7-21-2020

Cluster Analysis of Opioid Accessibility in the Carolinas Using Data from the ARCOS Database and an Enhanced Two-Step Floating Catchment Area Method

Michael T. Bunker
Coastal Carolina University

Cluster Analysis of Opioid Accessibility in the Carolinas Using Data from the ARCOS Database and an Enhanced Two-Step Floating Catchment Area Method

By

Michael T. Bunker

Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Information Systems Technology with a
Concentration in Information Security and Data Analytics

Gupta College of Science
Coastal Carolina University
2020

____
Dr. William Jones
Department of Computing Sciences

____
Dr. Jean French
Chair, Department of Computing Sciences

____
Dr. Susan Bergeron
Department of Anthropology and Geography

____
Dr. Michael Roberts
Dean, Gupta College of Science

____
Dr. Stephen Borders
Department of Health Sciences

ACKNOWLEDGMENT

This thesis could not have been completed without the advice and assistance of several people. Dr. William Jones stepped in and provided support under unexpected circumstances. My committee members Dr. Susan Bergeron and Dr. Stephen Borders provided tremendous support and guidance throughout this project. Dr. Borders gave me invaluable insight into the perspective of a public health official and the best ways to communicate technical details to policy-oriented personnel. Dr. Bergeron's knowledge of GIS and geographic principles helped to inform my thinking and opened up new ways of interpreting my results that otherwise would not have occurred.

I also want to thank Dr. Lindsey Bell and Dr. Keshav Jagannathan for reintroducing me to the world of statistics in their classes, and for being available, often with only a moment's notice, to answer questions and talk through the nuances of statistical models.

A big thank you goes to my wife, Dr. Sara Hottinger, for her love and support throughout this project. Our late-night conversations were essential to my steady progress and success. Thank you also to her for reading and editing multiple drafts and for providing lots of suggestions for improvement. Another big thank you goes to my kids, Amelia and Emmeline, for their patience. I disappeared into the library during many weekends over the last two years, and I plan on making up for those weekends for many years to come.

ABSTRACT

This study takes advantage of transaction level data from the U.S. Drug Enforcement Administration's (DEA) Automation of Reports and Consolidated Orders System (ARCOS) database made newly available under court order by *The Washington Post* in July 2019. This data details individual shipments of pharmaceutical opioid analgesics from wholesalers to retail distributors. Using the Enhanced 2-Step Floating Catchment Area (E2SFCA) method, this study calculated access to opioid morphine milligram equivalents (MME) per capita for census tracts in North Carolina and South Carolina during the year 2009. This study demonstrated that outlier volumes of opioid analgesics at individual pharmacies are not always co-located with census tracts that have access to outlier per capita opioid volumes. In addition, this study used 5-year average American Community Survey (ACS) data to identify distinct populations and compare their access to opioid analgesics using a k-medoids clustering algorithm. While opioid access for most clusters corresponded to previous research, a rural, socially vulnerable African American population in the Low Country of both states was identified with high access to opioid analgesics. This finding is contrary to previous research, indicating the need for further investigation.

Contents

List of Tables

List of Figures

I. INTRODUCTION

The opioid epidemic in the United States is a public health crisis. From the 1990's until 2010, the rates of opioid prescribing and overdose deaths linked to prescription opioid analgesics steadily rose. In 2010, 75% of all prescription drug overdose deaths were caused by opioids, and the number of those deaths were 400% greater than in 1999. In 2009, 256.9 million opioid prescriptions were filled at pharmacies [1].

The details of all shipments of prescription opioids from manufacturers and wholesalers to pharmacies and prescribing practitioners are recorded in a U.S. Drug Enforcement Administration (DEA) database, the Automation of Reports and Consolidated Orders System (ARCOS). Historically, that data has been aggregated and reported on a regular basis to the public, but access to individual records has never been allowed despite the details of distribution that might lead to insights related to the opioid epidemic [2]. In 2019, *The Washington Post* was granted the right via court order to publish original, unaggregated data from the DEA's ARCOS database for the years 2006 through 2012. The public availability of this data is an unprecedented opportunity for researchers to understand where, when, and by whom opioid analgesics were distributed throughout the United States during this time period. Transactional data from ARCOS makes it much easier to compare opioid analgesic distribution with smaller geographic units like the U.S. Census Bureau's census tracts and provides an opportunity to address the limitations outlined by Modarai et al. [3].

Evidence for the increase in sales of opioid analgesics over time is present in the ARCOS database. From 2006 to 2010 opioid sales increased over 50% from 66.89 million grams to 102.79

million grams nationally[1]. Volume of opioid sales during this time varied across states and within states as demonstrated by the DEA's quarterly publications of aggregated ARCOS data. Data submitted to this database exists as transactional shipment data. It is released by the DEA to the public with all transactional data aggregated to the 3-digit ZIP code which for many portions of the country represent multiple counties [4].

Researchers have studied the differential distribution of opioids at the state-, county-, and 3-digit ZIP code level and found associations between both increasing volumes of opioids sold and numbers of prescriptions written with an increase in the number of prescription opioid-related hospitalizations and overdose deaths [5]–[7]. Researchers have also found differentiation among demographic and socioeconomic groups and the odds ratios of their risk for overdose deaths due to misuse of opioid analgesics [5]. Modarai et al. [3] studied opioid distribution, opioid poisoning hospitalization, and OP-related deaths between 2008 and 2010 in North Carolina in order to identify variation of risk for different demographic groups at a geographic level smaller than the state level. They asked whether this variation could be identified at the sub-state level. One of their sources of data was the ARCOS database, however they argued that their results were limited by the 3-digit ZIP code aggregation that the DEA performed before releasing this data.

This thesis considers data from the ARCOS database covering the area of North Carolina and South Carolina for the year 2009. A brief description of the origins of the opioid epidemic and its progression is provided. The literature review explores previous research of opioid distribution and prescription rates and the health impacts on different sub-populations. It also explores floating catchment areas (FCA), a method for calculating a population's access to an available opioid

[1] These numbers are limited to sales of oxycodone and hydrocodone, because this study only examined those two opioid analgesic drugs.

supply. The methodology section explains the data sources and methodology used in this study. The results section provides an exploratory data analysis, a description of clusters formed with a k-medoids clustering algorithm, and the results of an OLS regression model using opioid accessibility calculated from an FCA method as the dependent variable. The discussion section reviews the study's results in the context of past research.

This study has several goals. The new availability of individual records from the ARCOS database is an opportunity to explore that data spatially and its relationship with the characteristics of the population in which opioid shipments were distributed. This will be accomplished using a k-medoids clustering algorithm to identify similar populations based on demographic and socioeconomic variables. Modarai et al. [3] argued that aggregated ARCOS data limited the results of their study; the public release of unaggregated ARCOS data provides an opportunity to address that limitation. Additionally, having access to more granular data in combination with available demographic and socioeconomic variables is an opportunity to test whether the relationship between those variables and geographic variability in opioid distribution remains the same at a smaller geographic unit. In this study, I hypothesize that the distribution of medical use opioid analgesics was non-random in relation to demographic and socioeconomic characteristics of the population. I used an ordinary least squares (OLS) regression model to test this hypothesis, however the basic assumption of homoscedasticity was not met indicating that other types of models should be considered for future research.

II. BACKGROUND

In the late 1980s, specialists in pain management began publishing studies that examined the use of opioid analgesics for treating chronic, non-malignant pain [8]. Despite acknowledging the potential for addiction, they argued that the risk of addiction was outweighed by the poor state of pain management in the United States [7], [9]. Just like today, many American patients were suffering from chronic pain from a host of different ailments, but the medical community was averse to pursuing treatments in which the side effects were potentially worse than the conditions being treated. In 1995, the president of the American Pain Society advocated in his annual address to the Society that a fifth vital sign[2], the indication of pain, be introduced as a measurement of patient well-being [11]. In response, regulatory guidelines were revised to encourage assessment of pain during doctors' visits [9]. Historically the uses of opioid analgesics had been limited to cancer patients or those suffering acute pain in a post-operation environment due to the addictive nature of opioids. During this same period of time, state regulators began to loosen the regulations governing when non-cancer patients could be prescribed opioid analgesics [12].

Around the same time, the semi-synthetic opioid, Oxycontin, was introduced by pharmaceutical manufacturer, Purdue Pharma. Oxycontin's claim to fame was that unlike any other opioid-based drug, its manufacturer had purportedly created a formula that was less addictive, and the potential for abuse of the drug was minimal [13]. The company's marketing department and sales representatives marketed it aggressively. A common practice was to personally visit physicians to promote the drug and to leave behind free samples for physicians to distribute to patients. Personal incentives for prescribing the drug were also provided to physicians

[2] Traditionally there have been four acknowledged vital signs: blood pressure, body temperature, pulse rate, and respiration rate. These signs are all externally identifiable and measurable by a physician [10].

by the company. Oxycontin sales grew rapidly. From 1996 to 2012; global annual sales grew from $48 million to $2.4 billion. Nationally, the number of prescriptions written for all opioid analgesic drugs rose drastically. Between 1997 and 2002, prescriptions for all oxycodone-based analgesics rose 402%. Concurrently, emergency department visits for oxycodone-related overdoses rose 346% [14]. Overall opioid prescriptions in the United States rose 300% between 1991 to 2009 [12].

Globally, opioid analgesic usage rates have increased from the 1990's into the first two decades of the 21st Century, but it is the United States that dominates the rest of the world in terms of consumption. For example, prior to 1990, the total global annual consumption of hydrocodone, a semi-synthetic opioid analgesic, was four tons. In 2009, annual global consumption of hydrocodone had increased to 39 tons and 99% of that consumption occurred in the United States. During the same time period, the annual consumption rate of oxycodone rose from three tons to 77 tons of which 81% was consumed in the United States. Two factors account for this increase in total weight consumed. First, between 2000 and 2009 the number of prescriptions per 100 persons issued rose from 61.9 to 83.7[3]. During the same time period, the average size of a prescription had also increased: by 69.4% for hydrocodone prescriptions and by 69.7% for oxycodone prescriptions [16].

Levy et al. [17] identified that a broad spectrum of medical specialists increased the rate of prescriptions written between 2007 and 2010. After 2010, prescriptions from most specialist categories of physicians leveled off as physicians began to pull back on the amount of opioids they were willing to prescribe. Some physicians, like those involved in rehabilitation, continued to

[3] These numbers appear to come from CDC sources, however they do not match the national prescription rate data that is currently available on the CDC website [15]. It should be noted, however, that this website excludes prescription rates previous to 2006.

prescribe opioids at increasing rates until 2012. This was a national trend, and regions within the country experienced different trends through time. This regional variation can be seen at multiple levels of geographic organization. In 2009, North Carolina and South Carolina opioid prescription rates were approximately the same: 89.3 and 95.8 prescriptions per 100 persons, respectively [18]. These rates put both states in the upper third of states' prescription rates. However, if one looks at county-level detail, there is much more nuance in the variation of prescription rates. Prescription rates in North Carolina ranged from 22.3 to 175.9 prescriptions per 100 persons for Currituck County and Columbus County, respectively. South Carolina's counties had a similar range: from 24.6 to 151 prescriptions per 100 persons for McCormick County and Oconee County, respectively [19].

None of this would necessarily be concerning if opioids were not so addictive to users, and if the potential for overdose-related deaths were not so great. However, both of these are true. The originator of Oxycontin, Purdue Pharma, had long claimed that its product was not addictive and that it was not until years after the drug went on the market that they became aware of abuse by addicted patients. However, in 2018 the New York Times published a confidential report from the U.S. Department of Justice. It indicated that in 2006 federal prosecutors had found that Purdue Pharma had become aware of substantial levels of abuse of the drug in 1996, the same year that the drug became commercially available. The drug, sold in tablet form, was being stolen or otherwise diverted from pharmacies and crushed for the purposes of snorting it or injecting it intravenously. In the late 1990's, the street value of the drug was more than 2000% higher than a legal prescription cost, approximately $35 per tablet. Despite becoming aware of the situation, the company hid this information and continued to manufacture and market the drug [13].

Once a person becomes addicted to opioids, their body physically needs the drug to feel good. A common result of this is the illegal diversion of prescription opioids like hydrocodone or oxycodone for non-medical uses. When prescription opioids are not available, research shows that opioid abusers seek out illegal opioids like heroin, which is often laced with fentanyl, an extremely powerful, fully synthetic opioid [9]. These drugs, which are classified as narcotics, can have an overwhelming effect on the human body's systems. Primarily, they impact areas of the brain that control both voluntary and involuntary breathing as well as the body's ability to sense a build-up of carbon dioxide in the bloodstream. The gag reflex is also suppressed, and irregular heart rates can be triggered. All of these effects accumulate into overwhelming a person, requiring them to be hospitalized due to opioid poisoning, or worse, causing their death [20].

Researchers have found evidence that increasing a population's exposure to opioids for legitimate medical uses also increases the amount of opioids diverted for non-medical abuse purposes and the consequences of that abuse. Powell et al. [21] found that a 10% increase in the distribution of opioids for medical uses leads to a subsequent 7.4% increase in opioid-related overdose deaths of people who had not been prescribed opioids. Ghertner [22] found a positive association between increasing opioid availability and hospitalizations due to opioid poisoning. Across 32 states, for every 1 morphine kilogram equivalents (MKE) of opioids sold per 10,000 persons there was a 9% increase in hospitalizations. Urban counties experienced a smaller rate than rural counties, but the difference was not statistically significant.

Mortality in North Carolina associated with prescription opioid poisonings began to rise in 1999 and continued to do so until 2008. Since that time, the number of deaths in the state have maintained a flat trend. In 2017, there were 659 overdose deaths of this nature – a rate of 6.5 deaths per 100,000 persons [23]. 2017 was also the last year that prescription OP-related mortality was greater than synthetic opioid-related deaths. South Carolina has seen similar trends with prescription OP-related mortality, although its trend flattened in 2006 and remained flat until increasing dramatically again in 2014. In 2017, 345 persons died of prescription opioid-related overdoses, a rate of 7.1 deaths per 100,000 persons. Like North Carolina, South Carolina prescription OP-related deaths were greater than those from synthetic opioids until 2017 [24]. Despite widespread acknowledgement of the addictive nature of opioids and the evidence from associated mortality rates, physicians continue to prescribe opioid analgesics at high rates (Fig. 1). Although both North Carolina and South Carolina prescription rates have fallen in more recent years (61.5 and 69.2 per 100 persons, respectively, in 2018), they are still much higher than the national average [25].

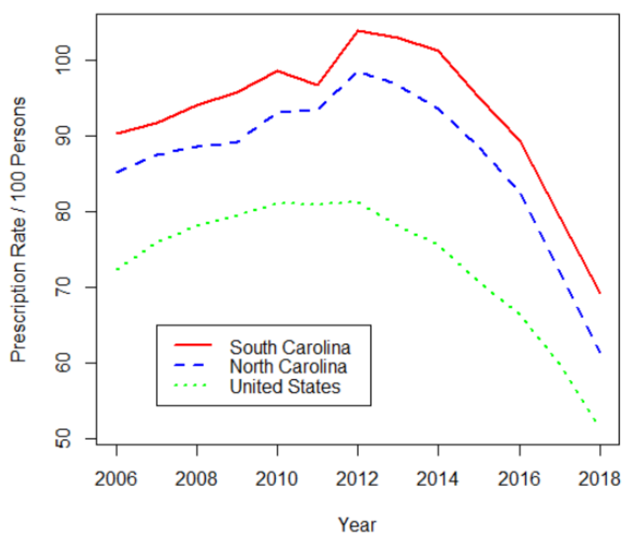

Fig. 1. Comparison of Opioid Analgesic Prescription Rates, 2006-2018. National prescription rate in 2018 was 51.4 per 100 persons.

Opioid analgesics are classified as controlled substances by the U.S. DEA [26]. As controlled substances, the DEA mandates that all shipments of opioids be documented. This documentation is preserved in the DEA's ARCOS database. Within this database are all transactional shipments of controlled substances among wholesale and retail distributors. The data within has long been considered proprietary by both drug manufacturers and the DEA. Pharmaceutical companies have argued in court that revealing this data would harm competitive advantages; the DEA has argued that revealing this data to the public would hurt ongoing investigations. Past settlements with pharmaceutical companies have included stipulations that transactional data from ARCOS not be published [2]. The DEA regularly publishes data from ARCOS; however, the data has been aggregated to the 3-digit ZIP code prior to publication. Aggregation of data removes much of the original detail. Not only are finer geographic details like the addresses of buyers and sellers removed but also the details regarding the timing of sales, and the volume and dosage strength of the drugs shipped [27]. However, for the general public and researchers it is this transactional level of detail that ARCOS contains that is vital for understanding the who, what, when, and where of all opioid shipments in the United States. It has the potential to reveal associations between opioid exposure and opioid-related death at a much more granular level of detail than has ever been available previously.

Over the course of one year, *The Washington Post* and HD Media, publisher of the *Charleston Gazette-Mail* in West Virginia, fought in court to have transactional-level data from the ARCOS database published. This lawsuit originated when the DEA refused to comply with a Freedom of Information Act request that was filed by *The Washington Post* in 2016. In July 2019, a U.S. District Judge, Dan Polster, allowed for data from the ARCOS database for the years 2006 through 2012 to be made publicly available [2]. Two more years of ARCOS data, 2013 and 2014, were subsequently released under a separate ruling in January 2020 [28].

Access to ARCOS transactional-level data opens new research opportunities for those interested in understanding the distribution of medical use opioid analgesics during the years data is now available [29]. Previously, aggregated ARCOS data could only be associated with 3-digit

ZIP codes which frequently encompass multiple counties within a single state. Any researcher interested in geodemographic variation associated with exposure to opioids has much of the spatial variability eliminated when working at the level of the 3-digit ZIP code, as demonstrated by Modarai et al. [3]. Now that transactional-level data is available, researchers can analyze data at a much more detailed level: spatially, temporally, and by buyer or seller or type of drug. In this study, I perform an exploratory spatial data analysis of the newly available ARCOS transactional data in combination with data from the U.S. Census Bureau's American Community Survey (ACS). I focus on North Carolina and South Carolina. I use the Enhanced 2-Step Floating Catchment Area (E2SFCA) method proposed by Luo and Qi [30] to assign an opioid accessibility score for all census tracts in the Carolinas. I use a k-medoids clustering method to explore demographic and socioeconomic differences, and I create an OLS regression model to identify explanatory variables that can explain the variation of opioid accessibility scores. Based on studies of demographic and socioeconomic variation in the distribution of opioid analgesics, including elderly populations in Powell et al. [21], African-American and public insurance recipient populations in Basak et al. [31], median age in Piper et al. [32], non-Hispanic Whites, rural residence and public insurance recipients in Guy et al. [33], and economically disadvantaged and non-Hispanic White populations in Anderson et al. [34], I hypothesize that there was a non-random distribution of opioid analgesics in relation to the demographic and socioeconomic characteristics of the populations in the census tracts of the Carolinas.

III. Literature Review

Researchers have access to different sources of data that can be used to analyze the distribution of medical-use opioids. Aggregated data from the ARCOS database has been one of those sources. However, shipments of opioid drugs to pharmacies is not equivalent to actual prescriptions prescribed by doctors and fulfilled at pharmacies. Individual states maintain Prescription Drug Monitoring Programs (PDMP) databases that record and preserve details of opioid prescriptions. At least one study [32] has compared prescriptions recorded in a state's PDMP to aggregated data available from the ARCOS database. The study found high correlation – over 98% - between the databases although the authors noted that data from the PDMP only represented 78% of the total opioid analgesic volume distributed to the state. While data from the ARCOS database is not equivalent to the individual actions of prescribing physicians, it represents an important step in the supply chain from manufacturer to consumers. In this study, ARCOS data serves as a proxy for a population's access to opioids.

A. Relationship between Opioid Availability and Related Mortality

There is a significant body of research demonstrating whether there is a relationship between increasing opioid analgesic sales and overdose-related mortality. Nationally, Paulozzi et al. [7] found similar positive growth of opioid analgesics sales (76%) and opioid-related poisonings (95%) between 1999 and 2002. Along with [7], other studies have also identified similar trends. A study of 32 states determined that a 1 MKE increase in opioid analgesic sales per 10,000 persons predicted a 9% increase in opioid poisoning hospitalizations [22]. Powell et al. [21] compared states with high and low Medicare Part D rates of usage. They concluded that a 10% increase in the distribution of opioid analgesics increased the mortality rate of a state's non-Medicare eligible population by 7.4%. Modarai et al. [3] found spatial relationships between increasing opioid analgesic distribution and increasing OP-related mortality, which will be expanded upon later in this section.

The increased usage of opioid analgesics and its addictive nature has resulted in large increases in overdose deaths. In a national study, the Centers for Disease Control and Prevention reported that state-level variation in prescription opioid-related mortality could not be explained by variation in demographic characteristics and could only be explained by variation among states in opioid prescribing practices [35]. In 2010, total drug overdose deaths had increased 11 years in a row. In that year, 75% of all prescription drug-related overdose deaths involved opioids. Deaths related to prescription opioids rose 400% between 1999 and 2010. In their survey of prescription opioid-related mortality, King et al. [1] identified determinants of mortality including demographic characteristics, user behavior, and prescribing doctor behavior that included increasing numbers of prescriptions and volume of dosages. Agnoli et al. [5] hypothesized that short-term mortality of patients is higher among those exposed to opioid analgesics. The authors took a nationally representative sample of patients. When adjusted for sociodemographic variables, those patients who had been exposed to at least one prescription of opioid analgesics had a statistically significant higher odds ratio of short-term mortality. Increasing rates of OP-related mortality associated with increasing prescription rates of opioid analgesics have been found in New York State [36], in states considered part of the Deep South region of the country [6], and in a national cohort study [37].

B. Opioid Exposure, Mortality, and Demographic Variation

Men are consistently reported to die from OP-related causes at higher rates than women [6], [35], [38]–[40], although Brandenburg [6] identified a flat trend over time for male mortality and a rising trend for female mortality in the Deep South from 1999 to 2015. These trends corresponded with a strong correlation between males and flat hydrocodone sales and between females and spiking oxycodone sales. However, many studies have found that women tend to be prescribed opioid analgesics for chronic pain at a higher rate than men [5], [7], [41], [42]. Those women tended to live in Southern or Midwestern states and to be older, publicly insured, and with lower incomes [5]. This corresponds with research that shows increased rates of OP-related hospitalizations or mortality among non-Hispanic Whites and persons with lower educational attainment, lower income, and eligible for public insurance [37], [39]–[41].

Much of the research done to understand the role of opioid analgesic prescriptions in the opioid crisis has come in the form of assessing what factors have a positive or negative relationship to OP-related hospitalization or mortality. From the larger body of literature several relevant studies are discussed below. The studies cited have contributed to the demographic and socioeconomic factors used in this study. The present study examines a list of variables that are reflective of the demographic determinants identified by a survey of studies conducted by King et al. [1]. Consistently found within the literature studying prescription opioid-related mortality is the inclusion and the examination of gender, age, race and ethnicity, urban vs. rural residence, and socioeconomic status. For example, McDonald et al. [43] found positive relationships with size of county population, urban residence, and proportions of Non-Hispanic Whites, African-Americans, poverty, and low levels of health insurance to rates of opioid prescribing. Their study also found that the factor contributing the most explanation to the variation in those rates was the abundance of prescribing physicians in a county. This study does not have access to physician-to-population data, so it cannot be included in this study's models. The variables that McDonald et al. included in their study only accounted for a third of all variation in prescribing rates.

Brandenburg [6] used a time series analysis to find a temporal relationship between opioid sales in Deep South states, including South Carolina, and medical-cause mortality (which is death due to a medically defined disease). The author identified a temporal association between the sales of hydrocodone and oxycodone and medically-caused mortality in Non-Hispanic whites between ages 45 and 54. In particular, he identified spikes in opioid sales in 2007 and 2013 that were followed by a spike in medically-caused mortality approximately one year later. The author postulated that the first spike was related to the implementation of Medicare Part D which greatly expanded access to prescription drugs, including opioid analgesics which has been found to be over-prescribed to public insurance beneficiaries. While the author acknowledged that other factors such as prevalence of smoking and obesity in the Deep South played a role in mortality, the lack of wide fluctuations year to year in these other factors pointed to the fluctuation in year to year opioid sales as a predictor of shifting medical-cause mortality. This was the first study to examine opioid exposure and medical-cause mortality and has not yet been replicated, but it points to yet another detrimental repercussion of the use of opioids for pain management.

Modarai et al. [3] were interested in understanding whether demographic variation of risk at the state level would continue to be identifiable at a sub-state level. They analyzed opioid sales data for North Carolina from the DEA’s ARCOS database that had been aggregated to the 3-digit ZIP code level. Sales of each type of opioid were calculated as MKE per 10,000 persons. This data was analyzed using correlation analysis and linear regression in combination with emergency department visits and unintentional overdose deaths. Based on this analysis, an 839% increase in the sale of hydrocodone and a 224% increase in the sale of oxycodone was identified state-wide for the years between 1997 and 2010. At the sub-state level, spatial relationships were identified between areas that had higher rates of opioid sales and increased rates of overdose-related deaths. This association was most pronounced in the southern and western portions of the state in the years 2008 through 2010. Additionally, the 3-digit ZIP codes that could be considered rural tended to have both higher sales and higher overdose-related mortality. The authors also used several spatial clustering tools in ArcGIS including the Local Moran I and the Local Indicators of Spatial Association (LISA) tests, however none of those tools identified statistically significant clusters. When considering the limitations of their study, the authors argued that aggregated ARCOS data smoothed away too much local variability, and they recommended that analysis at the census tract level could produce significant results.

Over the years, researchers have worked to improve the granularity of their understanding of aspects of the opioid crisis and how it varies spatially. The smallest geographic units have often been either 3-digit ZIP codes or counties. Marotta et al. [36] looked at geographic variability of opioid overdoses in New York State. They were able to identify outlier counties on both the low and high end of the spectrum using the LISA test. McDonald et al. [43] identified wide variation in the volume of opioids prescribed among counties nationally. Oxycodone, in particular, had wide variability; the top quarter of counties prescribed seven to 10 times more morphine equivalents per capita than the average county. Rossen et al. [44] used Global Moran's I to test for spatial autocorrelation of drug poisoning deaths at the county level between 2007-2009. More than three out of four of those deaths were associated with opioid analgesics. Their results suggested that drug poisoning mortality clustered among counties more frequently than by random chance. They were also able to identify hot spots and cold spots of drug poisoning mortality. Their findings corresponded with Modarai et al. [3]; the western third of North Carolina and the northwestern corner of South Carolina were part of a larger hotspot of mortality that included a large swath of Appalachia. Rossen et al. [44] also identified a single county hot spot of drug poisoning mortality in the southeast corner of North Carolina near Wilmington. It was smaller in geographic scope than the hot spot found by Modarai et al. [3], although this could be related to a difference in geographic units; Modarai et al. used 3-digit ZIP codes and Rossen et al. used counties.

C. Healthcare Accessibility

There is an extensive body of research that assesses whether a population has adequate access to a particular type of healthcare service. Such services can range from everyday healthcare needs like visiting a primary care physician or obtaining prescriptions from a pharmacy to obtaining specialty care like mammography screenings or cancer treatment [30], [45]–[48]. Part of the assessment of accessibility is measuring physical distance between a population in need and the location of the healthcare service. How physical distance as a proxy for accessibility is measured has evolved over time. Accessibility was once measured ‘as the crow flies’: a Euclidean distance radius from a service location. Researchers set a distance limit and any population that was beyond that limit might be considered to be living in a healthcare desert [49]. Only under ideal circumstances, like in a city with neatly gridded blocks, is this a reasonable measure of distance traveled, and even then, use of Manhattan distance over Euclidean would be more apropos. Technology has enabled researchers to develop more nuanced methods for measuring distances between a service and the population it serves. A Geographic Information System (GIS) provides the functionality to measure distance as a person typically experiences it: traversed along a road network. Many studies have utilized distance traveled as a measurement of accessibility in which a required greater distance traveled to obtain a service is equated with less access to that service [47], [50]–[52].

While GIS has played a role in measuring distance traveled along road networks, it also has played a role in measuring accessibility not via the distance traveled but the time required to travel to a destination. Details of a road network such as speed limits or average traffic conditions can provide a more nuanced understanding of how much time is required for a person to reach his or her destination. In their review of national travel survey data, Probst et al. [53] identified the average distance and time persons traveled to reach healthcare services. Unsurprisingly, there is a considerable difference in distance traveled between urban residents (8.3 miles) and rural residents (17.5 miles) without a substantial difference in time traveled (23.5 and 27.2 minutes, respectively). Travel time is a better predictor of access to healthcare services. There is a growing body of healthcare accessibility literature that advocates for and uses travel time. For example, Schuurman et al. [54] studied hospital accessibility in the rural regions of British Columbia, Canada. They demonstrated that in a rural mountainous environment Euclidean distance does not have much bearing on distances people travel to access services. They demonstrated that a road network was necessary to measure the distance traveled but that time traveled as an impedance should be a factor in measuring accessibility. Other studies that have used travel time to measure accessibility include [30], [48], [55]–[57].

D. Defining Pharmacy Catchment Areas

In order to identify a pharmacy's presumed catchment area and resulting customer base, a catchment area needs to be defined. Catchment areas are geographic areas in which the resident population is able to access and utilize the service of a business or institution. Oftentimes, catchment areas are exclusive and do not overlap, such as school districts. However, in situations where businesses compete for customers, catchment areas can overlap [54]. This study used overlapping catchment areas. Catchment areas can be defined by the distance a person will travel to access a pharmacy or the time a person is willing to travel to access a pharmacy. A primary consideration in choosing distance travelled or travel time is whether a population lives in a rural or an urban area. The research regarding whether there is a difference is mixed. Probst et al. [53]

used the 2001 US National Household Travel Survey to identify medical and dental accessibility differences based on residence and race. Across all races, they found that rural residents traveled 17.5 miles (27.2 minutes) on average compared to urban residents who traveled 8.3 miles (23.5 minutes). Syed et al. [46] surveyed 61 previous studies related to healthcare accessibility and race and residence. Their results were mixed when reviewing studies that focused on differences between rural and urban residents' healthcare accessibility and utilization. They also saw mixed results in their survey of studies that examined the effect of travel time and travel distance on healthcare accessibility. McGrail & Humphreys [58] studied differences in accessibility between rural and urban regions in Victoria, Australia. They concluded that accessibility could be more accurately measured when differences in population density between rural and urban areas were accounted for during the parameterization of the catchment areas: namely that the size of catchment areas needed to be different for rural and urban areas and that a decay function be implemented to account for decreasing accessibility with increasing travel. A discussion on the nuances of distinguishing between urban and rural regions follows in Subsection E: Defining Rural and Urban Catchment Areas.

Researchers have worked progressively over many years to identify a model that can accurately represent the accessibility of a supply to a local populace. A popular means to achieve this accuracy is the family of FCA methods. Generally speaking, these methods measure the ratio of supply to the demand of the local populace at a given location. The method chosen to define the catchment area of a location, whether it is a fixed radius, a defined time travelled along a road network, or any other method, is used to calculate accessibility for each source of demand. Thus the catchment area 'floats' from one location to another [59]. Catchment areas are premised on the supply and demand gravity model introduced by Weibull [60]. The earlier versions of FCAs treated access dichotomously [59]. Persons within a catchment area were considered to have access, and persons without a catchment area were considered to have no access. This approach does not consider the effect of travel cost; a person ¼ mile away from a supply and another person ten miles away are considered to have equal access which is not a reasonable assumption to make. Early FCA methods also tended to rely on distance travelled as the cost incurred to access a supply [47], [49], [61], [62], however travel time has been identified as a more accurate cost measurement [30], [54], [59], [63]. Luo & Qi [30] proposed a new FCA method that addressed the problem of dichotomous accessibility: the E2SFCA method. They considered hospitals as their supply locations and the populations of census tracts as their units of demand. In this method two sets of catchment areas are created. The first set includes the catchment areas for locations that have a desired supply. Instead of a single travel time applied to the entire catchment area, multiple zones of travel time are used. For instance, catchment areas might be broken into four zones: 0-5, 5-10, 10-15, and 15-20 minutes. Each of those zones have a weight applied to those zones that represent decreasing access to a supply as the cost of travel time increases. All census tract centroids that fall within a catchment area are considered part of that supply's catchment area; tract centroids are grouped based on each zone of a catchment area. The populace of the tracts in each zone are summed and multiplied by the weight of that zone. The weighted populations of all zones are summed again, and the ratio of available supply to residents of the catchment area is calculated. The second set of catchment areas represent the length of time the populace of each census tract is willing to travel to reach a supply. Catchment areas in this set have the same zones of travel time and the same weights. Any supply location that falls within a tract's catchment area is considered part of the supply for that tract's populace. Supply locations are grouped based on each zone of a catchment area, and the available supply for that location is multiplied by the zone's weight. The weighted supplies of all zones are summed again, and the ratio of supply available per capita in the census tract is calculated. This summed ratio represents the supply accessible for each person in that tract. Higher ratios represent a greater supply per capita for that tract's populace [30].

The weights that are assigned to the zones of a catchment area are based on a distance decay function [30]. The values of the function used represent increasing impedance that time of travel creates when attempting to reach a supply. Various functions have been suggested, including the inverse power, the exponential, and the Gaussian functions, however Wang [64] demonstrated that the Gaussian function is the best function to use when studying access to healthcare services. Other functions that have been proposed have a very steep initial decline and so produce too high of an impedance too near to a supply location. The initial gradual decline of the Gaussian function as well as its accelerating decline makes it a better suited function to generate distance decay weights for healthcare services.

Distance decay functions have long been a tool used by city planners and transportation experts to model travel behavior by persons in a variety of settings. Distance decay is a flexible enough concept that it can be applied to any mode of travel: public transit, biking, private automobile, or walking and can be applied to any reason for traveling: work, retail shopping, etc. Distance decay functions can be based on an unconstrained gravity model as represented in Equation (1) [67, eq. (1)]:

$$T_{ij} = k{v_i}^{\mu}{w_j}^{\alpha}{c_{ij}}^{-\beta} \tag{1}$$

in which T_{ij} is the number of trips between zone i and zone j, the variables v and w describe the level of attractiveness that an origin or a destination has, and the variable c is the cost that is incurred to travel between zones i and j. The term β is extremely important because it represents the amount of impedance that a traveler incurs to travel between two locations. β can also be viewed as the willingness of someone to travel to a given location. While the gravity model represented in Equation (1) uses a power function to represent this impedance, other functions are equally legitimate to use, including a Gaussian function [65].

Distance decay functions have also been applied to measuring accessibility, or the ability of a person to reach and utilize the services at a given location. Modeling accessibility has taken three forms: gravity-based models, behavioral models, and cumulative-opportunity models. When modeling accessibility of a location based on the gravity model, one is measuring the opportunity that a location provides to a traveler weighted against the cost, or impedance, to travel to that location. Equations for gravity-based accessibility models include an impedance function such as $f_{ij} = \exp(-bC_{ij})$ in which b is a non-negative parameter and C is a generalized cost variable that traditionally has represented distance traveled, but in more recent literature has also come to represent travel time [65].

While adding weights to catchment areas to represent distance decay can create a more accurate model of supply and demand over distance, implementing this solution creates an additional challenge because those weights ideally should be based on actual travel behavior. Unless previous surveys for the study area exist or researchers produce data from their own surveys, weights need to be borrowed from other studies with similar conditions. Ikram [45] found that the average travel time to pharmacies in Baton Rouge was 8.11 minutes; 86% of the population lived within 10 minutes, and 96% of the population lived within 15 minutes. Because these researchers did not have survey data, they produced multiple-sized catchment areas until the entire population was covered. Wang and Ramroop [66] studied pharmacy accessibility in the Greater Toronto Area. They chose catchment area zones of 5-, 10-, and 15-minute travel times with corresponding weights of 1.00, 0.42, and 0.09, respectively. Those weights were originally used by Luo and Qi [30] in which two sets of weights were applied to the same area of study. The weights above represented a faster rate of decay than the other set of weights: 1.00, 0.68, and 0.22. Luo and Qi recommended that faster decays should be used for commonly available healthcare services like pharmacies. McGrail [57] tested catchment areas of variable size for urban and rural areas versus same-sized catchment areas to assess the best method for calculating healthcare accessibility in Victoria, Australia. They also used two sets of weights (1.00, 0.60, 0.25, 0.05 and 1.00, 0.80, 0.55, 0.15) to test stepwise decay functions versus continuous decay functions. Their results found only minor differences between stepwise and continuous decay functions. Fast and slow sets of weights had an impact on the accessibility scores in both urban and rural areas, but they argued that it is most important that variable catchment area sizes be used to most accurately reflect accessibility scores especially on the outskirts of metro areas and the nearby rural areas. Studies in the United States have made use of McGrail's weight selections [48], [56] and variable catchment size areas [56] in their own assessment of healthcare accessibility.

E. Defining Rural and Urban Regions

Identifying an appropriate travel distance or travel time in the United States based on rural and urban residences is complicated by the lack of a consistent definition for what it means to be rural. The U.S. Census Bureau does not define rural. However, it has two definitions of what it means to be urban: an Urbanized Area has 50,000 or more people, and an Urbanized Cluster has between 2,500 and 49,999 people. Neither definition follows city or county boundaries. Any area with less people is considered rural [67]. The Office of Management and Budget identifies counties as either metropolitan, micropolitan, or rural. Rural counties have less than 10,000 people [68].

When neither definition works well, researchers identify an alternate definition that meets the needs of their study.

For this study, the geographic focus was North Carolina and South Carolina, which has a mix of both urban and rural regions. Different travel time parameters needed to be applied to the different regions within this geographic area, which meant a method was required to identify whether an area is urban or rural. As noted above, even within the federal government there is no consistent definition of what qualifies as rural in the United States, and different researchers and organizations use different methods based on their goals. Examining a single locale can reveal the complexity of assigning a rural designation to a location. For example, consider Aynor, South Carolina, which is located in the northwestern corner of Horry County. Because the town of Aynor is not located in an Urbanized Area it is given rural status by the Rural Health Clinics Program. However, according to the Federal Office of Rural Health Policy grant program Aynor is not in a designated rural census tract or county so it is not eligible for that program [69]. While not being within an Urbanized Area, it is associated with one by the U.S. Department of Agriculture's (USDA) Rural-Urban Commuting Area (RUCA) codes because at least 30% of Aynor's populace commutes to an Urbanized Area [70].

As a result of this complexity and lack of clarity, researchers choose a rural definition based on the focus of their study. This study followed the example of Zahnd et al. [48]. They analyzed travel times and access to mammography screening services in the Lower Mississippi Delta Region. They chose to use the USDA's RUCA codes to distinguish urban and rural locations in their study area.

F. Review of Studies Using ARCOS Data

The U.S. DEA began collecting records of shipments of controlled substances after passage by Congress of The Controlled Substances Act of 1970 (§ 827) [71]. According to this law all manufacturers and distributors of controlled substances must notify the government of every shipment of a controlled substance they send to hospitals, retail pharmacies, and medical practitioners [21]. Oxycodone and hydrocodone have both been declared Schedule II substances which are defined as having a high possibility for abuse leading to physical or psychological dependence [26]. Records of shipments from distributors and manufacturers are collected in the DEA's ARCOS database. The DEA publishes quarterly reports of all shipments reported to the database, however this data is aggregated to the 3-digit ZIP code [4].

Despite the limitation of aggregated data, these reports have been a useful source of data for researchers studying the effects of changing availability of opioids on the American population [3], [7], [16], [21], [32], [72], [73], [74]. Modarai et al. [3], however, pointed to the aggregation of ARCOS data as limiting their ability to find statistical significance in some of their work. They suggested future work be done to interpolate ARCOS data at the census tract level to improve the spatial variability that is lost due to data aggregation.

Address-level data has only been available since July 2019; as of February 2020, a comprehensive literature review did not identify any published work that utilized the address-level transactional data from the ARCOS database that *The Washington Post* published.

IV. Methodology

A. Data Sources

Demographic and socioeconomic variables came from the U.S. Census Bureau's ACS 2012 5-year average survey [75]. This study included the demographic and socioeconomic variables related to sex, age, racial identity, ethnic identity, income, housing and related housing expenses, health insurance, disability and veteran status, and poverty. The U.S. Census Bureau was also the source of census tract polygonal data that was used for work in a GIS software application. Data from the U.S. DEA's ARCOS database for North Carolina and South Carolina was downloaded from *The Washington Post*'s published database [76].

1) ACS Data: The Census Bureau conducts the ACS nationally every year. Approximately 3.5 million households are sampled throughout the entire year. The feedback from respondents is collated into a single year average that represents population, demographics, economic, social, housing, and other factors of life in the United States. Additionally, each year a five-year rolling average of the past five years of survey results is also published. The 2012 ACS included the results from the 2008-2012 surveys. Because the ACS takes a sample of the population each year unlike the full national census that occurs decennially, estimates for variables of the actual population have a margin of error. That margin of error is greater for the one-year average than for the five-year average because it accounts for a smaller sample size of the total population. The five-year average surveys are more statistically reliant, especially for smaller geographic units and smaller populations, but that comes with the downside that the data is not current to the year it is published because survey results from years of collection are averaged together [77].

In this study, a five-year average study was chosen for its reduced margin of error and because the Census Bureau publishes data for a small geographic unit – the census tract – and small populations only for the five-year average surveys. The Census Bureau defines small populations as less than 65,000 persons [77]. This study used the 2012 survey, because it was the first year that questions from the ACS regarding health insurance coverage and disability status were published by the Census Bureau. Due to the association found between exposure to opioid analgesics and health insurance coverage and disability status, it was imperative that this study include data from the 2012 5-year average [41]. Table I lists all demographic and socioeconomic variables used in this study, however not all variables were used for all models created in this study and some variables were only used for descriptive purposes as indicated in later sections. In addition, some variables are derived from other ACS variables; calculation details are also described in Table I [78].

Table I
Demographic and Socioeconomic Variables from American Community Survey. ACS Table indicates the DP table source of the variable. ACS Variable Name or Field Calculation is the combination of variables from the source DP table that each variable in this study is derived from.

Variable Description	ACS Variable Name or Field Calculation	ACS Table
Educational Attainment: Percent of population with High School degree or less	HC03_VC85 + HC03_VC86 + HC03_VC87	DP02
Disability: Percent of total population with a disability	HC03_VC104	DP02
Disability: Percent of under 18 population with a disability	HC03_VC107	DP02
Disability: Percent of 18 to 64 population with a disability	HC03_VC110	DP02
Disability: Percent of over 65 population with a disability	HC03_VC113	DP02
Median household income	HC01_VC85	DP03
Per capita income	HC01_VC115	DP03
Health Insurance: Percent of population with public health insurance	HC03_VC130	DP03
Health Insurance: Percent of population with no health insurance (public or private)	HC03_VC131	DP03
Poverty: Percent of population whose income in last 12 months was below the poverty level	HC03_VC166	DP03
Housing: Percent of housing units that are renter-occupied	HC03_VC64	DP04
Housing: Percent of housing units with no vehicle available	HC03_VC82	DP04
Housing: Percent of owner-occupied housing units with a mortgage	HC03_VC130	DP04
Housing: Percent of housing units with a mortgage in which housing costs are 35% or more of income	HC03_VC160	DP04
Housing: Percent of housing units without a mortgage in which housing costs are 35% or more of income	HC03_VC171	DP04
Housing: Percent of rented housing units with rental costs 35% or more of income	HC03_VC197	DP04
Median age in years	HC01_VC21	DP05
Sex & Age: Percent of population 65 and over that is male	HC03_VC33	DP05
Sex & Age: Percent of population 18 to 64 that is male	(HC01_VC29 – HC01_VC33)/(HC01_VC28 – HC01_VC32)	DP05
Race & Ethnicity: Percent population Hispanic or Latino – all races	HC03_VC82	DP05
Race & Ethnicity: Percent population not Hispanic or Latino – white only	HC03_VC88	DP05
Race & Ethnicity: Percent population not Hispanic or Latino – black only	HCO3_VC89	DP05
Race & Ethnicity: Percent population not Hispanic or Latino – all other races	((HC01_VC03 – HC01_VC82) – HC01_VC88 – HC01_VC89) / HC01_VC03	DP05

The variables in this study can be categorized by their unit of measurement: individuals, households, or housing units. Variables regarding disability status are based on the civilian noninstitutionalized population in which respondents indicated some combination of hearing, ambulatory, self-care, cognitive, independent living, or visual difficulty. This study combined educational attainment variables pertaining to all persons ages 18 and older that obtained no more than a high school diploma, including passing the General Educational Development (GED) test, but did not attend college for any length of time. Per capita income is the aggregate income of an entire census tract divided by its total population. Variables related to health insurance pertain only to comprehensive health insurance. Persons with public health insurance coverage have at least one of the following coverage plans, Medicaid, Medicare, VA Health Care, Children's Health Insurance Program, or a state health care plan. Persons with no health insurance have neither public nor private coverage. Private health insurance coverage includes plans from an employer, a private insurance company, or some form of military plan such as TRICARE. The Census Bureau began asking respondents about health insurance coverage in 2008; 2012 was the first year that a five-year average survey included the results from this question.

This study considers poverty at the individual and family level and not at the household level. Determination of poverty status is based on the Census Bureau's monetary thresholds that vary based on family size and age of householders. Poverty status is also determined based on total income from the previous 12 months from which respondents took the survey. Median age is the calculated middle age of all persons residing in a census tract. This study considered sex and age in tandem. Variables used from the ACS include males between the ages of 18 and 64 and males age 65 and above. Percentage of a census tract population that is in one of those age brackets and is female is derived by subtracting the relevant value from one (1). Race and Hispanic ethnicity were considered in tandem in this study. Any persons self-identified as having Hispanic, Latino, or Spanish origins, of any race, are considered 'Hispanic' in this study. Persons who self-identified as non-Hispanic (NH) were grouped into three distinct racial categories: White, Black or African American, and Other. The Other category includes American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, and any persons who self-identify as two or more races. (All racial categories follow federal guidelines from the Office of Management and Budget.) The four variables Hispanic, NH White, NH Black or African American, and NH Other represent the total population of a census tract with no overlapping of persons [78]. Population density was calculated by dividing the total population of a census tract as reported in the ACS by the square miles of the tract. This calculation was performed in ArcGIS Pro; square miles values were derived in ArcGIS Pro.

Renter-occupied housing units are all housing units that are occupied by anyone other than the unit's owner regardless of whether rent is paid. The percentage of owner-occupied housing units can be inferred by calculating the inverse. Housing units with a mortgage only include owner-occupied housing units in which the said property is security for payment of a debt. Variables related to housing costs and income are used in this study to gauge economic stress. For renter-occupied housing units, this study looked only at units where the gross rent paid is at least 35% of income for the past 12 months. Gross rent is defined as the contractual rent plus utilities (fuel, water, electricity). For owner-occupied housing units, this study looked only at units where the housing costs paid is at least 35% of income for the past 12 months. Housing costs are any costs that must be paid by the owner to maintain ownership and occupancy of the unit: mortgage, taxes, insurance, and utilities. Housing units with and without a mortgage are distinguished as two separate variables. The variable 'no vehicle available' captures the percentage of housing units in a census tract in which occupants of the unit do not have access to any vehicle for personal use. Median household income is the middle income of all households in a census tract including households with no income [78].

2) RUCA Codes: Census tracts were categorized as either rural or urban based on the USDA Economic Research Service's (ERC) RUCA codes [79].

3) ARCOS data: Data from the ARCOS dataset for North Carolina and South Carolina was downloaded from *The Washington Post*'s published database [76]. Each record represents a single shipment of one opioid analgesic drug from any given reporting entity to any given purchasing entity. Reporting entities are either pharmaceutical manufacturers or wholesale distributors. Purchasing entities are other distributors, storefront pharmacies, both chain and retail, or individual practitioners. If the same reporting entity sent multiple shipments to the same purchasing entity on the same day, then there is a separate record for each of those shipments in the data. *The Washington Post* made alterations to the data they received from the DEA prior to publishing it. While the ARCOS database contains shipment records of all controlled substances, *The Washington Post*'s database is limited to only shipments of opioid analgesic in tablet form containing either oxycodone or hydrocodone. Shipments for other opioids like morphine were removed because they represented a small percentage of total opioids shipped and because the data showed that those other opioids did not have dramatic increases in distributed volumes during the years represented in their database: 2007 – 2012 [29].[4] While measuring opioid sales is not the same as measuring the number of prescriptions written, there is a fairly approximate equivalency. Piper et al. [32] found a statistically significant positive correlation in Maine between the volume

[4] Note that *The Washington Post* published an additional two years of ARCOS data, 2013 and 2014, on January 17, 2020. Because this year's data became available after this study began, those years were not considered as part of the methodology [76].

of opioids reported in the ARCOS database and the number of prescriptions written as reported by the state's prescription drug monitoring program (PDMP). Notably, the PDMP reflected only 78% of the total opioid volume distributed to the state as reported in ARCOS. The authors speculated that the difference was due to opioids prescribed at Veterans Affairs clinics which are not represented in the state's PDMP database but are represented in the ARCOS database as shipments to the clinics.

B. ARCOS Data Preparation

Opioid buyer addresses were compiled from an assortment of ARCOS data fields. Opioid buyer names and addresses were standardized so that they were consistently represented across all data records. Buyer addresses were geocoded and reviewed for quality. All addresses that were geocoded with a confidence score less than 8 on a scale of 1 (low) through 10 (high) were manually reviewed and corrected as needed. The volume of opioids shipped to a single geocoded location was standardized by calculating the morphine milligram equivalents (MME) present in each shipment. This is common practice among healthcare researchers [21], [32], [74]. The calculation was based on Equation (2):

$$MME_s = T_s D_s C_o \tag{2}$$

where MME_s is the total MME in a shipment; T is the number of opioid analgesic tablets in shipment, s; D is the dosage strength, in milligrams (mg) of the tablets in the shipment; and C is the standard conversion factor for the opioid analgesic, o, present in the shipment. The hydrocodone conversion factor is 1.0; the oxycodone conversion factor is 1.5 [3].

Analysis of ARCOS data was confined to records of shipments to pharmacies in 2009. There are two types of pharmacies categorized in the data: chain and retail. Chain pharmacies are nationally recognized chains with dozens if not hundreds of retail outlets like Wal-Mart, CVS, and Kerr Drugs. Retail pharmacies are locally owned pharmacies that do not have a national chain affiliation. Because so many retail pharmacy locations changed ownership throughout the year 2009, either with another retail pharmacy or with a chain pharmacy, the name of pharmacies and the type of pharmacy was removed from the analysis. Records of shipments to practitioners were also eliminated from the analysis, because they represented a small minority of records in the data. This is true both for the total number of shipments (0.39%) and the total shipment MME (0.26%). Despite these small percentages, practitioner locations also represented over half of all shipment destinations (52.7%). Shipments to practitioners were also eliminated because any given practitioner's potential customer base is only a subset of a region's population whereas the potential customer base for a pharmacy can be considered a region's entire population.

C. Defining Rural and Urban Regions

The USDA ERC's RUCA codes [79] were developed to classify commuting flow among differently sized population centers, however they have proven to be a useful classifier for distinguishing between rural and urban regions [80]. This study used the Rural Health Research Center's guidance for classifying census tracts using RUCA codes [70]. Just as Zahnd et al. [48] used the Center's guidelines for classifying rural and urban census tracts in their study of mammography services accessibility, this study uses the Rural Health Research Center's Categorization C guideline which creates a binary distinction between rural and urban census tracts. Census tracts with RUCA codes 1.0, 1.1, 2.0, 2.1, 3.0, 4.1, 5.1, 7.1, 8.1, and 10.1 were labeled as 'urban' in this study. Census tracts with RUCA codes 4.0, 5.0, 6.0, 7.0, 7.2, 8.0, 8.2, 9.0, 10.0, 10.2, and 10.3 were labeled as 'rural' (Fig. 2). Any census tract with the RUCA code value 99 had zero population according to the U.S. Census Bureau and was eliminated from this study. See Table II for definitions of the RUCA codes.

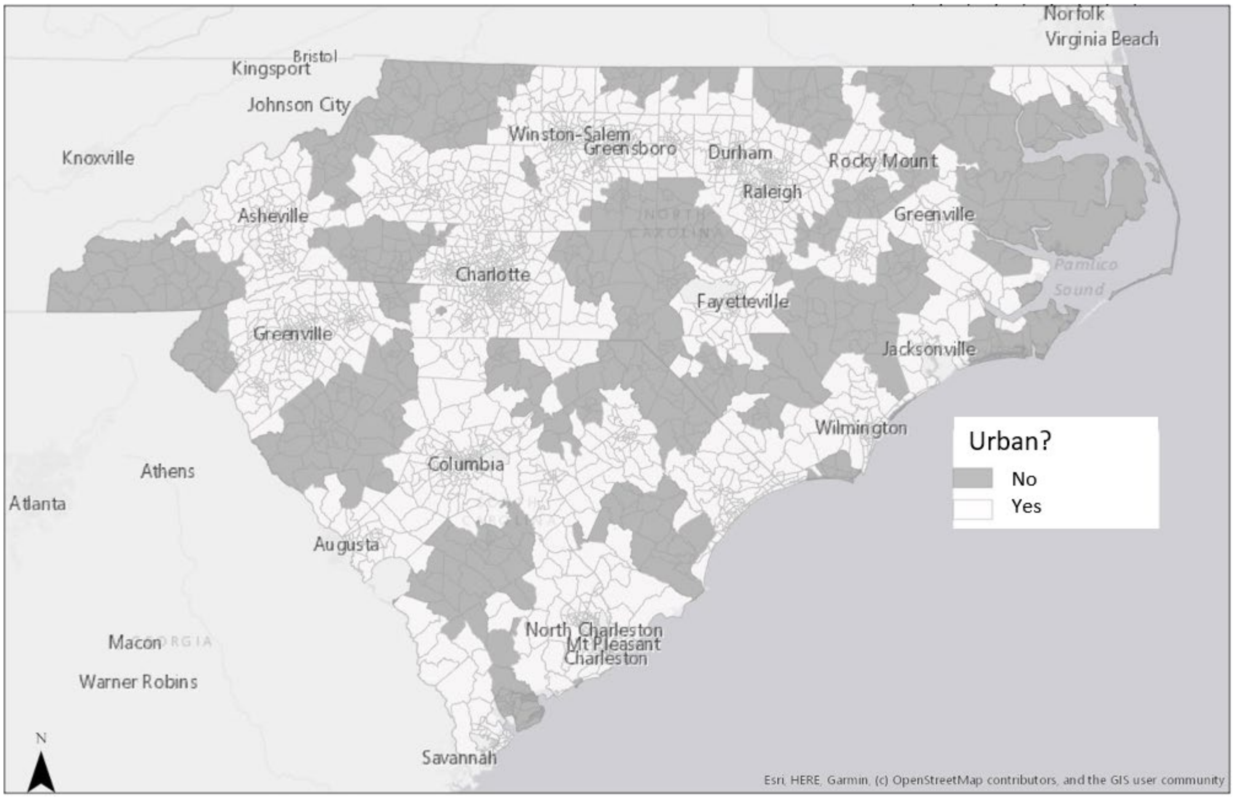

Fig. 2. Map of Urban and Rural Census Tracts in North Carolina and South Carolina. Classification based on Rural Health Research Center's Categorization C of USDA ERC's RUCA codes.

TABLE II
RUCA Codes and Definitions with Rural Health Research Center Classifications. The United States Department of Agriculture's Economic Research Service created and defined the Rural-Urban Commuting Area codes. UA refers to an Urbanized Area with a population of 50,000 or greater. UC refers to an Urbanized cluster with a population between 10,000 and 49,999. The Rural Health Research Center's classification used in this study Categorization C.

RUCA Code	**USDA ERC Definition**	**Rural Health Research Center Classification**
1.0	Metropolitan area core: primary flow within an urbanized area (UA)	Urban
1.1	Secondary flow 30% to 50% to a larger UA	Urban
2.0	Metropolitan area high commuting: primary flow 30% or more to a UA	Urban
2.1	Secondary flow 30% to 50% to a larger UA	Urban
3.0	Metropolitan area low commuting: primary flow 10% to 30% to a UA	Urban
4.0	Micropolitan area core: primary flow within an Urban Cluster of 10,000 to 49,999 (large UC)	Rural
4.1	Secondary flow 30% to 50% to a UA	Urban
5.0	Micropolitan high commuting: primary flow 30% or more to a large UC	Rural
5.1	Secondary flow 30% to 50% to a UA	Urban
6.0	Micropolitan low commuting: primary flow 10% to 30% to a large UC	Rural
7.0	Small town core: primary flow within an Urban Cluster of 2,500 to 9,999 (small UC)	Rural
7.1	Secondary flow 30% to 50% to a UA	Urban
7.2	Secondary flow 30% to 50% to a large UC	Rural
8.0	Small town high commuting: primary flow 30% or more to a small UC	Rural
8.1	Secondary flow 30% to 50% to a UA	Urban
8.2	Secondary flow 30% to 50% to a large UC	Rural
9.0	Small town low commuting: primary flow 10% to 30% to a small UC	Rural
10.0	Rural areas: primary flow to a tract outside a UA or UC	Rural
10.1	Secondary flow 30% to 50% to a UA	Urban
10.2	Secondary flow 30% to 50% to a large UC	Rural
10.3	Secondary flow 30% to 50% to a small UC	Rural

D. Generating Catchment Areas

The E2SFCA method, as defined by Luo & Qi [30], was used to generate catchment areas. For the first step of this method, catchment areas of pharmacies were created, and in the second step of this method, catchment areas of population centers represented by census tract centroids were created. Due to the concentration of pharmaceutical services in urban areas and the paucity of similar services in rural areas, catchment areas were generated with different parameters for rural and urban regions on the basis that less travel time is required to reach a pharmacy within an urban region than in a rural region. Catchment areas in urban regions were limited to a maximum drive time of 20 minutes with four zones: 1) 0-5 minutes, 2) >5-10 minutes, 3) >10-15 minutes, and 4) >15-20 minutes. The maximum travel time and the zone travel time ranges were based on Ikram et al. [45]. Literature on travel times to pharmacies in rural regions is sparse, so a reasonable approximation was made. A national survey of travel behavior in 2007 indicated that in a rural environment the mean travel time for a person to reach a healthcare service is 27 minutes [53]. Luo & Qi [30] created 30 minute catchment areas when they proposed the E2SFCA model to calculate spatial accessibility to primary care physicians in northern Illinois. This catchment area size was used for both urban and rural regions in their study area. In light of these studies, rural catchment areas were defined with a maximum travel time of 36 minutes and four zones of travel: 1) 0-9 minutes, 2) >9-18 minutes, 3) >18-27 minutes, and 4) >27-36 minutes. A distance decay function was used to model increasing friction (declining accessibility) the greater the time spent traveling. For assessing healthcare accessibility, a Gaussian function has become a de facto standard [30], [57], [48]. McGrail [57] proposed two sets of distance decay values, one with slow decay (1.0, 0.80, 0.55, 0.15) and one with fast decay (1.0, 0.60, 0.25, 0.05). Based on Luo & Qi's [30] opinion that fast distance decay functions should be used for common healthcare services like primary care physicians and pharmacies and slow distance decay functions should be used for specialty, or rarer, healthcare services, this study used McGrail's [57] fast distance decay function values. These values were used for both rural and urban catchment areas.

The ratio of opioid supply volume in MME to the accessing population for a pharmacy catchment area was calculated using Equation (3) [30, eq(4)] in which R_j is that ratio. Parameters in Equation (3) include j, the pharmacy location; S_j, the supply volume in MME at location j; k, the census tract centroids that fall within the pharmacy's catchment area travel time zone, D_r; P_k, the population of census tracts, k; and W_r, the distance decay weight applied to the distance decay breakpoint, r.

$$R_j = \frac{S_j}{\sum_{k \in \{d_{kj} \in D_r\}} P_k W_r} \tag{3}$$

The ratio of accessible opioid volume in MME to the population of a census tract was calculated using Equation (4) [30, eq. (5)] in which A_i^F is that ratio. It is the sum of the supply-to-population ratio, R_j, for each pharmacy, j, that lies within the census tract, i, catchment area, weighted by the distance decay value, W_r, in which r is the distance decay breakpoint. The term A_i^F is referred to in this paper as the opioid accessibility score of a census tract's population. Its unit of measurement is MME. Larger values indicate that the population had greater access to opioid analgesics per capita, and smaller values indicate that the population had lesser access to opioid analgesics per capita.

$$A_i^F = \sum_{j \in \{d_{ij} \in D_r\}} R_j W_r \tag{4}$$

E. Exploratory Data Analysis

Across North Carolina and South Carolina, there were 3,210 census tracts used in this study. There were 34 census tracts that were removed from the study because they had an estimated population of zero people. They primarily represented wilderness areas, airports or small uninhabited barrier islands or narrow areas along coastlines. There were also 53 census tracts that had no data for at least one ACS variable used in the study. These tracts were also removed from the study.

Several variables were removed from this study or excluded from certain models. The Veteran Affairs mail-order pharmacy in Charleston, South Carolina, serves veterans in multiple states beyond North Carolina and South Carolina and thus received a massive volume of opioids in 2009 (656.1 million MME). Because it does not exclusively serve a local population, that data point was removed from this study as was the corresponding ACS variable, percent of population that is a veteran. Collinear variables (>0.70) were dropped from both the k-medoids cluster analysis and the OLS regression model (Table III). Variables with low variability were dropped from the k-medoids cluster analysis.

TABLE III
Collinear ACS Variables

X Variable	Y Variable	R^2 Value
NHB	NHW	-0.91
Below Poverty	Med HH Income	-0.75
Per Cap Inc	HS Grad or Less	-0.74
Per Cap Inc	Med HH Income	0.80
HC03_VC104	HC03_VC110	0.92
Foreign Born	Hisp_Ltnx	0.76
Below Poverty	No Vehicle	0.71
HC03_VC104	Public Ins	0.74

F. Model Building

Cluster analysis was performed in RStudio. The pam (Partitioning Around Medoids) function in the cluster library was used to produce K-medoids clusters. Given the number of variables used in this study with skewed data distributions, this algorithm was chosen as a more robust alternative to the more frequently used K-means cluster methodology. K-medoids clustering requires that the parameter, *k,* be chosen ahead of time as the number of clusters to be created. Determining the optimal number of clusters to use in the analysis is an iterative process in which within cluster variation is measured as *k* is increased incrementally. When *k* is small, the variation drops precipitously. As *k* increases in value the change in variation diminishes. Researchers choose a value for *k* based on when the within cluster variation no longer decreases appreciably. This is commonly referred to as the elbow method [81]. Based on this method, seven clusters were created. Because the variables used for the clustering analysis have inconsistent scales, all variables were rescaled.

An OLS regression model was created in RStudio. In order to produce a valid model, assumptions were checked to confirm they could be met. Independent variables that showed high collinearity were removed from the OLS regression model (Table III). The response variable, the opioid accessibility score, and many of the independent variables from the ACS were skewed; the assumption of normality had been met, however, due to the large number of observations in the study (3,210). Heteroscedasticity was reviewed visually and statistically with the studentized Breusch-Pagan Test and the Non-constant Variance Score Test. All variables, response and independent, were power transformed with lambda based on results from the Box-Cox method. The model was checked for high-leverage outliers based on their standardized variance (>4) and hat values (> 0.014). Hat values were calculated using Equation (5):

$$\frac{2(p+1)}{n} \tag{5}$$

where p is the number of variables in the regression model and n is the number of observations[5]. Three methods were used for feature selection: forward stepwise, backward stepwise, and Bayesian model averaging. Stepwise methods were assessed using Bayesian Information Criteria (BIC) which prioritizes parsimonious models.

G. Software

ARCOS data was cleansed and aggregated using an Anaconda distribution, version 1.9.7, using Jupyter Notebooks, version, 6.0.1, and Python, version 3.7.4. Rural and urban classification, pharmacy address geocoding, and the E2SFCA method was implemented using Esri's GIS software, ArcGIS Pro, version 2.3.2. Geocoding used Esri's online World Geocoding Service. K-medoids cluster analysis and OLS Regression models were built using RStudio v1.1.456.

[5] Equation (5) and recommended values for standardized variance and hat values come from personal communication with Dr. Lindsey Bell.

V. Results

Catchment areas were generated in this study in order to calculate the MME volume of opioid analgesic drugs accessible to each census tract's population on per capita basis. Travel time cost weighted by distance decay values affected the contribution of each pharmacy that fell within a catchment area. The opioid analgesic MME volume to person ratio is referred to as the opioid accessibility score in this study (Fig. 3).

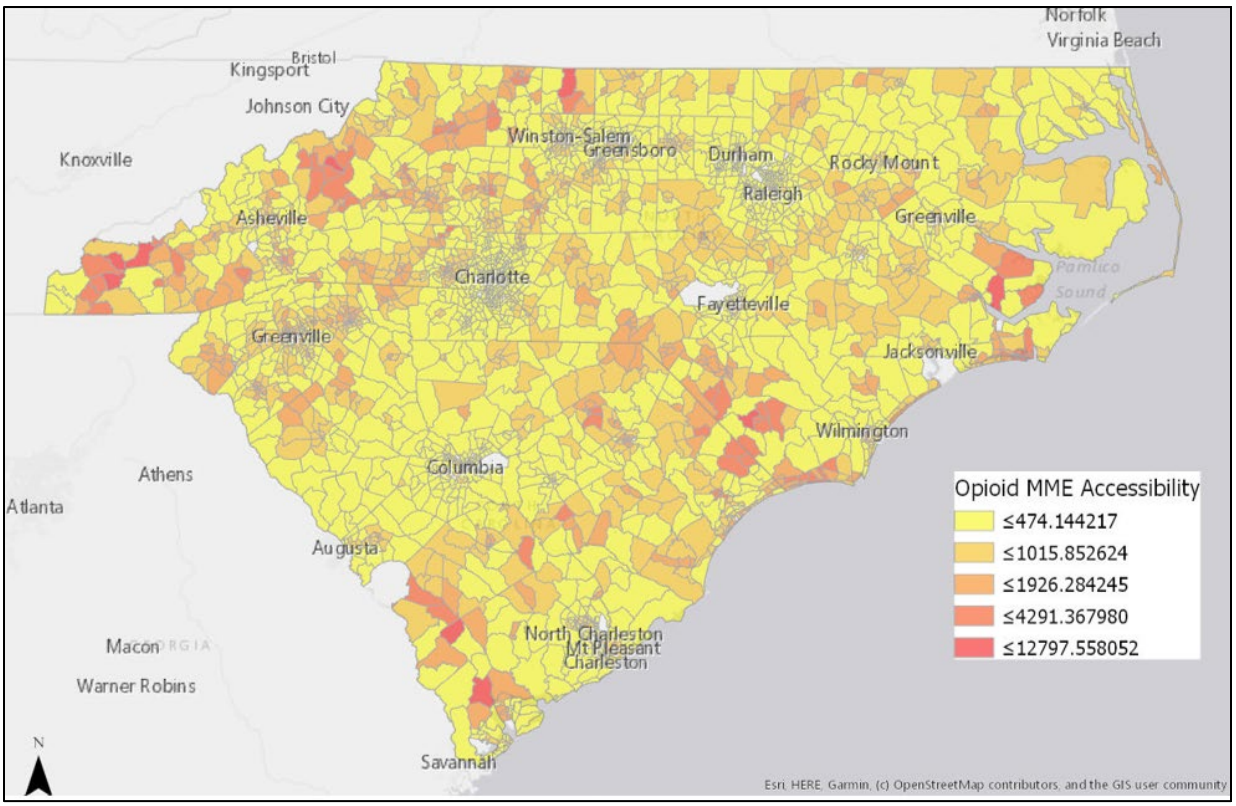

Fig. 3. Choropleth Map of Opioid Accessibility Scores for Census Tracts in North Carolina and South Carolina. Choropleth values are based on Natural Jenks.

A. Descriptive Analysis

In 2009 there were 3,003 pharmacies in North Carolina and South Carolina that received shipments of hydrocodone or oxycodone-based opioid analgesics. The median pharmacy received 1,557,550 MME (Table IV). Outlier pharmacies were identified as receiving greater than 1.5 times the value of the interquartile range (IQR) of the 3rd quartile. There were no outliers below the lower fence. There were 177 outlier pharmacies above the upper fence ($\geq 6{,}012{,}687.5\ MME$). There were 2,262 (75.3%) pharmacies located in census tracts classified in this study as urban. There were 107 (60.5%) pharmacies located in urban census tracts that were identified as outliers.

TABLE IV
Summary of Distribution of Opioids to All Pharmacies. Values are total MME distributed during 2009.

Min	**1st Quartile**	**Median**	**3rd Quartile**	**Max**
500 MME	772,375 MME	1,557,550 MME	2,868,500 MME	20,884,000 MME
IQR	**Lower Fence**	**Upper Fence**		
2,096,125 MME	< 0 MME	6,012,687.50 MME		

Census tracts used in this study were based on the 2010 decennial census. Excluding tracts with zero population or with missing ACS variables, there were 3,210 tracts included in this study of which 2,652 (82.6%) were classified as urban. The volume of opioids available to each person in any given census tract was calculated using the E2SFCA method. The median volume of opioid analgesic accessible per capita in the census tracts was 442.9 MME (Table V). Outlier census tracts were identified as having access to a volume of opioids greater than 1.5 times the value of the IQR of the 3rd quartile. There are no outliers below the 1st quartile. There were 167 census tracts above the upper fence ($\geq 1601.85\ MME$ per capita). Fifty-five (32.9%) of those outliers were urban census tracts.

TABLE V
Summary of Distribution of Opioids in All Census Tracts. Values are opioid accessibility scores: the ratio of MME volume accessible per capita in a census tract.

Min	**1st Quartile**	**Median**	**3rd Quartile**	**Max**
0.0 MME	246.1 MME	442.9 MME	788.4 MME	12,797.4 MME
IQR	**Lower Fence**	**Upper Fence**		
542.29 MME	< 0 MME	1601.85 MME		

There were 89 census tracts that had an opioid accessibility score of zero. This means that those tracts' centroids did not fall within the catchment areas of any of the pharmacies in the study area. Twelve of these tracts were defined as rural for this study. Seven of them represent either coastal edges, barrier islands, or remote wilderness in Appalachia; all of them have no population. The other five rural tracts had populations, but they are remote coastal or Appalachia tracts that are too distant from pharmacies to fall within a pharmacy's catchment area. Of the 77 urban tracts, 11 lie at the edges of the study area. That artificial barrier may have played a role in impacting accessibility for those tracts. Nine tracts have at least one pharmacy that fell within them. A lack of roads that come close to the geographic centroid of these tracts may have played a role in those tracts not falling within any pharmacy's catchment area. The vast majority of the other urban tracts with an accessibility score of zero lie at the fringe between urban and rural regions. They represent portions of the study area that lie at the furthest areas that are drawn towards an urban center and yet are also too far away from the small rural population centers where rural pharmacies tend to be located. Of the urban tracts that are surrounded by other urban tracts and yet do not have access to pharmacies, a combination of geographic barriers, distance from pharmacies, and distance from tract centroids to the road network account for their lack of accessibility. There is one census tract in Charleston County, South Carolina, with an accessibility score of zero, but was within travel time of one pharmacy. That pharmacy's catchment area did not reach any tract centroids and thus had a weighted population sum of zero. This effected its contribution to all tracts (six) that could reach that pharmacy. A second pharmacy in Madison County, North Carolina, also did not have any tract centroids that fell within its catchment area. Unlike the pharmacy in Charleston County, this pharmacy was not reachable by any tract centroids.

B. Cluster Analysis

A k-medoids clustering algorithm was used to produce seven clusters of census tracts based on ACS variables (Fig. 4). A subset of ACS variables was used in the analysis. Four variables were removed for high collinearity – > 0.70 – with other variables: percent of population that is non-Hispanic Black, percent of population below the poverty level, per capita income, and percent of population disabled. After the results of an initial cluster analysis were analyzed, an additional five variables were removed from a second cluster analysis because they showed little variability among the clusters: percent of population under 21 with a disability, percent of housing units without a mortgage and housing costs greater than 35% of income, percent of population over age 65 that is male, percent of population between ages 18 and 64 that is male, and percent of population that is non-Hispanic: other races. Only the results of the second cluster analysis are reported here.

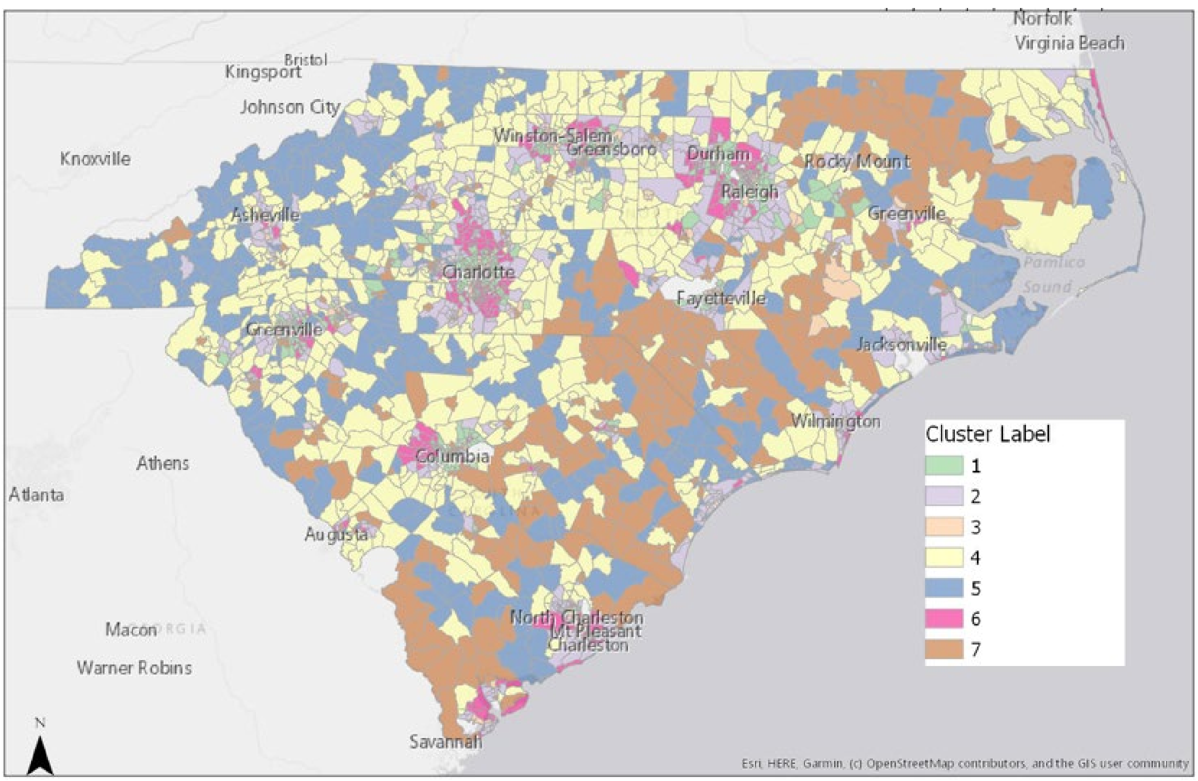

Fig. 4. Map of K-Medoids Clusters.

Each cluster is described below in terms of its predominant characteristics. All characteristics are based on median variable values; however, many clusters' variables show a wide range of values – in some cases between 0% to 100% - meaning there is much overlap variable by variable amongst the clusters. It is the distribution of variables' values that distinguish one cluster from another.

Details regarding outliers of opioid accessibility are based on the interquartile range of the variable's distribution as described in Equation (6) in which O is the volume of opioid analgesics in MME, $Q3$ and $Q1$ are the 75th and 25th percentiles of the distribution, respectively, and IQR is the interquartile range, calculated as $Q3 - Q1$.

$$O = \begin{cases} Q3 + (IQR * 1.5) \\ Q1 - (IQR * 1.5) \end{cases} \tag{6}$$

Some variables showed little differentiation in median values. The variable percent of population between ages 18 and 64 that is male ranged between 47.2% and 49.6%. The variable percent of population over age 65 that is male ranged between 38.5% and 45.2%. Due to this small differentiation, these variables are not referenced in the descriptions of clusters below. All median variable values are presented in Table VI.

TABLE VI
Median Variable Values of Clusters

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Educational Attainment: Percent of population with High School degree or less	0.31	0.356	0.503	0.531	0.555	0.1765	0.6145
Disability: Percent of total population with a disability	0.085	0.116	0.114	0.147	0.198	0.071	0.186
Disability: Percent of under 18 population with a disability	0.0255	0.029	0.03	0.037	0.044	0.019	0.045
Disability: Percent of 18 to 64 population with a disability	0.07	0.09	0.10	0.13	0.18	0.05	0.17
Disability: Percent of over 65 population with a disability	0.325	0.359	0.427	0.379	0.4455	0.263	0.487
Median household income	$53,190.50	$52,111	$32,621	$45,246	$37,286.50	$86,002.50	$28,767.50
Per capita income	$26,701	$27,057.50	$17,642	$21,815	$20,169.50	$40,562	$15,679.50
Health Insurance: Percent of population with public health insurance	0.227	0.265	0.32	0.332	0.385	0.17	0.4455
Health Insurance: Percent of population with no health insurance (public or private)	0.143	0.137	0.268	0.155	0.1845	0.064	0.208
Poverty: Percent of population whose income in last 12 months was below the poverty level	0.1115	0.1115	0.277	0149	0.1795	0.042	0.30
Housing: Percent of housing units that are renter-occupied	0.3525	0.284	0.625	0.198	0.242	0.125	0.426
Housing: Percent of housing units with no vehicle available	0.0315	0.037	0.114	0.042	0.061	0.015	0.143
Housing: Percent of owner-occupied housing units with a mortgage	0.7815	0.721	0.686	0.595	0.52	0.789	0.5325
Housing: Percent of housing units with a mortgage in which housing costs are 35% or more of income	0.2055	0.247	0.292	0.215	0.2915	0.1925	0.326
Housing: Percent of housing units without a mortgage in which housing costs are 35% or more of income	0.07	0.08	0.096	0.088	0.1005	0.068	0.1405
Housing: Percent of rented housing units with rental costs 35% or more of income	0.345	0.41	0.46	0.41	0.41	0.31	0.50
Median age in years	34	39	31	41	44	40	38
Sex & Age: Percent of population 65 and over that is male	0.419	0.4275	0.388	0.447	0.441	0.452	0.385
Sex & Age: Percent of population 18 to 64 that is male	0.474	0.486	0.491	0.496	0.493	0.484	0.472
Race & Ethnicity: Percent population Hispanic or Latino – all races	0.077	0.0425	0.186	0.033	0.026	0.027	0.043
Race & Ethnicity: Percent population not Hispanic or Latino – white only	0.6	0.772	0.367	0.813	0.8335	0.849	0.3775
Race & Ethnicity: Percent population not Hispanic or Latino – black only	0.2515	0.1215	0.351	0.102	0.086	0.058	0.51
Race & Ethnicity: Percent population not Hispanic or Latino – all other races	0.055	0.035	0.396	0.0175	0.0188	0.04	0.024
Population Density (Persons per Square Mile)	1610.3	846.6	2808	134	97.8	1156.7	454.8
Opioid Accessibility Score (MME)	328.535	483.873	477.424	364.997	628.426	280.424	704.103

1) Cluster 1: Cluster 1 can be categorized as an urban and diverse population with a higher educational attainment and lower economic stress (Fig. 5). Its census tracts are primarily centered around the metropolitan areas of both states: Greenville, Spartanburg, and Columbia in South Carolina, and Charlotte, Fayetteville, Winston-Salem, Greensboro, Durham, and Raleigh in North Carolina. While the majority of census tracts are categorized as an urban metropolitan core based on the USDA's RUCA codes [70], they tend to encircle the very center of these cities. Persons in these areas have the greatest access to pharmaceutical services – 78 pharmacies per census tract catchment area, but they have the second lowest accessibility score of opioid analgesics (328.5 MME). There are six tracts that are considered outliers for opioid accessibility, however none of them are urban core tracts nor are they located near any of the cities referenced above. This population is one of the youngest with a median age of 34 years. It can also be described as the most diverse: median percentages of non-Hispanic Whites and non-Hispanic African Americans are 60% and 25.15%, respectively, and the median percentage of Hispanics and non-Hispanics of all other races are also high: 7.7% and 5.6%, respectively. Housing occupancy is mixed with about one-third of housing units rented and the remainder owner-occupied. Of those owned, a large proportion of owner have a mortgage (78.2%). Less than one-third of persons have no more than a high school diploma (31.0%), and median household income is one of the highest ($53,190.50). Per capita income ($26,701) and the median age suggest that households in general include a lot of families. This cluster shows fewer signs of economic stress than other clusters. Lack of access to a personal vehicle is low (3.2%), and housing stress is lower than most other clusters: housing expenses greater than 35% of income for those with a mortgage is 20.6% and renting expenses greater than 35% of income is 35.4%. Poverty rates are also low (11.2%). Use of public health insurance is lower than most clusters (22.7%); disability is likewise lower than most other clusters (8.5%).

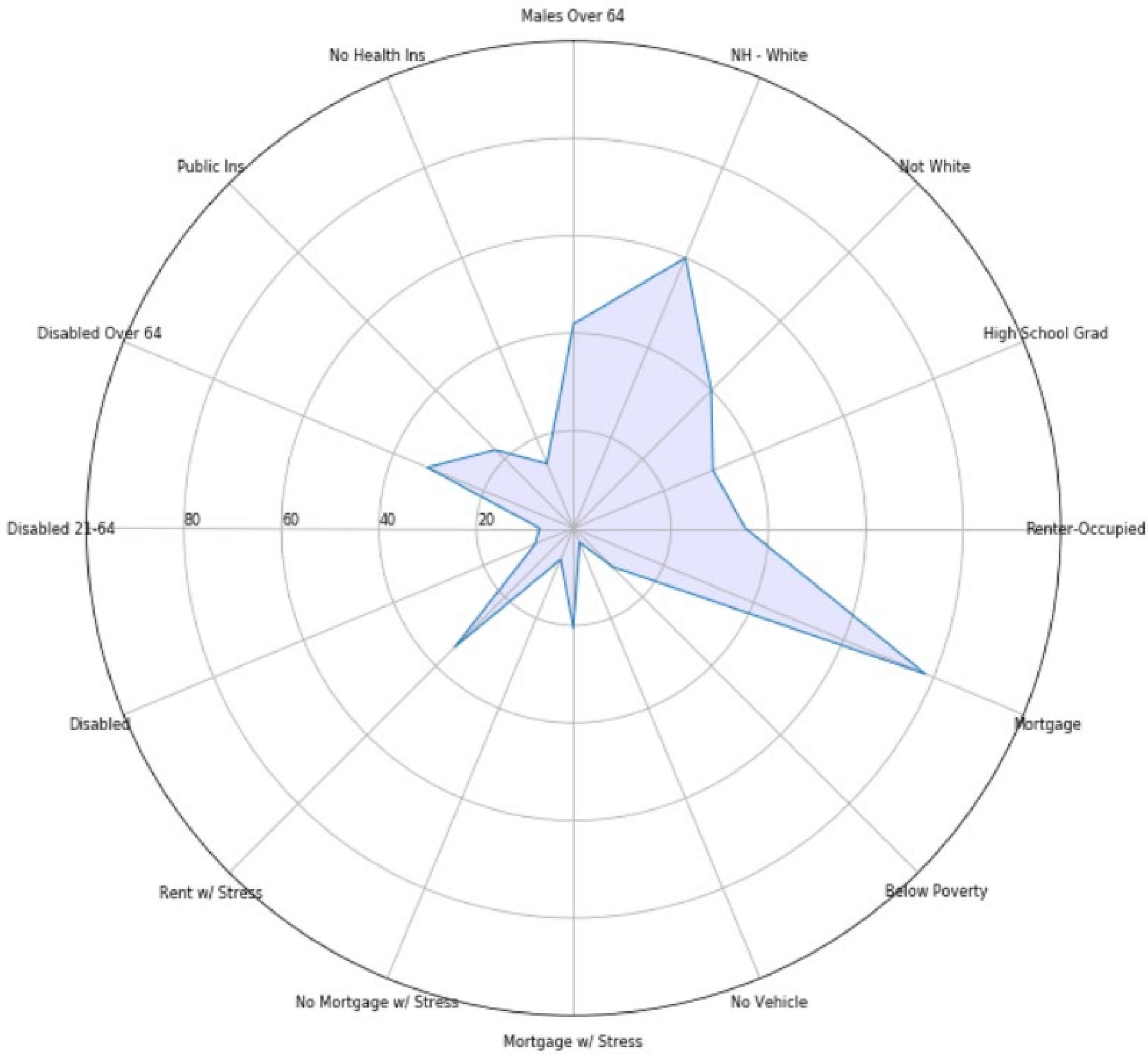

Fig. 5. Radar Plot of Cluster 1. All values listed are median values for all census tracts in this cluster. Age in years: 34, household income: $53,191, per capita income: $26,701, persons per square mile: 1610, pharmacy count: 78, and opioid MME volume: 328.53.

2) Cluster 2: Many of the tracts in this cluster are part of the urbanized areas described for cluster 1, or they have high levels of commuting into those same urbanized areas (Fig. 6). This cluster also includes tracts in the center of many other urban areas: Asheville, Greenville, Rocky Mount, and Wilmington, among others, in North Carolina, and Charleston, Florence, and Myrtle

Beach in South Carolina. Persons in these areas have lower access to pharmaceutical services than cluster 1 (32 pharmacies per census tract catchment area), but a higher volume of opioids (483.9 MME). There are 18 census tracts that are outliers in terms of the available volume of opioids. Outlier tracts are outside the suburban areas that surround the larger metro areas. Rather, they are found in smaller towns like Florence, Hartsville, and Conway in South Carolina and Statesville, Lumberton, and New Bern in North Carolina. It is a modestly older population – median age is 39, and it is less diverse than the population in cluster 1: non-Hispanic Whites (77.2%), non-Hispanic Blacks (12.2%), and Hispanics (4.3%). Economically, this population is similar to the population in cluster 1. Educational attainment is similar (35.6%), and median household income is also similar ($52,111). Homeownership is greater (28.4%), although more households own their home outright (72.1%). There may be signs of slightly more economic stress than for the population in cluster 1. While the percentage of the population below the poverty level is the same (11.2%), housing costs take up a large share of income – greater than 35% - for more households: with a mortgage (24.7%) and for renters (41.0%). About one-quarter of the population is on public health insurance (26.5%); 13.7% have no health insurance. There is a slightly larger proportion of the population that is disabled (11.6%) than in cluster 1.

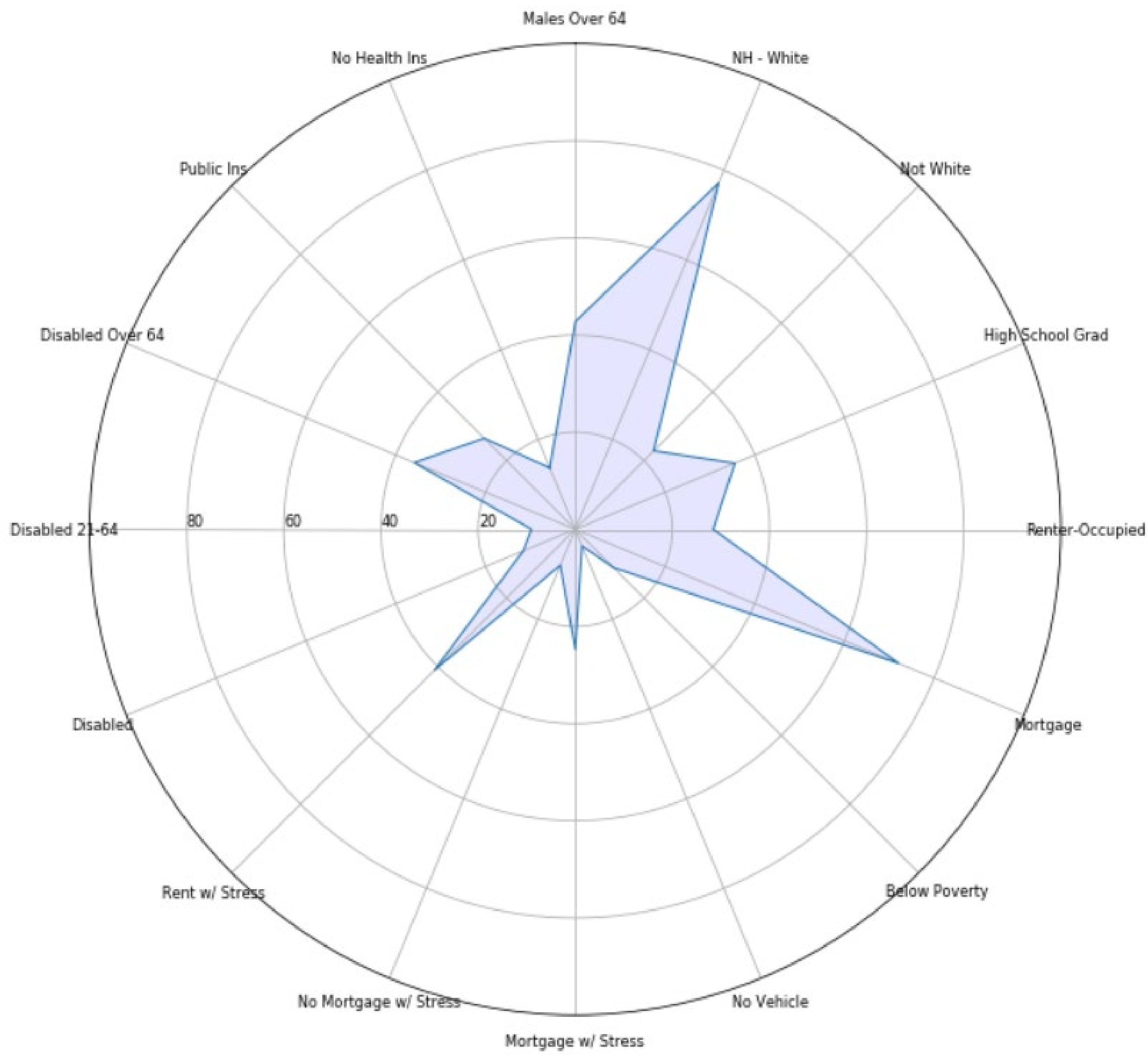

Fig. 6. Radar Plot of Cluster 2. All values listed are median values for all census tracts in this cluster. Age in years: 39, household income: $52,111, per capita income: $27,058, persons per square miles: 847, pharmacy count: 32, and opioid MME volume: 483.87.

3) Cluster 3: This populace is impoverished, poorly educated, highly diverse, and living in close proximity to each other (Fig. 7). The population of this cluster lives in the most densely populated census tracts of the study area's urban cores. Median population density is 2,808 persons per square mile. Like the populace in cluster 1, this population has high access to pharmaceutical services – median pharmacy count is 72 per catchment area, however the volume of opioids is more similar to cluster 2 (477.3 MME). Diversity is high. The percent of those who identify as Hispanic is highest (18.6%) and the non-Hispanic White (36.7%) and non-Hispanic Black (35.1%)

populations are equivalent. Like cluster 7, this population has one of the highest percentages of females over 64 years (61.2%). Just over half the population has no more than a high school education (50.3%); the level of poverty is also the second highest among the clusters (27.7%). Percentage of households that rent is the highest (62.5%). No access to a personal vehicle is also high (11.4%). This provides evidence of economic stress as do higher percentages of households with housing (29.2%) and rental (46.0%) costs above 35% of income. This population has one of the highest percentages of those with a disability and over 64 (42.7%) and the highest percentage with no health insurance coverage (26.8%).

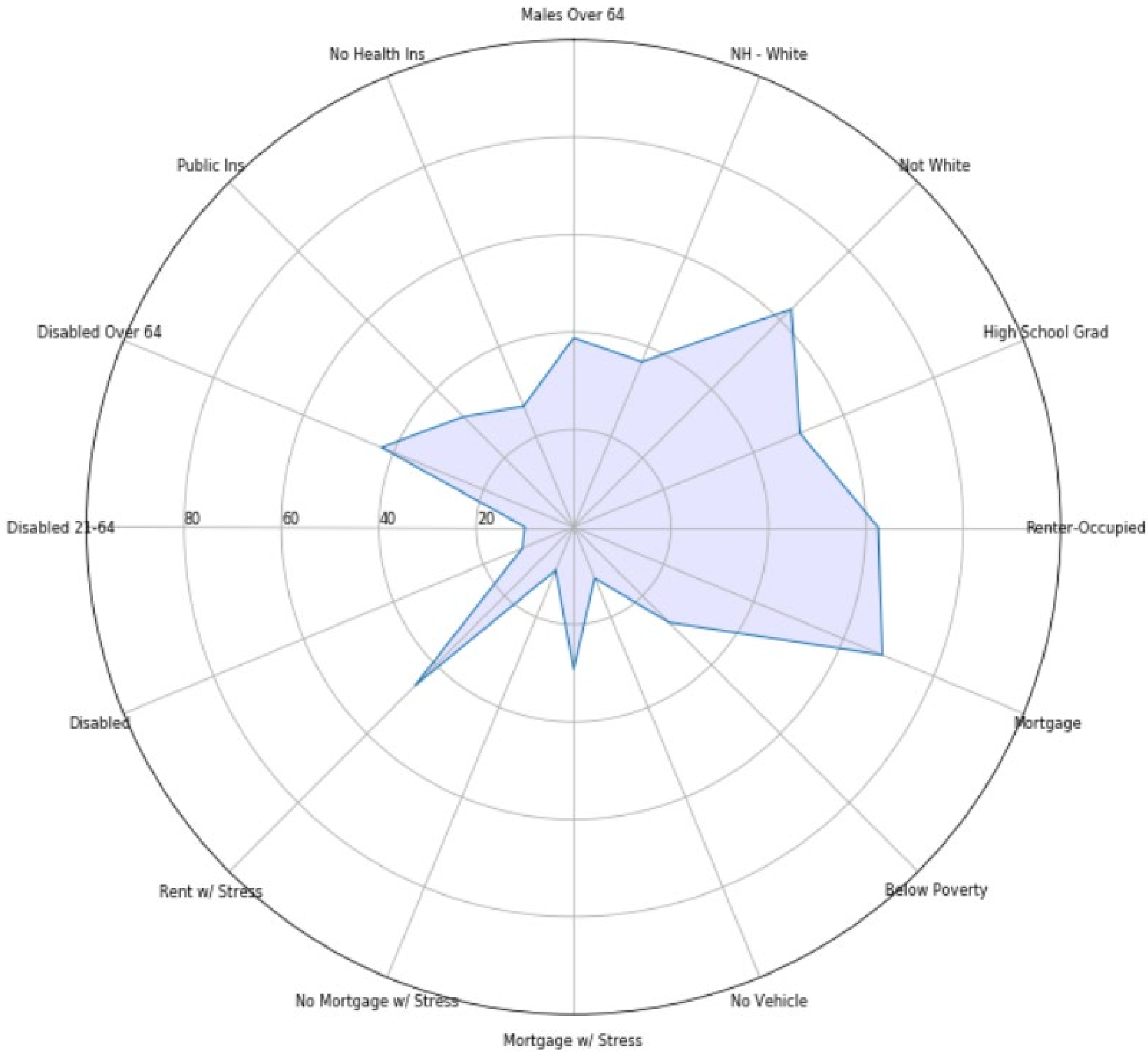

Fig. 7. Radar Plot of Cluster 3. All values listed are median values for all census tracts in this cluster. Age in years: 31, household income: $32,621, per capita income: $17,642, persons per square miles: 2,808, pharmacy count: 72, and opioid MME volume: 477.42.

4) Cluster 4: This cluster is the most spatially dispersed. It represents areas surrounding all the larger urban centers in the study area: Charlotte, Columbia, Greenville, Winston-Salem, and others (Fig. 8). It also represents remote portions of Appalachia and coastal areas in both states. Over two-thirds of the tracts represent outer portions of urbanized areas or areas beyond urbanized areas that have a higher percentage of commuters traveling into an urbanized area (RUCA = 2), however this cluster also represents all other categories of the RUCA system from micropolitans to small towns and rural areas. It is one of the least densely populated of the clusters (134 persons per square mile). It also represents some of the lowest access to pharmaceutical services (18 per census tract catchment area) and access to opioids (365 MME). Despite this low access, 35 of the census tracts are outliers. Most of these 35 tracts are in the Appalachian region; only seven lie outside that region. It is also one of the least diverse populations: non-Hispanic Whites account for 81.3% of the populace. It is the second oldest cluster with a median age of 41 years, and it has one of the highest proportions of men over 64 years (44.7%). The median percentage of renter-occupied housing is rather low (19.8%), yet rates of homeowner-occupied housing has one of the lowest median percentages of mortgages (59.5%). Those who do rent have greater housing stress than those who pay a mortgage (41.0% vs 21.5%, respectively). In terms of health and insurance coverage, this cluster fits about in the middle among the clusters: median percentage of disabled population (14.7%), public health insurance coverage (33.2%), and no health insurance coverage (15.5%).

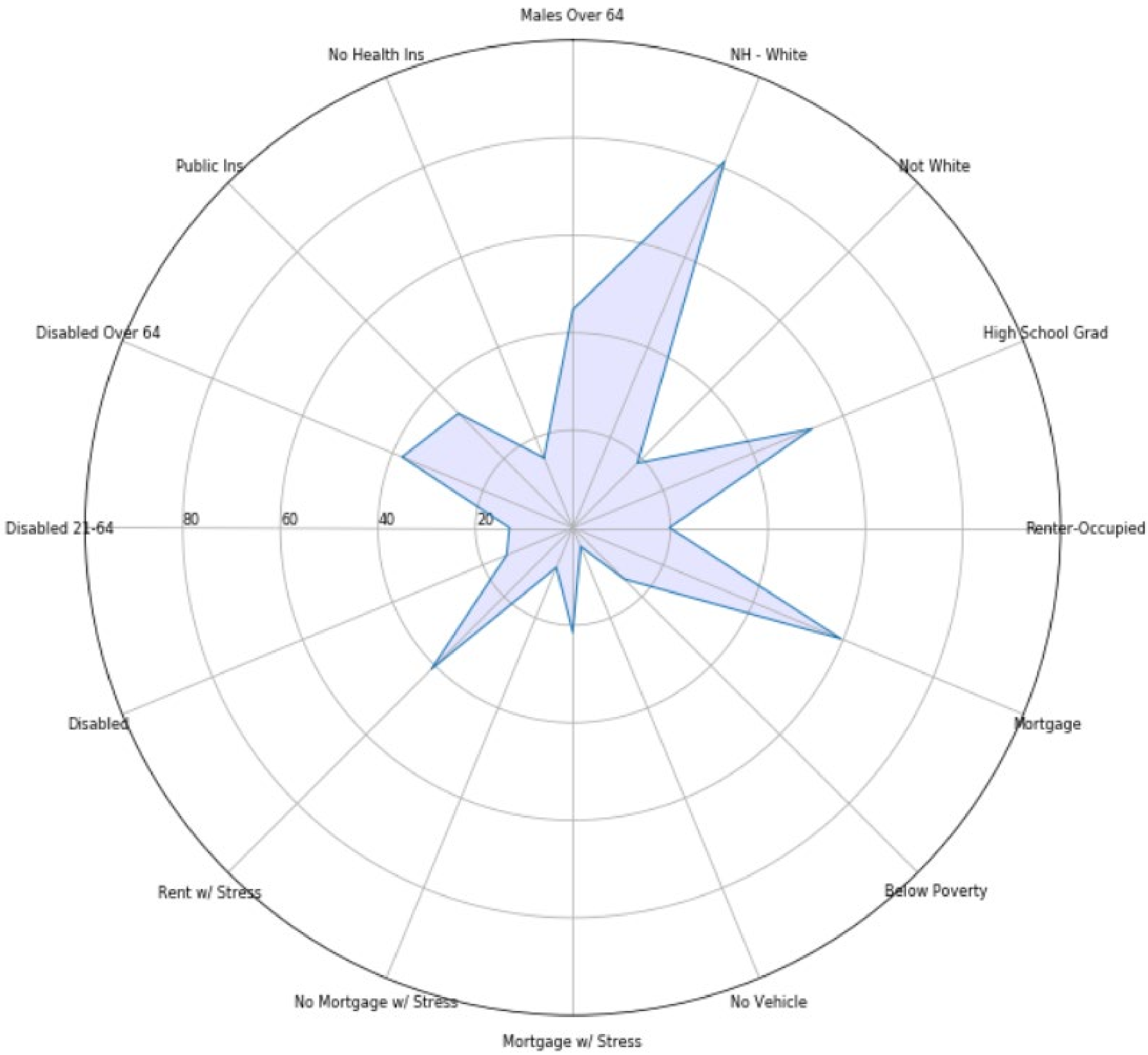

Fig. 8. Radar Plot of Cluster 4. All values listed are median values for all census tracts in this cluster. Age in years: 41, household income: $45,246, per capita income: $21,815, persons per square miles: 134, pharmacy count: 18, and opioid MME volume: 365.

5) Cluster 5: Cluster 5 is the least densely populated of all the clusters with a median 97 persons per square mile (Fig. 9). Much of Appalachia is represented in this cluster, although one-third of its census tracts are in the Low Country of both states and the remainder of its census tracts are spread out between the coast and Appalachia. Like clusters 4 and 6, this is a racially homogenous population. The median non-Hispanic White percentage of the population is 83.4%; census tracts that drop below 50% non-Hispanic White are all east of a line that runs from Greensville, South Carolina, to Winston-Salem, North Carolina. The population has the oldest median age (44 years), and it has one of the highest proportions of men over 64 (44.1%). Educational attainment is the second lowest of all clusters (55.5%). Median household income is one of the lowest as well ($37,286.50). However, homeownership is fairly high; median percentage of rental occupancy is 24.2%, and the low median percentage of owner-occupied housing units with a mortgage (52.0%) suggest a stable, non-mobile populace. Poverty is more prevalent (18.0%) and is evident in the percentage of households that spend more than 35% of income on housing costs: with a mortgage (29.2%), without a mortgage (10.1%), and renting households (41.0%). Use of public health insurance (38.5%) and lack of any health insurance coverage (18.5%) are among the highest of all the clusters. The median percentage of the population with a disability is the highest (19.8%) and the second highest for the over 64 population (44.6%). Despite having the lowest access to pharmaceutical services (median pharmacy count = 17), there is access to a high volume of opioids per capita (628.4 MME), second only to cluster 7. There is also an equivalent number of census tracts that are outliers for opioid access in cluster 7 – 53. They are evenly split spatially between Appalachia and the Low Country.

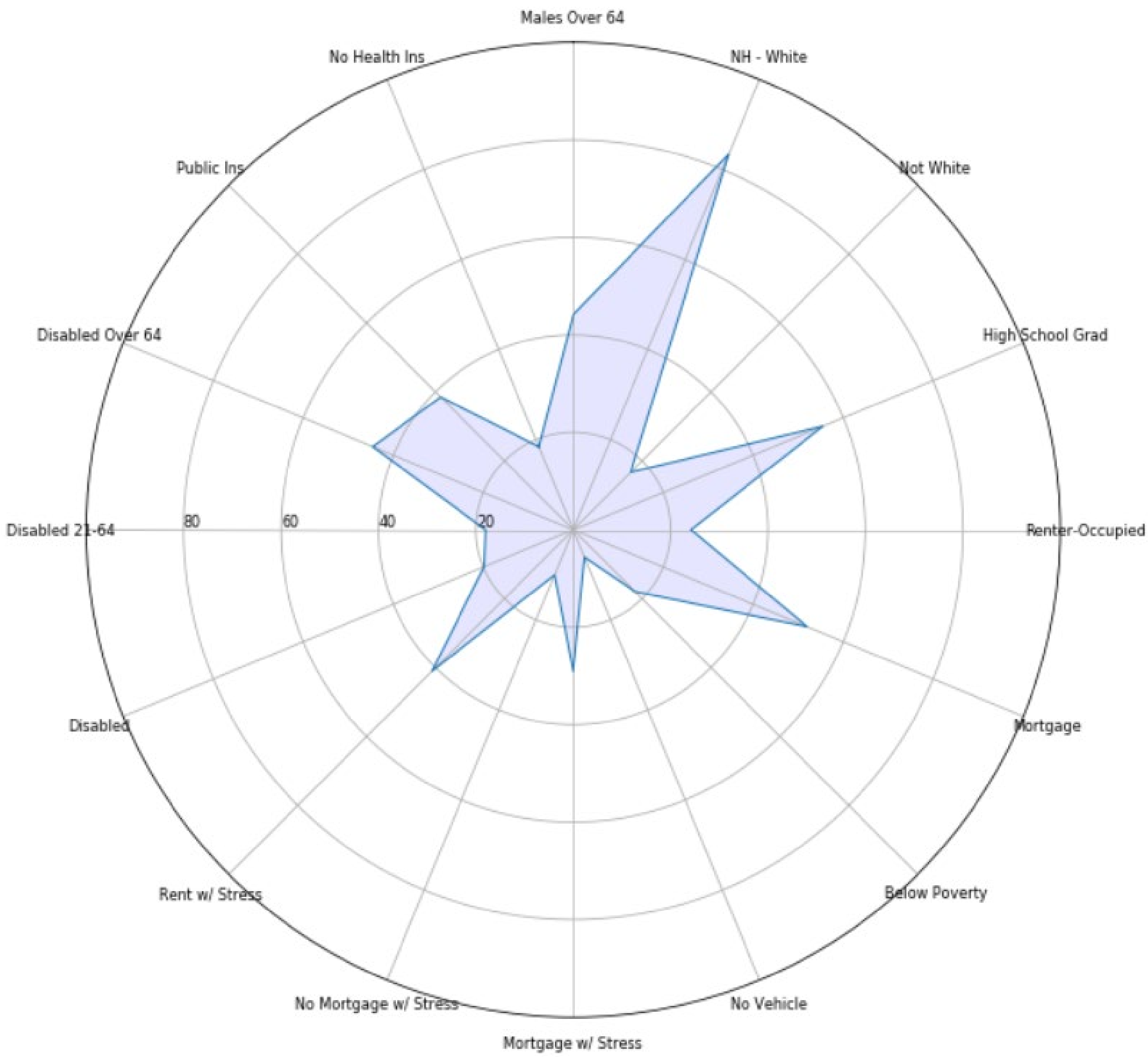

Fig. 9. Radar Plot of Cluster 5. All values listed are median values for all census tracts in this cluster. Age in years: 44, household income: $37,287, per capita income: $20,170, persons per square miles: 98, pharmacy count: 17, and opioid MME volume: 628.43.

6) Cluster 6: The populace in cluster 6 is best characterized as homogenously non-Hispanic White living in suburban neighborhoods with high levels of educational and economic attainment (Fig. 10). Census tracts in this cluster are primarily located between census tracts in cluster 1 and census tracts in cluster 2 in urbanized areas. This is more or less the case in Columbia, Spartanburg, and Greenville in South Carolina and in Charlotte, the Winston-Salem-Greensboro region, and the Durham-Raleigh region in North Carolina. This cluster is also present in Charleston and Beaufort, South Carolina, but the concentric nature of the clusters is not present as it is in other metro areas. It represents a populace that is the least diverse among the clusters: the median percentage of non-Hispanic Whites is 84.9% with even the 10th percentile of census tracts having a non-Hispanic White population over 55%. Like other tracts with a median percentage of non-Hispanic Whites over 80%, it has a higher median percentage of males over 64 (45.2%). This is the most highly educated of all the clusters; the median percentage who have no more than a high school diploma is 17.7%. Median household income is also the highest among the clusters: $86,002.50. Home ownership is the highest (87.5%) as is the median percentage of owner-occupied housing units with a mortgage (78.9%) suggesting a mobile populous that has the capital to purchase a home but does not stay in any one place long enough to pay off a mortgage. There are almost no signs of economic stress; all stress indicators are the lowest median values of all the clusters. It is the only cluster with a median percentage of households below the poverty level in the single digits (4.2%), however nearly one-third of renting households pay more than 35% of income for housing costs (31.0%). In line with the educational and economic success evident, this is a population with the lowest median levels of disability and access to private health insurance. Median percentages of public health insurance coverage (17.0%) and no health insurance coverage (6.4%) are both the lowest among the clusters. The populace has one of the highest rates of access to pharmaceutical services – median pharmacy count is 58 per catchment area, and yet it also has the lowest volume of opioids accessible from those pharmacies – 280.4 MME per capita. There are just two outliers in this cluster, but they are not located in the suburban neighborhoods that this cluster best represents.

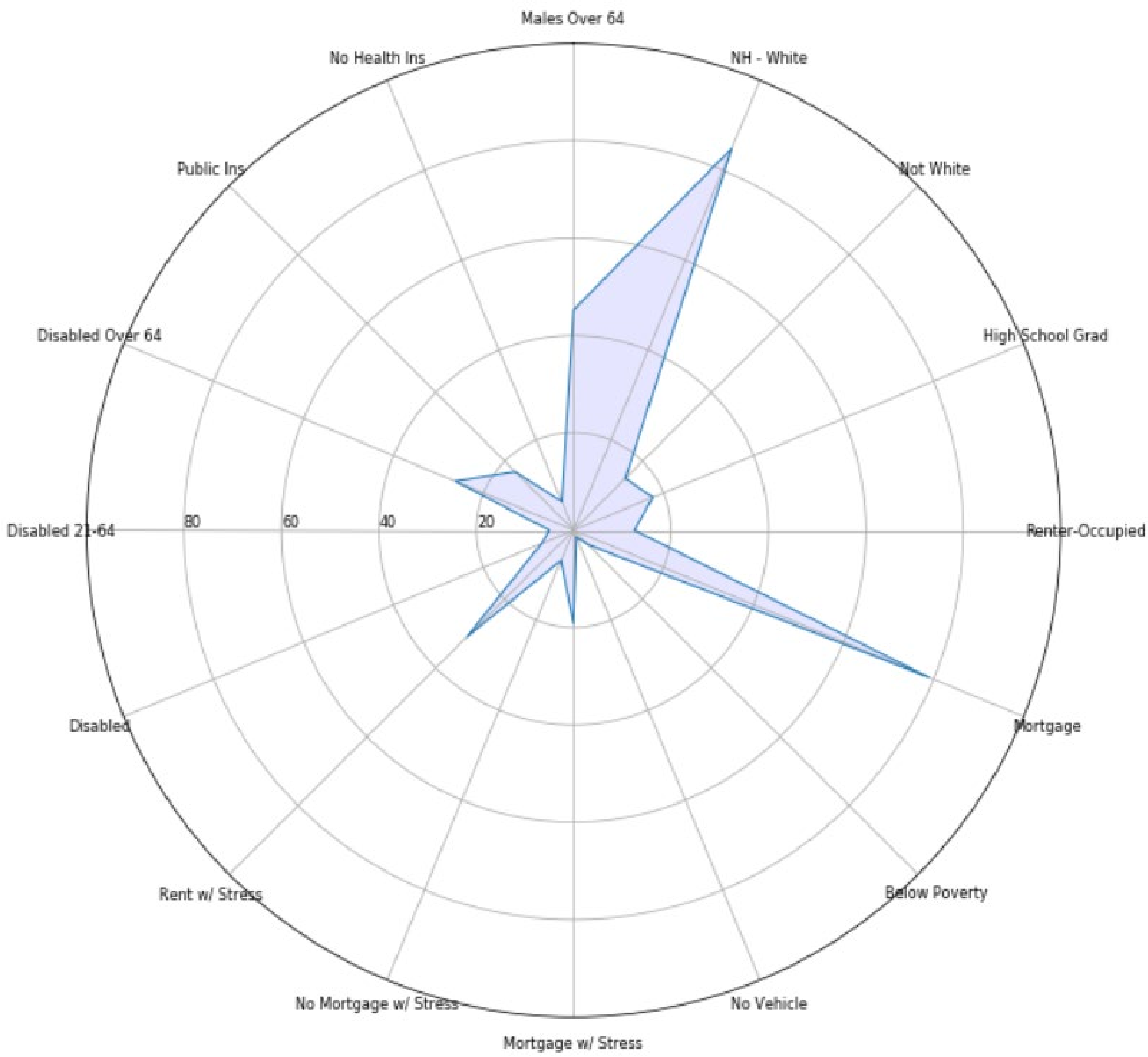

Fig. 10. Radar Plot of Cluster 6. All values listed are median values for all census tracts in this cluster. Age in years: 40, household income: $86,003, per capita income: $40,562, persons per square miles: 1157, pharmacy count: 56, and opioid MME volume: 280.42.

7) Cluster 7: This cluster can be categorized as being the most vulnerable among all the clusters in terms of socio-economic stress, health care, and educational attainment (Fig. 11). The vast majority of the census tracts grouped in this cluster are located in a non-contiguous swath from the Virginia border in the north to the Georgia border in the south and bounded on the west by the metropolitan areas of Durham-Raleigh, Charlotte, and Columbia. It is a mix of urban (60%) and rural (40%) census tracts in which the tracts classified as urban represent smaller towns or areas where many commuters feed into more densely populated urban centers, but it also includes neighborhoods in the heart of several urban cores such as Charleston, Columbia, and Greenville in South Carolina and Wilmington, Charlotte, Durham, and Winston-Salem in North Carolina. East of Columbia, Charlotte, and Raleigh, the population is predominantly Non-Hispanic African American (51.0%), while west of those metro areas, census tracts have a more even mix of non-Hispanic Whites and African Americans. Women are also more prominent in the 18-64 demographic (52.8%), but even more so in the over 64 demographic (61.5%). This cluster shows signs of economic stress. It has the lowest median household income ($28,768) of all the clusters, and the highest median rate of households below the poverty level (30.0%). It also has the highest median percentage of persons whose educational attainment is no greater than a high school diploma (61.5%).

Economic stress is evident in vehicle ownership and housing costs. While it may be reasonable to anticipate that persons living in an urban core can manage without access to a vehicle, it is much more difficult to manage in rural or suburban areas where commuting via personal vehicle is a must. This cluster has the highest median proportion of households without access to a vehicle (14.3%). Urban tracts have much higher percentages – greater than 50% for several tracts in Charleston. However, throughout the rural swath of this cluster described earlier, inaccessibility to a vehicle ranges between 9% and 22% of all households. In terms of housing stress, this cluster shows consistently high percentages of households that spend more than 35% on income to meet housing costs (50.0% for renters, 32.6% for homeowners with a mortgage). This is consistent for both rural and urban tracts.

This cluster has the highest median percentage of persons over 64 with a disability (48.7%); it also has the second highest median percentage of persons with no health insurance (20.8%). It has, however, the highest median percentage of person on public health insurance (44.6%) with its range between 24.3% and 74.5%.

This cluster has a higher median volume of opioids accessible to the cluster's population than the median for the entire study area (704.1 MME per capita vs 442.9 MME per capita). This is the highest median MME volume of all the clusters. Nearly one-third of outliers in the study are present in this cluster – 52. In South Carolina, they are mostly scattered throughout the Low Country. Along the border of North Carolina, outliers in Loris and Lake View, South Carolina are spatially grouped with several outlier census tracts in North Carolina stretching from Tabor City northward to Lumberton and Elizabethtown. Elsewhere in North Carolina there is a separate grouping of outliers in New Bern as well as several small towns strung along the Interstate 40 highway between Ashville and Winston-Salem.

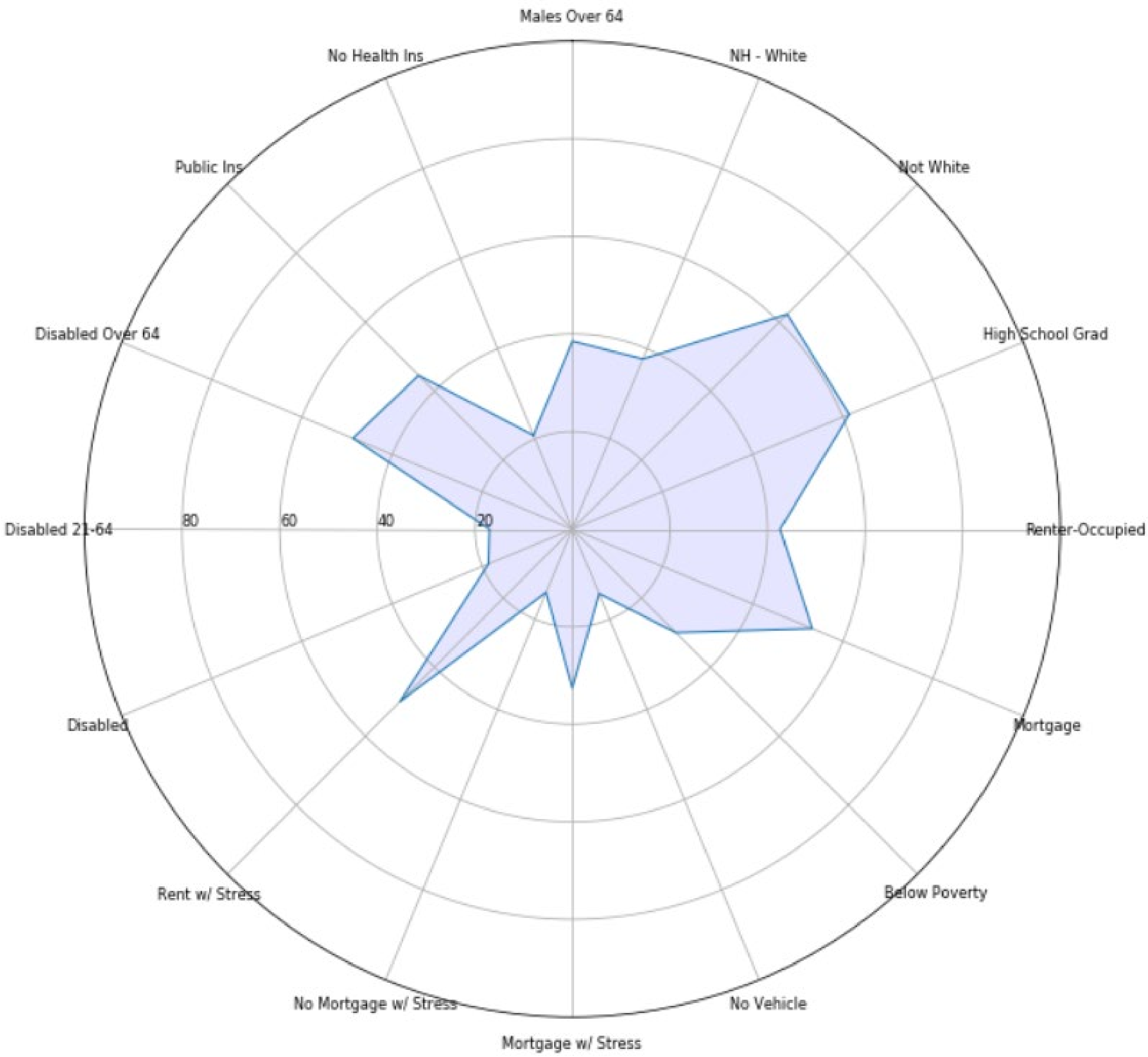

Fig. 11. Radar Plot of Cluster 7. All values listed below are median values for all census tracts in this cluster. Age in years: 38, household income: $28,768, per capita income: $15,680, persons per square mile: 455, pharmacy count: 31, and opioid MME volume: 704.1.

C. OLS Regression Model

Both forward stepwise and backward stepwise methods returned an identical criterion (BIC = 5495.856). The Bayesian model averaging method identified the mean number of regressors as 8.6 with a Posterior Model Probability (PMP) of 0.9102, although the most parsimonious model had seven regressors. The best model identified had the same eight regressors as the results of the forward and backward stepwise methods (Table VII).

TABLE VII
OLS Regression Coefficients and Significance

Independent Variable	Estimate	Standard Error	t-value	p-value
Percent Disabled	0.96	0.19	4.99	6.22e-07
Pharmacy Count	1.33	0.05	27.90	<2e-16
Percent with Public Insurance Coverage	1.19	0.18	6.71	2.24e-11
Percent Non-Hispanic White	1.32	0.07	18.24	<2e-16
Median Household Income	-0.08	0.01	-5.21	1.99e-07
Percent Housing Units with Mortgage	-1.10	0.18	-5.97	2.72e-09
Percent Housing Units Renter Occupied	0.36	0.10	3.78	0.000163
Percent Housing Units with Mortgage and Housing Costs >= 35% of Income	0.29	0.09	3.02	0.002564

The OLS regression model was checked to determine if assumptions for the model had been met. While the response variable, opioid accessibility scores, and many of the independent variables were skewed (Fig. 12), the large number of observations in the study, 3,210, meant that the assumption of normality was met. Independent variables identified as collinear (correlation coefficient > |0.7|) were removed from the model (Table III). Inconstant variance of the residuals – heteroscedasticity – is evident in the residual plot [Fig. 13(a)] and was confirmed statistically with the studentized Breusch-Pagan Test (BP = 64.364, p=2.731e-16). In an attempt to studentized the residuals, all variables, response and independent, were power transformed by lambda (λ). Lambda was calculated using the BoxCox method ($\lambda = 0.3434$). After recreating the regression model, it was tested for heteroscedasticity again. The statistical test, studentized Breusch-Pagan Test (BP = 354.17, p-value = 2.2e-16) showed no improvement in the variance of the residuals [Fig. 13(b)]. A single high-leverage outlier was identified: a census tract located in Appalachia in the western most corner of North Carolina.

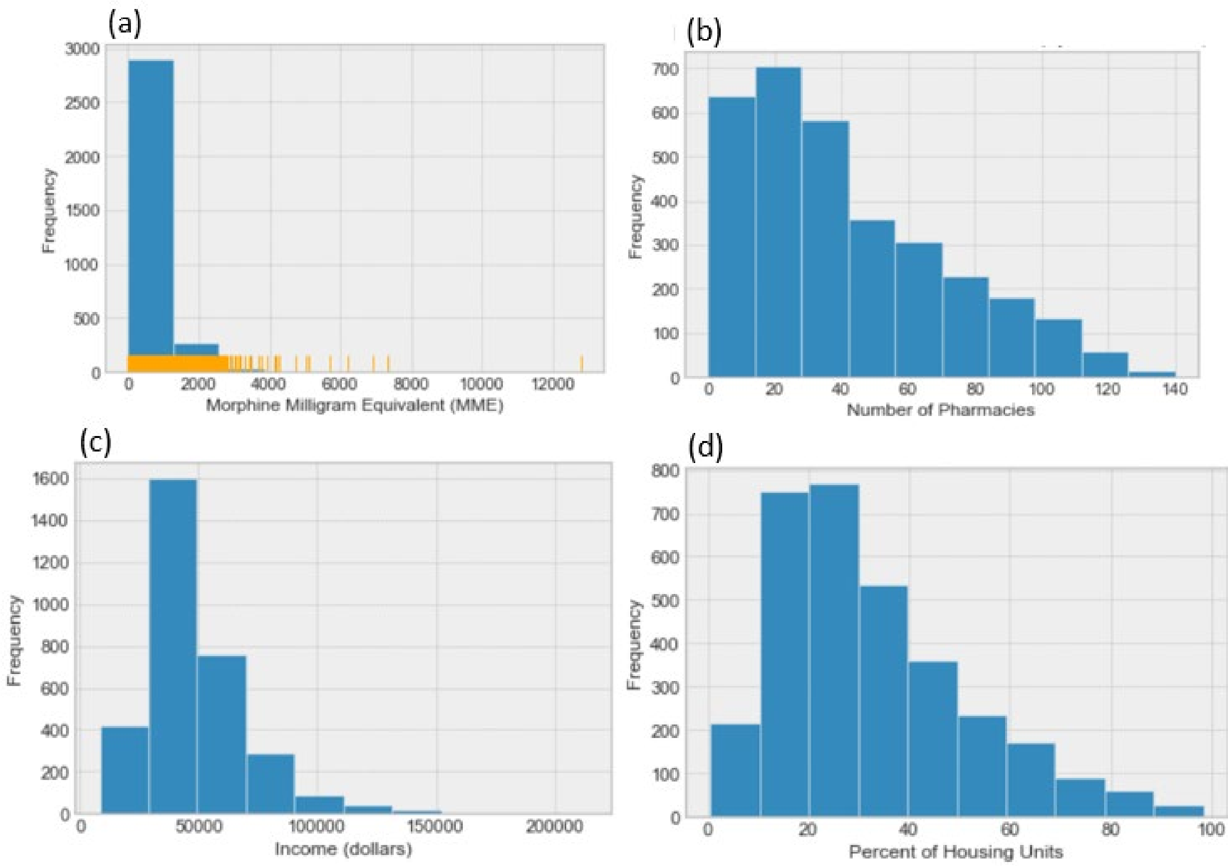

Fig. 12. Histograms of Variables Demonstrating Skewness of Data. (a) Distribution of Opioid Accessibility Scores. (b) Distribution of Accessible Pharmacies in a Census Tract Catchment Area. (c) Distribution of Median Household Income. (d) Distribution of Renter-Occupied Housing Units.

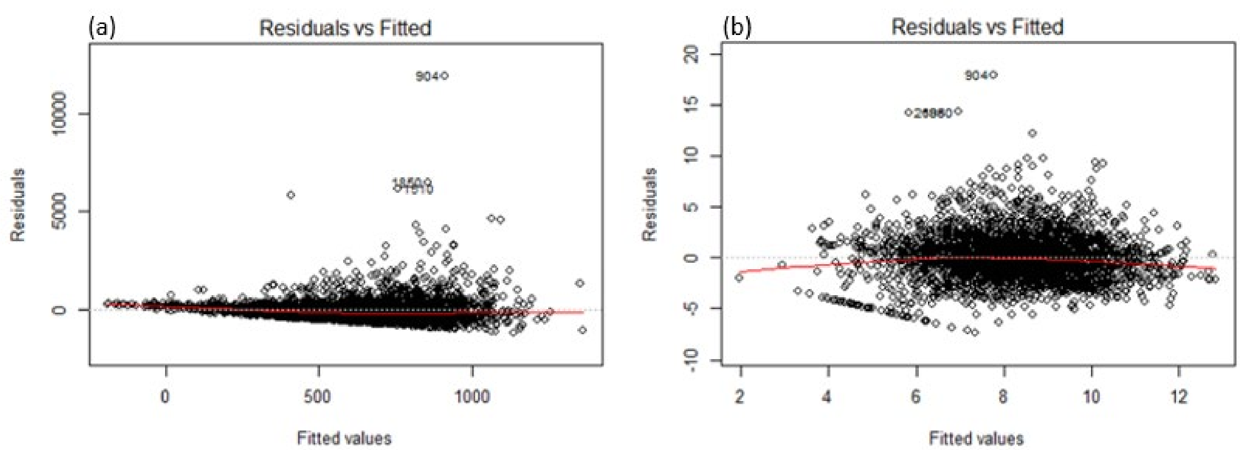

Fig. 13. Residual Plots from OLS Regression Models. (a) Original Data – Residuals vs. Fitted. The cone shaped pattern of residual values is indicative of heteroscedasticity. (b) Power Transformed Data – Residuals vs. Fitted. Heteroscedasticity has been reduced, but not removed entirely.

VI. DISCUSSION

A. *Opioid Volume Outliers*

Because pharmacy ownership changed for so many pharmacies in the study area in 2009, certain attributes like type of pharmacy ownership – chain or retail – could not be kept for any of the models. However, during the analysis of the results outlier pharmacies' type of ownership was reviewed. No outlier pharmacies changed ownership between retail and chain owners, so the rate of outliers among an ownership type can be compared to the general population. Retail pharmacies, those that are independently owned, are overrepresented in the population of outlier pharmacies (61% of outliers vs. 36% of all pharmacies).

Access to transaction-level data from the U.S. DEA's ARCOS database allows one to study the volume of medical-use opioid analgesics distributed to North Carolina and South Carolina. One approach is to look at the total volume distributed to each pharmacy and identify those locations that are outliers within that distribution. Based on the interquartile range of this distribution (Fig. 14), there are 177 pharmacies that can be considered outliers (> 6,012,687.5 MME). These outliers visually follow a few spatial patterns (Fig. 15). One group is clustered in the southeast corner of North Carolina and the northeast corner of South Carolina. There are many outliers that broadly follow the path of Interstate Highway 85 from the Virginia-North Carolina border to the Georgia-South Carolina border. Another set of outliers follow Interstate Highway 40 from Durham, North Carolina, to Asheville, North Carolina. There are a very small number of outliers that follow Interstate Highway 95 from the Virginia-North Carolina border southward to Florence, South Carolina, but they are sparsely spread out. It is worthwhile to note that in some portions of the study area, outlier pharmacies are closely aligned with major arterials of the highway system while other areas have no such evident patterns (e.g. Interstate Highways 20 and 26 in South Carolina).

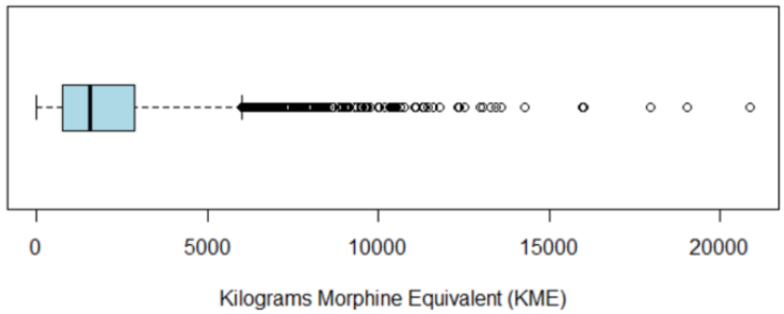

Fig. 14. Box and Whisker Plot of Total KME Delivered to Each Pharmacy in 2009. Each circle represents a pharmacy that received an outlier volume of opioid analgesics.

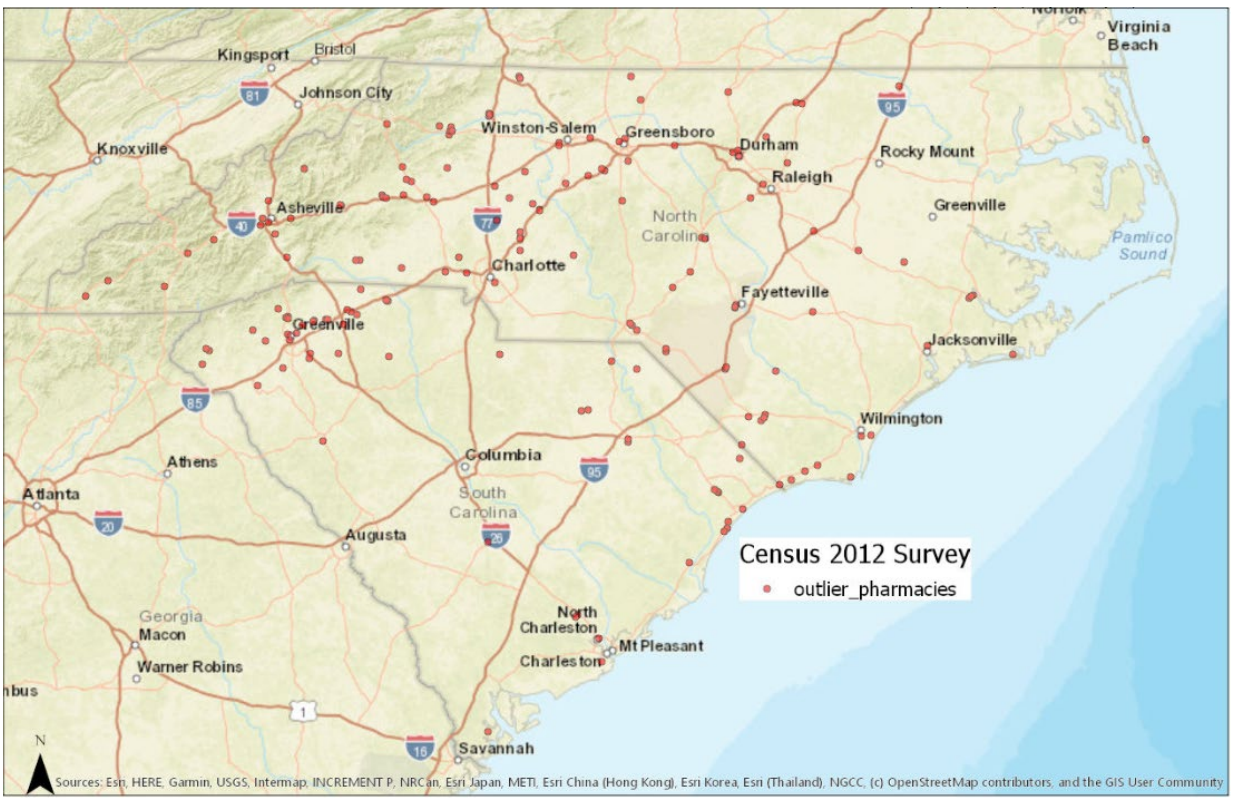

Fig. 15. Pharmacy Outlier Locations in North Carolina and South Carolina. Outlier pharmacies received in excess of 6,012,687.5 MME opioid analgesics during the year 2009.

This approach, however, only takes into consideration the location and supply of opioids without considering variation in the demand for opioids. If we assume that demand is constant throughout the population, then a greater density of people is equivalent to a greater demand.

However, the supply of opioids is relegated to specific locations, and the demand from a populace is spread out across a large geographic area. The E2SFCA method is a valid method to bridge that gap. It can be used to quantify the accessibility of a supply to a given population with the caveat that accessibility diminishes as the time required to reach a supply's location increases. When this method was applied to geocoded pharmacies located in North Carolina and South Carolina, the volume of opioids accessible to each census tract's population was calculated. Based on the interquartile range of this distribution (Fig. 16), there are 167 census tracts that can be considered outliers (> 1,602 MME per capita).

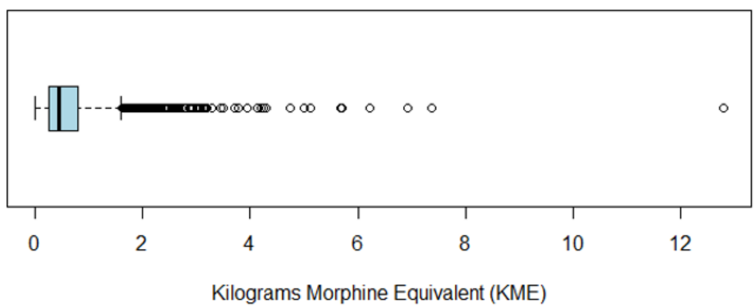

Fig. 16. Box and Whisker Plot of Census Tract Opioid Accessibility Scores. Each circle represents a census tract population that had access to an outlier volume of opioid analgesics.

There are considerable differences between the spatial distribution of outlier pharmacies and outlier census tracts (Fig. 17). One can see three types of areas with outliers: areas with only outlier pharmacies (like Greenville, Spartanburg, and Charleston in South Carolina), areas with only outlier census tracts (central South Carolina), and areas with both types of outliers (much of Appalachia, New Bern, and between Lumberton and Wilmington in North Carolina). It is only when accessibility of the population to the available volume of opioid analgesics is calculated can areas with suspect volumes of opioids be identified. For example, because there are no outlier pharmacies near the towns of Ridgeland, Allendale, and others in southwestern South Carolina (see red circle in outlier map, Fig. 17), one would be unlikely to suspect that the local populaces have an extremely high volume of opioids accessible to them (between approximately 2,400 and 6,200 MME per capita in those census tracts). It should be noted as well that what is considered extreme is somewhat arbitrary. The outliers identified in this study are based on the univariate IQR method. This does not take into consideration domain expertise of what ought to be considered extreme for any given population. Census tracts identified as outliers have accessibility scores in excess of 1,600 MME per capita. This is equivalent to 160 10-mg pills of hydrocodone or 107 10-mg pills of oxycodone per person for the year 2009 alone. In this researcher's estimation a quantity far smaller than 1,600 MME per capita could still be considered extreme. Based on whatever number that is, census tracts with extreme quantities accessible to the population can be identified.

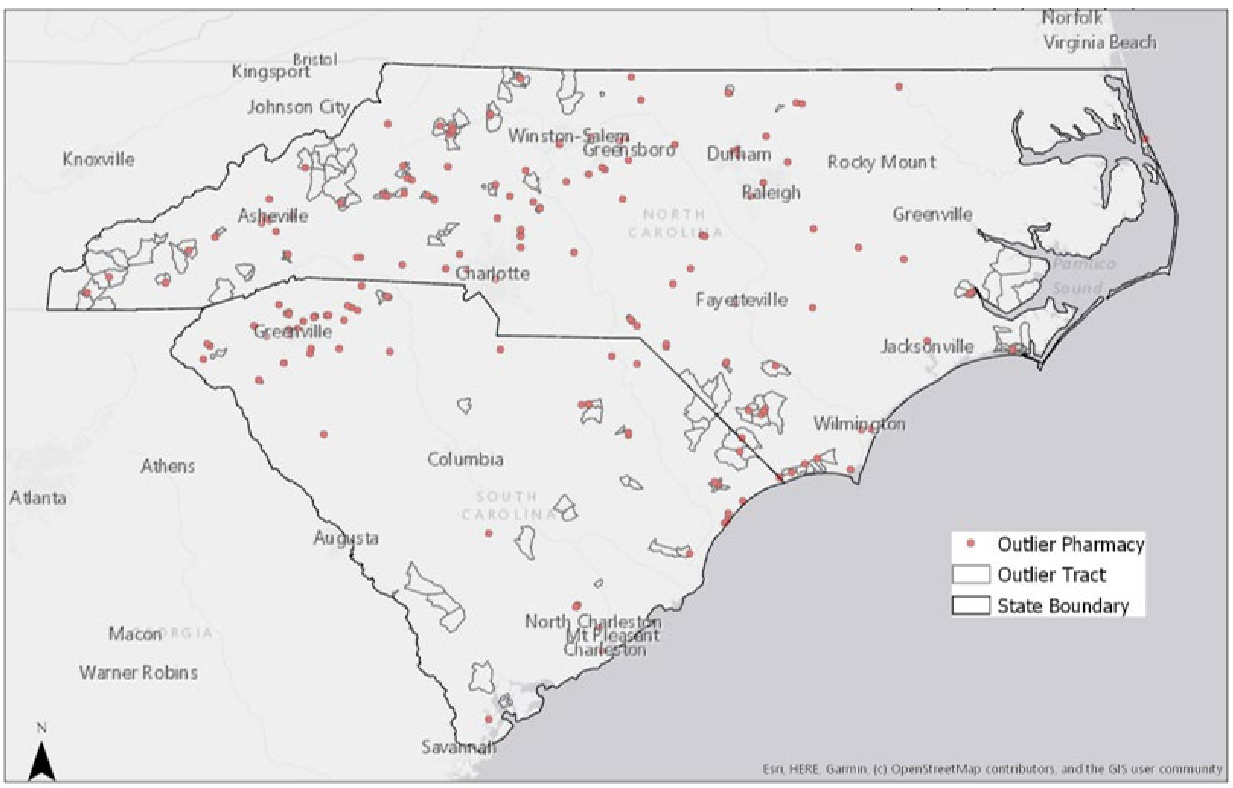

Fig. 17. Pharmacy and Census Tract Outlier Locations in North Carolina and South Carolina. Grey polygons are census tracts. Outliers were identified using the IQR outlier method. Pharmacies labeled as an outlier had at least 6,012,687.5 MME opioids distributed to it in 2009. Census tracts labeled as an outlier had an opioid accessibility score in excess of 1602 MME.

B. Comparison to Modarai et al.

Modarai et al. [3] used data from the ARCOS database that was aggregated to the 3-digit ZIP code to evaluate comparisons between opioid sales, prescription opioid-related hospitalizations, and prescription-opioid-related overdose deaths. At the end of their paper they cited the data aggregation as a limitation that reduced geographic specificity and hindered obtaining statistical significance from their spatial clustering analysis. Their recommendation to overcome this limitation was to use a technique called kriging that allows researchers to impute values from nearby values. With the availability of transactional data from the ARCOS database, geographic specificity is improved without the need to rely on any imputation methods. When comparing outlier census tracts from this study (Fig. 17) to regions of higher opioid sales (Fig. 18) in [3, Fig. 3], there is broad similarity between the two maps.[6] Modarai et al. demonstrated above average opioid sales throughout the Appalachian region and in the southeastern region around Fayetteville and Wilmington. A similar spatial pattern is reflected in this study's results with the quantity of outlier census tracts in both of those regions. Modarai et al. also saw higher than average sales in Winston-Salem. Although this study does not identify any outlier tracts in that city, there are several outliers to the north and west. Around New Bern, the data was too unreliable for their study, but in this study, there is evidence of several outliers both in New Bern as well as in the areas to the east and south along the coastline. Beyond where this study and the work of Modarai et al. [3] correspond or contradict, this study demonstrates that access to transactional level data combined with the E2SFCA method addresses the limitation that they experienced.

[6] Data visualized in [3] is from 2010. However, the authors found similar spatial patterns in the 2008 and 2009 data. "Spatial relationships existed between high rates of sales and overdoses in specific regions, particularly in the southern and western regions of the state in 2010 (similar patterns were noted in 2008 and 2009, however data are not presented)." [3, p. 82]

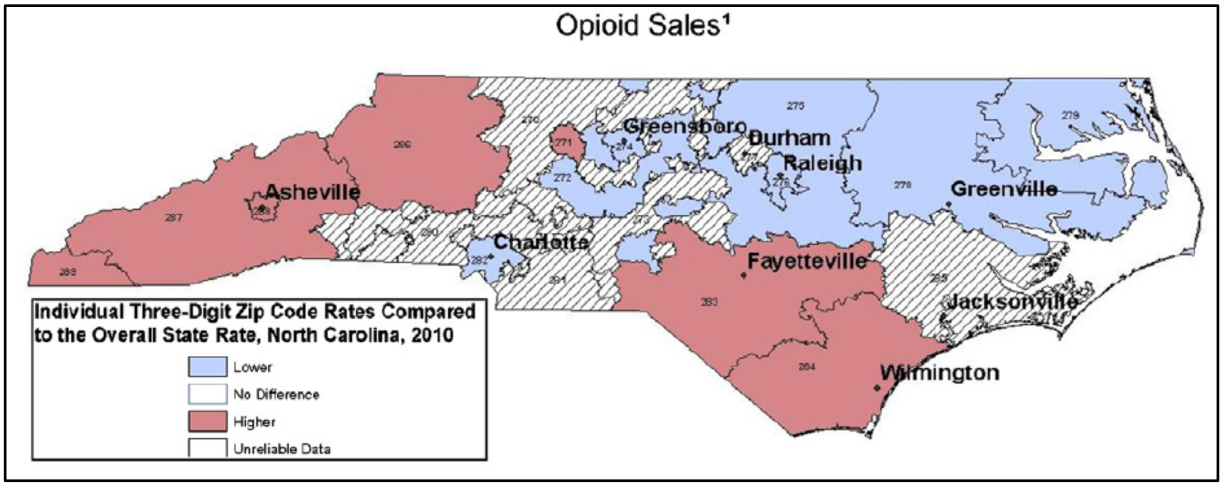

Fig. 18. Reprint of Fig. 3 from Modarai et al. [3] This figure shows relative volume of opioid sales in North Carolina during 2010 compared to the statewide average.

C. Review of Cluster Analysis

Qualitatively, the clusters produced using the K-medoids clustering method make intuitive sense given prior knowledge of the study area. There is strong racial segregation throughout large portions of both states (clusters 4, 5, and 6 vs. cluster 7); urban populations (clusters 1, 3, and 6) are easily distinguished from the most rural populations (clusters 4, 5, and 7). When one considers the medians of each cluster, many of the demographic and socioeconomic factors show strong linear correlation with increasing accessibility to opioid analgesics in line with past researchers' results (Table VIII). As has been demonstrated in previous studies [48], [58], [74], [82], there is a dichotomy between rural and urban areas in terms of accessibility of opioids. Rural census tracts consistently have a much higher median opioid accessibility score than their urban census tract counterparts (Table IX).

TABLE VIII
Bivariate fits with MME Accessibility Scores of Clusters. Asterisks indicative of strength of statistical significance.

Variable	Adjusted R^2	Coefficient	Standard Error	P-value
Educational Attainment: Percent of population with High School degree or less	0.62	7.85	2.72	0.0344
Veteran Status: Percent of civilian 18 and older population that is a veteran	-0.20	-6.01	45.47	0.8999
Disability: Percent of total population with a disability	0.70	28.11	7.21	0.0114
Disability: Percent of under 18 population with a disability	0.68	139.17	37.39	0.0137
Disability: Percent of 18 to 64 population with a disability	0.69	27.35	7.25	0.0130
Disability: Percent of over 65 population with a disability	0.80	18.75	3.70	0.0039
Place of Birth: Percent of population foreign born	-0.08	-10.95	14.96	0.4971
Median household income	0.51	-0.01	0.002	0.0427
Per capita income	0.46	-0.014	0.01	0.0550
Health Insurance: Percent of population with public health insurance	0.76	14.90	3.33	0.0066
Health Insurance: Percent of population with no health insurance (public or private)	0.25	15.10	8.65	0.1414
Poverty: Percent of population whose income in last 12 months was below the poverty level	0.50	12.80	4.87	0.0468
Housing: Percent of housing units that are renter-occupied	-0.03	3.59	3.90	0.3992
Housing: Percent of housing units with no vehicle available	0.51	25.49	9.45	0.0429
Housing: Percent of owner-occupied housing units with a mortgage	0.59	-11.31	3.63	0.0265
Housing: Percent of housing units with a mortgage in which housing costs are 35% or more of income	0.88	28.97	4.37	0.0012**
Housing: Percent of housing units without a mortgage in which housing costs are 35% or more of income	0.75	56.40	12.94	0.0073
Housing: Percent of rented housing units with rental costs 35% or more of income	0.67	20.02	6.23	0.0236
Median age in years	-0.16	6.85	15.71	0.6807
Sex & Age: Percent of population 65 and over that is male	0.14	-30.81	21.72	0.2152
Sex & Age: Percent of population 18 to 64 that is male	-0.19	-16.06	75.84	0.8406
Race & Ethnicity: Percent population Hispanic or Latino – all races	-0.20	-0.99	12.21	0.9388
Race & Ethnicity: Percent population not Hispanic or Latino – white only	0.01	-3.05	3.00	0.3555
Race & Ethnicity: Percent population not Hispanic or Latino – black only	0.15	5.00	3.51	0.2144
Race & Ethnicity: Percent population not Hispanic or Latino – all other races	0.17	-63.59	42.43	0.1942

TABLE IX
Breakdown of Cluster Descriptive Statistics by Rural and Urban Residence

Cluster	% Urban	Urban Pharma Count	Rural Pharma Count	Urban MME	Rural MME	Urban Pop Density	Rural Pop Density	Urban Disability	Rural Disability
1	90.52	75	93	312.67	1046.04	1793.85	332.60	8	13.1
2	90.36	33	25	454.09	868.63	897.99	378.43	11.5	12.4
3	93.81	72	63	465.96	817.62	2884.45	856.34	11.4	14.2
4	71.57	13	26	269.78	694.22	150.21	104.80	14.4	15.4
5	53.50	13	21	426.03	841.29	138.93	80.83	19.2	20.5
6	96.79	57	25	279.73	634.59	1218.21	220.50	7.1	12.2
7	59.14	31	31	558.71	858.37	959.44	169.32	17.9	19.3

Cluster 7 produced unexpected results; it had the highest median opioid accessibility score, but it also represents the predominantly rural African American population of both states. These findings are contrary to studies demonstrating that African Americans receive opioid analgesic prescriptions at far lower rates than non-Hispanic Whites [41], [83]. Qato et al. [47] found disparities between segregated African American communities and segregated White or integrated communities when it comes to access to pharmacies. Several surveys of racial disparity in pain management contemporaneous with this study's data consistently identified African American communities as suffering from more pain and receiving less treatment in the form of opioid analgesics [84]–[86]. So, there must be something else going on in this cluster – some other combination of factors that drove up the accessibility scores. It could be a combination of other factors that result in higher accessibility regardless of the high percentages of African Americans in the population. A combination of high disability, high public insurance usage, and low percent of males over 64 could suggest that there are many individuals in this cluster who are participants in Medicare Part D. Powell et al. [21] found that Part D expanded the usage of prescription drugs, including opioid analgesics, in states with larger elderly populations. Other studies have noted that females, especially elderly females, are more likely to be prescribed opioid analgesics than their male counterparts [5], [43]. This cluster has the highest median percent usage (44.55%) of public insurance which includes Medicare Part D, the federal prescription drug coverage plan. It also has the lowest median percentage of males in the 65 and over age category (38.5%) indicating that the elderly population has a much higher proportion of females than the other clusters. However, this cluster has one of the younger median ages (38), and unlike most other clusters, cluster 7 has no census tracts with truly high median age – the maximum median age for any tract is 55. This means that a relatively younger population is making use of public insurance plans at much higher rates than older populations in other clusters.

When considering the unexpectedly high median opioid accessibility score for cluster 7, it is possible that it is a result of the proximity of that cluster with a large portion of cluster 5. While the majority of cluster 7 is confined to the Low Country of the Carolinas, cluster 5 is split between the same area of cluster 7 and the Appalachian region. Given the known racial disparities in pain management [83]–[85], [87], it is possible that the proximity of many census tracts in these two clusters increased opioid accessibility scores in cluster 7 tracts. Sorting through whether this may have occurred is beyond the scope of this research. This may be better understood utilizing additional data on physician prescribing practices and on details of who received prescriptions [43], [88]. Additionally, a spatial cluster analysis to test spatial autocorrelation between the two clusters may also provide insight.

D. Review of OLS Regression Model

This study's OLS regression model variates have similarities to many studies that have also examined opioid analgesic volumes and demographic and socioeconomic variation. Other studies have identified correlations between available opioid analgesic volumes and prescription rates per capita and rates of public insurance beneficiaries [5], [21], [31], [33], non-Hispanic Whites [33],

[37], [41], density of nearby pharmacies or pharmacists [88], proportions of populations with a disability [33], and economically disadvantaged neighborhoods [5]. The renter-occupied housing units variable has a 63% correlation with the percent poverty variable; to a limited extent it can be considered a proxy for poverty which has been correlated in past studies to increasing accessibility to opioids [5]. The variable percent of housing units that are owner-occupied with a mortgage does not match up with the results from any studies in this study's literature review. There were also some surprises about variables that were dropped out by the stepwise and BMA methods used to choose the most parsimonious model. higher poverty rates [5], [43] and lack of health insurance [43] have been associated with greater access to opioids, but these variables were not included in the final OLS regression model. In fact, they were only identified as useful in a very small percentage of models generated using the BMA method. Many variables chosen for this study were collinear. This may have played a role in the variable selection process. As stated above, the poverty rate variable is correlated 63% with renter-occupied housing units, and percent with no health insurance coverage is correlated with lower educational attainment (60%) and median household income (62%). In each case, a dropped variable that is relevant based on past studies is more strongly correlated (>60%) with a variable that remained in the model.

With the high-leverage outlier removed, the OLS regression model produced an adjusted R^2 value of 0.3084 meaning that nearly 70% of all the variation of the opioid MME accessibility scores is accounted for by factors other than the eight variables included in the model. Other researchers have found similarly low percentages of explanation of their response variables [33], [43]. There are several factors that were not part of the current study that may account for some of the remaining variability. McDonald et al. [43] studied geographic variation of volumes of opioid prescribing; they found the strongest correlation to opioid prescribing volumes to be the numbers of prescribing physicians within a county. In Guy et al. [33], their linear regression model identified higher rates of unemployment led to higher rates of opioid analgesic prescribing at the county level. In retrospect, including unemployment data from the ACS may have had a beneficial impact on this study's OLS regression model. This study also considered age groups (18-64 and over 64) coupled with sex. However, other studies [1], [33] have found links between increased availability and usage based on age which this study did not consider.

An alternative factor for a low R^2 value is that it is possible that the geographic unit used in this study – the census tract – has introduced a scale problem that resulted in a diminished R^2 value. The Modifiable Areal Unit Problem (MAUP) can occur when spatial aggregation occurs. Data aggregated to larger geographical units, like 3-digit ZIP codes or counties, has a smoothing effect on variability. Aggregated data mask extreme data values as it is represented by measurements of its central tendency [89].

One of the benefits of producing a linear regression model is that it can be used to infer outcomes when the response variable is unknown. However, to do so, the model relies on assumptions that support its mathematical framework. One of those assumptions is the normal distribution of the error terms – known as homoscedasticity. In this study, the residuals were found to be heteroscedastic, and efforts to remove the heteroscedasticity via a power transform of response and independent variables was not successful. Because the homoscedastic assumption could not be met, it seems evident that a linear regression model that relies on this assumption is not ideal for modeling this data. Any future researcher attempting to model the data should examine non-parametric models that do not rely on this assumption as a possible candidate.

VII. Limitations And Future Work

This study explored the uses of the E2SFCA methodology to identify variation of opioid accessibility at the level of the census tract. The family of FCA methodologies have experienced iterations of improvements by a variety of researchers. There are recommendations for tweaks in this study's E2SFCA methodology for future researchers. Boundaries can be physical and real like a coastline, but they can also be artificial. State boundaries are political creations that do not affect the movement of persons. This study limited ARCOS data to the states of North Carolina and South Carolina. This potentially limited opioid supplies from pharmacies in bordering states. It is recommended that future studies include a buffer of ARCOS data from neighboring states to be included for the creation of catchment areas. Buffer size should be guided by future studies' catchment area sizes. This study generated geographic centroids of census tracts as part of the E2SFCA methodology. Rural census tracts are typically much larger than urban census tracts and their geographic centroids may not correspond well with its population center. During routine inspection of catchment areas, some examples were found in which the distance of a large census tract's centroid from the road network may have impacted opioid accessibility calculations. It is recommended that future studies that use the E2SFCA methodology calculate population-weighted centroids of census tracts.

This study's methodology included k-medoids cluster analysis and OLS regression based on data from the ACS and the ARCOS database. The variables chosen were strictly non-spatial data. However, there is a rich array of models available to researchers that can combine non-spatial data with spatial data. Given the amount of variability of opioid accessibility unexplained in the OLS regression model, it would be worth exploring spatial regression algorithms to determine how much more variability is explained once the spatial nature of the data has been incorporated into the model. Additionally, the high opioid accessibility scores assigned to census tracts in cluster 7 were unexpected and contrary to previous research. Exploring the spatial relationship between clusters 7 and 5, as discussed in the results section with spatial correlation tools like LISA and Global Moran I may provide some insight into those results for future researchers.

Variable selection for this study was guided by the literature review. Based on the results of the OLS regression model it is evident that there are other factors that could help to provide additional explanation of the variability of opioid accessibility scores. One possible additional variable is the percentage of workforce population that is unemployed. This variable is available in the 5-year average ACS data. There are other variables that previous studies have found significant in explaining variation of opioid accessibility: the number of prescribing physicians available to a population [43], differences in prescribing patterns by diagnosed condition and by type of physician specialty [17], and variation of opioid usage by type of opioid [43]. The first two of these variables required accessing databases not available for this study and so was considered out of scope. The ARCOS data available from *The Washington Post* includes two opioid analgesics: oxycodone and hydrocodone. This study's methodology grouped all types of opioid drugs together, however given the amount of variation not explained in the OLS regression model and challenges related to heteroscedastic variance, exploring variation of opioid accessibility based on opioid drug type may provide fruitful results. Future researchers may also want to consider models that do not have the assumption of homoscedasticity. For example, a logistic regression model may be a preferable alternative to a linear regression model.

VIII. CONCLUSION

The abuse of opioids continues to inflict harm on the lives of Americans. Opioids accounted for 70% of the 46,802 drug overdose deaths in 2018. While in recent years the synthetic opiate fentanyl has been the primary cause of opioid-related overdose deaths and continues to be a major concern, the good news is that overdose deaths related to prescription opioids has been declining. Nationally, there was a 13.5% relative decrease in prescription opioid-related deaths between 2017 and 2018. As has been discussed in this paper, however, there is always significant variation evident when data is explored at smaller geographic units. North Carolina experienced a 27.7% relative decrease in deaths between 2017 and 2018 while South Carolina experienced a 4.2% relative increase in deaths during the same time period. While the overall decline is good news, there were still over 17,000 deaths related to prescription opioid abuse [90]. Clearly, more work is needed to prevent future deaths.

The publication of data from the U.S. DEA's ARCOS database in 2019 provided an opportunity for researchers to gain a far more nuanced and detailed understanding of how opioid analgesics were distributed to pharmacies across the country. The goals of this study were to engage in an initial exploration of the data for North Carolina and South Carolina, assess whether more granular data could capture greater detail of variability in the distribution of opioids in comparison to the research done in Modarai et al. [3] in 2009, and to model the relationship between demographic characteristics of the population and opioid volumes accessible to that population. In order to achieve these goals, this study used the E2SFCA method advocated by Luo and Qi [30] to measure healthcare accessibility on a per capita basis. Based on this study's literature review, using the E2SFCA method to measure prescription opioid accessibility has not been used in previous studies. In comparing the results to Modarai et al. [3], this study has demonstrated that the E2SFCA method can successfully transform point-based data from the ARCOS database to areal data that can be used to reveal spatial variability in smaller geographic units than past studies have been able to achieve.

Finally, while this study produced population clusters from the k-medoids clustering algorithm that matched much of past literature in terms of opioid volume accessibility, the nature of cluster 7 leaves some questions unanswered. Cluster 7 represented a rural, socially vulnerable, predominantly African American population that had access to the highest volumes of opioid analgesics and included 31% of all outlier census tracts but less than 17% of all census tracts in the study area. Given the findings in this study's literature review that African Americans lacked access to opioids for pain management, this result was surprising. However, cluster 7 is geographically intertwined with cluster 5 which represents a rural, socially vulnerable, predominantly non-Hispanic White population with a similar overrepresentation of outlier census tracts in its cluster. An obvious question is whether the proximity of these two clusters inflated the opioid accessibility scores for the African American population against expectations or if there are other factors at play that are more heavily influencing opioid volumes in these areas. It could be that working with spatial correlation tools might provide some insight into the spatial relationship between these two clustered populations.

REFERENCES

[1] N. B. King, V. Fraser, C. Boikos, R. Richardson, and S. Harper, "Determinants of Increased Opioid-Related Mortality in the United States and Canada, 1990–2013: A Systematic Review," *Am J Public Health*, vol. 104, no. 8, pp. e32–e42, Aug. 2014, doi: 10.2105/AJPH.2014.301966.

[2] S. Higham, S. Horwitz, and S. Rich, "76 billion opioid pills: Newly released federal data unmasks the epidemic," *Washington Post*. https://www.washingtonpost.com/investigations/76-billion-opioid-pills-newly-released-federal-data-unmasks-the-epidemic/2019/07/16/5f29fd62-a73e-11e9-86dd-d7f0e60391e9_story.html (accessed Feb. 22, 2020).

[3] F. Modarai *et al.*, "Relationship of opioid prescription sales and overdoses, North Carolina," *Drug and Alcohol Dependence*, vol. 132, no. 1–2, pp. 81–86, Sep. 2013, doi: 10.1016/j.drugalcdep.2013.01.006.

[4] "ARCOS Retail Drug Summary Reports." https://www.deadiversion.usdoj.gov/arcos/retail_drug_summary/index.html (accessed Feb. 23, 2020).

[5] A. Agnoli, A. Jerant, W. Becker, and P. Franks, "Opioid Prescriptions and Short-Term Mortality: a U.S. National Study," *J GEN INTERN MED*, Oct. 2019, doi: 10.1007/s11606-019-05501-w.

[6] M. A. Brandenburg, "Prescription Opioids Are Associated With Population Mortality in US Deep South Middle-Age Non-hispanic Whites: An Ecological Time Series Study," *Front. Public Health*, vol. 7, p. 252, Sep. 2019, doi: 10.3389/fpubh.2019.00252.

[7] L. J. Paulozzi, D. S. Budnitz, and Y. Xi, "Increasing deaths from opioid analgesics in the United States," *Pharmacoepidem. Drug Safe.*, vol. 15, no. 9, pp. 618–627, Sep. 2006, doi: 10.1002/pds.1276.

[8] R. K. Portenoy and K. M. Foley, "Chronic use of opioid analgesics in non-malignant pain: Report of 38 cases:," *Pain*, vol. 25, no. 2, pp. 171–186, May 1986, doi: 10.1016/0304-3959(86)90091-6.

[9] A. E. Alpert, W. N. Evans, E. M. J. Lieber, and D. Powell, "Origins of the Opioid Crisis and its Enduring Impacts," National Bureau of Economic Research, 26500, Nov. 2019. Accessed: Jan. 17, 2020. [Online]. Available: http://www.nber.org/papers/w26500.

[10] "Vital Signs (Body Temperature, Pulse Rate, Respiration Rate, Blood Pressure)." https://www.hopkinsmedicine.org/health/conditions-and-diseases/vital-signs-body-temperature-pulse-rate-respiration-rate-blood-pressure (accessed Feb. 19, 2020).

[11] W. N. Evans, E. M. J. Lieber, and P. Power, "How the Reformulation of OxyContin Ignited the Heroin Epidemic," *The Review of Economics and Statistics*, vol. 101, no. 1, pp. 1–15, Mar. 2019, doi: 10.1162/rest_a_00755.

[12] T. Lyapustina and G. C. Alexander, "The prescription opioid addiction and abuse epidemic: how it happened and what we can do about it," *The Pharmaceutical Journal*, 2015, doi: 10.1211/PJ.2015.20068579.

[13] B. Meier, "Origins of an Epidemic: Purdue Pharma Knew Its Opioids Were Widely Abused.," *The New York Times*, New York, NY, May 29, 2018.

[14] A. Van Zee, "The Promotion and Marketing of OxyContin: Commercial Triumph, Public Health Tragedy," *Am J Public Health*, vol. 99, no. 2, pp. 221–227, Feb. 2009, doi: 10.2105/AJPH.2007.131714.

[15] "U.S. Opioid Prescribing Rate Maps | Drug Overdose | CDC Injury Center," Mar. 12, 2020. https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html (accessed Mar. 26, 2020).
[16] K. Kenan, K. Mack, and L. Paulozzi, "Trends in prescriptions for oxycodone and other commonly used opioids in the United States, 2000-2010," *Open Med*, vol. 6, no. 2, pp. e41-47, 2012.
[17] B. Levy, L. Paulozzi, K. A. Mack, and C. M. Jones, "Trends in Opioid Analgesic–Prescribing Rates by Specialty, U.S., 2007–2012," *American Journal of Preventive Medicine*, vol. 49, no. 3, pp. 409–413, Sep. 2015, doi: 10.1016/j.amepre.2015.02.020.
[18] "U.S. State Prescribing Rates, 2009." https://www.cdc.gov/drugoverdose/maps /rxstate2009.html (accessed Feb. 22, 2020).
[19] "U.S. County Prescribing Rates, 2009." https://www.cdc.gov/drugoverdose/maps /rxcounty2009.html (accessed Feb. 22, 2020).
[20] L. Sanders, "Opioids kill. Here's how an overdose shuts down your body." https://www.sciencenews.org/article/opioid-crisis-overdose-death (accessed Feb. 19, 2020).
[21] D. Powell, R. L. Pacula, and E. Taylor, "How Increasing Medical Access to Opioids Contributes to the Opioid Epidemic: Evidence from Medicare Part D," National Bureau of Economic Research, Cambridge, MA, w21072, Apr. 2015. doi: 10.3386/w21072.
[22] R. Ghertner, "U.S. county prevalence of retail prescription opioid sales and opioid-related hospitalizations from 2011 to 2014," *Drug and Alcohol Dependence*, vol. 194, pp. 330–335, Jan. 2019, doi: 10.1016/j.drugalcdep.2018.10.031.
[23] N. I. on D. Abuse, "North Carolina Opioid Summary," Mar. 30, 2019. https://www .drugabuse.gov/opioid-summaries-by-state/north-carolina-opioid-summary (accessed Mar. 20, 2020).
[24] N. I. on D. Abuse, "South Carolina Opioid Summary," Mar. 29, 2019. https://www .drugabuse.gov/opioid-summaries-by-state/south-carolina-opioid-summary (accessed Mar. 20, 2020).
[25] "U.S. State Prescribing Rates, 2018 | Drug Overdose | CDC Injury Center," Mar. 09, 2020. https://www.cdc.gov/drugoverdose/maps/rxstate2018.html (accessed Mar. 26, 2020).
[26] "Controlled Substance Schedules." https://www.deadiversion.usdoj.gov/schedules/ (accessed Feb. 23, 2020).
[27] "ARCOS Retail Drug Summary Report - 2009 Reporting Period." https://www .deadiversion.usdoj.gov/arcos/retail_drug_summary/2009/index.html (accessed Feb. 22, 2020).
[28] S. Rich, S. Higham, and S. Horwitz, "More than 100 billion pain pills saturated the nation over nine years," *Washington Post*. https://www.washingtonpost.com/investigations/more-than-100-billion-pain-pills-saturated-the-nation-over-nine-years/2020/01/14/fde320ba-db13-11e9-a688-303693fb4b0b_story.html (accessed Feb. 22, 2020).
[29] S. Rich, M. S. Díez, and K. Vongkiatkajorn, "How to download and use the DEA pain pills database," *Washington Post*. https://www.washingtonpost.com/national/2019/07/18/how-download-use-dea-pain-pills-database/?arc404=true (accessed Feb. 9, 2020).
[30] W. Luo and Y. Qi, "An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians," *Health & Place*, vol. 15, no. 4, pp. 1100–1107, Dec. 2009, doi: 10.1016/j.healthplace.2009.06.002.

[31] A. Basak, J. Cadena, A. Marathe, and A. Vullikanti, “Detection of Spatiotemporal Prescription Opioid Hot Spots With Network Scan Statistics: Multistate Analysis,” *JMIR Public Health Surveill*, vol. 5, no. 2, p. e12110, Jun. 2019, doi: 10.2196/12110.
[32] B. J. Piper, D. T. Shah, O. M. Simoyan, K. L. McCall, and S. D. Nichols, “Trends in Medical Use of Opioids in the U.S., 2006–2016,” *American Journal of Preventive Medicine*, vol. 54, no. 5, pp. 652–660, May 2018, doi: 10.1016/j.amepre.2018.01.034.
[33] G. P. Guy *et al.*, “Vital Signs: Changes in Opioid Prescribing in the United States, 2006–2015,” *MMWR Morb Mortal Wkly Rep*, vol. 66, no. 26, pp. 697–704, Jul. 2017, doi: 10.15585/mmwr.mm6626a4.
[34] T. L. Anderson, X. Zhang, S. S. Martin, Y. Fang, and J. Li, “Understanding Differences in Types of Opioid Prescriptions Across Time and Space: A Community-Level Analysis,” *Journal of Drug Issues*, vol. 49, no. 2, pp. 405–418, Apr. 2019, doi: 10.1177/0022042618815687.
[35] Centers for Disease Control and Prevention, “Vital signs: overdoses of prescription opioid pain relievers—United States, 1999–2008.,” *Morbidity and Mortality Weekly Report*, vol. 60, no. 43, pp. 1487–1492, Nov. 2011.
[36] P. L. Marotta, T. Hunt, L. Gilbert, E. Wu, D. Goddard-Eckrich, and N. El-Bassel, “Assessing Spatial Relationships between Prescription Drugs, Race, and Overdose in New York State from 2013 to 2015,” *Journal of Psychoactive Drugs*, vol. 51, no. 4, pp. 360–370, Aug. 2019, doi: 10.1080/02791072.2019.1599472.
[37] A. S. B. Bohnert *et al.*, “Association Between Opioid Prescribing Patterns and Opioid Overdose-Related Deaths,” *JAMA*, vol. 305, no. 13, p. 1315, Apr. 2011, doi: 10.1001/jama.2011.370.
[38] D. Nitcheva, “Opioid Deaths in South Carolina,” South Carolina Department of Health and Environmental Control.
[39] C. A. Schalkoff *et al.*, “The opioid and related drug epidemics in rural Appalachia: A systematic review of populations affected, risk factors, and infectious diseases,” *Substance Abuse*, vol. 41, no. 1, pp. 35–69, Jan. 2020, doi: 10.1080/08897077.2019.1635555.
[40] E. R. Schoenfeld *et al.*, “Geographic, Temporal, and Sociodemographic Differences in Opioid Poisoning,” *American Journal of Preventive Medicine*, vol. 57, no. 2, pp. 153–164, Aug. 2019, doi: 10.1016/j.amepre.2019.03.020.
[41] M. M. Ali *et al.*, “Factors associated with potentially problematic opioid prescriptions among individuals with private insurance and medicaid,” *Addictive Behaviors*, vol. 98, p. 106016, Nov. 2019, doi: 10.1016/j.addbeh.2019.06.005.
[42] L. J. Paulozzi *et al.*, “A History of Being Prescribed Controlled Substances and Risk of Drug Overdose Death,” *Pain Med*, vol. 13, no. 1, pp. 87–95, Jan. 2012, doi: 10.1111/j.1526-4637.2011.01260.x.
[43] D. C. McDonald, K. Carlson, and D. Izrael, “Geographic Variation in Opioid Prescribing in the U.S.,” *The Journal of Pain*, vol. 13, no. 10, pp. 988–996, Oct. 2012, doi: 10.1016/j.jpain.2012.07.007.
[44] L. M. Rossen, D. Khan, and M. Warner, “Hot spots in mortality from drug poisoning in the United States, 2007–2009,” *Health & Place*, vol. 26, pp. 14–20, Mar. 2014, doi: 10.1016/j.healthplace.2013.11.005.
[45] S. Z. Ikram, Y. Hu, and F. Wang, “Disparities in Spatial Accessibility of Pharmacies in Baton Rouge, Louisiana,” *Geographical Review*, vol. 105, no. 4, pp. 492–510, Oct. 2015, doi: 10.1111/j.1931-0846.2015.12087.x.

[46] S. T. Syed, B. S. Gerber, and L. K. Sharp, “Traveling Towards Disease: Transportation Barriers to Health Care Access,” *J Community Health*, vol. 38, no. 5, pp. 976–993, Oct. 2013, doi: 10.1007/s10900-013-9681-1.

[47] D. M. Qato, M. L. Daviglus, J. Wilder, T. Lee, D. Qato, and B. Lambert, “‘Pharmacy Deserts’ Are Prevalent In Chicago’s Predominantly Minority Communities, Raising Medication Access Concerns,” *Health Affairs*, vol. 33, no. 11, pp. 1958–1965, Nov. 2014, doi: 10.1377/hlthaff.2013.1397.

[48] W. E. Zahnd, S. L. McLafferty, R. L. Sherman, H. Klonoff-Cohen, S. Farner, and K. A. Rosenblatt, “Spatial Accessibility to Mammography Services in the Lower Mississippi Delta Region States,” *The Journal of Rural Health*, vol. 35, no. 4, pp. 550–559, Sep. 2019, doi: 10.1111/jrh.12349.

[49] M. M. Casey, J. Klingner, and I. Moscovice, “Pharmacy Services in Rural Areas: Is the Problem Geographic Access or Financial Access?,” *J Rural Health*, vol. 18, no. 3, pp. 467–477, Jun. 2002, doi: 10.1111/j.1748-0361.2002.tb00911.x.

[50] E. B. Lamont *et al.*, “Is Patient Travel Distance Associated With Survival on Phase II Clinical Trials in Oncology?,” *JNCI Journal of the National Cancer Institute*, vol. 95, no. 18, pp. 1370–1375, Sep. 2003, doi: 10.1093/jnci/djg035.

[51] B. Littenberg, K. Strauss, C. D. MacLean, and A. R. Troy, “The use of insulin declines as patients live farther from their source of care: results of a survey of adults with type 2 diabetes,” *BMC Public Health*, vol. 6, no. 1, p. 198, Dec. 2006, doi: 10.1186/1471-2458-6-198.

[52] K. Strauss, C. MacLean, A. Troy, and B. Littenberg, “Driving distance as a barrier to glycemic control in diabetes,” *J Gen Intern Med*, vol. 21, no. 4, pp. 378–380, Apr. 2006, doi: 10.1111/j.1525-1497.2006.00386.x.

[53] J. C. Probst, S. B. Laditka, J.-Y. Wang, and A. O. Johnson, “Effects of residence and race on burden of travel for care: cross sectional analysis of the 2001 US National Household Travel Survey,” *BMC Health Serv Res*, vol. 7, no. 1, p. 40, Dec. 2007, doi: 10.1186/1472-6963-7-40.

[54] N. Schuurman, R. S. Fiedler, S. C. Grzybowski, and D. Grund, “Defining rational hospital catchments for non-urban areas based on travel-time,” *Int J Health Geogr*, vol. 5, no. 1, p. 43, 2006, doi: 10.1186/1476-072X-5-43.

[55] E. K. Cromley and S. McLafferty, *GIS and public health*, 2nd ed. New York: The Guilford Press, 2012.

[56] J. Donohoe, V. Marshall, X. Tan, F. T. Camacho, R. Anderson, and R. Balkrishnan, “Evaluating and comparing methods for measuring spatial access to mammography centers in Appalachia,” *Health Serv Outcomes Res Method*, vol. 16, no. 1–2, pp. 22–40, Jun. 2016, doi: 10.1007/s10742-016-0143-y.

[57] M. R. McGrail, “Spatial accessibility of primary health care utilising the two step floating catchment area method: an assessment of recent improvements,” *Int J Health Geogr*, vol. 11, no. 1, p. 50, 2012, doi: 10.1186/1476-072X-11-50.

[58] M. R. McGrail and J. S. Humphreys, “Measuring spatial accessibility to primary care in rural areas: Improving the effectiveness of the two-step floating catchment area method,” *Applied Geography*, vol. 29, no. 4, pp. 533–541, Dec. 2009, doi: 10.1016/j.apgeog.2008.12.003.

[59] F. Wang, *Quantitative methods and socio-economic applications in GIS*, Second edition. Boca Raton: CRC Press, 2015.

[60] J. W. Weibull, "An axiomatic approach to the measurement of accessibility," *Regional Science and Urban Economics*, vol. 6, no. 4, pp. 357–379, Dec. 1976, doi: 10.1016/0166-0462(76)90031-4.
[61] S. E. Russell and C. P. Heidkamp, "'Food desertification': The loss of a major supermarket in New Haven, Connecticut," *Applied Geography*, vol. 31, no. 4, pp. 1197–1209, Oct. 2011, doi: 10.1016/j.apgeog.2011.01.010.
[62] C. D. Yeager and J. D. Gatrell, "Rural food accessibility: An analysis of travel impedance and the risk of potential grocery closures," *Applied Geography*, vol. 53, pp. 1–10, Sep. 2014, doi: 10.1016/j.apgeog.2014.05.018.
[63] S. Farber, M. Z. Morang, and M. J. Widener, "Temporal variability in transit-based accessibility to supermarkets," *Applied Geography*, vol. 53, pp. 149–159, Sep. 2014, doi: 10.1016/j.apgeog.2014.06.012.
[64] L. Wang, "Immigration, ethnicity, and accessibility to culturally diverse family physicians," *Health & Place*, vol. 13, no. 3, pp. 656–671, Sep. 2007, doi: 10.1016/j .healthplace.2006.10.001.
[65] M. Iacono, K. Krizek, and A. El-Geneidy, "Access to Destinations: How Close is Close Enough? Estimating Accurate Distance Decay Functions for Multiple Modes and Different Purposes," Minnesota Department of Transportation Research Services Section, Final Report MN/RC 2008-11, May 2008.
[66] L. Wang and S. Ramroop, "Geographic disparities in accessing community pharmacies among vulnerable populations in the Greater Toronto Area," *Can J Public Health*, vol. 109, no. 5–6, pp. 821–832, Dec. 2018, doi: 10.17269/s41997-018-0110-1.
[67] U. C. Bureau, "2010 Census Urban and Rural Classification and Urban Area Criteria," *The United States Census Bureau*. https://www.census.gov/programs-surveys/geography—/guidance/geo-areas/urban-rural/2010-urban-rural.html (accessed Apr. 04, 2020).
[68] U. C. Bureau, "2010 Office of Management and Budget (OMB) Standards," *The United States Census Bureau*. https://www.census.gov/programs-surveys/metro-micro/about/omb -standards.html (accessed Apr. 04, 2020).
[69] "Am I Rural? Tool - Rural Health Information Hub." https://www.ruralhealthinfo.org/am-i-rural# (accessed Feb. 23, 2020).
[70] "Using RUCA Data," *RUCA Rural Health Research Center*. https://depts.washington.edu /uwruca/ruca-uses.php (accessed Feb. 10, 2020).
[71] "Section 827." https://www.deadiversion.usdoj.gov/21cfr/21usc/827.htm (accessed Apr. 06, 2020).
[72] F. F. Cabrera *et al.*, "Opioid distribution trends (2006–2017) in the US Territories," *PeerJ*, vol. 7, p. e6272, Jan. 2019, doi: 10.7717/peerj.6272.
[73] K. A. Mack, C. M. Jones, and R. J. McClure, "Physician Dispensing of Oxycodone and Other Commonly Used Opioids, 2000–2015, United States," *Pain Medicine*, vol. 19, no. 5, pp. 990–996, May 2018, doi: 10.1093/pm/pnx007.
[74] E. O. Ighodaro, K. L. McCall, D. Y. Chung, S. D. Nichols, and B. J. Piper, "Dynamic changes in prescription opioids from 2006 to 2017 in Texas," *PeerJ*, vol. 7, p. e8108, 2019, doi: 10.7717/peerj.8108.
[75] "American FactFinder," *American FactFinder*. https://factfinder.census.gov/faces/nav/jsf /pages/searchresults.xhtml?refresh=t (accessed Feb. 07, 2020).

[76] "DEA database: Where the pain pills went - Washington Post." https://www.washingtonpost.com/graphics/2019/investigations/dea-pain-pill-database/ (accessed Oct. 27, 2019).
[77] U.S. Census Bureau, "Understanding and Using American Community Survey Data: What All Data Users Need to Know," U.S. Government Printing Office, Washington, DC, 2018. Accessed: Mar. 13, 2020. [Online]. Available: https://www.census.gov/programs-surveys/acs/guidance/handbooks/general.html.
[78] U.S. Census Bureau, "American Community Survey and Puerto Rico Community Survey 2012 Subject Definitions," U.S. Census Bureau, 2012. Accessed: Mar. 15, 2020. [Online]. Available: http://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/2012_ACSSubjectDefinitions.pdf?#.
[79] J. Cromartie, "Rural-Urban Commuting Area Codes," Jul. 03, 2019. https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx (accessed Feb. 06, 2020).
[80] J. Cordes, "Spatial Trends in Opioid Overdose Mortality in North Carolina : 1999–2015," *Southeastern Geographer*, vol. 58, no. 2, pp. 193–211, 2018.
[81] P. Bholowalia and A. Kumar, "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN," *International Journal of Computer Applications*, vol. 105, no. 9, pp. 17–24, Nov. 2014.
[82] J. P. Prunuske *et al.*, "Opioid prescribing patterns for non-malignant chronic pain for rural versus non-rural US adults: a population-based study using 2010 NAMCS data," *BMC Health Serv Res*, vol. 14, no. 1, p. 563, Dec. 2014, doi: 10.1186/s12913-014-0563-8.
[83] J. R. Gaither *et al.*, "Racial disparities in discontinuation of long-term opioid therapy following illicit drug use among black and white patients," *Drug and Alcohol Dependence*, vol. 192, pp. 371–376, Nov. 2018, doi: 10.1016/j.drugalcdep.2018.05.033.
[84] K. O. Anderson, C. R. Green, and R. Payne, "Racial and Ethnic Disparities in Pain: Causes and Consequences of Unequal Care," *The Journal of Pain*, vol. 10, no. 12, pp. 1187–1204, Dec. 2009, doi: 10.1016/j.jpain.2009.10.002.
[85] V. L. Shavers, A. Bakos, and V. B. Sheppard, "Race, Ethnicity, and Pain among the U.S. Adult Population," *Journal of Health Care for the Poor and Underserved*, vol. 21, no. 1, pp. 177–220, Feb. 2010, doi: 10.1353/hpu.0.0255.
[86] S. H. Meghani, E. Byun, and R. M. Gallagher, "Time to Take Stock: A Meta-Analysis and Systematic Review of Analgesic Treatment Disparities for Pain in the United States," *Pain Med*, vol. 13, no. 2, pp. 150–174, Feb. 2012, doi: 10.1111/j.1526-4637.2011.01310.x.
[87] D. J. Burgess *et al.*, "Racial Differences in Prescription of Opioid Analgesics for Chronic Noncancer Pain in a National Sample of Veterans," *The Journal of Pain*, vol. 15, no. 4, pp. 447–455, Apr. 2014, doi: 10.1016/j.jpain.2013.12.010.
[88] E. R. Wright, H. E. Kooreman, M. S. Greene, R. A. Chambers, A. Banerjee, and J. Wilson, "The iatrogenic epidemic of prescription drug abuse: County-level determinants of opioid availability and abuse," *Drug and Alcohol Dependence*, vol. 138, pp. 209–215, May 2014, doi: 10.1016/j.drugalcdep.2014.03.002.
[89] D. Wong, "The Modifiable Areal Unit Problem (MAUP)," in *The SAGE Handbook of Spatial Analysis*, London, UK: SAGE Publications, Ltd, 2009, pp. 104–123.
[90] N. Wilson, M. Kariisa, P. Seth, H. Smith, and N. L. Davis, "Drug and Opioid-Involved Overdose Deaths — United States, 2017–2018," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 69, no. 11, pp. 290–297, Mar. 2020, doi: 10.15585/mmwr.mm6911a4.