poi}(\boldsymbol{\theta}) / L_{nb}(\boldsymbol{\theta})]$, where $L(\boldsymbol{\theta})$ represents the value of the likelihood function for the model. The likelihood of a model is calculated by taking the product of the probability density function for a given distribution at all of the observed data points with the given parameter values.

The univariable relationship between death counts in Pennsylvania and its adjacent states and each SVI variable was modeled using negative binomial regression. Death count is assumed to follow a negative binomial distribution with probability density function (pdf), $f(y;\, \theta,\, p) = \binom{y+\theta-1}{\theta-1} \cdot p^{\theta} \cdot (1-p)^{y}$ and mean, or expected number of counts, $E(y) = \theta \cdot (1-p) / p = \mu$, variance, $V(y) = \mu + 1/\theta \cdot \mu^2$.

In the generalized linear model (GLM) regression framework, we have $y = \mu + \varepsilon$ with log link function, to give $\log(\mu) = \beta_0 + \beta_1 \cdot x$, where x is the predictor variable. Incidence rate ratio (IRR) can also be calculated from the regression equation, by taking $\exp(\beta_1)$, which represents the proportional change in outcome based on a one-unit increase in variable x. This means, for example, that if the coefficient, β_1, is equal to zero, then $\exp(\beta_1 = 0) = 1$, and there is no proportional change in the outcome for an increase in x.

First, univariable negative binomial regression was applied to regress each SVI variable on outcome variable death count using *glm* function from R package *stats*.[19] Offset log(Population) for each county was included in the model. Incidence rate-ratio (IRR) was calculated for each county. The parameter θ was estimated using the maximum likelihood method implemented by the function *glm.nb* in R package *MASS*.[20]

The results of the univariable regression and the Spearman correlation between pairs of SVI variables was used to select a representative subgroup of SVI variables that were non-collinear (i.e., Spearman correlation ≤ 0.50) and strong univariable predictors of the outcome (i.e., univariable p-value < 0.05). These variables were used to create the full model that was used in a backwards selection model-building process, where variables were removed based on an $\alpha = 0.05$ rule.

2.2.3 Geographically Weighted Negative Binomial Regression

Geographically weighted regression (GWR) is statistical technique used to model local relationships between outcome and predictor variables. It uses a weighted ordinary least squares (OLS) regression framework that builds on the traditional model equation, $y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$. Distance-based weights are incorporated into the estimation of $\hat{\boldsymbol{\beta}}$ using a Gaussian weight function, such that $\hat{\boldsymbol{\beta}} = (X^T W X)^{-1} X^T W Y$ and $W_{(i)j} = \exp(-1/2 \cdot (d_{(i)j} / b)^2)$, where $d_{(i)j}$ is the distance between points i and j, and b is the kernel bandwidth.[15] The Gaussian kernel comes from the Gaussian pdf, without the normalization constant, $(2\pi\sigma^2)^{-1/2}$. Distance, $d_{(i)j}$, is the input value of the random variable into the function that has mean zero, and bandwidth, b, is the standard deviation. As distance increases, the weight trends towards zero, so that the nearest neighbors will contribute more to the estimation of a particular location. The bandwidth is estimated for each model using cross validation to minimize the root mean square prediction error, employed by the *ggwr.sel* function from the *spgwr* package in R.[10] The root mean square error (RMSE) is found by taking the square root of the mean square error (MSE) for the geographically weighted regression model. The MSE of the model is calculated by taking the average of the squared distance between actual data values and fitted values from the model. The bandwidth is chosen in a way that minimizes the RMSE for the given data and model parameters. The same bandwidth is then used to estimate the parameter values for each local coefficient, so there is no variability in the bandwidth between parameters or for different regions in the study area. If a study area is at the perimeter of the study region, only those areas included in the study area will contribute to its estimated parameters.

The negative binomial GWR model is similar to the negative binomial generalized linear model described 2.2.2, but with the distance-based weight matrix used in the estimation of $\hat{\boldsymbol{\beta}}$**.** The parameter θ was estimated using the maximum likelihood method for nonspatial models, as described above.

Univariable GWR was implemented for each of the SVI variables, using outcome variable death count with a log(Population) offset. A multivariable GWR model was also fit using the same variables that were selected by the negative binomial generalized linear multivariable model as described in section 2.2.2. The bandwidth was estimated using the maximum likelihood method for each regression model. The local coefficients for each regression model were exponentiated to give local IRRs for each county for visual analysis of geospatial trends.

2.2.4 Hypothesis Testing for Local Coefficients

The results of the multivariable negative binomial GWR provide a set of local coefficients for each SVI variable for each county. The typical regression null hypothesis is that the coefficients are uniformly equal to zero. To test this null hypothesis, 95% confidence intervals were constructed for each local coefficient to confirm if the null coefficient, zero, was included in the interval. The covariance matrix of the coefficients was calculated using $\text{cov}(\hat{\beta}) = (X^TWX)^{-1}\sigma^2$, where $\sigma^2 = V(y)$, defined in section 2.2.2, was estimated using the maximum likelihood estimate of θ and the mean of the outcome variable, deaths, as an estimate for μ. The weight matrix, $W_{(i)}$ was calculated using command *gw.weight* from R package *GWmodel* [11]. The diagonal values from matrix $\text{cov}(\hat{\beta})$, $\text{diag}(\hat{\beta})_i$ for i = 1,2,3,4, were used to construct conservative 95% confidence intervals in the usual way: $\hat{\beta} \pm 2 \cdot (\mathrm{diag}(\hat{\beta})_i \ / \ n)^{1/2}$. A second test was performed with the null hypothesis that the coefficients are uniformly equal to the global coefficients from the non-spatial multivariable negative binomial GLM Test statistics following a t-distribution were also calculated by taking $\hat{\beta} - \hat{\beta}_{global} \ / \ (\mathrm{diag}(\hat{\beta})_i \ / \ n)^{1/2}$, from which p-values were calculated and adjusted for multiple testing according to Benjamini and Yekutieli method for controlling the false discovery rate when test statistics are dependent. This null hypothesis was tested using a two-sided alternative at the $\alpha = 0.05$ level.[4]

3.0 Results

There are 320 counties included in this analysis, with no missing data for death counts or social vulnerability variables. Death counts by county in Pennsylvania and its adjacent states are shown in Figure 1. These were examined because this analysis uses a count-based regression model for the accidental poisoning mortality deaths. Suppression is enforced per the National Center for Health Statistics (NCHS) guidelines, so death counts of fewer than 10 deaths during the 5-year time period are shown in white.

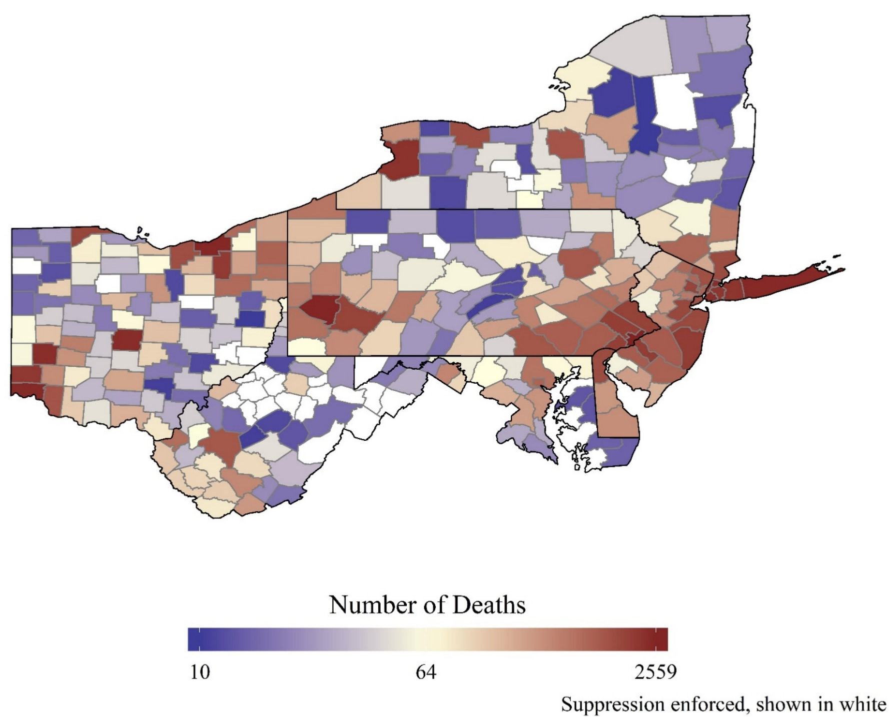

Figure 1 Accidental Poisoning Mortality Death Counts (2012-2016)

The color scale used in Figures 1 and 2 is based on percentile, so that the yellow (64 deaths, 16.5 per 100,000 rate) represents the 50th percentile, and the darkest red (2559 deaths, 79.9 per 100,000 rate) represents the 100th percentile, or maximum number of deaths and rate, respectively. The highest death counts are clustered around the major cities of each state, which also represent the largest population areas.

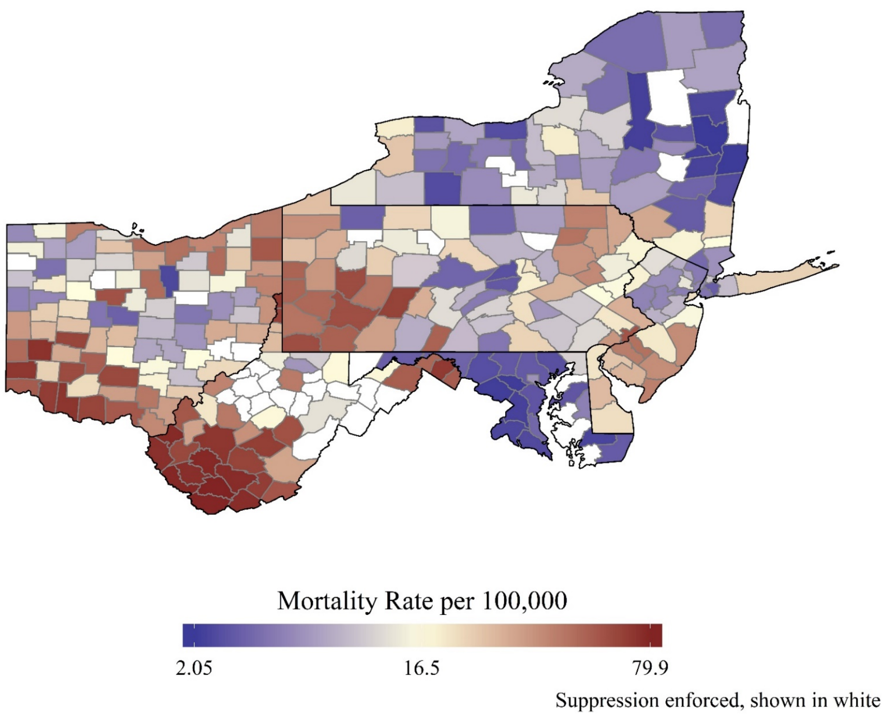

Figure 2 Accidental Poisoning Mortality Rate (2012-2016)

In Ohio, the darkest red counties are the areas containing its cities, Columbus, Cincinnati, Toledo, and Cleveland. In Pennsylvania, the highest death counts surround Pittsburgh and Philadelphia. In New York, New York City and Buffalo have the highest death counts. It is notable that there are no dark red areas in Maryland, despite the high mortality rate of Baltimore[9], but the whole of its neighbor state, Delaware, is near the 50th to 60th percentile of death counts in this region. This unexpected difference is explained by differences in reporting. Maryland classifies the majority of its poisoning mortality with undetermined intent (IDC-10 codes Y10-Y14), rather than accidental intent (ICD-10 codes X40-X44) as is more typical for the other states.[16] The southwest region of West Virginia is also shown in red, and its highest death counts can be found in the county containing its capital, Charleston.

In Figure 2, the counties are shown with their representative mortality rates per 100,000 population. In this graph, the denominator, population, is taken into consideration, so that those areas with higher mortality rates proportional to their population size are visible. The highest overall mortality rates are found in southern West Virginia, southern Ohio, and southwest Pennsylvania. The aggregation of those areas is commonly known as Appalachia, which extends from central Pennsylvania down to the northern tip of Georgia, Alabama, and Mississippi.

Table 1 Social Vulnerability Index Variable Summary

Social Vulnerability Index Variable	Mean	Standard Deviation	Min.	Median	Max.
Housing units estimate (× 1,000 units)	82.01	135.46	3.24	32.27	1021.75
Percentage of persons below poverty	14.4	4.8	4.5	14.1	37.6
Percentage of civilian (age 16+) unemployed	7.2	2.0	2.7	7.0	16.2
Per capita income	$27,172	$6,815	$13,283	$25,438	$66,522
Percentage of persons with no high school diploma (age 25+)	12.4	4.6	3.4	11.4	41.5
Percentage of persons aged 65 and older	17.3	2.9	10.7	17.2	26.9
Percentage of persons aged 17 and younger	21.3	2.5	3.9	21.4	33.0
Percentage of civilian noninstitutionalized population with a disability	15.4	4.4	7.2	14.9	33.3
Percentage of single parent households with children under 18	8.3	2.0	1.9	8.3	20.0
Percentage minority (all persons except white, non-Hispanic)	14.6	15.4	0.4	9.0	90.4
Percentage of persons (age 5+) who speak English 'less than well'	1.3	2.3	0.0	0.6	15.9
Percentage of housing in structures with 10 or more units	6.3	8.1	0.0	4.0	89.6
Percentage of mobile homes	9.0	7.2	0.1	7.4	35.3
Percentage of occupied housing units with more people than rooms	1.6	1.3	0.3	1.4	12.2
Percentage of households with no vehicle available	8.8	6.8	2.8	7.7	77.3
Percentage of persons in institutionalized group quarters	3.6	4.6	0.0	2.4	59.3
THEME: Socioeconomic (× 100)	43.9	23.3	0.03	43.1	99.7
THEME: Minority Status & Language (× 100)	39.2	24.8	0.22	37.3	96.1
THEME: Housing & Transportation (× 100)	39.5	28.1	0.06	33.8	99.8
THEME: Household Composition & Disability (× 100)	52.1	25.7	1.11	52.3	99.4

Table 1 describes the distribution of the social vulnerability variables from the SVI for counties in Pennsylvania and its adjacent states, only. Housing units and per capita income are expressed as units and dollars, respectively, and the rest are expressed as percentage of the population that belongs to that category.

3.1 Correlation

The most strongly correlated SVI variables were percentage of persons (age 5+) who speak English 'less than well' and percentage of occupied housing units with more people than rooms (Spearman correlation coefficient, $\rho = 0.84$). Limited English was also strongly correlated with percentage minority (all persons except white, non-Hispanic) ($\rho = 0.83$). The least correlated pair of SVI variables was percentage of households with no vehicle available and percentage of persons in institutionalized group quarters ($\rho = 0.004$). Of the 120 unique pairs of SVI variables, 77 (64%) had absolute correlation coefficient less than 0.40, which was considered a weak correlation.

Figure 3 shows the Spearman correlation matrix for all 20 variables. It showed positive correlations in the block containing percentage of persons in households with no vehicle available, occupied housing units with more people than rooms, persons (age 5+) who speak English 'less than well', minority (all persons except white, non-Hispanic), housing in structures with 10 or more units, and the housing units estimate. There was also a block of positive correlation between percentage of persons with no high school diploma, persons below poverty, and civilian (age 16+) unemployed. Per capita income was negatively correlated with percentage of persons with no high school diploma (age 25+), persons below poverty, mobile homes, and civilian non-institutionalized population with a disability. Percentage of persons in institutionalized group quarters was negatively associated with percentage of persons aged 17 and younger.

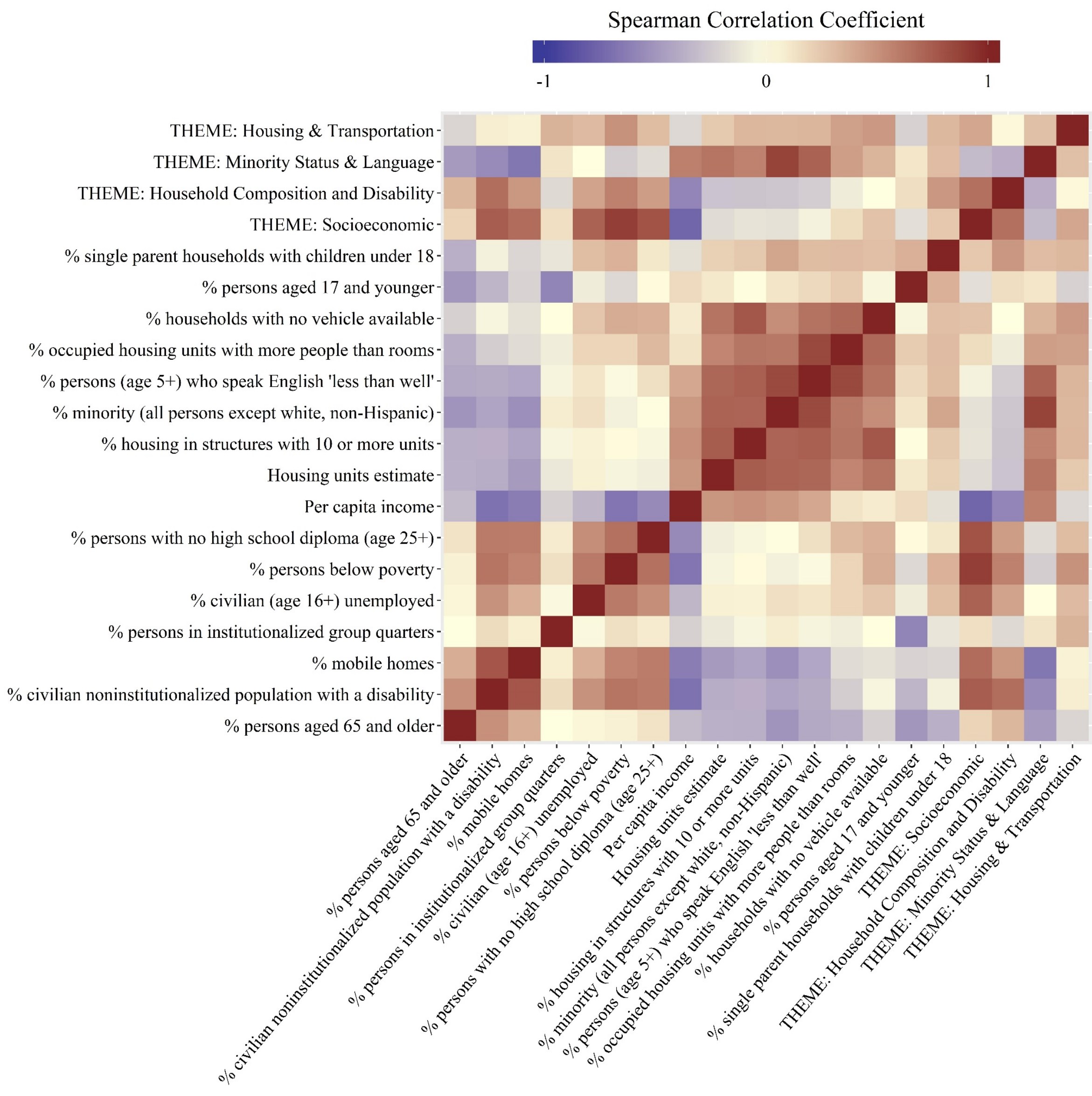

Figure 3 SVI Variable Correlation Matrix

The Spearman correlation coefficient was also examined for the SVI variables and themes with accidental poisoning mortality rate. The resulting correlation coefficients are shown in Table 2.

Table 2 SVI Correlation with Accidental Poisoning Mortality Rate

Social Vulnerability Index Variable	Spearman Correlation Coefficient
Housing units estimate (× 1,000 units)	-0.04
Percentage of persons below poverty	0.39
Percentage of civilian (age 16+) unemployed	0.36
Per capita income	-0.27
Percentage of persons with no high school diploma (age 25+)	0.29
Percentage of persons aged 65 and older	0.08
Percentage of persons aged 17 and younger	0.03
Percentage of civilian noninstitutionalized population with a disability	0.53
Percentage of single parent households with children under 18	0.10
Percentage minority (all persons except white, non-Hispanic)	-0.17
Percentage of persons (age 5+) who speak English 'less than well'	-0.18
Percentage of housing in structures with 10 or more units	-0.12
Percentage of mobile homes	0.31
Percentage of occupied housing units with more people than rooms	-0.11
Percentage of households with no vehicle available	0.02
Percentage of persons in institutionalized group quarters	-0.13
THEME: Socioeconomic	0.36
THEME: Minority Status & Language	0.41
THEME: Housing & Transportation	-0.23
THEME: Household Composition & Disability	0.06

3.2 Univariable Negative Binomial Regression

The results of the univariable negative binomial regression is shown in Table 3.

Table 3 Univariable Negative Binomial Regression Results

Social Vulnerability Index variable	Coefficient	Standard Error	IRR	p-value
Housing units estimate (× 1,000 units)	-0.0002	0.0003	1.00	0.42
Percentage of persons below poverty	0.05	0.0067	1.05	<0.01*
Percentage of civilian (age 16+) unemployed	0.11	0.0174	1.12	<0.01*
Per capita income (× $1,000)	-0.03	0.005	0.97	<0.01*
Percentage of persons with no high school diploma (age 25+)	0.04	0.0074	1.04	<0.01*
Percentage of persons aged 65 and older	0.03	0.0133	1.03	0.03*
Percentage of persons aged 17 and younger	0.01	0.0155	1.01	0.70
Percentage of civilian noninstitutionalized population with a disability	0.07	0.0068	1.07	<0.01*
Percentage of single parent households with children under 18	0.04	0.0194	1.04	0.05
Percentage minority (all persons except white, non-Hispanic)	-0.01	0.0024	0.99	<0.01*
Percentage of persons (age 5+) who speak English 'less than well'	-0.06	0.0158	0.94	<0.01*
Percentage of housing in structures with 10 or more units	-0.01	0.0045	0.99	0.01*
Percentage of mobile homes	0.02	0.0048	1.02	<0.01*
Percentage of occupied housing units with more people than rooms	-0.07	0.0287	0.93	0.01*
Percentage of households with no vehicle available	0.002	0.0055	1.00	0.72
Percentage of persons in institutionalized group quarters	-0.03	0.0100	0.98	0.01*
THEME: Socioeconomic (× 100)	0.01	0.0014	1.01	< 0.01*
THEME: Household Composition & Disability (× 100)	0.01	0.0013	1.01	< 0.01*
THEME: Minority Status & Language (× 100)	-0.01	0.0013	0.99	< 0.01*
THEME: Housing & Transportation (× 100)	0.002	0.0015	1.00	0.29

* indicates significance at the 0.05 level

Percentage of civilian (age 16+) unemployed had the largest regression coefficient and IRR (0.11 and 1.12, respectively), so that in this model a one-unit increase in %-unemployment resulted in a 12% increase in expected number of deaths, holding population constant. The only SVI variables that were not found to be statistically significant predictors of accidental poisoning mortality at the 0.05-level were housing units estimate, percentage of persons aged 17 or younger, percentage of single parent households with children under 18, and percentage of households with no vehicle available.

The theme variables socioeconomic, household composition & disability, and minority status & language were statistically significant predictors of accidental poisoning mortality at the 0.05-level, but the housing & transportation theme was not. The themes were not considered for inclusion in the multivariable model due to collinearity and lack of interpretability.

3.3 Multivariable Negative Binomial Regression

Statistical significance in the univariable models, the magnitude of the coefficient (IRR), and its correlation with other SVI variables were all considered when determining which variables would be included in the full model. The statistically significant variables are noted with an asterisk (*) in Table 1 and were first considered for the multivariable model. A cut-point of absolute Spearman correlation coefficient of 0.50 was used to determine if the variables would be considered concurrently in the full model. To represent the strong positively correlated block containing percentage of occupied housing units with more people than rooms, persons (age 5+) who speak English 'less than well', minority (all persons except white, non-Hispanic), housing in structures with 10 or more units, only percentage of occupied housing units containing more people than rooms was selected for the multivariable model because it has the largest effect size. In the correlated group containing percentage of persons with no high school diploma, persons below poverty, and percentage (age 16+) unemployed, percentage of civilian (age 16+) unemployed was included in the full model because of its effect size. Percentage of civilian noninstitutionalized population with a disability was chosen to represent the group of variables with strong negative correlations with per capita income. Persons in institutionalized group quarters was also included in the full model and did not have any collinearity issues with the other SVI variables.

The full model used for backwards selection included percentage of civilian (age 16+) unemployed, civilian noninstitutionalized population with a disability, occupied housing units with more people than rooms, persons in institutionalized group quarters, and persons aged 65 and older (Table 4).

Table 4 Multivariable Negative Binomial Regression Results

	Coefficient	Standard Error	IRR	p-value
Percentage of civilian (age 16+) unemployed	0.043	0.018	1.04	0.022*
Percentage of civilian noninstitutionalized population with a disability	0.060	0.008	1.06	<0.001*
Percentage of occupied housing units with more people than rooms	-0.059	0.024	0.94	0.014*
Percentage of persons in institutionalized group quarters	-0.036	0.008	0.96	<0.001*

* indicates significance at the 0.05 level

Unemployment and disability were positively related with accidental poisoning mortality, while housing units with more people than rooms, a measure of overcrowding, and persons in institutionalized group quarters, like nursing homes or correctional facilities, were negatively related to accidental poisoning mortality. These model predictors were assessed for the presence of collinearity, and the variance inflation factors (VIF) were all less than 2. VIF less than 10 is considered unproblematic, therefore we conclude that there is no collinearity issue in this regression model.

A likelihood ratio test was performed to confirm the use of a negative binomial model over Poisson regression in this case. The null hypothesis of the test is that there is no difference between the models. Here, the chi-square test statistic with one degree of freedom was 8196.31, $p < 0.001$, so there was strong evidence that the negative binomial model was more appropriate for these data.

Figure 4 shows a plot of the actual vs. predicted deaths from the multivariable negative binomial model. The data is more concentrated in lower numbers of deaths, and the prediction shows that the model performs well in counties with lower death counts but underestimates the high death counts. The data is sparse in areas with high death count, and so the model tends to underestimate deaths more often than it overestimates deaths.

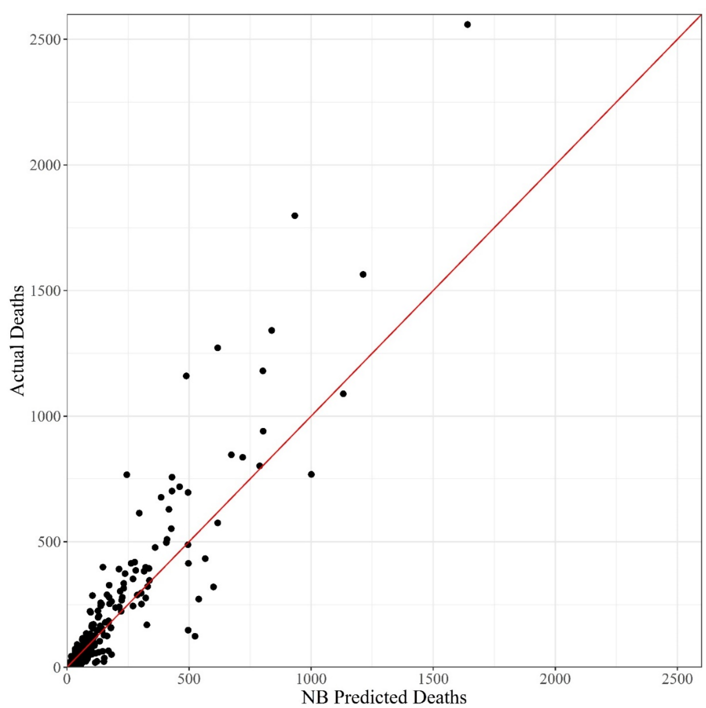

Figure 4 Actual vs. Multivariable Negative Binomial Predicted Deaths

3.4 Univariable Geographically Weighed Negative Binomial Regression

The local IRR results of the univariable SVI GWR models are shown in Figure 5. Some SVI variables showed strong geographical heterogeneity in their relationship with accidental poisoning mortality, while others had a homogeneous IRR across the whole region or showed little variability.

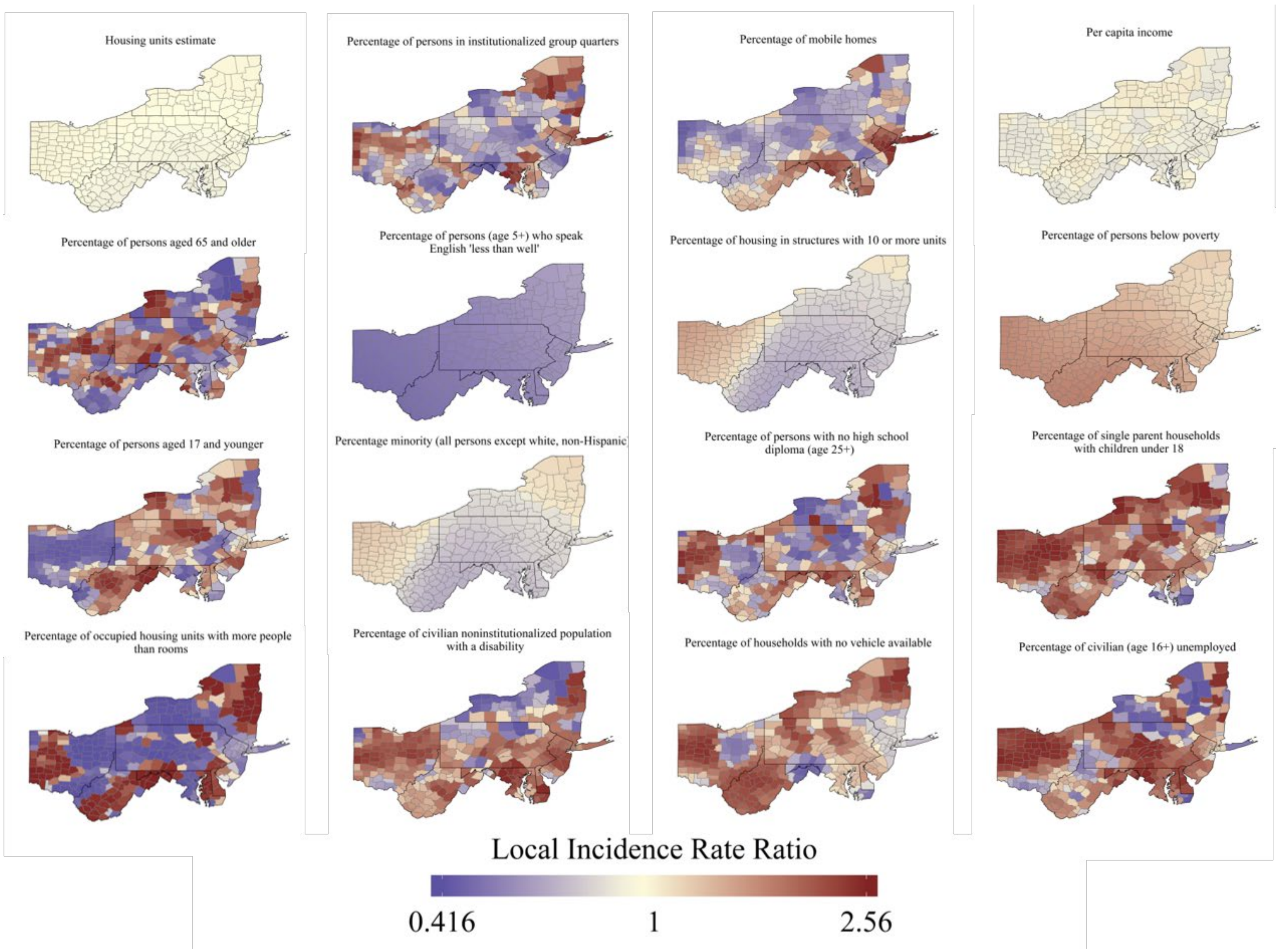

Figure 5 Univariable GWR Results

The SVI variables housing units estimate, per capita income, percentage of persons (age 5+) who speak English 'less than well', housing in structures with 10 or more units, persons below poverty, and minority (all persons except white, non-Hispanic) showed little change in IRR across the region. The local IRR for percentage of persons below poverty are all one or greater, indicating that there were no regions where poverty had a negative association with accidental poisoning mortality, or where increased poverty may have decreased mortality. The remaining SVI variables showed varying IRRs across the region and did not show any stark contrasts across state lines or other administrative boundaries. Maryland was distinguishable for having differences in reporting practices, however the univariable GWR results do not indicate that these differences in reporting are influencing the geographical relationship with SVI variables.

The estimated bandwidths used in the univariable GWR models had mean(standard deviation) 114.46(139.47) km and ranged from 22.29 km to 467.45 km, meaning there was significant variability in the bandwidth, and therefore the regional input, across the variables.

3.5 Multivariable Geographically Weighed Negative Binomial Regression

The multivariable GWR and the non-spatial negative binomial GLM models were fit with the following covariates: percentage of civilian (age 16+) unemployed, civilian noninstitutionalized population with a disability, occupied housing units with more people than rooms, and persons in institutionalized group quarters. The estimated bandwidth for this model was 56.49 km. The distance matrix and weight matrix are shown in Appendix A. Figure 6 shows the resulting local IRRs for each of GWR model's variables. The distribution of local IRRs across the region indicate the presence of heterogeneity in a global sense, with some counties clustered together having similar IRRs.

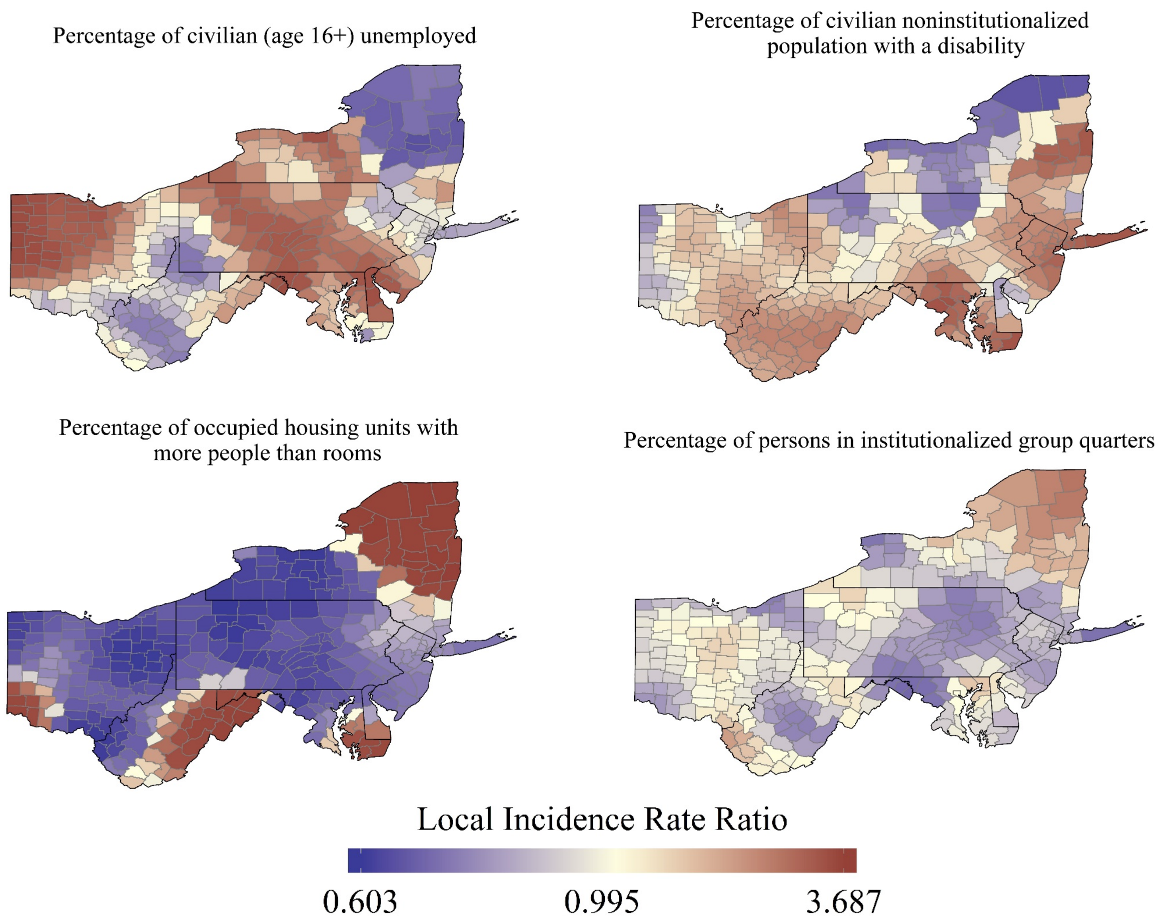

Figure 6 Multivariable GWR Results

3.6 Hypothesis Testing GWR Coefficients

The null hypothesis for the hypothesis test for the GWR coefficients was that the local coefficients are identical to the global coefficients resulting from the non-spatial multivariable negative binomial model, seen in Table 3. The results of these tests are shown in Figure 7. The covariates whose coefficients differ from the global value at the 0.05 significance level, after adjusting for multiple comparisons, are percentage of civilian noninstitutionalized population with a disability and occupied housing units with more people than rooms. There were areas of departure from the global IRR of accidental poisoning mortality by these variables in multiple areas. Some of these counties overlapped between the SVI predictors in northern New York.

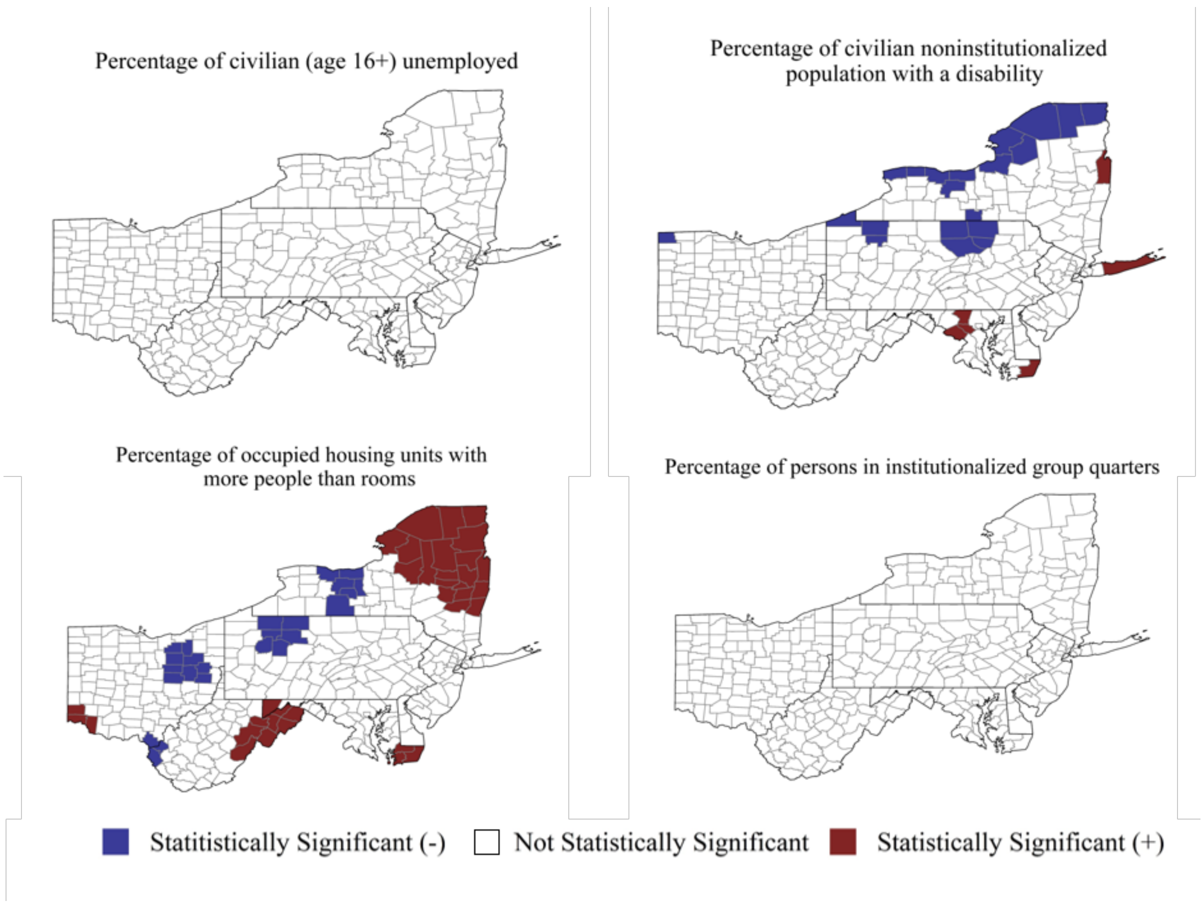

Figure 7 Multivariable GWR Hypothesis Testing Results

4.0 Discussion

The goal of this study was to identify the occurrence of spatial heterogeneity in the relationship between social determinants of health and accidental poisoning mortality. The results indicate that there are counties within Pennsylvania and neighboring states that have significantly different associations between SVI variables and overdose mortality than their global estimate.

Our results show that percentage of civilian noninstitutionalized population with a disability and occupied housing units with more people than rooms may be important in some areas as risk factors or protective features of the socioeconomic landscape. Measures of disability and overcrowding may be directly related to accidental poisoning mortality or may serve as a proxy for other unknown features of the population. A similar interpretation of these results could be that the clusters of counties whose rate ratios vary significantly from their global estimate have higher vulnerability to changes in those social determinants as they relate to accidental poisoning mortality.

The above claims support the original assertion that the opioid epidemic is made up of smaller sub-epidemics that lack one cohesive driving force. Heterogeneity of social relationships to accidental poisoning mortality is supported by other research studies. A recent study of Ohio counties found that there were unmeasured risk factors for opioid related deaths in some parts of the state and unmeasured protective factors in others.[14] This study also suggests that counties in Ohio experience heterogeneity in their relationship between overcrowding and overdose mortality (Figure 6). Substantial urban-rural variation has been shown in the relationship between prescription opioid and heroin overdose with social variables such as poverty, unemployment, high-school graduation rates, and median household income.[17] We did not control for urban rural variation in this study, however it is possible that some of the SVI variables may serve as a proxy for this metric, for example the measure of overcrowding may be interpreted to indicate urbanicity, or may indicate the occurrence of urbanization. It has previously been shown that there are fluctuations over space and time in drugs attributable to accidental poisoning mortality, which is supported by the claims made in this study.[12]

Figure 7 shows that much of the heterogeneity in the studied relationship can be attributed to random fluctuation, however some areas show a high level of deviation from their global estimates. It is interesting to note that several of the statistically significant clusters occur along the perimeter of the study area. This finding may be an artifact of limiting the study area, and future work on this topic may seek to determine if a cluster exists across the boundaries of those state lines in counties that were excluded here. An analysis that included all counties of the continental United States may be better poised to address the presence of these potential clusters.

There are a few limitations present in this study. First, the mortality data was limited to accidental, or unintentional, poisoning mortality at the county level. This method of classification is not atypical in overdose studies,[3; 5; 12] but it may undercount deaths in those counties whose overdose mortality is frequently assigned to ICD-10 codes Y10-Y14, poisoning of undetermined intent, or if other unknown reporting variability exists. A difference in reporting in Maryland was noted, however without national or state-level regulation for overdose death reporting, there may be other unknown variability across counties that is less trivial and unaccounted for in this study and other studies of this kind. These data are county level, which requires areal analysis instead of point data that indicate the exact geographic location of the deaths. Data at this level lose some specificity because they rely on a county centroid that may poorly that represent the location of the actual deaths.

Next, although modeling spatial variation through GWR has been shown to perform similarly to and have computational benefits over spatial random effects models, it is limited in its options for statistical inference.[21] The present study is intended to be exploratory and so strong methods of inference are not necessary, however in predictive studies supplementary methods of inference should be explored. An additional limitation of the chosen method is that the bandwidth is fixed by the model specifications, rather than allowing it to vary based on the variation in each covariate or across the study region. The optimal bandwidth chosen for the model may not have been fully optimized for each variable in the model, despite being calibrated for the model overall. Future work on this method could allow the bandwidth to vary over space and by predictor. There may be more interaction between counties in some regions than others, and therefore the optimal bandwidth may be wider in those interacting areas, and smaller in more isolated counties. There may also be differences in how each predictor variable interacts with nearby counties, and some social factors may have higher impact geographically than others.

Finally, this analysis did not fully explore variations in demographic variables by refined age or race groups, nor did it explore other measures of social or economic variability that are not contained in the CDC's SVI. In this analysis, the minority population was broadly defined as all excluding only non-Hispanic whites, despite evidence that there may also be differences between blacks and those of other races.[2; 12] Urban-rural classification was also not considered.

There is much work left to be done in researching spatial variation in the opioid epidemic. As mentioned above, it may be quite useful to consider the spatial heterogeneity when controlling for racial and age distribution, or alternatively, to reconduct a similar analysis based on death counts stratified by various race and age groups. Trends in overdose mortality vary across demographic groups, and therefore it is also likely that spatial difference exist. It may also be interesting to examine the spatial distribution of drug-specific mortality trends and how social factors are impacting those deaths. Demographic characteristics of a population are correlated with drug-specific usage and mortality[13], and therefore the social variables or recorded trends may be serving as a proxy measure for drug-specific trends that are not captured by the ICD-10 codes X40-X44 for accidental poisoning mortality. A way to expand this analysis could be to examine the drug-specific mortality trends and their relationship with socio-demographic attributes within the region.

5.0 Conclusion

This analysis was able to discover spatial heterogeneity in the relationship of social factors and accidental poisoning mortality. It identified higher and lower risk areas than the global rates for many of the SVI variables. The SVI variables that were found to be the strongest predictors of accidental poisoning mortality were percentage of unemployment, disability, overcrowding and population living in institutionalized group quarters. There were also statistically significantly differences from the global IRRs for variables measuring population percentages of disability and overcrowding in regional clusters. The relationship between disability and addiction is complex[18], however the difference in resource availability across the study region may impact the mortality rate in this population. If overcrowding is interpreted as a measure of past or current urbanization then the relationship that is seen may reflect the changing population and social dynamics in a community.

The present study showed that the spatial variation is not limited by administrative boundaries or dissimilar reporting practices, despite being important features in the data. It highlights the need to further explore the geographical variation in social and demographic variables to gain a clearer understanding of the factors driving the opioid epidemic.

Appendix A Distance and Weight Matrices for GWR

Sample values from the distance matrix (Table 5) and corresponding weight matrix (Table 6) are shown below.

Table 5 Distance Matrix (km)

	Hancock County, OH	Allen County, OH	Beaver County, PA	Lake County, OH	Pocahontas County, WV	Steuben County, NY	Forest County, PA	Fayette County, OH	Tioga County, PA	Butler County, OH
Hancock County, OH	0	45.0	282.0	225.6	431.7	541.9	375.6	161.1	543.0	190.1
Allen County, OH	45.0	0	317.5	270.3	444.2	585.0	417.0	145.5	584.6	153.3
Beaver County, PA	282.0	317.5	0	155.9	262.6	303.9	131.4	292.7	286.3	386.1
Lake County, OH	225.6	270.3	155.9	0	411.2	322.4	173.4	320.6	332.3	392.8
Pocahontas County, WV	431.7	444.2	262.6	411.2	0	490.6	359.3	328.5	448.4	414.9
Steuben County, NY	541.9	585.0	303.9	322.4	490.6	0	175.1	593.2	56.1	682.7
Forest County, PA	375.6	417.0	131.4	173.4	359.3	175.1	0	418.1	167.6	507.9
Fayette County, OH	161.1	145.5	292.7	320.6	328.5	593.2	418.1	0	578.9	97.2
Tioga County, PA	543.0	584.6	286.3	332.3	448.4	56.1	167.6	578.9	0	671.4
Butler County, OH	190.1	153.3	386.1	392.8	414.9	682.7	507.9	97.2	671.4	0

Table 6 Weight Matrix

	Hancock County, OH	Allen County, OH	Beaver County, PA	Lake County, OH	Pocahontas County, WV	Steuben County, NY	Forest County, PA	Fayette County, OH	Tioga County, PA	Butler County, OH
Hancock County, OH	1	0.72809	3.88E-06	0.00034	2.07E-13	1.04E-20	2.52E-10	0.01711	8.65E-21	0.00348
Allen County, OH	0.72809	1	1.38E-07	1.06E-05	3.71E-14	5.15E-24	1.47E-12	0.03629	5.57E-24	0.02513
Beaver County, PA	3.88E-06	1.38E-07	1	0.02218	2.03E-05	5.20E-07	0.06690	1.48E-06	2.64E-06	7.19E-11
Lake County, OH	0.00034	1.06E-05	0.02218	1	3.10E-12	8.47E-08	0.00900	1.01E-07	3.07E-08	3.16E-11
Pocahontas County, WV	2.07E-13	3.71E-14	2.03E-05	3.10E-12	1	4.15E-17	1.63E-09	4.52E-08	2.07E-14	1.93E-12
Steuben County, NY	1.04E-20	5.15E-24	5.20E-07	8.47E-08	4.15E-17	1	0.00820	1.14E-24	0.61081	1.94E-32
Forest County, PA	2.52E-10	1.47E-12	0.06690	0.00900	1.63E-09	0.00820	1	1.27E-12	0.01227	2.79E-18
Fayette County, OH	0.01711	0.03629	1.48E-06	1.01E-07	4.52E-08	1.14E-24	1.27E-12	1	1.56E-23	0.22731
Tioga County, PA	8.65E-21	5.57E-24	2.64E-06	3.07E-08	2.07E-14	0.61081	0.01227	1.56E-23	1	2.13E-31
Butler County, OH	0.00348	0.02513	7.19E-11	3.16E-11	1.93E-12	1.94E-32	2.79E-18	0.22731	2.13E-31	1

The distance matrix gives of distance between county centroids in kilometers. Table 6 shows the weight matrix that was calculated using the Gaussian kernel with a bandwidth measure of 56.49 km. Counties near one another are weighted much more heavily than those counties that are further apart.

Appendix B Social Vulnerability Index Variables

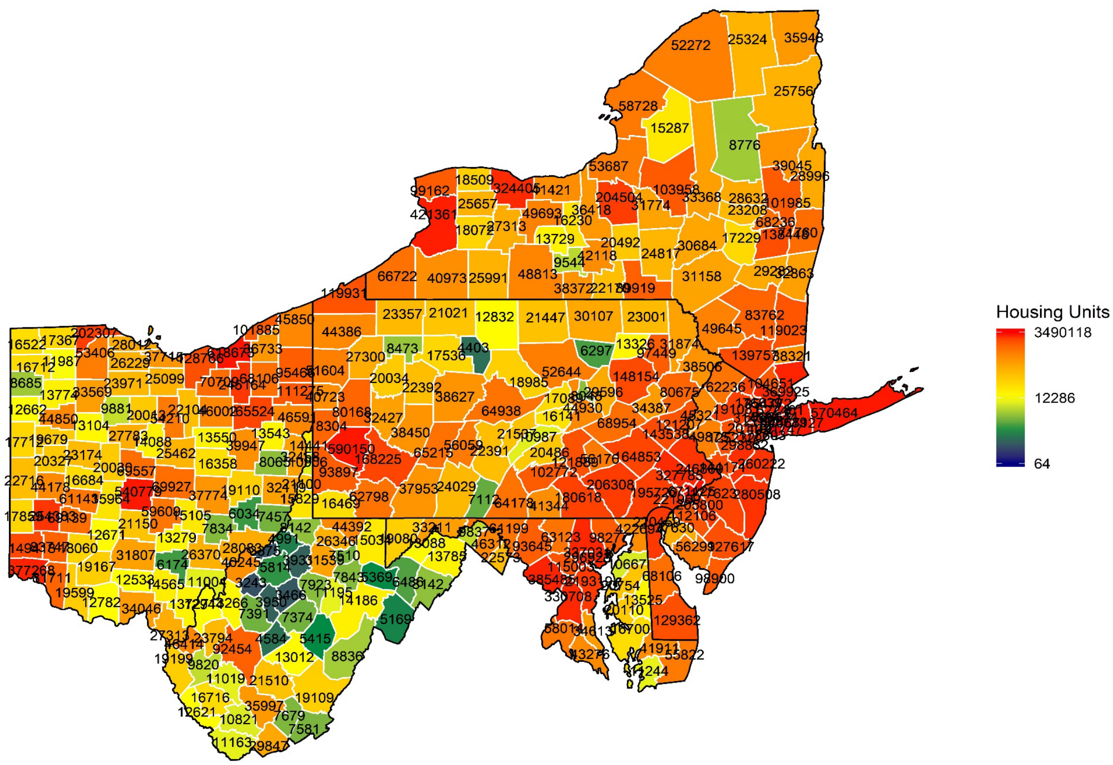

Figure 8 Housing Units Estimate by County

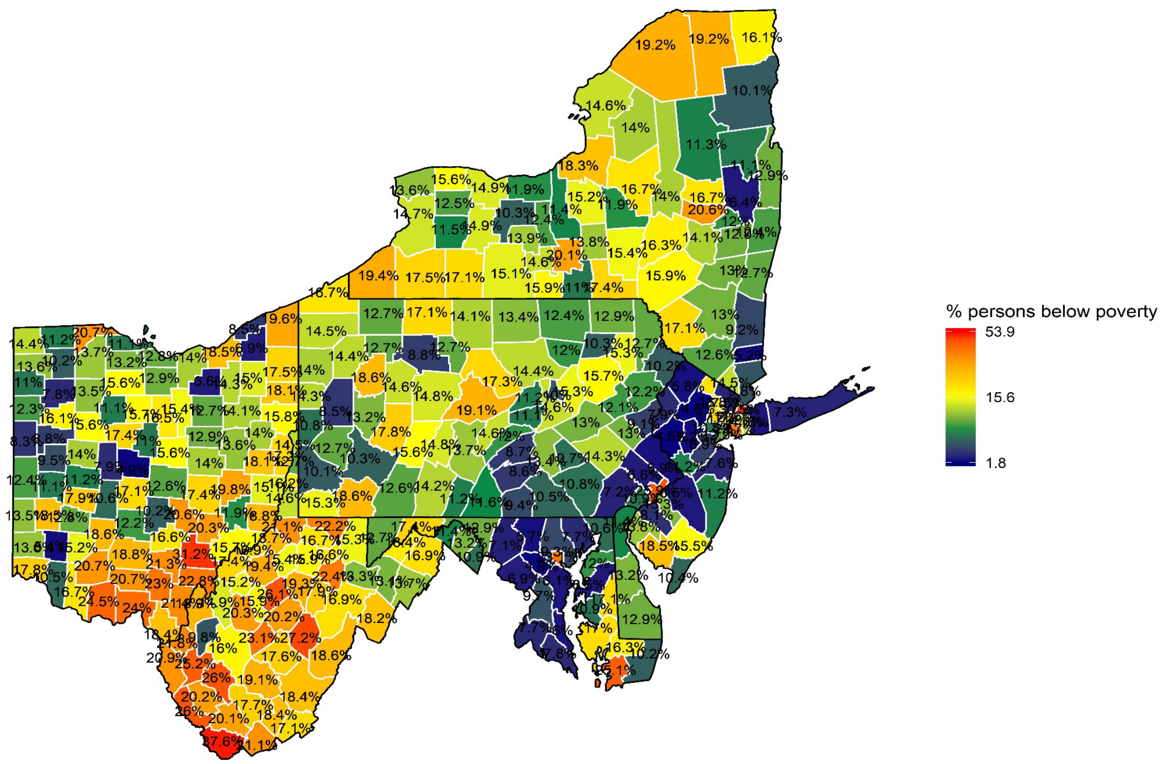

Figure 9 Percentage of Persons Below Poverty by County

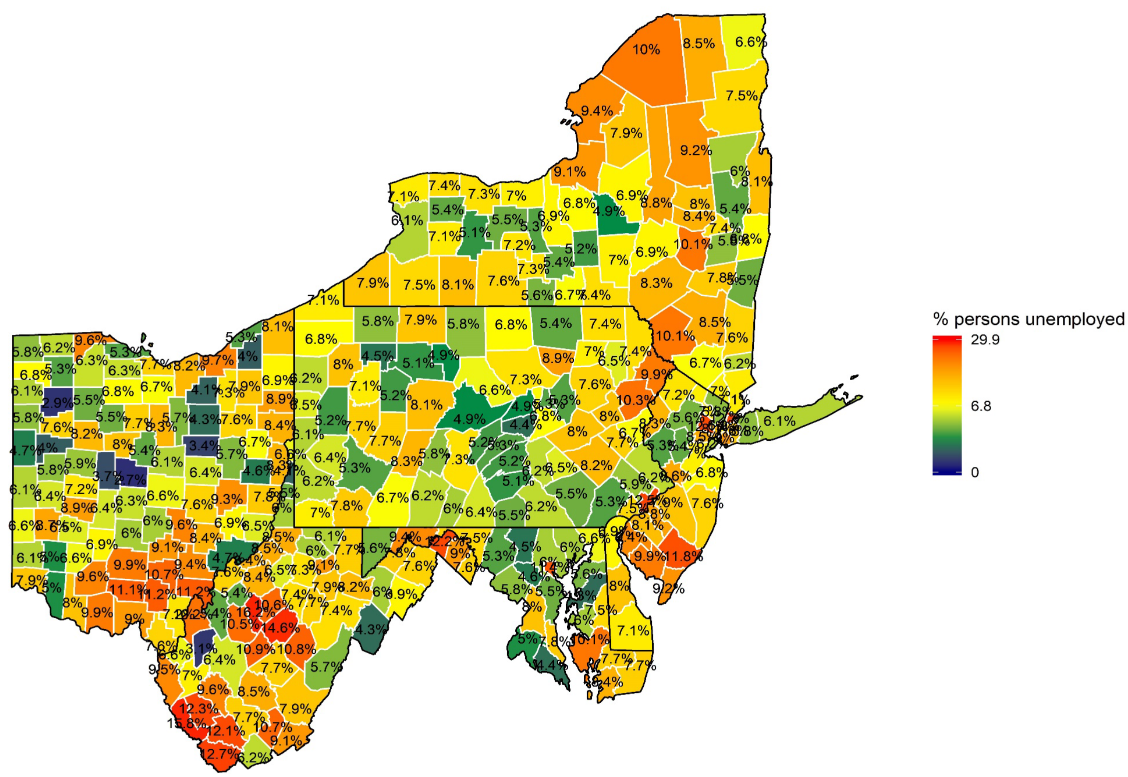

Figure 10 Percentage of Persons Unemployed by County

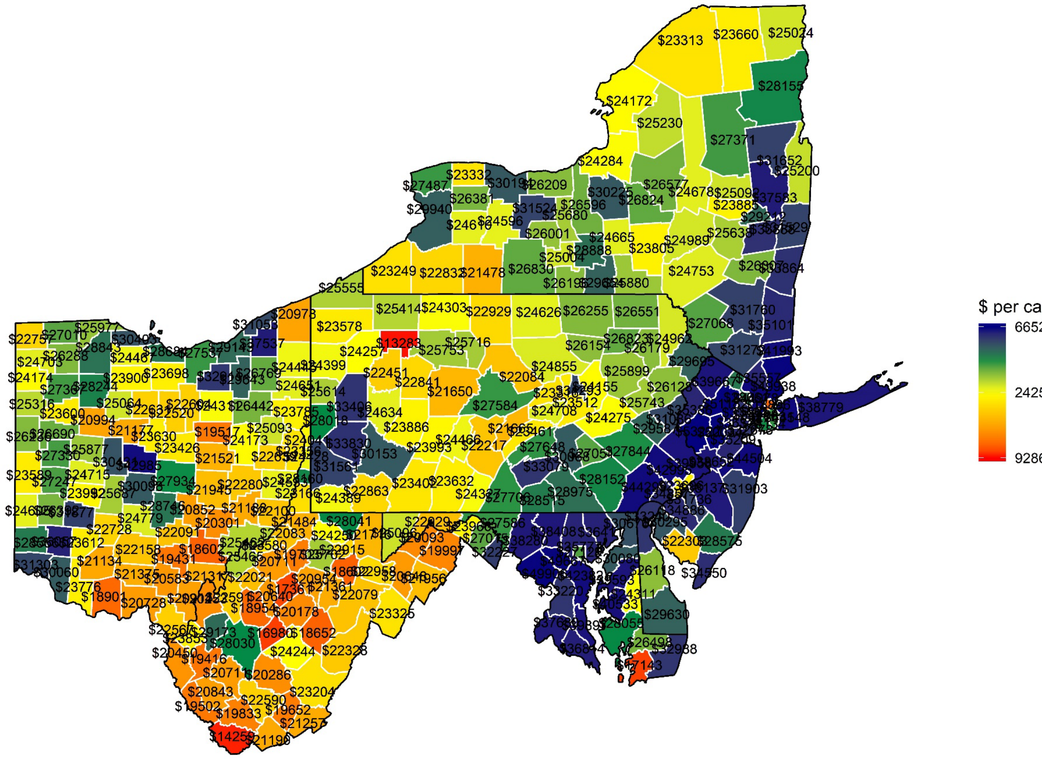

Figure 11 Per Capita Income by County

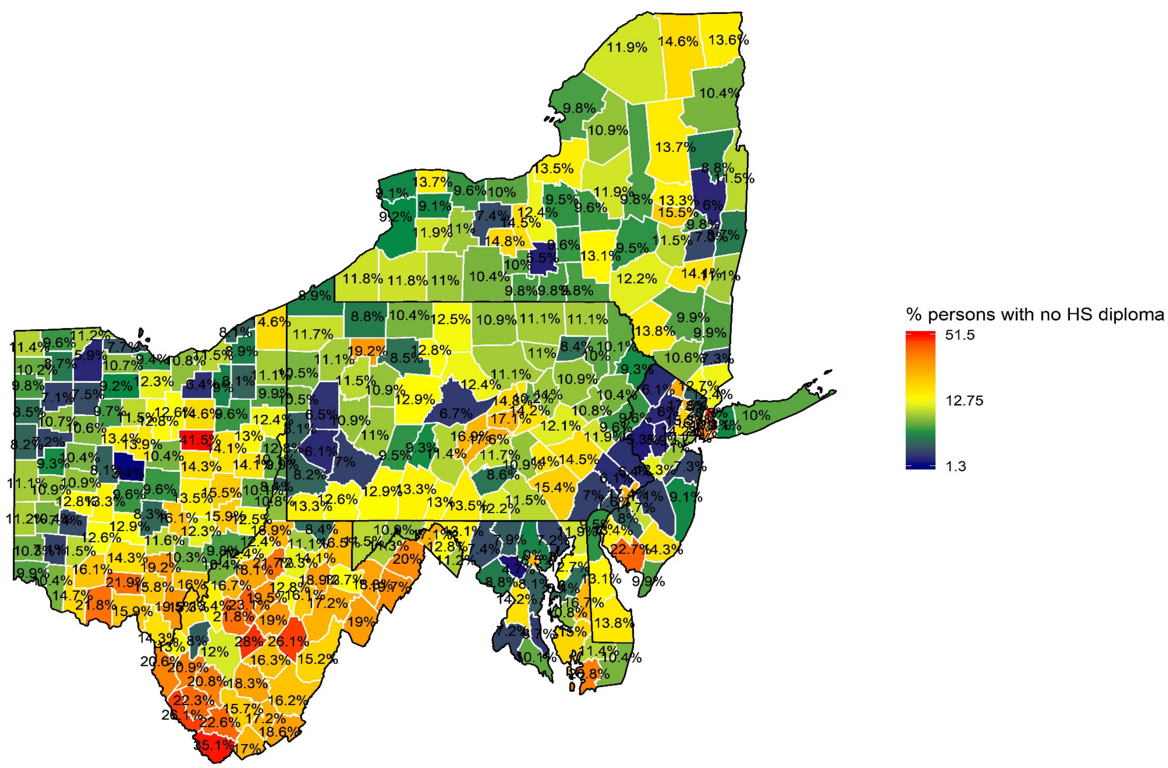

Figure 12 Percentage of Persons with No High School Diploma (Age 25+) by County

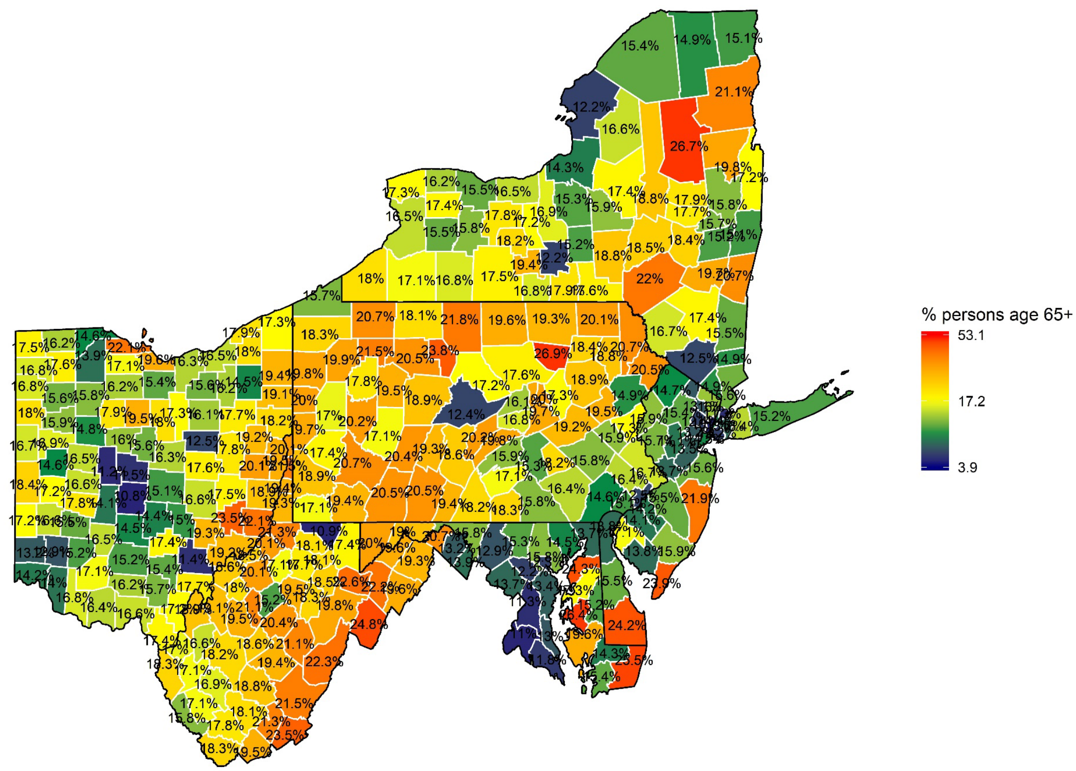

Figure 13 Percentage of Persons Aged 65 and Over by County

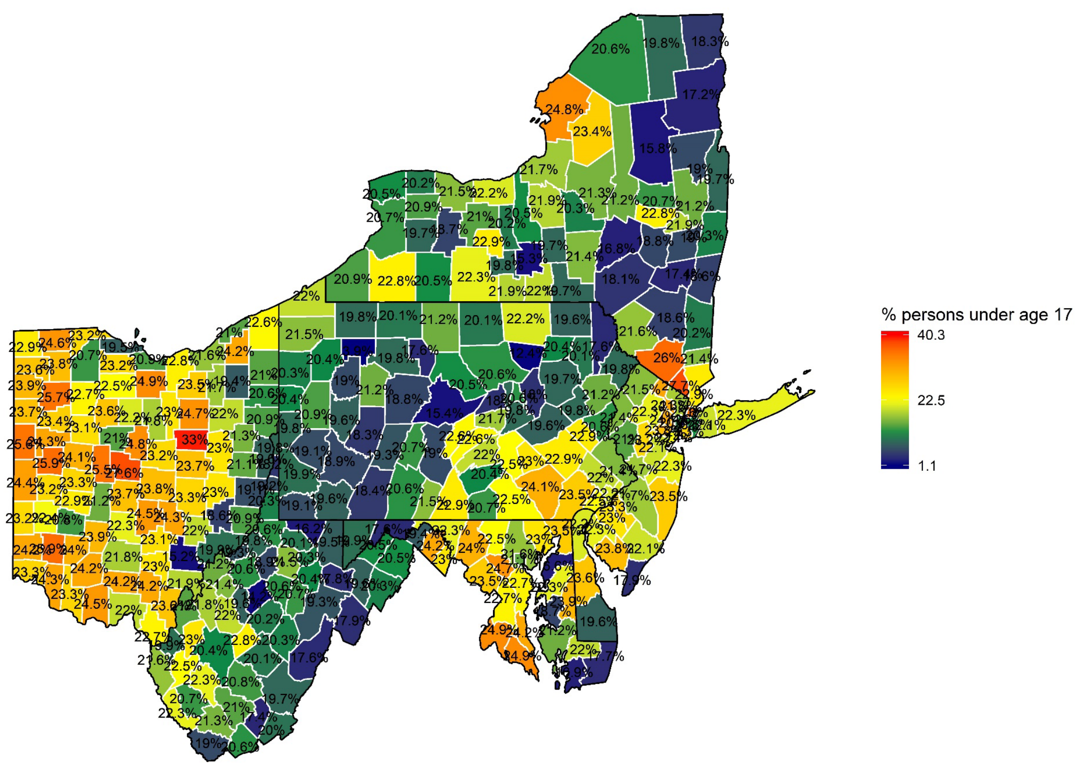

Figure 14 Percentage of Persons Aged 17 and Younger by County

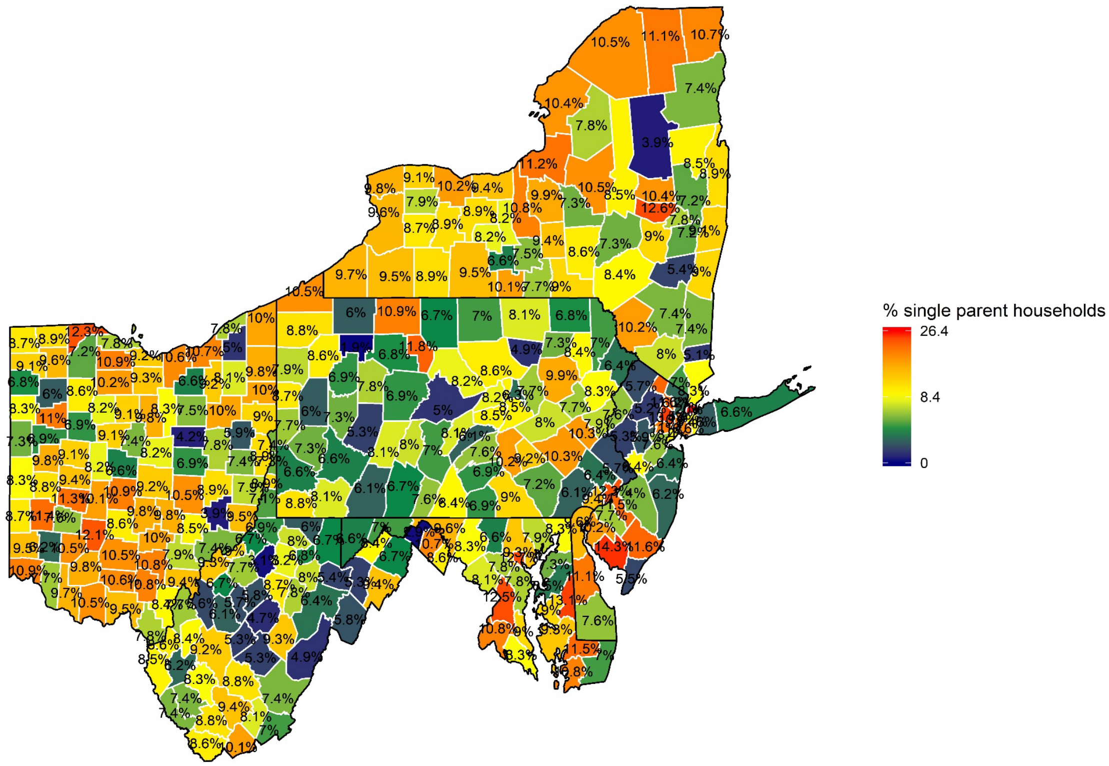

Figure 15 Percentage of Single Parent Households with Children Under 18 by County

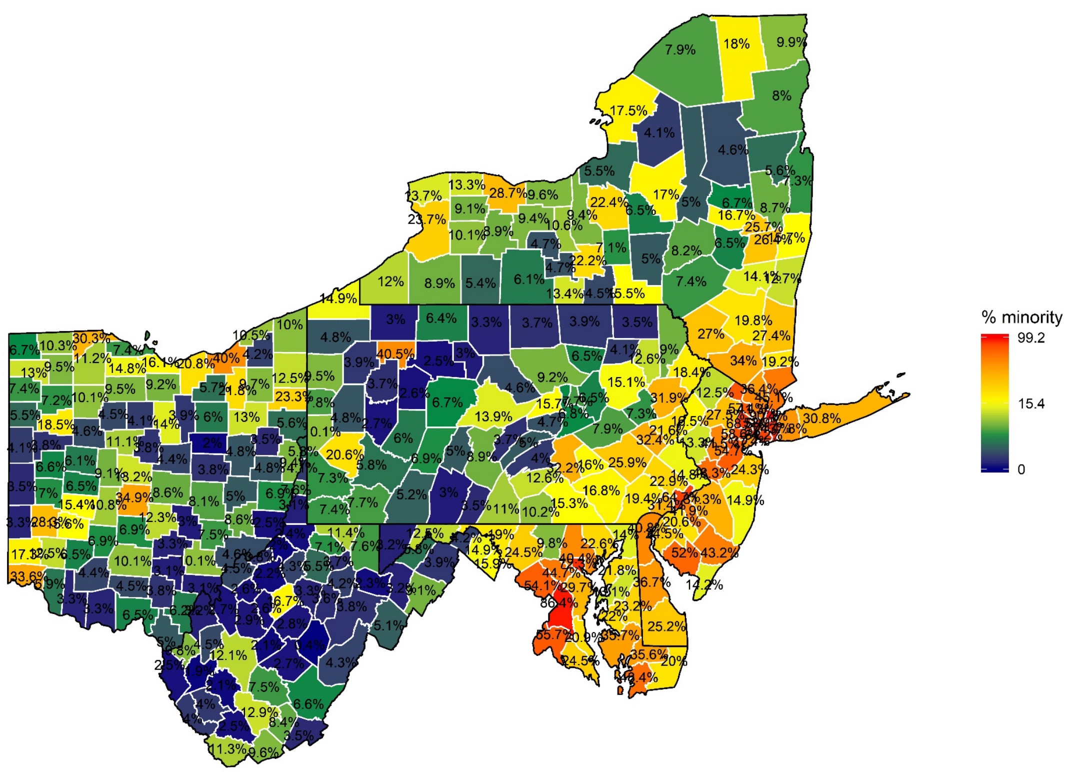

Figure 16 Percentage Minority (All Persons Except White, Non-Hispanic) by County

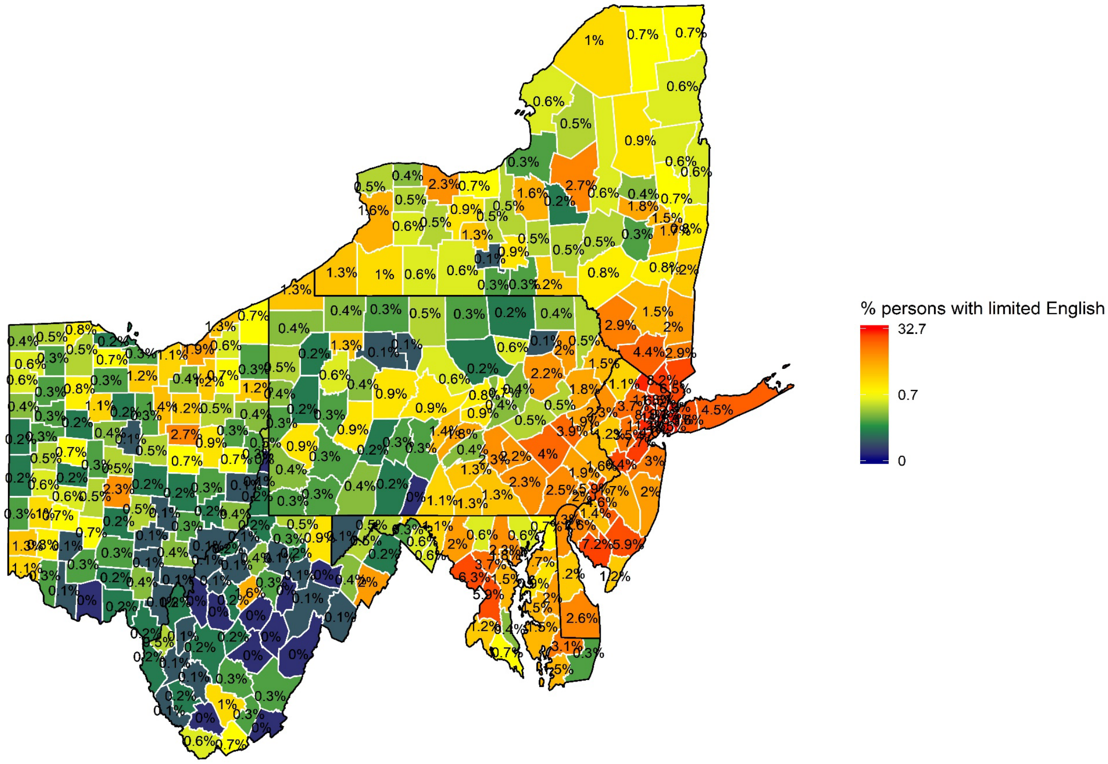

Figure 17 Percentage of Persons (Age 5+) Who Speak English 'Less Than Well' by County

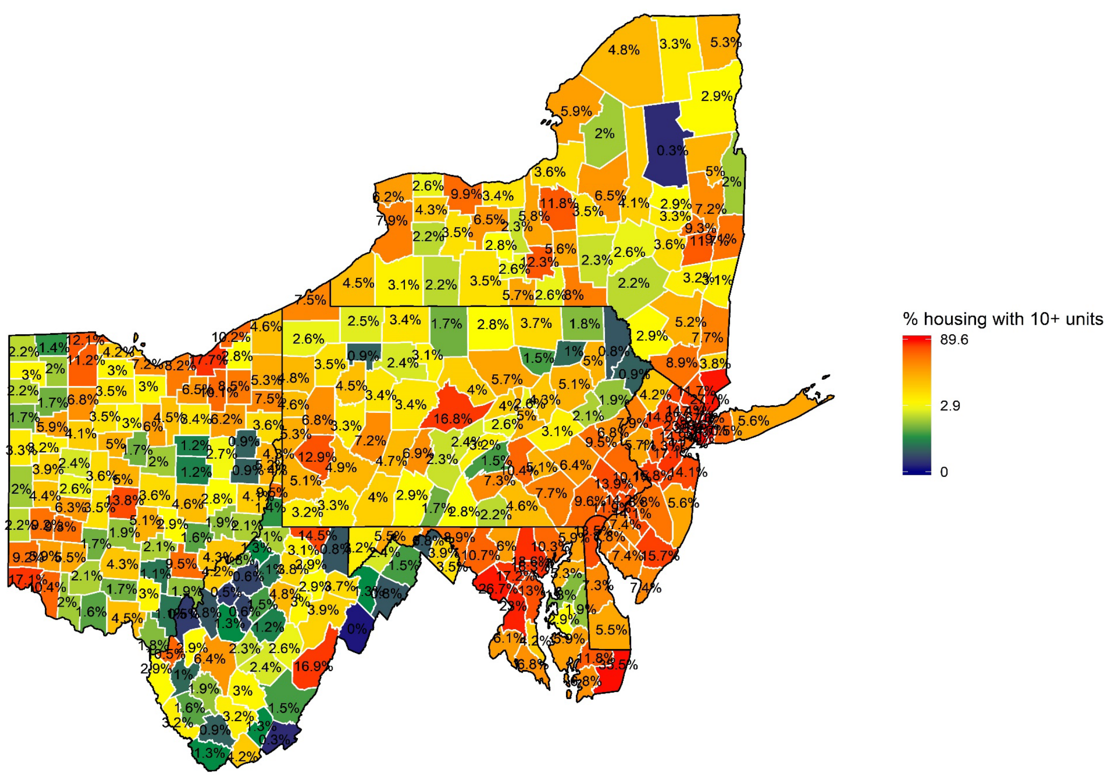

Figure 18 Percentage Of Housing in Structures with 10 or More Units by County

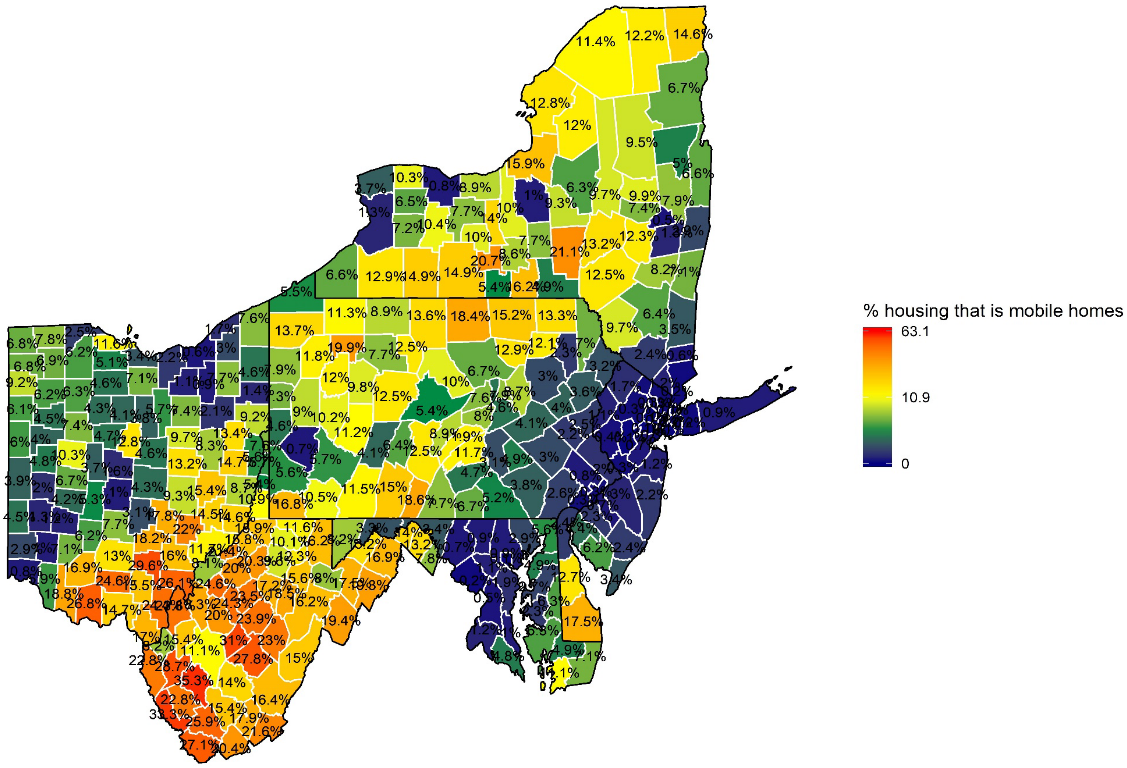

Figure 19 Percentage of Mobile Homes by County

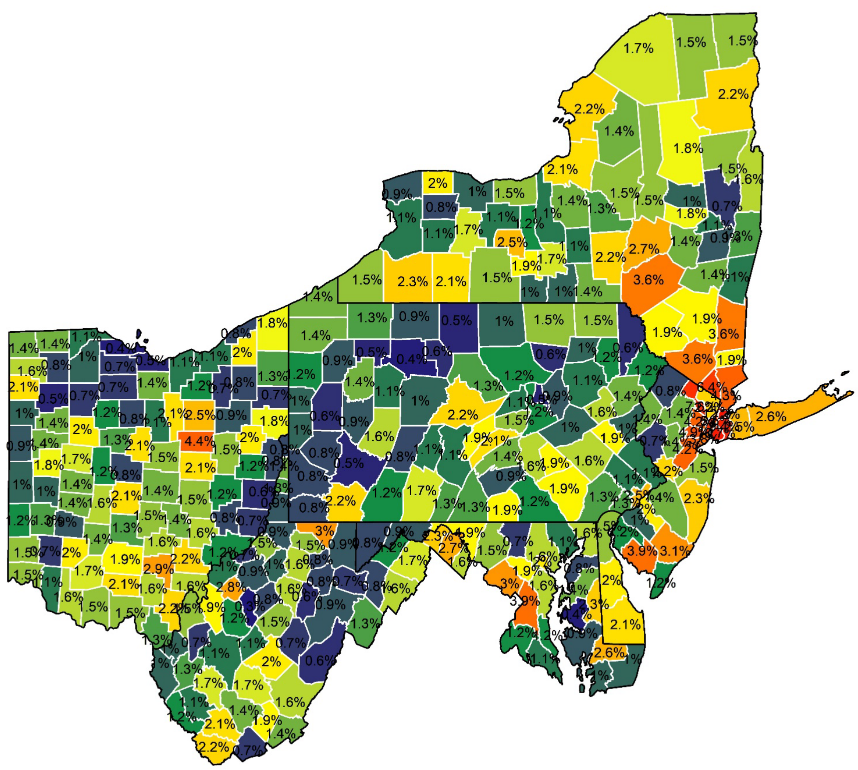

Figure 20 Percentage of Housing in Structures with 10 or More Units by County

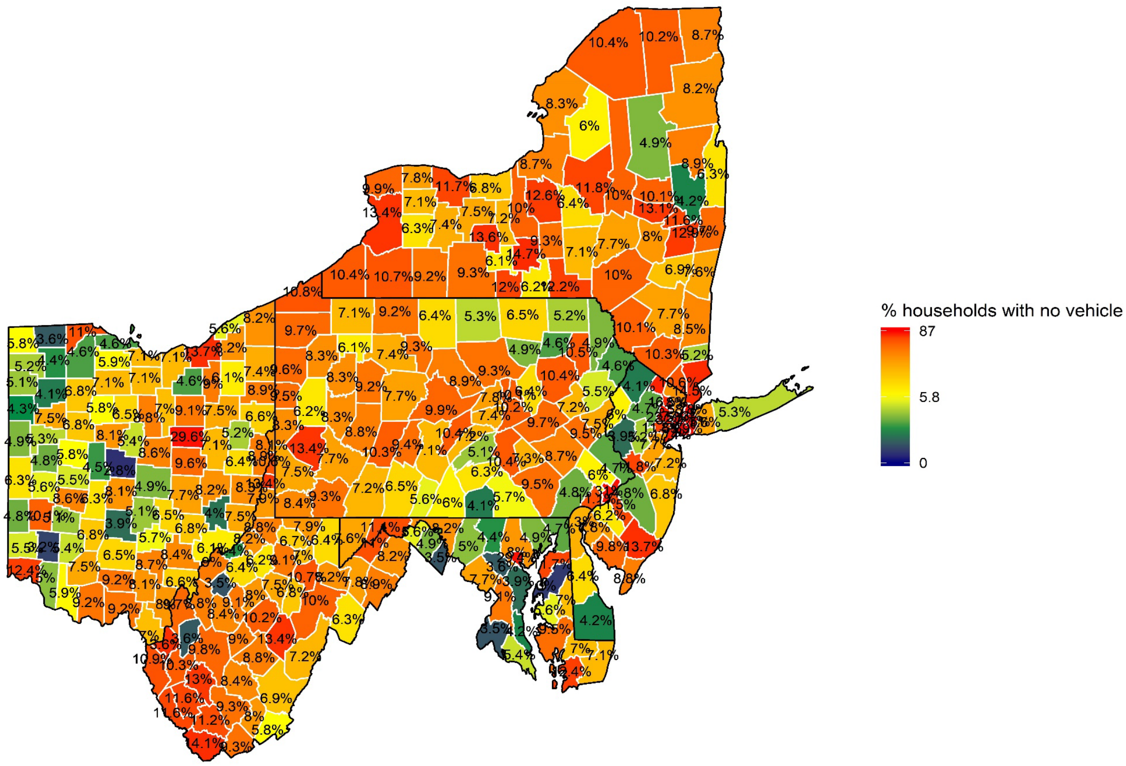

Figure 21 Percentage of Households with No Vehicle Available by County

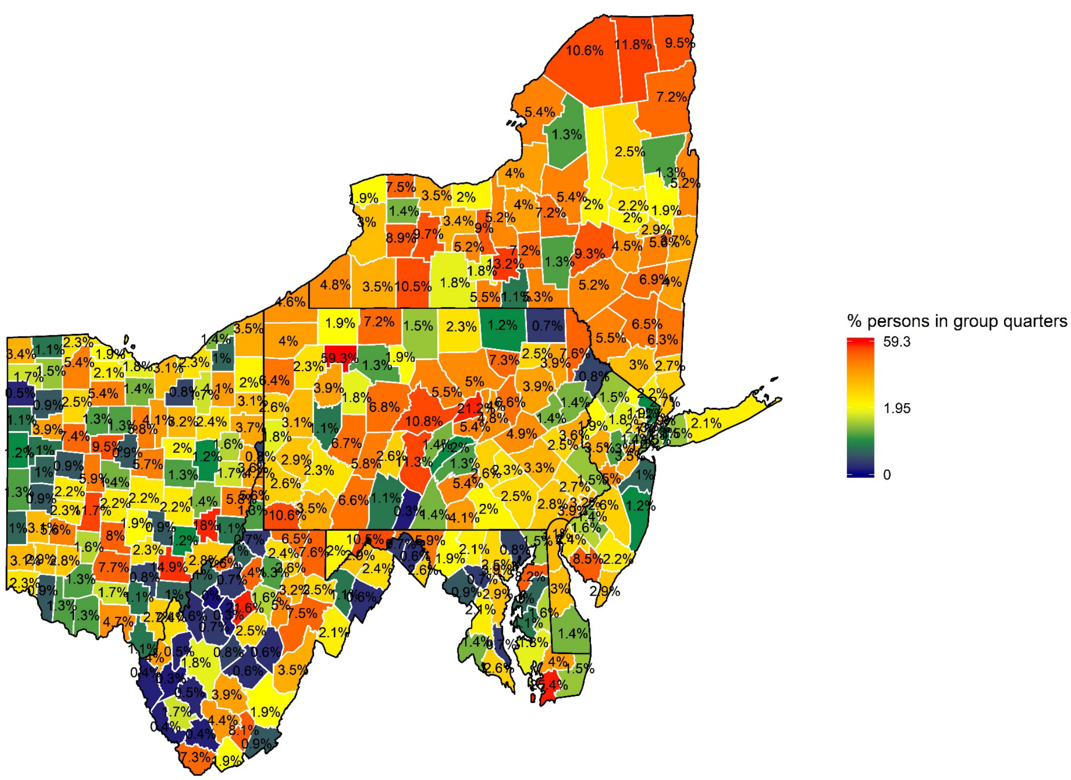

Figure 22 Percentage of Persons in Institutionalized Group Quarters by County

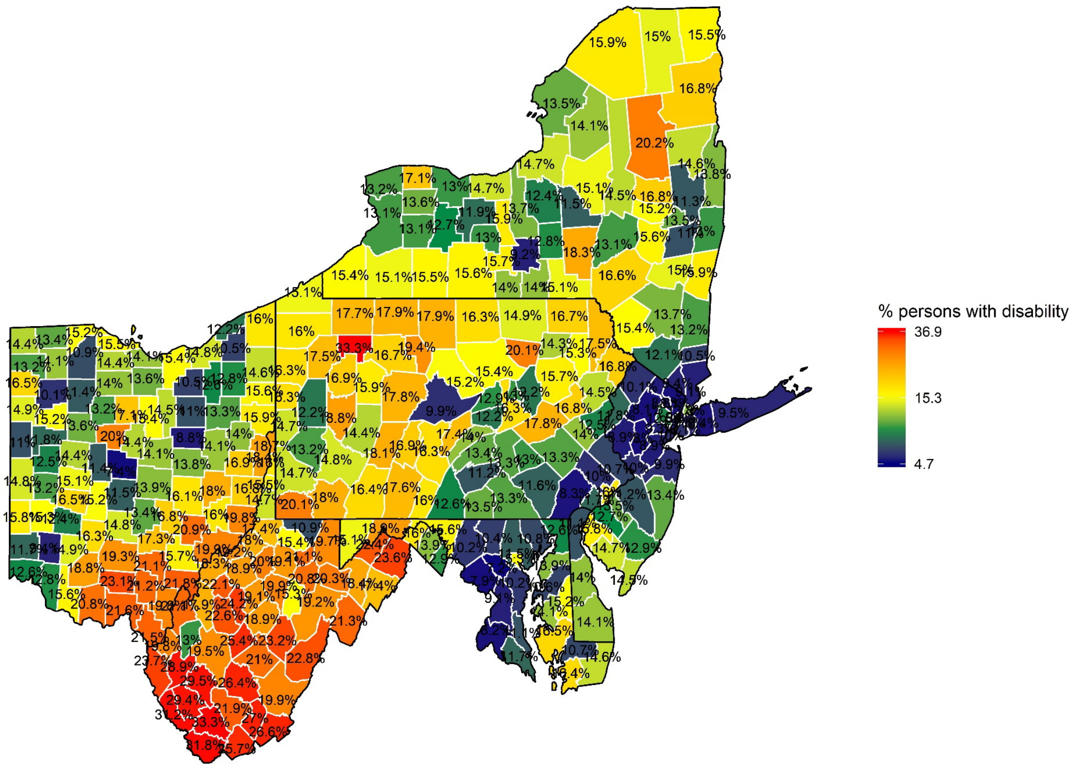

Figure 23 Percentage of Civilian Noninstitutionalized Population with a Disability by County

Bibliography

1. Agency for Toxic Substances and Disease Registry. (2016). Social Vulnerability Index. from Center for Disease Control https://svi.cdc.gov/data-and-tools-download.html

2. Alexander, M. J., Kiang, M. V., & Barbieri, M. (2018). Trends in Black and White Opioid Mortality in the United States, 1979–2015. *Epidemiology, 29*(5).

3. Balmert, L. C., Buchanich, J. M., Pringle, J. L., Williams, K. E., Burke, D. S., & Marsh, G. M. (2016). Patterns and Trends in Accidental Poisoning Deaths: Pennsylvania's Experience 1979-2014. *PLoS ONE, 11*(3), e0151655. doi:10.1371/journal.pone.0151655

4. Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist., 29*(4), 1165-1188. doi:10.1214/aos/1013699998

5. Buchanich, J. M., Balmert, L. C., Williams, K. E., & Burke, D. S. (2018). The Effect of Incomplete Death Certificates on Estimates of Unintentional Opioid-Related Overdose Deaths in the United States, 1999-2015. *Public Health Reports, 133*(4), 423-431. doi:10.1177/0033354918774330

6. Burke, D. S. (2016). Forecasting the opioid epidemic. *Science, 354*(6312), 529.

7. Case, A., & Deaton, A. (2017). *Mortality and Morbidity in the 21st Century*. Retrieved from https://www.brookings.edu/wp-content/uploads/2017/03/6_casedeaton.pdf

8. Dasgupta, N., Beletsky, L., & Ciccarone, D. (2017). Opioid Crisis: No Easy Fix to Its Social and Economic Determinants. *American Journal of Public Health, 108*(2), 182-186. doi:10.2105/AJPH.2017.304187

9. Department of Biostatisics. (2019). MOrtality Informatics and Research Analytics. Available from University of Pittsburgh, Graduate School of Public Health MOrtality Informatics and Research Analytics, from University of Pittsburgh, Graduate School of Public Health https://moira.pitt.edu

10. Fotheringham, A., Brunsdon, C., & Charlton, M. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. *John Wiley & Sons, 13*.

11. Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *Journal of Statistical Software; Vol 1, Issue 17 (2015)*. doi:10.18637/jss.v063.i17

12. Jalal, H., Buchanich, J. M., Roberts, M. S., Balmert, L. C., Zhang, K., & Burke, D. S. (2018). Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016.

13. Jalal, H., Buchanich, J. M., Roberts, M. S., Balmert, L. C., Zhang, K., & Burke, D. S. (2018). Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. *Science, 361*(6408), eaau1184. doi:10.1126/science.aau1184

14. Kline, D., Hepler, S., Bonny, A., & McKnight, E. (2019). A joint spatial model of opioid-associated deaths and treatment admissions in Ohio. *Annals of Epidemiology, 33*, 19-23. doi:https://doi.org/10.1016/j.annepidem.2019.02.004

15. Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science, 28*(4), 660-681.

16. National Center for Health Statistics. Underlying Cause of Death 1999-2016 on CDC WONDER Online Database. Retrieved September 9, 2019, from Centers for Disease Control and Prevention https://wonder.cdc.gov/ucd-icd10.html

17. Pear, V. A., Ponicki, W. R., Gaidus, A., Keyes, K. M., Martins, S. S., Fink, D. S., . . . Cerda, M. (2019). Urban-rural variation in the socioeconomic determinants of opioid overdose. *Drug Alcohol Depend, 195*, 66-73. doi:10.1016/j.drugalcdep.2018.11.024

18. Peters, J. L., Durand, W. M., Monteiro, K. A., Dumenco, L., & George, P. (2018). Opioid Overdose Hospitalizations among Medicare-Disability Beneficiaries. *The Journal of the American Board of Family Medicine, 31*(6), 881-896. doi:10.3122/jabfm.2018.06.180152

19. R Development Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

20. Venables, W. N., Ripley, B. . (2002). *Modern Applied Statistics with S* (Fourth ed.). New York, NY: Springer.

21. Waller, L., Zhu, L., Gotway, C., Gorman, D., & Gruenewald, P. (2007). Quantifying Geographic Variations in Associations between Alcohol Distribution and Violence: A Comparison of Geographically Weighted Regression and Spatially Varying Coefficient Models. *Stochastic Environmental Research and Risk Assessment, 21*, 573-588. doi:10.1007/s00477-007-0139-9

22. Woolf, S. H., Chapman, D. A., Buchanich, J. M., Bobby, K. J., Zimmerman, E. B., & Blackburn, S. M. (2018). Changes in midlife death rates across racial and ethnic groups in the United States: systematic analysis of vital statistics. *British Medical Journal, 362*, k3096. doi:10.1136/bmj.k3096