TRCLC 17-09
September 12, 2019

"ASSESSING THE IMPACT OF AIR POLLUTION ON PUBLIC HEALTH ALONG TRANSIT ROUTES"

FINAL REPORT

Stephen Mattingly and Kyung (Kate) Hyun
with
Shirin Kamali Rad, Mehrdad Arabi and Sina Famili

The Department of Civil Engineering
University of Texas at Arlington

Transportation Research Center
for Livable Communities

Western Michigan University | University of Texas at Arlington | Utah State University | Wayne State University | Tennessee State University

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
TRCLC 17-09	N/A	N/A

4. Title and Subtitle Assessing the Impact Of Air Pollution on Public Health Along Transit Routes	**5. Report Date** September 12, 2019 **6. Performing Organization Code** N/A
7. Author(s) Stephen Mattingly and Kyung (Kate) Hyun with Shirin Kamali Rad, Mehrdad Arabi and Sina Famili	**8. Performing Org. Report No.** N/A
9. Performing Organization Name and Address Department of Civil Engineering, University of Texas at Arlington Arlington, TX 76019	**10. Work Unit No. (TRAIS)** N/A
11. Contract No. 17-09	
12. Sponsoring Agency Name and Address Transportation Research Center for Livable Communities (TRCLC) 1903 W. Michigan Ave., Kalamazoo, MI 49008-5316	**13. Type of Report & Period Covered** Final Report 9/01/2017-6/30/2019
14. Sponsoring Agency Code N/A	
15. Supplementary Notes	

16. Abstract

Transportation sources account for a large proportion of the pollutants found in most urban areas. Also, transportation activity and intensity appear likely to contribute to the risk of respiratory disease occurrence. This research investigates the impacts of transportation, urban design and socioeconomic characteristics on the risk of air pollution-related respiratory diseases in two of the biggest MSAs (Metropolitan Statistical Areas) in the US, Dallas-Fort Worth (DFW) and Los Angeles at the block group (BG) level, by considering the US Environmental Protection Agency's respiratory hazard quotient (RHQ) as the dependent variable. The researchers identify thirty candidate indicators of disease risk from previous studies and use them as independent variables in the model. The study applies a three-step modeling including Principal Component Analysis (PCA), Ordinary Least Square (OLS) and Geographically Weighted Regression (GWR) to reach the final model. The results of this study demonstrate strong spatial correlations in the variability in both MSAs which help explain the impact of the indicators such as socioeconomic characteristics, transit access to jobs, and automobile access on the risk of respiratory diseases. The populations living in areas with higher transit access to jobs in urbanized areas and greater automobile access in more rural areas appear more prone to respiratory diseases after controlling for demographic characteristics.

17. Key Words **Transportation and Health, RHQ, Respiratory Disease, PCA, OLS, GWR**	**18. Distribution Statement** No restrictions.

19. Security Classification - report	20. Security Classification - page	21. No. of Pages	22. Price
Unclassified	Unclassified	84	N/A

Disclaimer

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the information presented herein. This publication is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. This report does not necessarily reflect the official views or policies of the U.S. government, or the Transportation Research Center for Livable Communities, who assume no liability for the contents or use thereof. This report does not represent standards, specifications, or regulations.

Acknowledgments

This research was funded by the US Department of Transportation through the Transportation Research Center for Livable Communities (TRCLC), a Tier 1 University Transportation Center at Western Michigan University.

Table of Contents

List of Tables

List of Figures

Executive Summary

The impact of air pollution on public health has represented a common concern for over fifty years; however, as population and economic activity continue to increase air pollution often worsens. Careful strategic planning and management of transportation systems must occur to prevent this worsening of air pollution in the face of this growth. This study analyzes the combined impacts of transportation, land use and socioeconomic factors on public respiratory health using respiratory hazard quotient as an indicator of public health; the study specifically investigates the role transit access plays in health risk. The study investigates two large metropolitan statistical areas (MSAs) in the US (Dallas-Fort Worth and Los Angeles).

This study applies Principal Component Analysis (PCA), Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) to investigate the impacts of the selected demographic, land use and transportation factors on the occurrence risk of respiratory diseases in two of the biggest MSAs in the US, Dallas-Fort Worth and Los Angeles, by considering respiratory hazard quotient as the dependent variable. As many of these variables cause multicollinearity problem within the model, the study applies PCA to eliminate multicollinearity and group the initial variables into fewer components, which could be used as OLS and GWR inputs.

The results of the PCA explain about 73 percent of the variation in the dependent variable in both the DFW and Los Angeles MSA using nine components. The OLS model results indicate one of the components appears insignificant for each MSA (old adults in DFW and employment density in Los Angeles), and spatial autocorrelations appear significant. As this study seeks to estimate the impacts of selected indicators locally and evaluate their effects in different locations of an MSA, a GWR to addresses the spatial autocorrelations observed in the OLS. The results of GWR in both MSAs show a good fit between the final independent variables and risk of respiratory diseases, while demographic and transit access to job represent the most significant variables. The GWR results show an overall positive effect of all variables on the independent variable with a median R^2 value of 0.83, compared to 0.48 from OLS in DFW and 0.79 (GWR) and 0.48 (OLS) in Los Angeles.

While demographic characteristics appear the most important determinant of aggregate respiratory disease risk in both MSAs, transit access to jobs represents the second most important component. This indicates that after controlling for demographic effects, higher transit access to jobs clearly indicates a greater risk of respiratory disease, which directly confirms the research question and hypothesis. Those living along transit corridors and likely in transit-oriented development face a greater risk of respiratory disease. While other components experience greater spatial variations in both MSAs, the transit access to jobs displays a clear pattern and significance.

While the specific variables in the components vary slightly between the DFW and Los Angeles MSAs, the components largely measure the same effects as can be noted in their descriptions. The importance of similar effects in both MSAs indicates that large MSAs may experience similar impacts related to transit access to jobs, automobile access, and vehicle miles traveled. The results of the GWR also show the varying effect of chosen variables on the risk of respiratory disease in different area of DFW and Los Angeles. This can be explained by the local characteristics of each factor in different block groups or areas within the MSAs. Analyzing and comparing the results of GWR maps in these two MSAs show that the population living in rural areas of the metropolitan area appear more affected by transportation and land use factors. Demographic and socioeconomic characteristics appear to also play a significant role in risk, especially in urban and suburban BGs.

The respiratory risks in high transit areas may indicate the need for new policies and building codes to provide greater protection to the residents living in those areas. This study also suggests that departments of transportation and local environmental agencies can use the results of a GWR model rather than global models to analyze the key factors and indicators (i.e. land use and transportation) that impact the risk of health issues in different locations.

1 Chapter 1: Research Overview

1.1. Background and Significance

In most countries, increased economic activity and population growth result in an increased number of cars and higher levels of air pollution from vehicle emissions. If the built environment stimulates increased vehicular travel, this may increase per capita vehicle emissions, and these may increase exposure to pollutants and the risk of respiratory and cardiovascular ailments (Frank et al., 2006). With the world's population estimated to reach 10 billion people by 2050, and 75% of this population living in cities, (UNFPA, 2011) policy makers must understand the impacts of urban and transport planning and design decisions on public health (Giles-Corti et al., 2016). Through more effective planning and design decisions, policymakers and elected officials may encourage economic development while reducing its negative societal costs.

Exposure to air pollutants varies significantly based on a household's location within an urban area. Cities around the world deal with the consequences of changing population socioeconomics and strategies that have failed to effectively manage the relationship between land use, mobility, and population health (Giles-Corti et al., 2016). In addition, community design influences the residents' dependence on automobiles (Ewing et al., 2002), and air pollution from automotive sources commonly represents the single largest source of regional air pollution in urban areas. Air pollutant concentrations close to major traffic routes often increase much higher than background regional levels (Zhu et al., 2002), which endangers nearby populations and disproportionately exposes them to traffic-related air pollution (Samaranayake et al., 2014). The local concentrations of pollutants may disproportionately impact particular communities and contribute to higher rates of morbidity and mortality in these communities.

Not all citizens can afford to select locations to live and work based on the health risks imposed by nearby traffic. Studies indicate that populations living, working, or going to school near major roads may be subjected to an increased risk for several adverse health effects such as respiratory, cardiovascular, low birth weight and cancer (Adar and Kaufman, 2007). The adverse health impact correlated with air pollution (Samaranayake et al., 2014) varies depending on the type of pollutant, the magnitude, the exposure duration and frequency, and the associated toxicity (Vallero, 2014). According to the Centers for Disease Control and Prevention (CDC), in 2009 more than three thousand people died due to asthma (CDC, 2017) and asthma affects about 25 million people in the United States including 7 million children (NHLBI, 2017). Thus, a need to track changes in the public health impacts of urban and transport planning, and prioritize policies and infrastructure investments by considering public health outcomes exists because communities benefit from accurate and timely localized knowledge of air pollution levels to identify potential responses and long term mitigation strategies (Samaranayake et al., 2014). Determining the air pollution exposures at a community level allows policy makers and elected officials to ensure that particular groups do not appear to be disproportionately impacted.

In the past, this disproportionate impact often falls on environmental justice populations. Recent transportation innovations often promise a transportation revolution that eliminates road deaths, serious injury, and congestion through connected-automated vehicles and advanced software. However, these solutions fail to address the broader health and environmental consequences such as air pollution related to land use, the transport system, and rapid motorization (Health Effects Institute (HEI), 2010), and without careful planning the "revolution" may worsen the public health impacts of motor vehicles. Control technologies have reduced emissions per vehicle-mile, but motor vehicle pollution remains a major health risk because reduced emission rates are often offset by increased vehicle travel (HEI, 2010) and all vehicular emission levels appear to have a detrimental impact on public health. Motor vehicle air pollution probably causes a similar number of premature deaths as do traffic crashes (Krzyzanowski, 2005). Previous studies indicate that subjects living adjacent to major roads more likely suffer adverse health effects respiratory diseases such as asthma, and cardiovascular diseases (HEI, 2010). Since the traffic represents the major source of pollutants, an investigation of the public health impacts on communities due to traffic-related air pollution remains crucial.

1.2. Research Gap

While some studies have investigated the relationship between socioeconomic or land use variables and developed models to quantify the effect of these variables, the impact of all these factors on health status while considering geographic influence in a metropolitan statistical area (MSA) remains rarely studied. Consequently, these factors require further examination to document the factors that indicate an area may experience greater health risks. Figure 1 describes the knowledge gap and the contribution of this study.

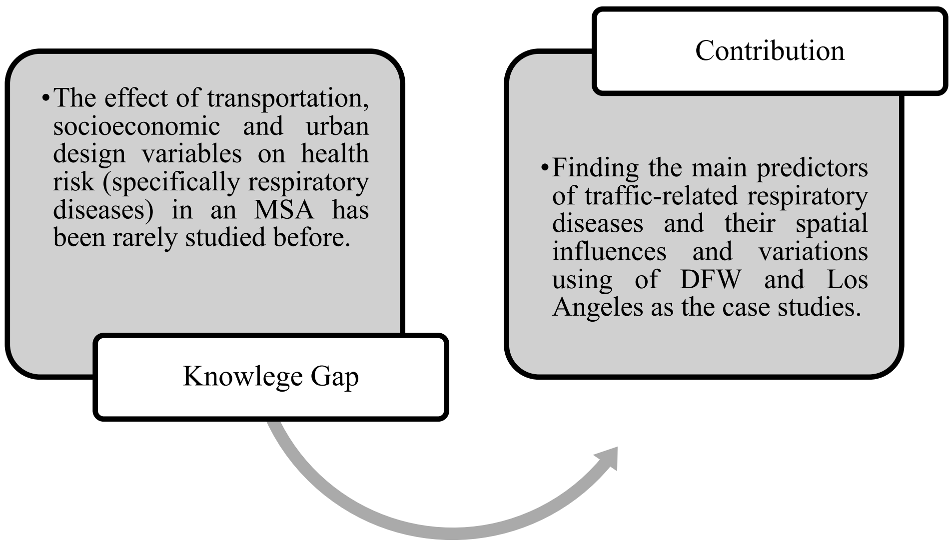

Figure 1. Research Elements

1.3. Research Goals

The main goals of this study are to:

- Understand the overall impact of different groups of variables which contribute to air pollution on public health
- Provide a model which can estimate the impact of selected variables on public health
- Finding the locations in metropolitan areas which are more susceptible to respiratory diseases as a result of air pollution

1.4. Research Questions

- What transportation and land use factors contribute to air pollution?
- Will there be a meaningful and significant relationship between transportation, land use and demographic variables and risk of respiratory disease?
- How to the dependent variables change in different geographical location in subject MSAs?

1.5. Methodology

The methodology utilized in this study encompasses the following tasks:

(1) Literature Review: A comprehensive literature on the association transportation, urban design and traffic- related health impacts have been reviewed. The current literature identifies potential respiratory health indicators and categorizes them in four different groups of transportation, urban design, socioeconomics and health.

(2) Data Collection: This step gathers data for the Dallas-Fort Worth and Los Angeles MSA study areas. The data includes respiratory diseases caused by pollutants, block group geocodes, and the transportation, urban design (5Ds) variables and socio-economic and health factors impacting respiratory health. All data is collected from publicly available data sources. The two MSAs have different data available to characterize motor vehicle use; therefore, the characterization of motor vehicle use differs slightly for each site.

(3) Modeling and Data Analysis: The study uses a GIS framework to aggregate all data layers and determine the geocodes for the spatial boundaries of each corridor and system. The study uses a three-step modeling approach; this strategy starts with a principal component analysis (PCA) that reduces the dimensionality of the independent variables. The ordinary least squares model (OLS) examines the importance of the principal components as a predictive model and determines the presence of spatial autocorrelation. The final model, geographically weighted regression (GWR) seeks to address the spatial autocorrelation and improve the model explanatory power.

(4) Results and Findings: Finally, researchers conducted these ultimate outcomes:

(a) Transportation and land use factors that contribute to air pollution, (b) The relationship between transportation, land use and demographic variables and risk of respiratory disease (c) How the dependent variables change in different geographical location in subject MSAs. This study only characterizes two MSAs which may limit its applicability to large automobile dominated MSAs. Figure 2 shows the methodology steps in this research.

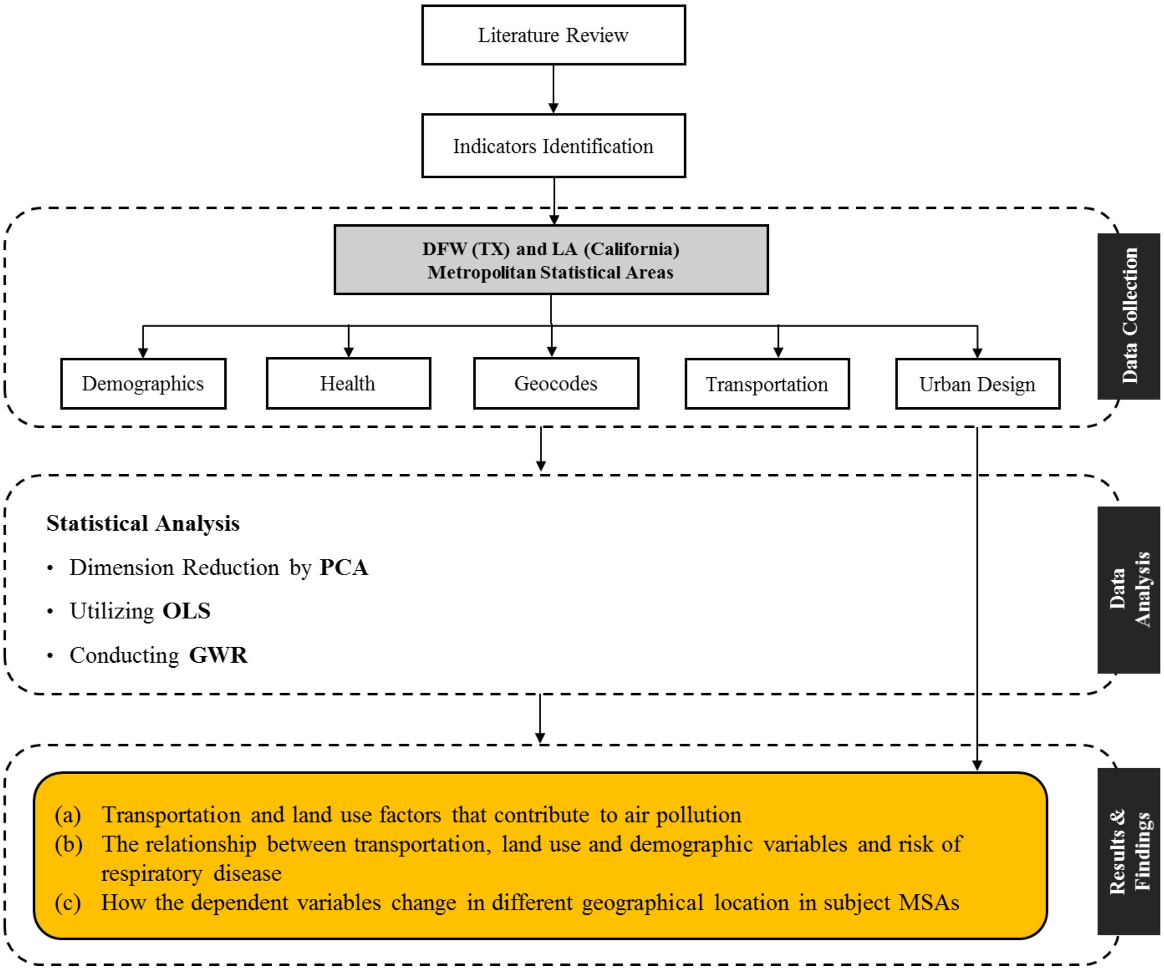

Figure 2. Methodology chart

1.6. Report Formation

The rest of the report consists of seven sections. Section 2 reviews the literature and previous studies associated with effective indicators that can affect air pollution and subsequently public health. Section 3 describes the data collection process and section 4 lays out the methods and models used in this study. The next sections (section 5 and 6) discuss the results in the DFW and LA MSAs. Then, section 7 summarizes the study's findings as the conclusion.

2 Chapter 2: Literature Review

Transportation systems in most urban areas of the world have a negative impact on public health outside of providing more effective access to healthcare. Decisions about housing, food, water, energy, transport, social services, and health care locations within an urban area profoundly impact the health, well-being, and safety of the growing and aging urban populations (Badland et al., 2014). Economic growth often causes private car use to significantly increase (Kopits, 2003); this usually reduces physical activity and increases air pollution, noise, and the risk of motor vehicle crashes (Stevenson et al., 1995). The reduction in physical activity, increase in crash risk and increase in exposure to air pollution caused by modern motor vehicle focused transportation systems pose a significant threat to public health.

The impacts of land use and transport mode choice on public health remain unclear because they happen against a backdrop of complex, interacting, and dynamic environmental, technological, and population conditions that evolve over extended temporal periods. Current research tends to focus on the aggregate impacts of transportation and land use rather than disaggregate impacts due to the nature of the air pollution exposure and other data being fused to health outcomes. Recent transportation innovations promise a transportation revolution that eliminates or significantly reduces the crash risk mentioned in the previous paragraph; however, the health impacts associated with this revolution remain difficult to quantify because the amount of air pollution may increase or decrease depending on the market behaviors that occur.

This literature review focuses on public health outcomes associated with transportation planning and operations (Sallis et al., 2016) while controlling for land use and air quality effects as confounding factors. The literature on the relationship between transportation and health explores three principal mechanisms where the transportation system in an MSA can influence regional public health. The first and most widely investigated set of interactions concerns the linkages between transportation, land use, and health. The second set of interactions investigates the linkages between transportation, regional air quality, and health. In addition, the third mechanism investigates all four dimensions (land use, transportation, air quality and health) together. Figure 3 displays these three sets of interactions schematically.

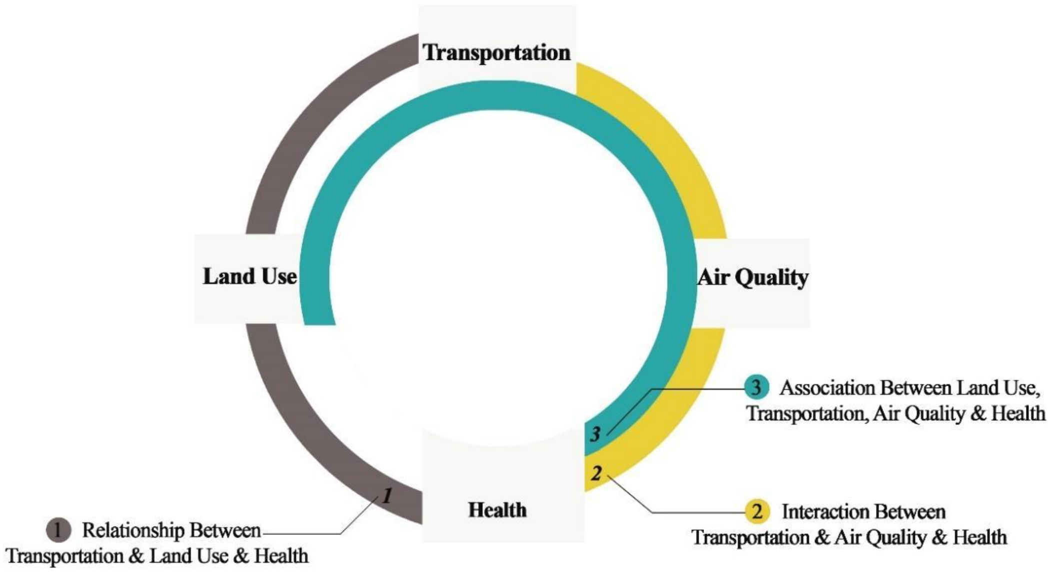

Figure 3. Three Main Mechanisms in Literature Review

2.1 Relationship Between Transportation, Land-Use and Health

This section discusses transport strategy and planning decision impacts on public health. Recent urban growth and rapid changes in motorized transport increase the geographic size of urban areas; this places transport mobility at the forefront of city planning. Past land-use and transport strategies currently have widespread negative effects on health through reduced physical activity, prolonged sitting, injuries, air pollution, social isolation, noise, stress, compromised personal safety, unhealthy diets, urban-heat-island effects, and greenhouse gas emissions. These negative consequences often result from the high priority given to motor vehicles in land-use and transport planning (Sallis, 2016). The need for economic development and the resulting transportation activity makes meeting the World Health Organization (WHO) recommendation, "placing health and health equity at the heart of [city] governance and planning" difficult to achieve (WHO, 2005). Air quality analysis performs an increasingly significant role in the planning of new urban development and when seeking solutions to correct the current problems of metropolitan areas. In the United States, the 1990 Clean Air Act Amendments and the 1992 Intermodal Surface Transportation Efficiency Act (ISTEA) put air quality in the forefront of planning priorities (Medina et al., 1994); ISTEA requires the planning of transportation improvements in urban areas, and the resulting plan must move a region toward conformity with National Ambient Air Quality Standards (NAAQS).

2.2 Interactions Between Transportation, Air-Quality and Health

A second category of transport-related studies excludes land use and focuses on the direct linkage between transportation activities, poor air quality and negative health impacts. The potential negative health effects related to living close to traffic sources include respiratory diseases such as asthma, and cardiovascular diseases (HEI, 2010).

Major freeways and major arterials pose a particular risk to nearby neighborhoods. Although control technologies have reduced emissions per vehicle-mile, motor vehicle pollution remains a major health risk because vehicle travel continues to increase (HEI, 2010). People living within 300 meter of busy roads expose to higher levels of pollutants, including particulate matter (PM), carbon monoxide (CO), and nitrogen oxide (NOx) (Zhu et al., 2002).

2.2.1 Health Impacts of Emissions

The adverse health impacts related to air pollution (Samaranayake et al., 2014) vary depending on the type of pollutant, its magnitude, the exposure duration and frequency, and the associated toxicity. The major air pollutants monitored by the Environmental Protection Agency (EPA) include nitrogen oxides (NO_x), carbon monoxide (CO), volatile organic compounds (VOC), ozone (O_3), sulfur dioxide (SO_2), and large and small particles. Oxidative stress, inflammation, and genetic defects represent some of the basic mechanisms where the vapor and particulate phases of pollutants induce negative health effects (Vallero, 2014). Cardiovascular and respiratory diseases (e.g. lung cancer and asthma), chronic obstructive pulmonary diseases (COPD), cancer, birth defects, low-birth weight and type 2 diabetes (Wang et al., 2014) denote some of the major diseases that may be caused by air pollution (NIEHS, 2016; HEI, 2010). Favarato et al. (2014) performed a meta-analysis of cohort studies to examine the association between long-term exposure to air pollution and the prevalence of asthma. According to the HEI (2010), long-term exposure to NO_2 has a positive incidence on asthma. Air pollution has the potential to contribute to many negative health outcomes beyond those already identified in research and the magnitude of the role that traffic-related air pollution plays in these outcomes requires further investigation.

While stationary and natural sources play an important role in air pollution, motor vehicle exhaust emission represents the single largest source of regional air pollution in urban areas and emits pollutants into the air due to the incomplete burning of fossil fuels (Colvile et al., 2001). Studies indicate that populations living, working, or going to school near major roads may be subjected to an increased risk for a number of adverse health effects such as respiratory, cardiovascular, premature mortality, low birth weight and cancer (Adar and Kaufman, 2007). Air quality monitoring studies have measured elevated concentrations of pollutants emitted directly by motor vehicles near large roadways-relative to overall urban background concentrations (Baldauf et al. 2008). Since traffic represents the major source of pollutants such as NO_2, CO, and $PM_{2.5}/PM_{10}$, an investigation of the incidence of health issues due to traffic-related air pollution remains crucial.

The scientific community has recognized the importance of monitoring and managing particulate emissions for many decades. Dockery et al. (1993) improves on several studies that found associations between mortality rates and particulate air pollution in U.S. metropolitan areas by estimating the effects of air pollution on mortality after controlling for other factors such as smoking status. Dockery et al. (1993) recognize that combustion products from transportation represent the main source of sulfate and fine-particulate air pollution. Studies by Barone-Adesi et al. (2015) and Gehring et al. (2013) investigate exposure to traffic-related air pollution based primarily on meteorological conditions and traffic activity. The number of vehicles, the fleet mix, and vehicle speed/operating pattern represent the major parameters for traffic activity that affect the concentration of near-road pollutants. Interpreting near-road air quality data and exposure levels require meteorological measurements, namely wind speed and direction, temperature, humidity, and atmospheric stability, to describe the dispersion speed and pattern (Venkatram et al. 2007). The EPA recently developed R-Line to model near-road air quality impacts.

2.2.2 Transportation Barriers to Healthcare Access

Transportation could be a significant barrier to healthcare access for at-risk population. Syed et al. (2013) who reviewed 61 articles found:

Patients with a lower socioeconomic status (SES) experience higher rates of transportation barriers to healthcare access than those with a higher SES since they have limited access to pharmacies and medication. This limited access forces patients miss the opportunity for evaluation and treatment of chronic disease states, which changes to treatment regimens, escalation or de-escalation of care and delay interventions that may reduce or prevent disease complications.

2.2.3 Public Transit Contribution to Health

Improving public transit service such as providing more routes, longer operating hours, and more frequent service, clean vehicles and stations, grade separation, and improved user information would attract choice riders (people who would otherwise drive) and reduce pollution emissions. Quality public transit encourages vehicle travel reductions and expedites transit-oriented development, which creates neighborhoods where residents own fewer cars, drive less, and rely more on walking, cycling, and public transit, providing additional health and safety benefits (Bailey et al., 2008).

2.3 Association Between Land Use, Transportation, Air Quality and Health

The third mechanism associates all four components (land use, transportation, air pollution and health) together. Transportation policy and planning decisions can affect health in various ways. The modes used for personal mobility affects physical and mental health outcomes, which include cancer, cardiovascular disease, vehicle crashes, and diabetes (Litman, 2013). Because the land use distribution, such as residential, industrial or commercial, over the urban area determines the locations of human activities such as living, working, shopping, education or leisure. The distribution of human activities in space requires spatial interactions or trips in the transport system to overcome the distance between the locations of activities (Wegener, 2004); therefore, the user's travel behavior depends on the land use. However, several steps between a policy or planning decision, its land use and travel behavior changes, and the ultimate consequences exist (Litman, 2013). Figure 4 shows these steps.

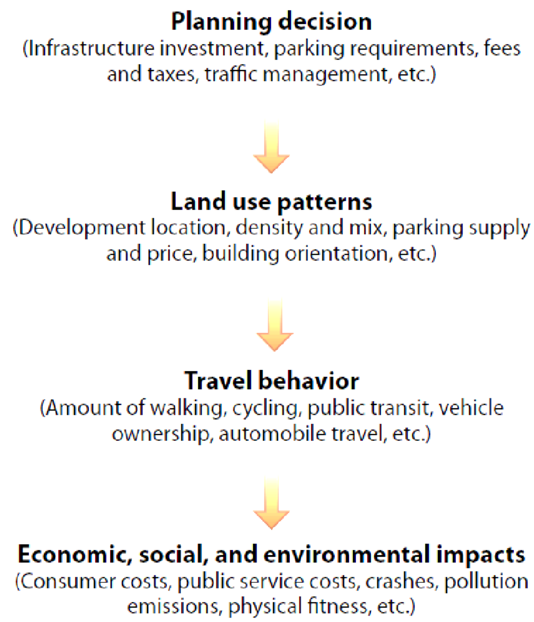

Figure 4. Steps between Planning Decisions and Ultimate Impacts

(Adopted from Litman, 2013)

Several leading transportation studies in the late 1980s and early 1990s began using land use as an input variable (Montgomery County, Maryland, 1989). Although earlier academic studies initially investigated the use of this approach (e.g., Edwards, 1976; Weiner, 1999), the later projects formalized the method and brought it into public planning and decision making. Assisted by substantially expanded computing capacity and methods, the practice developed substantially over the following two decades, and became ordinary enough to be considered state-of-the-practice (Ewing et al., 2006).

2.3.1 Urban and Transportation Planning Interventions to Promote Health

D Variables: Urban planning and transport planning academics have long sought to understand ways to reduce motor vehicle miles travelled and motivate the use of public transport and active transport modes such as walking and cycling to enhance health (Macintyre, 2003). In travel research, urban development patterns have come to be characterized by the "D" variables (Giles-Corti et al., 2016). The original "three Ds," created by Cervero and Kockelman (1997), are density, diversity, and design, followed later by destination accessibility and distance to transit (Ewing and Cervero, 2010). Figure 5 displays the 5Ds' definitions.

Density	*Density* is always measured as the variable of interest per unit of area. The area can be gross or net, and the variable of interest can be population, dwelling units, employment, building floor area, or something else.
Diversity	*Diversity* measures pertain to the number of different lands uses in each area and the degree to which they are represented in land area, floor area, or employment.
Design	*Design* includes street network characteristics within an area. Measures include average block size, proportion of four-way intersections, and number of intersections per square mile.
Destination	*Destination accessibility* measures ease of access to trip attractions. It may be regional or local. Regional accessibility is either distance to the central business district or the number of jobs or other attractions reachable within a given travel time, which tends to be highest at central locations and lowest at peripheral ones.
Distance	*Distance to transit* is usually measured as an average of the shortest street routes from the residences or workplaces in an area to the nearest rail station or bus stop. Alternatively, it may be measured as transit route density, distance between transit stops, or the number of stations per unit area

Figure 5. 5Ds' definitions

(Reference: Ewing and Cervero, 2010)

Academics identify five key built-form characteristics and related policies as the 5Ds. Building on this earlier work, the study identifies five integrated interventions needed to create cities that improve health. Table 1 shows potential pathways through which city planning decisions impact public health. Moving from left to right, the figure shows the role that the urban system policies play to build urban and transport planning and design interventions that directly and indirectly impact health by influencing daily living options and transport mode choices and demand. In turn, these interventions determine five risk exposures related to non- communicable diseases, road trauma, and other adverse health outcomes. Next, these risk exposures identify intermediary outcomes (eg, greenhouse gas emissions and chronic disease risk factors) as well as traffic injury and disease outcomes, which ultimately determine quality of life and health, social, and environmental equity (Giles-Corti et al., 2016).

Table 1. Health Impacts of D Variables

D Variable	Urban and Transport Planning Features	Examples	Health Impacts
Destination accessibility	Employment, facilities, and services conveniently accessible by public transport; destinations for daily living available locally	Jobs, facilities, and services within 30 min travel from home by public transport; daily living destinations within walking distance	• Tailpipe emissions are one of the major contributors to poor air quality and thus poor cardiovascular and respiratory health. • Studies found that increased negative health impacts from PM, NOx, hydrocarbons, and CO are found within 2 to 300 meters of busy streets, both inside and outside buildings.
Design	Urban design creates walkable catchments around activity centers and incorporates accessible public open space; street networks minimize distances between homes and daily destinations, reduce traffic exposure, and create safe pedestrian, cycling, and public transport networks	High street connectivity including ped-sheds ≥ 0.6 within 0.8- 1.2 km (i.e., 1–15 min walk) of activity centers, transport hubs, and schools; separated pedestrian and cycle paths; local public open space provided; housing overlooks streets and public open spaces	• Residents of a highly walkable/bikeable neighborhood are likely to exercise for at least 30 minutes one additional day per week and may increase activity on as many as three days a week. • Greater connectivity provides travelers with more route choices and reduces trip lengths and NOx and VOC emissions generated on a per household basis.

Continued

D Variable	Urban and Transport Planning Features	Examples	Health Impacts
Density	Residential densities enough to support the viability of local business and high-frequency public transport services Distribution of employment is an appropriate mix of employment available across a region	Multiunit housing built around activity centers with shops, services, and transport hubs A job–housing balance from 0.8 to 1.2 km	• Reductions in driving in terms of VMT, trip length and number of trips • Decreased need for automobile ownership • Increased walking, bicycling and transit use • Reduced VMT and trip generation into employment centers, also results in better traffic safety.
Distance to public transport	High-frequency public transport located within short walking distance from homes	Bus stops accessible ≤400 m; rail stops accessible ≤800 m from homes	• An accessible, frequent transit service may reduce car ownership, vehicle trips, miles traveled and emissions as well as increase walking and biking and thus improve cardiovascular and respiratory health and physical fitness.
Diversity	Residential areas built with different types of housing mixed with commercial, public, and recreational opportunities	Different types of housing available near, around, and on top of shops and services required for daily living	• A more diverse area facilitates pedestrian, bicycle, ridesharing or transit travel and reduces vehicle travel, thus decreasing overall vehicle emissions. • Land use mix may contribute to the formation of social capital. Diversity of population and income denotes prolonged life, better overall health, improved cardiovascular function, faster recovery from illness and improved mental health and reduced violent crime, less frequent binge drinking, lower birth rates and more leisure-time physical activity.

(**Resource:** Giles-Corti et al., 2016 and Ewing and Cervera, 2010)

2.4 Indicators

Previous studies recognize the need to benchmark and monitor progress on the implementation of policies, and to track changes in health effects. The researchers identify an indicator set to evaluate transportation's role in public health from previous studies. The indicators should reflect overall goals and consider data availability, understandability, and usefulness in decision making. Thus, Table 2 displays indicators that could be used to monitor progress towards the implementation of urban and transport policies, investment, and outcomes to create cities that enhance health and reduce non-communicable diseases.

Table 2. Transportation Indicators Contributing to Health Outcomes

1. Access to Health-Related Goods and Services (Litman, 2013)	17. Housing affordability in accessible locations (Litman, 2007)	33. Quality of transport for disadvantaged people (Litman, 2007)
2. Activities (Marquez and Smith, 1999)	18. Infectious diseases (Corti et al., 2016)	34. Residential Density (Badoe and Miller, 2000)
3. Auto Ownership (Badoe and Miller, 2000)	19. Land use mix (Ewing et al., 2002 and Stone, 2008)	35. Respiratory disease (Corti et al., 2016)
4. Birth defects (Samaranayake et al., 2014)	20. Link Loads (Geurs and Wee, 2004)	36. Road Network (Badoe and Miller, 2000)
5. Body mass index (Frank et al., 2007)	21. Rout Choice (Geurs and Wee, 2004)	37. Sprawl index (Stone, 2008)
6. Cancer (Corti et al., 2016)	22. Location and characteristics of infrastructure (Geurs and Wee, 2004)	38. Street Connectivity (Ewing et al., 2002)
7. Cardiovascular disease (Corti et al., 2016)	23. Low-birth weight (Samaranayake et al., 2014)	39. Traffic Assignment (Armstrong and Khan, 2004)
8. Respiratory diseases e.g. lung cancer and asthma (Samaranayake et al., 2014)	24. Mean daily grams of NOx, CO, PM2.5, PM10 (EPA)	40. Traffic Crashes (Litman, 2013)
9. Connectivity (Ewing et al., 2002 and Stone, 2008)	25. Mean daily VMT per person (Litman, 2007)	41. Transit affordability (Litman, 2007)
10. Demand management (Corti et al., 2016)	26. Minutes of active transportation last week (Frank et al, 2007)	42. Transit Service (Badoe and Miller, 2000)
11. 5Ds (Corti et al., 2016)	27. Modal Split (Armstrong and Khan, 2004)	43. Travel speed (Geurs and Wee, 2004)
12. Demographic and other covariates Miles to nearest bus stop (Frank, 2006)	28. Mode Choice, Destination Choice (Geurs and Wee, 2004)	44. Travel Times/ Distances/costs (Geurs and Wee, 2004)
13. Destination accessibility (Corti et al., 2016)	29. Neighborhood Design (Badoe and Miller, 2000)	45. Trip Distribution (Armstrong and Khan, 2004)
14. Distribution of employment (Corti et al., 2016)	30. Net residential density (Frank et al., 2007)	46. Vehicle hours lost in congestion (Geurs and Wee, 2004)
15. Employment Density (Badoe and Miller, 2000)	31. Physical Activity and Fitness (Litman, 2013)	47. Vehicle Pollution Exposure (Litman, 2013)
16. Food and health, service access (Corti et al., 2016)	32. Population (Marquez and Smith, 1999)	48. Walkability Index (Litman, 2007)

3 Chapter 3: Data Collection

The study explores aggregate data at the US Census block group level for transportation, urban design, health and demographic characteristics in two major MSAs (DFW and LA). Since the 1950 census, the Office of Management and Budget (OMB) has designated metropolitan areas for statistical purposes. Metropolitan areas are characterized by a central urban area surrounded by other urban areas that work together economically or socially. The central urban area must have a population of at least 50,000 people with a combined regional population of 100,000.

3.1 Study Sites

The study area was selected considering various socioeconomic and demographic characteristics of metropolitan areas with a population of more than 1 million. The purpose was to choose study areas with the optimum variation in such indicators. Among the candidates, the following areas were selected:

3.1.1 Dallas- Fort Worth (TX)

The Dallas–Fort Worth metroplex (officially designated the Dallas–Fort Worth–Arlington, TX Metropolitan Statistical Area by the U.S. Office of Management and Budget) encompasses 13 counties within Texas, and it is the largest inland metropolitan area in the United States. According to the 2018 U.S. Census estimate, the Dallas–Fort Worth metroplex's population is 7,539,711; this makes it the largest metropolitan area in both Texas and the south.

A comparison of the descriptive statistics for the demographic and transportation variables in the DFW MSA with the average values from the 53 MSAs in the US with a population of more than one million people shows relatively similar values; however, DFW has a larger Hispanic population than other MSAs. The white population represents 47% of the DFW population and 60% of the population in the large US MSAs; the Hispanic population represents 40% of the DFW population and only 17% of the population in the large US MSAs. The age distribution and private vehicle use appears similar between DFW and the large US MSAs. The proportion of the population from 18-65 is 63% in DFW and the large US MSAs. DFW residents complete 80% of their work trips using private vehicle while large US MSAs residents complete 78% of their work trips using private vehicle.

3.1.2 Los Angeles (CA)

The Los Angeles MSA, which is the second largest metroplex in the US, includes Los Angeles and Orange Counties. According to the 2018 U.S. Census estimate, the population of the Los Angeles MSA is 13,291,486. Similar to the DFW MSA, the demographic profiles of the Los Angeles MSA and the high share of private vehicle use as the prevailing mode of transportation, makes this MSA another good option to evaluate transportation and land use impacts on the risk of respiratory diseases.

A comparison of the descriptive statistics for the demographic and transportation variables in the Los Angeles MSA with the average values from the 53 MSAs in the US with a population of more than one million people shows relatively similar values; however, the demographics for Los Angeles appear significantly different from the other MSAs. The white population represents only 32% of the Los Angeles population but 60% of the population in the large US MSAs; the Hispanic population represents 43% of the Los Angeles population and only 17% of the population in the large US MSAs. The age distribution and private vehicle use appears similar between Los Angeles and the large US MSAs. The proportion of the population from 18-65 is 65% in Los Angeles and 63% in the large US MSAs. Los Angeles residents complete 75% of their work trips using private vehicle while large US MSAs residents complete 78% of their work trips using private vehicle. This places the DFW and Los Angeles MSAs above and below the mean automobile mode choice rate.

3.2 Variables

To evaluate the roles of transportation and urban design on respiratory health impacts, this study identifies a set of indicators that affect respiratory diseases. After extensive literature review, a comprehensive pool of factors has been found and the initial list with 48 factors has been reduced to 30 variables by merging similar indicators into one factor. The authors categorize the finalized factors into three groups including socioeconomic, urban design and transportation into a hierarchy of categories, attributes and indicators in Figure 6. This hierarchy provides a structure for the factors that appear likely to impact transportation related health outcomes either directly or indirectly.

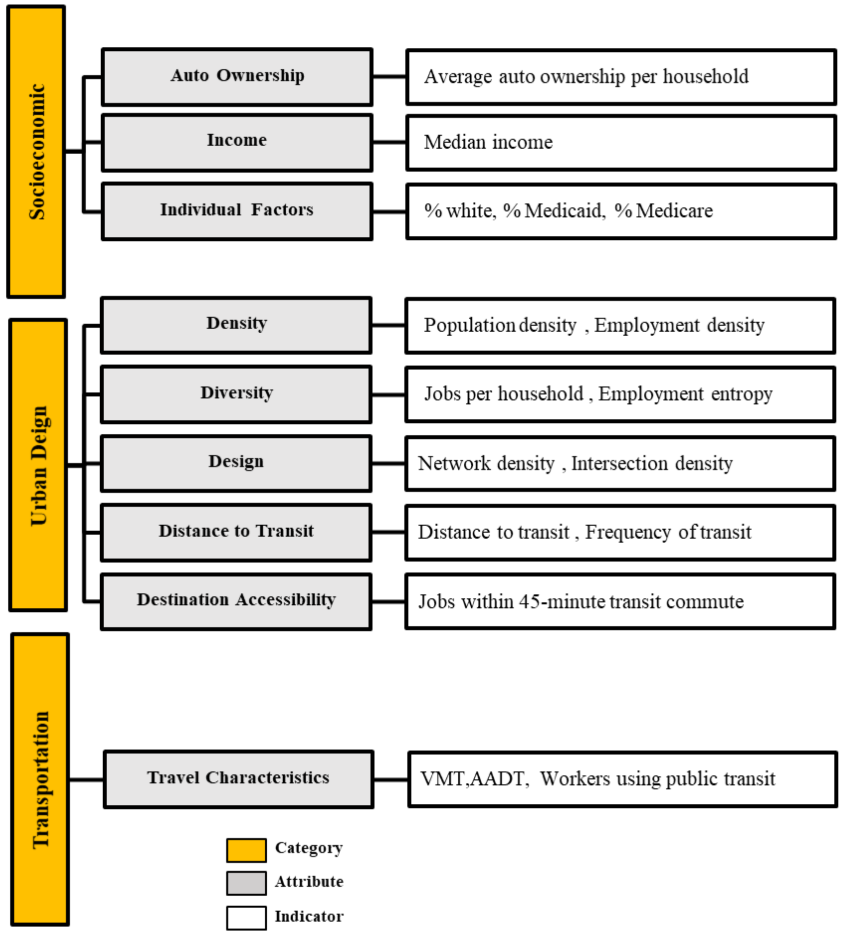

Figure 6. Hierarchy of Variables

3.2.1 Socioeconomic

Socioeconomic characteristics impact health risk and outcomes regardless of transportation and land use indicators; therefore, they must be included in the study to isolate the impacts of transportation and land use. Previous studies suggest that socioeconomic position (e.g. age groups and race) and auto ownership (Badoe and Miller, 2000) directly relate to traffic exposure and subsequently the risk of respiratory diseases (Frank, 2006 and Cesaroni et al., 2010). Low income or older adult residents also pose a concern because people with low socioeconomic status may suffer disproportionately from the detrimental consequences of transportation and land-use policies in their communities (Bullard et al., 1997) and experience more negative healthcare outcomes in general. The type of insurance held by a population directly impacts its preventative health and health outcomes. For example, Medicare provides health insurance for Americans aged

65 and over, and Medicaid provides coverage for people with limited income. Various socioeconomic characteristics such as race (percentage of White and Hispanic population), age (percentage of under 18 years, 18 to 64 years and 65 years and older) and median income also represent important indicators for health status. Since the population of the DFW and Los Angeles MSAs mainly consist of the White and Hispanic races, this study focuses on these two groups are selected as variables indicating racial distribution in the study area. All the data comes from the American Community Survey (ACS) database at the block group level and uses 5-year estimates (2013-2017).

3.2.2 Urban Design

Urban design and transport planning researchers have long sought to reduce motor vehicle travel and promote the use of public transport and active transport modes such as walking and cycling to enhance health (Ellaway et al, 2003). Researchers often characterize urban development patterns using "D" variables (Giles-Corti et al., 2016). Previous researchers found a significant relationship between the 5Ds and travel behavior (Corti et al., 2016), and this study seeks to determine if urban design and transportation impact the risk of respiratory disease. Building on the previous work, the study uses a total of 13 indicators to represent density, diversity, design, distance to transit, and destination accessibility, and their associated data collects from the Smart Location Database developed by the US EPA (Ramsey and Bell, 2014).

3.2.3 Transportation

Transportation activities represented by vehicle miles traveled (Bartholomew and Ewing, 2006; Litman, 2007; Ewing and Cervero, 2010), public transportation use (Badoe and Miller, 2000; Litman, 2007; Armstrong and Khan, 2004) and mode choice (Geurs and Wee, 2004) play an important role in understanding health impacts. This study uses daily vehicle miles traveled (DVMT) and daily truck vehicle miles traveled (DTRKVMT) to investigate the contribution of auto and truck traffic on the health risk. The authors aggregate the VMT data obtained from the 2017 Texas Department of Transportation (TxDOT) roadway inventory for each BG in DFW. This study also considers variables representing transportation mode to work including percentage of private vehicle use, carpooling use, public transit use, and active transport use in each BG obtained from the ACS 5-year estimates.

Since the VMT data on all roads in Los Angeles MSA is not publicly available, the study uses average annual daily traffic (AADT) data from Highway Performance Monitoring System (HPMS) database available from the Federal Highway Administration (FHWA). The HPMS provides the highway segment AADT values. As this indicator evaluates the impact of traffic intensity on the risk of respiratory diseases, either VMT or AADT appear to be viable indicators.

3.2.4 Respiratory Hazard Quotient (RHQ)

The correlation between adverse health impact and air pollution (Samaranayake et al., 2014) varies depending on the type of pollutant, the magnitude, the exposure duration and frequency, and the associated toxicity. The EPA developed the dependent variable in this study, respiratory hazard quotient (RHQ), as part of the 2011 National Air Toxics Assessment (NATA). This dataset provides the EPA's 2011 NATA ambient concentration, exposure concentration, and risk estimates across the US at the census tract level. The ambient concentrations generate exposure concentrations from an inhalation exposure model and then estimate hazard quotients based on health-benchmark information. The RHQ refers to the ratio of the potential exposure to a substance and the level at which no adverse effects are expected (calculated as the exposure divided by the appropriate chronic or acute value). In the study area, this number ranges from 0 to 0.7 with the average of 0.46. A hazard quotient of 0 means adverse health effects (respiratory disease) appear unlikely and pose no risk; for RHQs greater than 0 and closer to 1, the potential for adverse effects increases. Figure 7 displays the distribution of the RHQ in the DFW MSA, and Figure 8 displays the RHQ distribution in the LA MSA. Both figures use the RHQ quartiles from DFW to facilitate a comparison between the MSAs and demonstrate the significantly higher overall RHQ scores for LA. The RHQ severity differs significantly between DFW and Los Angeles. In DFW, higher RHQ values cluster north and northwest of the Dallas central business district (CBD) and just north of the Fort Worth CBD. In Los Angeles almost the entire LA basin south of the Angeles and Los Padres National Forests experience high RHQ values; Malibu, Palos Verdes and some areas of central and southern Orange County experience lower RHQ values. Because no significant difference in air pollutant exposure between block groups within a census tract exists, the researchers assume an equal hazard quotient for all BGs within each census tract.

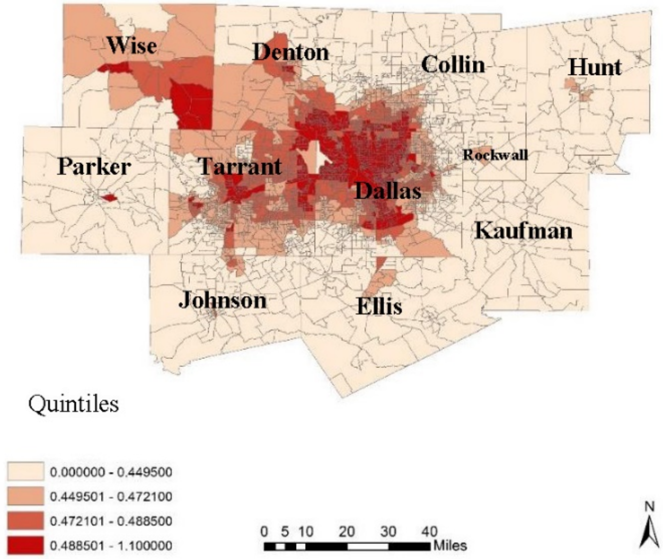

Figure 7. The distribution of the RHQ in the DFW

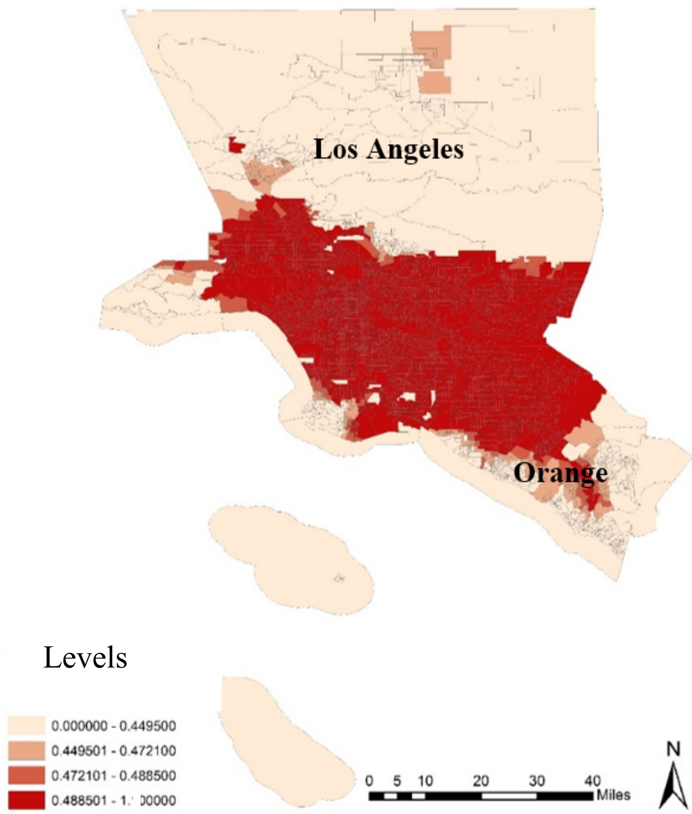

Figure 8. The distribution of the RHQ in Los Angeles MSA

3.3 Descriptive Statistics

Descriptive statistics have been used to describe the basic features of the data in the study. Table 3 compares the mean values for selected variables in the DFW and LA MSAs while Tables 4 and 5 provide more complete descriptive statistics for the candidate factors, which the study labels as Effective Indicators of Respiratory (EIR) diseases. The DFW MSA has much lower population and employment density than the Los Angeles MSA. The road network and intersection density appear similar in both MSAs, but on average, transit in the Los Angeles MSA appears more frequent and provides access to more jobs and population within 45 minute travel times because the transit system in Los Angeles is more extensive. The average socioeconomic characteristics appear similar between the two study sites. The Los Angeles and DFW MSAs have similar carpooling rates, but Los Angeles has higher average public transit usage rates and higher average walking and bicycling rates, which decreases the average private vehicle use rate.

Table 3. Comparison of selected indicators in the DFW and Los Angeles MSAs

Variables	Unit	Mean (DFW)	Mean (LA)
Gross population density	people/acre	8.1	20.21
Gross employment density	jobs/acre	2.5	5.22
Jobs within 45 minutes auto travel time	-	196901.0	472171.43
Total road network density	-	17	21.39
Frequency of transit within 0.25 miles	-	25.7	71.45
White population	percentage	47	32.06
Hispanic population	percentage	29.8	43.15
Population (18-64 years old)	percentage	63.1	64.27
Population over 65 years old	percentage	11.6	13.69
Average vehicle ownership	-	1.9	1.89
Medicare and Medicaid population	percentage	1.6	3.33
No insurance coverage population	percentage	18.2	12.01
Workers using their own vehicle	percentage	80	74.67
Workers using carpooling	percentage	10.2	9.53
Workers using public transit	percentage	1.7	5.18
Workers using bike or walking	percentage	1.5	3.29

Figures 10 and 9 show the population density in the DFW and LA MSAs; the significantly greater population density in the LA MSA appears throughout the urbanized areas of the MSA and even in more rural areas in Lancaster and Palmdale. In the DFW MSA, the population density appears much lower throughout the region and large portions of Dallas, Fort Worth and Arlington have relatively low population densities.

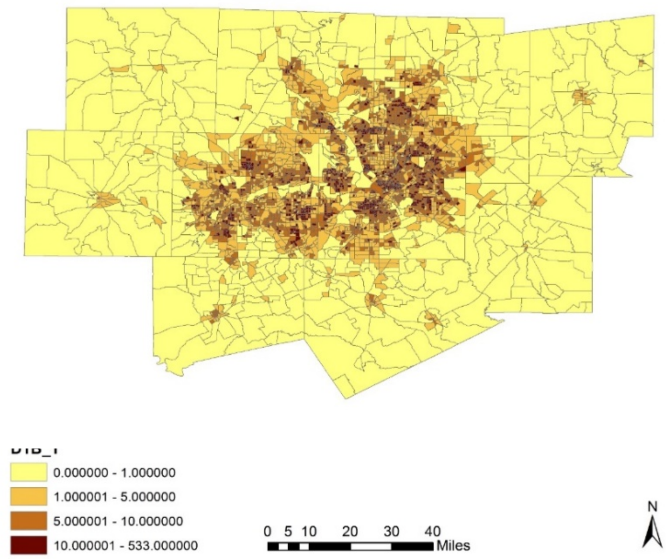

Figure 10. Population density in the DFW MSA

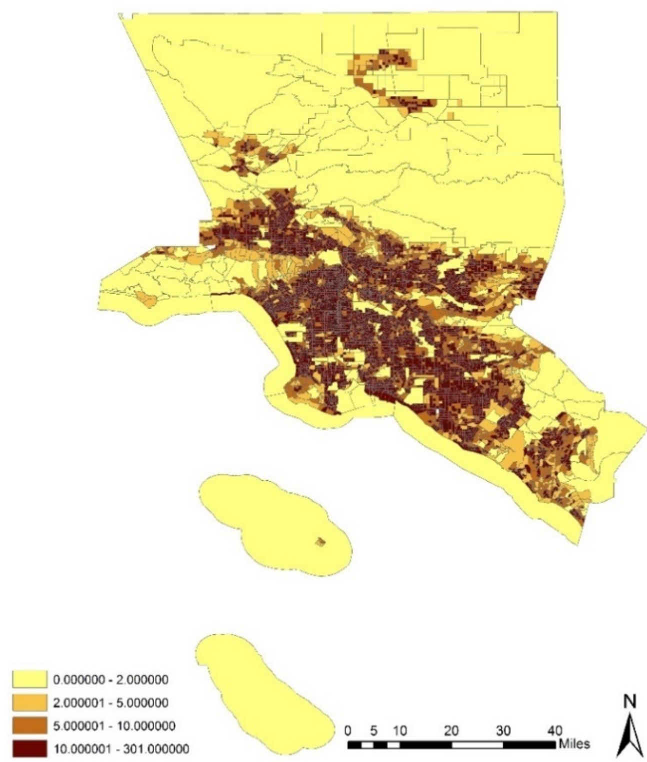

Figure 11. Population density in the LA MSA

Figures 12 and 13 show the employment density in the DFW and LA MSAs in these two MSAs. Both MSAs appear polycentric. In the DFW MSA, employment density coincides with the freeway network while the LA MSA has higher employment throughout and more distinct employment centers outside the CBD.

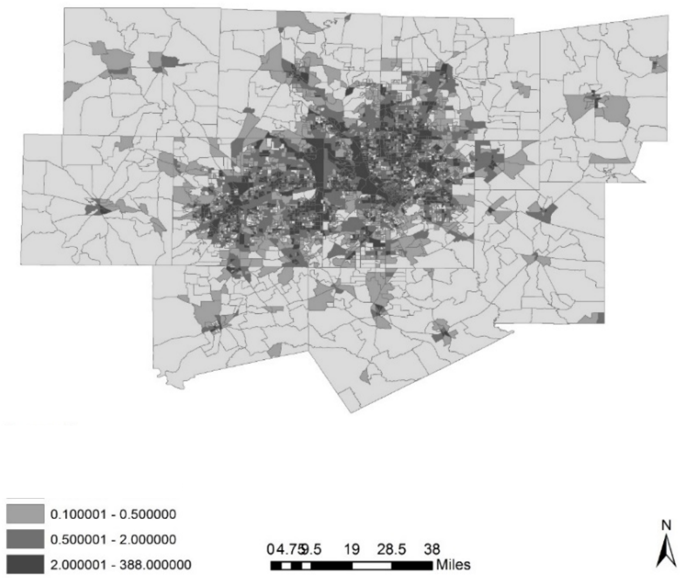

Figure 12. Employment density in the DFW MSA

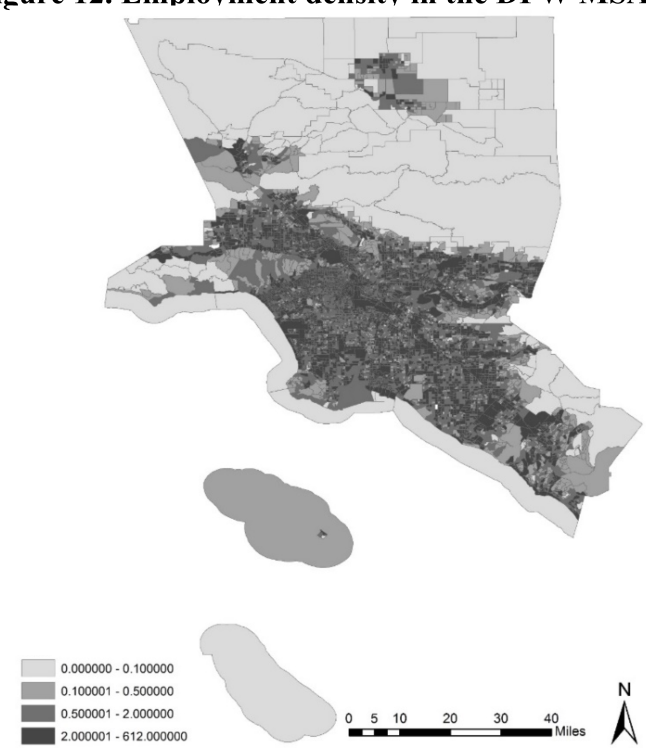

Figure 13. Employment density in the Los Angeles MSA

Table 4. Descriptive Statistics of Selected Indicators-DFW MSA

	Variables	Mean	SD	Min	Max
Urban Design	EIR1-Gross population density (people/acre)	8.1	11.5	0.0	532.6
	EIR2-Gross employment density (jobs/acre)	2.5	9.8	0.0	387.2
	EIR3-Jobs per household	7.7	229.2	0.0	12609.0
	EIR4-Employment and household entropy*	0.4	0.2	0.0	1.0
	EIR5-Total road network density	17.0	7.5	0.0	50.1
	EIR6-Intersection density per square mile	2.1	4.5	0.0	59.2
	EIR7-Distance from jobs to transit stop (meters)	0.5	0.5	0.0	1.0
	EIR8-Frequency of transit within 0.25 miles	25.7	53.2	0.0	960.3
	EIR9-Aggregate frequency of transit per square mile	124.7	288.7	0.0	5093.8
	EIR10-Jobs within 45 minutes auto travel time	196901.0	98282.5	2361.0	465185.7
	EIR11-Working age population -45 min travel time	254940.0	97522.5	5497.0	438184.0
	EIR12-Jobs within 45-minute transit commute	4819.0	8997.9	0.0	103282.7
	EIR13-Population within 45-min transit commute	2299.0	3639.3	0.0	42140.0
Socioeconomics	EIR14-White population (%)	47.0	28.8	0.0	100.0
	EIR15-Hispanic population (%)	29.8	25.3	0.0	100.0
	EIR16-Population under 18 years old (%)	25.2	9.3	0.0	68.1
	EIR17-Population 18-64 years old (%)	63.1	9.7	0.0	100.0
	EIR18-Population over 65 years old (%)	11.6	8.4	0.0	100.0
	EIR19-Average vehicle ownership	1.9	0.4	0.0	3.3
	EIR20-Medicare coverage population (%)	8.7	6.7	0.0	100.0
	EIR21-Medicaid population (%)	12.3	11.5	0.0	75.4
	EIR22-Medicare and Medicaid population (%)	1.6	2.8	0.0	68.1
	EIR23-No insurance coverage population (%)	18.2	13.2	0.0	78.2
	EIR24-Average median income	68803.9	40292.9	0.0	250000.0
Transportation	EIR25-Daily VMT by all vehicles	217092.1	415437.9	0.0	4957761.0
	EIR26-Daily VMT by trucks	15893.9	35423.9	0.0	476180.0
	EIR27-Workers using their own vehicle (%)	80.0	11.1	0.0	100.0
	EIR28-Workers using carpooling (%)	10.2	8.4	0.0	64.6
	EIR29-Workers using public transit (%)	1.7	3.9	0.0	60.9
	EIR30-Workers using bike or walking (%)	1.5	3.8	0.0	54.4

* Employment and household entropy calculations are based on trip production and trip attractions including employment categories. The vehicle trip productions and attractions are derived by multiplying average ITE vehicle trip generation rates by employment types and households (- [H(VT) +E(VT)]/ (ln (6))

Table 5. Descriptive Statistics of Selected Indicators-LA MSA

	Variables	Mean	SD	Min	Max
Urban Design	EIR1-Gross population density (people/acre)	20.21	16.41	0.00	300.03
	EIR2-Gross employment density (jobs/acre)	5.22	14.17	0.00	611.21
	EIR3-Jobs per household	17.58	491.26	0.00	32725.00
	EIR4-Employment and household entropy*	0.48	0.21	0.00	0.99
	EIR5-Total road network density	21.39	7.22	0.00	68.63
	EIR6-Intersection density per square mile	2.12	5.66	0.00	83.02
	EIR7-Distance from jobs to transit stop (meters)	0.79	0.40	0.00	1.00
	EIR8-Frequency of transit within 0.25 miles	71.45	140.12	0.00	4400.67
	EIR9-Aggregate frequency of transit per square mile	831.74	3042.66	0.00	209112.32
	EIR10-Jobs within 45 minutes auto travel time	472171.43	157447.44	0.00	916589.45
	EIR11-Working age population -45 min travel time	790607.33	254666.40	0.00	1598202.65
	EIR12-Jobs within 45-minute transit commute	15000.17	16735.94	0.00	159226.14
	EIR13-Population within 45-min transit commute	12281.94	11487.56	0.00	129098.16
Socioeconomics	EIR14-White population (%)	32.06	27.75	0.00	100.00
	EIR15-Hispanic population (%)	43.15	30.14	0.00	100.00
	EIR16-Population under 18 years old (%)	21.65	8.39	0.00	58.97
	EIR17-Population 18-64 years old (%)	64.27	9.82	0.00	100.00
	EIR18-Population over 65 years old (%)	13.69	9.05	0.00	100.00
	EIR19-Average vehicle ownership	1.89	0.49	0.00	3.41
	EIR20-Medicare Coverage Population (%)	9.20	7.33	0.00	100.00
	EIR21-Medicaid population (%)	18.40	15.16	0.00	83.15
	EIR22-Medicare and Medicaid Population (%)	3.33	3.54	0.00	62.61
	EIR23-No insurance coverage Population (%)	12.01	8.94	0.00	68.41
	EIR24-Average median income	71835.39	39064.52	0.00	249034.0
Transportation	EIR25-Average Annual Daily Traffic	943209.83	1453415.97	0.00	26957175.0
	EIR26-Workers using their own vehicle (%)	74.67	13.33	0.00	100.00
	EIR27-Workers using carpooling (%)	9.53	7.01	0.00	56.38
	EIR28-Workers using public transit (%)	5.18	7.36	0.00	80.40
	EIR29-Workers using bike or walking (%)	3.29	5.83	0.00	100.00

4 Chapter 4: Methodology

4.1 Overview

The methodology of this study relies on a three-step modeling process using a PCA-OLS-GWR approach to find the final model. After generating 30 EIRs, the authors reduce the number of indicators to nine main components using Principal Component Analysis (PCA). In the next step, OLS verifies the main predictors of traffic-related respiratory disease. Figure 14 describes the three-step modeling process in this study.

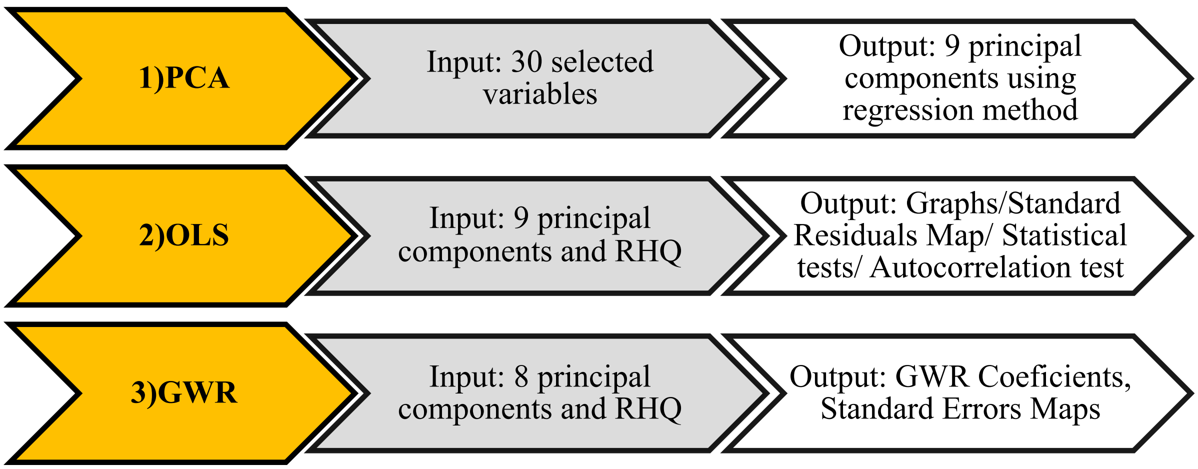

Figure 14. Modeling Procedure

4.2 Principal Component Analysis

This study applies a PCA to reduce the number of selected indicators and find the main components. PCA uses a linear combination of variables to explain the variance structure of a matrix that reduces the data into a few principal components (PC). According to Johnson and Wichern (1982), if there is a random vector X with a covariance matrix M with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$ and eigenvectors $v_1, v_2, \ldots$, the linear combination of the X matrix is as following:

$$X_i = v_i^t \mathrm{X} = v_{1\mathrm{i}}x_1 + v_{2\mathrm{i}}x_2 + \cdots + v_{ni}x_n \quad (1)$$

$$Var[Xi] = v_i^t K v_i \qquad \mathrm{i} = 1, 2, \ldots, \mathrm{n} \quad (2)$$

$$Cov[Xi, Xj] = v_i^t K\mathrm{j} \quad \mathrm{i} = 1, 2, \ldots, \mathrm{n} \quad \mathrm{j} = 1, 2, \ldots, \mathrm{n} \quad (3)$$

Uncorrelated linear combinations of $X_1, X_2 \ldots X_n$ are principal components and in the output, they will be ranked based on their variance in a descending order.

4.3 Ordinary Least Square

In the next step, an ordinary least square (OLS) model eliminates any irrelevant explanatory factors and investigates the model enhancement when considering spatial autocorrelations between variables. This study uses nine components of independent variables derived from the PCA (C1,…,C9) to estimate the RHQ. The researchers assess multicollinearity through the variance inflation factor (VIF) values, where VIFs greater than 10 indicate that multicollinearity exists (Menard, 2002).

4.4 Geographically Weighted Regression

The authors use Geographically Weighted Regression (GWR) to capture spatially varying relationships between RHQ and the final components from PCA and OLS. GWR appears advantageous to estimate parameters showing higher spatial correlations with neighboring regions because it captures spatial heterogeneity in the regression structure. (Chiou et al., 2015; Selby and Kockelman, 2013; Zhao and Park, 2004; Wang and Tenhunen, 2005). Compared to OLS, which estimates global relationships among variables, GWR produces a localized regression model for each geographic location to illustrate spatially varying relationships by estimating coefficient parameters using a weighted least squares method:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{n} \beta_{ik}(u_i, v_i)x_{ik} + \varepsilon_i \qquad (4)$$

Where y is the dependent variable, β_k is a coefficient of independent variable, x_k is the kth independent variable, β_0 is the intercept, (u_i, v_i) is the location of observation i and ε is an error term.

In GWR, BGs located within the pre-determined bandwidth are included in modeling and other elements outside of the bandwidth will have zero values. GWR typically uses two types of bandwidth, fixed and adaptive, and this study uses an Adaptive kernel bandwidth because the distribution of BGs is not homogeneous in the study area.

5 Chapter 5: Dallas-Fort Worth MSA Results

5.1 PCA

This study performs PCA to eliminate collinearity between the initial thirty independent variables. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy is 0.728 and Bartlett's Test of Sphericity is significant; these results are acceptable for PCA. PCA categorizes the initial variables into a relatively small number of factors, which demonstrate the relationships among interrelated variables. The PCA process produces nine components that explain 73.6% of the variation of the variables.

The first principal component, demographic characteristics, represents five factors including percent of Medicaid insurance coverage, percent of no insurance coverage, percentage of Hispanic population, percentage of white population and median income of BGs. The percentage of Medicaid insurance, no insurance coverage, and Hispanic population positively affect the component, while median income and percentage of white population negatively affect the component. The percentage of Medicaid insurance coverage is the most significant factor in this component and shows a loading value of 0.82. The second and third components represent transit and automobile access. These components include aggregate frequency of service, jobs within 45-minute transit or auto commute, and working age population within 45-minute transit or auto commute. The fourth component, older adults, includes percentage of over 65 years old population and percentage of population with Medicare coverage. The fifth, sixth, and seventh components relate to transportation such as DVMT, percentage of auto-oriented intersections per square mile, percentage of using bike or walk mode for commute, and average auto ownership. The eighth component includes percentage of using public transit to go to work and percentage of population with both Medicare and Medicaid insurance coverage, which both indicate low-income characteristics. Only one factor, jobs per household represents the last component. Table 6 shows the nine factors and their total variance explained by the components.

Table 6. Variance and Loadings explained by components obtained from PCA-DFW

Component	Factor	loadings	Initial Eigenvalues	Rotation Sums of Squared Loadings
1: Demographic Characteristics	EIR-21	0.82	7.013 (23.37 % Cumulative Variance)	4.014(13.37% Cumulative Variance)
	EIR-24	-0.79		
	EIR-23	0.77		
	EIR-15	0.76		
	EIR-14	-0.76		
2: Transit Access to Jobs	EIR-8	0.84	3.962 (36.58% Cumulative Variance)	3.838(26.17% Cumulative Variance)
	EIR-9	0.81		
	EIR-12	0.81		
	EIR-13	0.76		
	EIR-2	0.70		
3: Automobile Access	EIR-11	0.89	2.377 (44.50% Cumulative Variance)	3.348(37.33% Cumulative Variance)
	EIR-10	0.86		
	EIR-5	0.72		
	EIR-7	0.56		
	EIR-1	0.43		
4: Older Adults	EIR-18	0.93	2.175 (51.75% Cumulative Variance)	2.380(45.26% Cumulative Variance)
	EIR-20	0.89		
5: Miles Driven	EIR-25	0.90	1.670 (57.32% Cumulative Variance)	2.325(53.01% Cumulative Variance)
	EIR-26	0.85		
	EIR-6	0.58		
	EIR-4	0.47		
6: Active Population	EIR-16	-0.63	1.389 (61.95% Cumulative Variance)	1.890(59.31% Cumulative Variance)
	EIR-17	0.62		
	EIR-30	0.58		
	EIR-19	-0.53		
7: Auto Mode Use	EIR-27	-0.86	1.325 (66.37% Cumulative Variance)	1.656(64.83% Cumulative Variance)
	EIR-28	0.79		
8: Low Income and Older Population	EIR-29	0.69	1.157 (70.22% Cumulative Variance)	1.452(69.67% Cumulative Variance)
	EIR-22	0.51		
9: Jobs Per Household	EIR-3	0.83	1.018 (73.62% Cumulative Variance)	1.183(73.62% Cumulative Variance)

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization a.
a. Rotation converged in 11 iterations.

5.1.1 Demographic characteristics PC

Five factors (Medicaid insurance coverage, percent of no insurance coverage, percentage of Hispanic population, percentage of white population and median income comprise the first principal component. As expected, the percentage of Medicaid insurance, no insurance coverage, and Hispanic population all increase the component loading, and median income and percentage of white population decrease the component value. The percent of Medicaid insurance coverage represents the most significant factor in this component with a loading value of 0.82. Figure 15 displays the distribution of this factor in the DFW MSA, and this component appears to align with the regional income and segregation distribution.

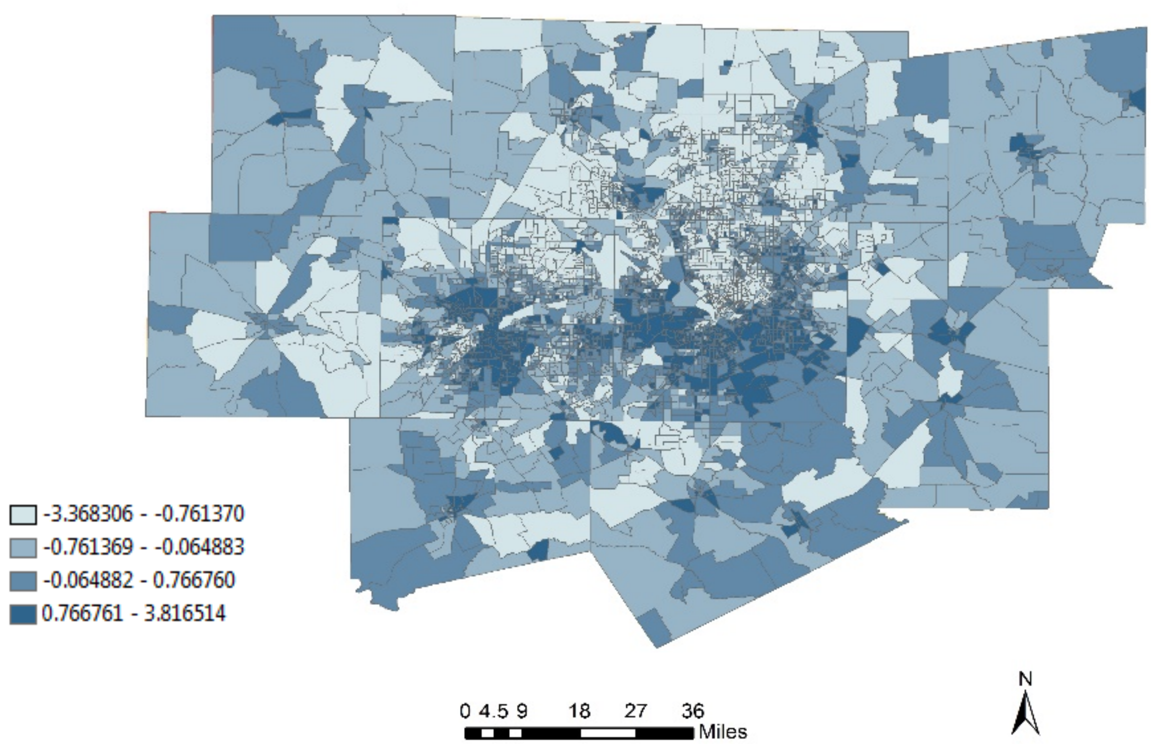

Figure 15. Demographic characteristics in the DFW MSA

5.1.2 Transit access to jobs PC

The second component includes aggregate frequency of transit service within 0.25 miles of block group boundary per hour, aggregate frequency of transit service per square mile, jobs within 45-minute transit commute, working age population within 45-minute transit commute and gross employment density; all of these variables appear positively correlated with the component. The most significant indicator is the aggregate frequency of transit service within 0.25 miles of block group boundary with a loading magnitude of 0.84. Figure 16 displays the distribution of this factor in DFW MSA, and its distribution aligns with the urban cores and unusual pockets of higher scores in rural areas.

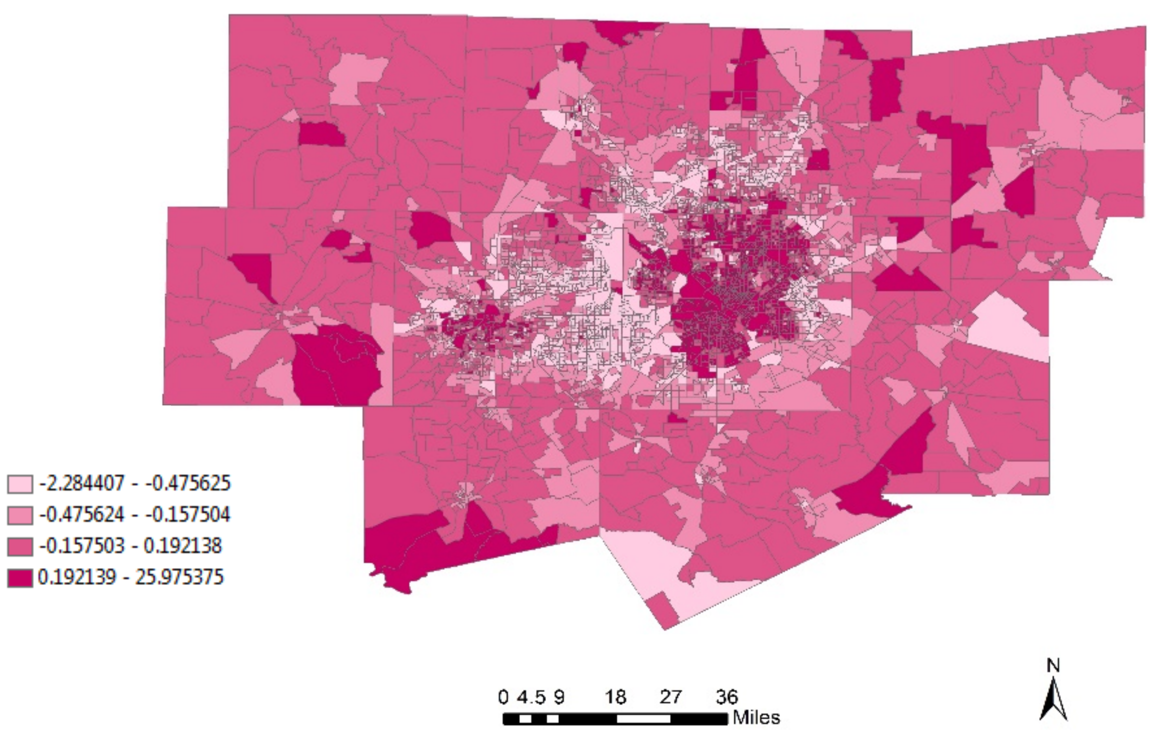

Figure 16. Transit access to jobs in the DFW MSA

5.1.3 Automobile access PC

The third component also contains 5 indicators (working age population within 45 minutes auto travel time, jobs within 45 minutes auto travel time, total road network density, existence of transit stop within ¾ miles of the population weighted centroid, gross population density). All indicators in automobile access have a positive correlation with the component. Working age population within 45 minutes auto travel time has the highest loading in this component (0.89). Figure 17 displays the distribution of this factor in the DFW MSA, and almost all block groups in the top three quantiles appear in Dallas and Tarrant Counties and southern Denton and Collin Counties.

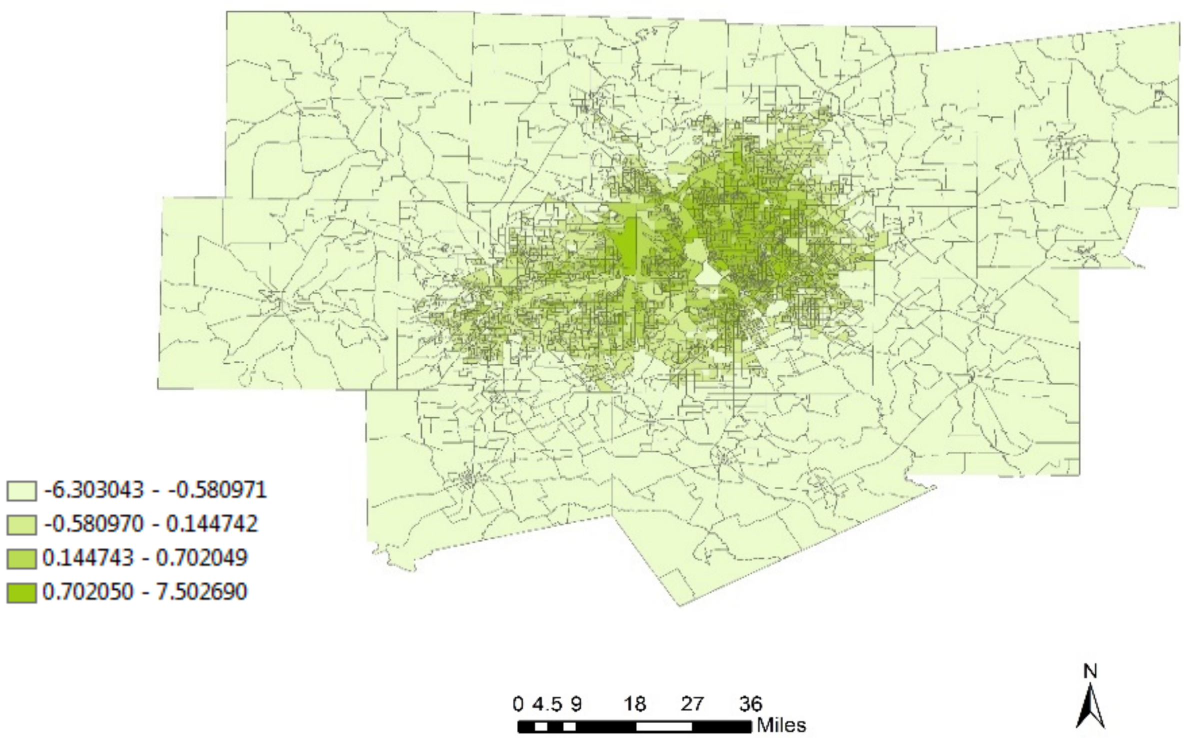

Figure 17. Automobile Access in the DFW MSA

5.1.4 Older adults PC

The fourth component includes percentage of population over 65 years old and with Medicare coverage. Both of these indicators have a positive correlation with the component. Figure 18 displays the distribution of this factor in the DFW MSA, and the distribution lacks a consistent pattern other than rural areas tend to fall in the top two quantiles.

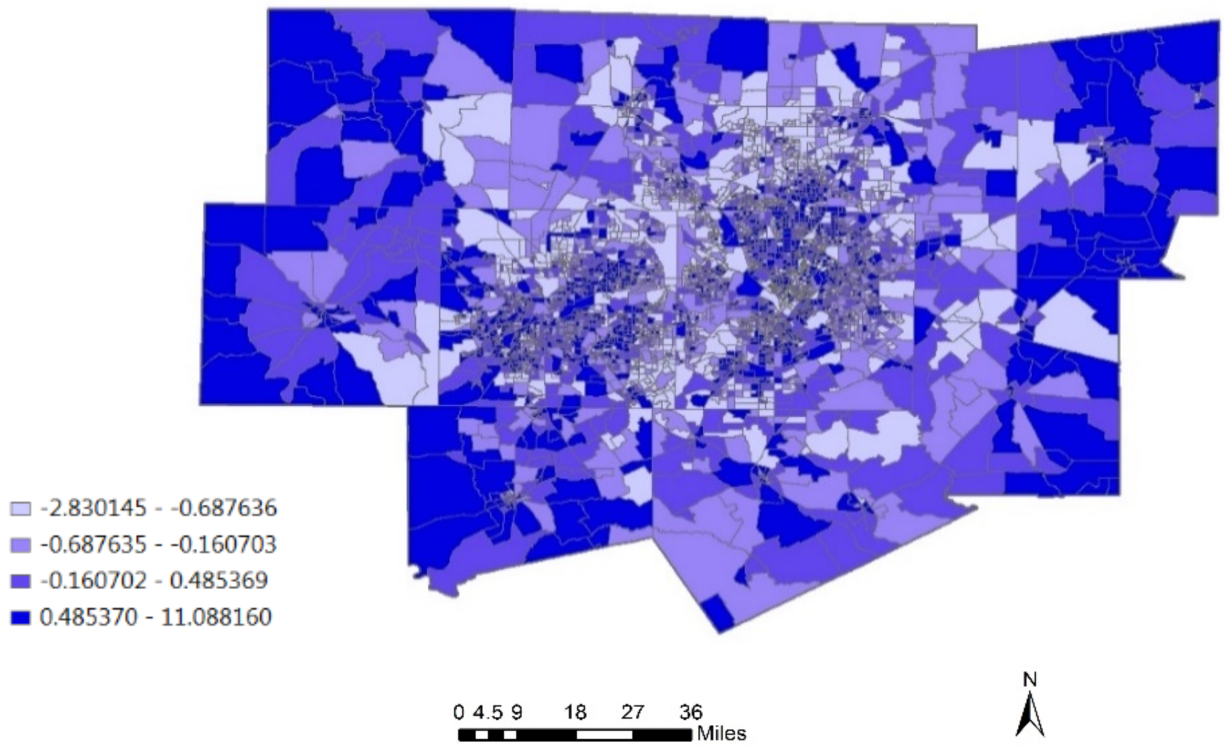

Figure 18. Older Adults in the DFW MSA

5.1.5 Miles driven PC

Fifth principal component includes DVMT, DTRKVMT, intersection density in terms of auto-oriented intersections per square mile and employment and household entropy. Employment and household entropy shows the spatial distribution of residential and business areas within a block group and has the same size as the other indicators; this indicates that all of them have a similar impact on traveled miles by vehicles in the area. DVMT has the highest loading value in this component with a magnitude of 0.9. Figure 19 displays the distribution of this factor in the DFW MSA which aligns well with the freeway network.

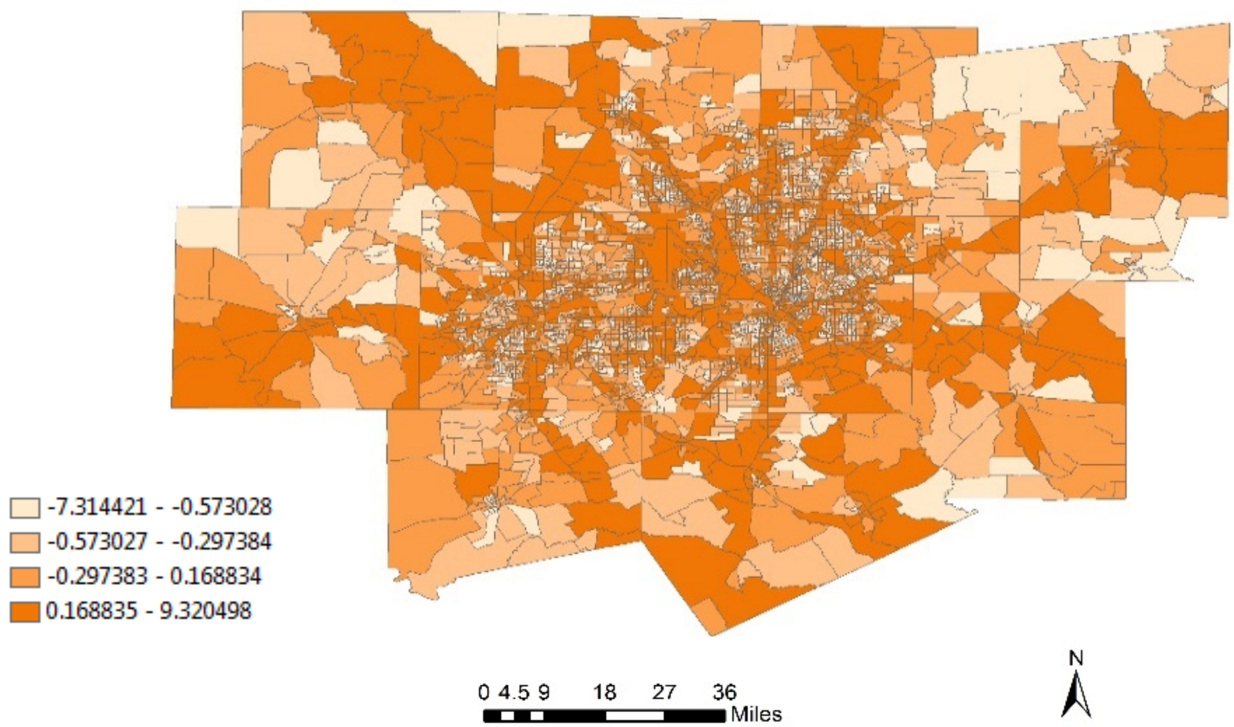

Figure 19. Miles Driven in the DFW MSA

5.1.6 Active population PC

The sixth component contains the percentage of the population under 18 years old from 18 to 65 years old, and using bike or walk mode the work trip, and average auto ownership. The sign of the first and last indicator is the opposite of other two indicators. This occurs because while the percentage of the active population who use the bike and walk mode to go to work appears likely to be negatively impacted by average auto ownership. Similarly, the percentage of the population under 18 appears negatively correlated with the percentage of the working aged population from 18-65. The loading values of the percentage of the population under 18 and from 18 to 65 years old indicators appear the most significant ones with values of 0.63 and 0.62. Figure 20 displays the distribution of this factor in DFW MSA.

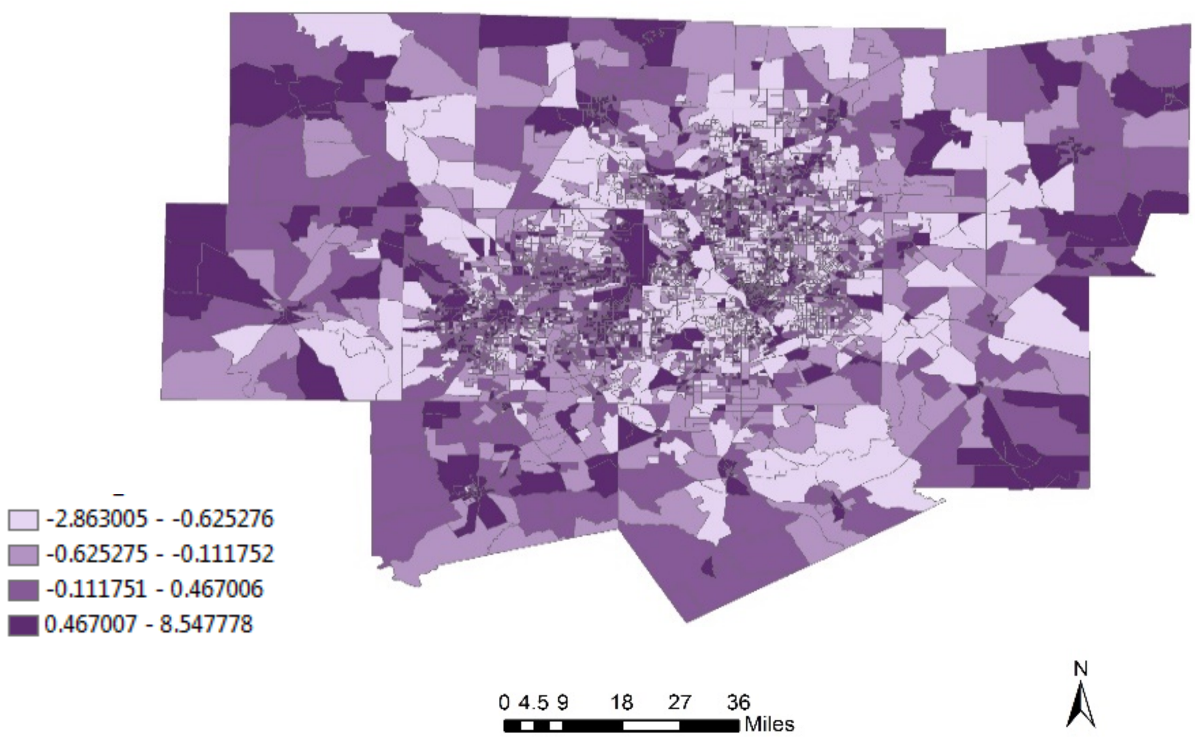

Figure 20. Active Population in the DFW MSA

5.1.7 Auto Mode Use PC

Seventh component includes percentage of population using private vehicle and using carpooling for work trips. They expectedly have different loading signs because as the carpooling rate increases, the share of other modes and specifically private vehicle reduces. The percentage of the population using private vehicle indicator has a higher loading value (0.86). Figure 21 displays the distribution of this factor in DFW MSA, and its regional distribution presents no clear pattern.

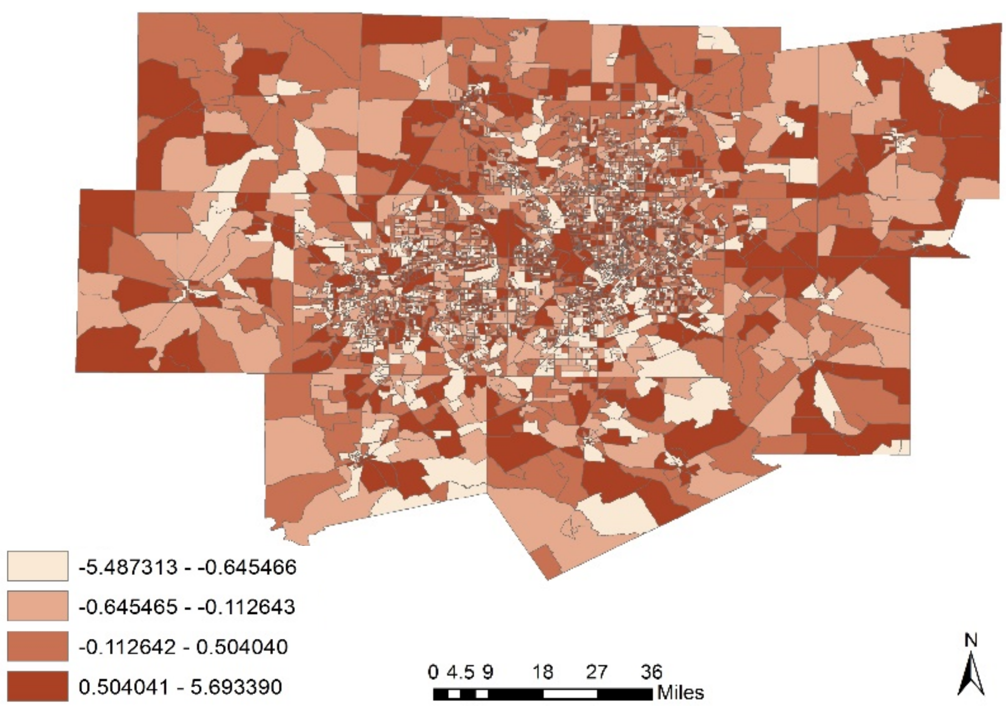

Figure 21. Auto Mode Use in the DFW MSA

5.1.8 Low income and older population PC

The eighth component includes percentage of population using public transit for work trip (loading value of 0.69) and with both Medicare and Medicaid insurance coverage. They both have the same sign because they indicate low-income and older adult population. Figure 22 displays the distribution of this factor in DFW MSA, and no clear regional pattern emerges.

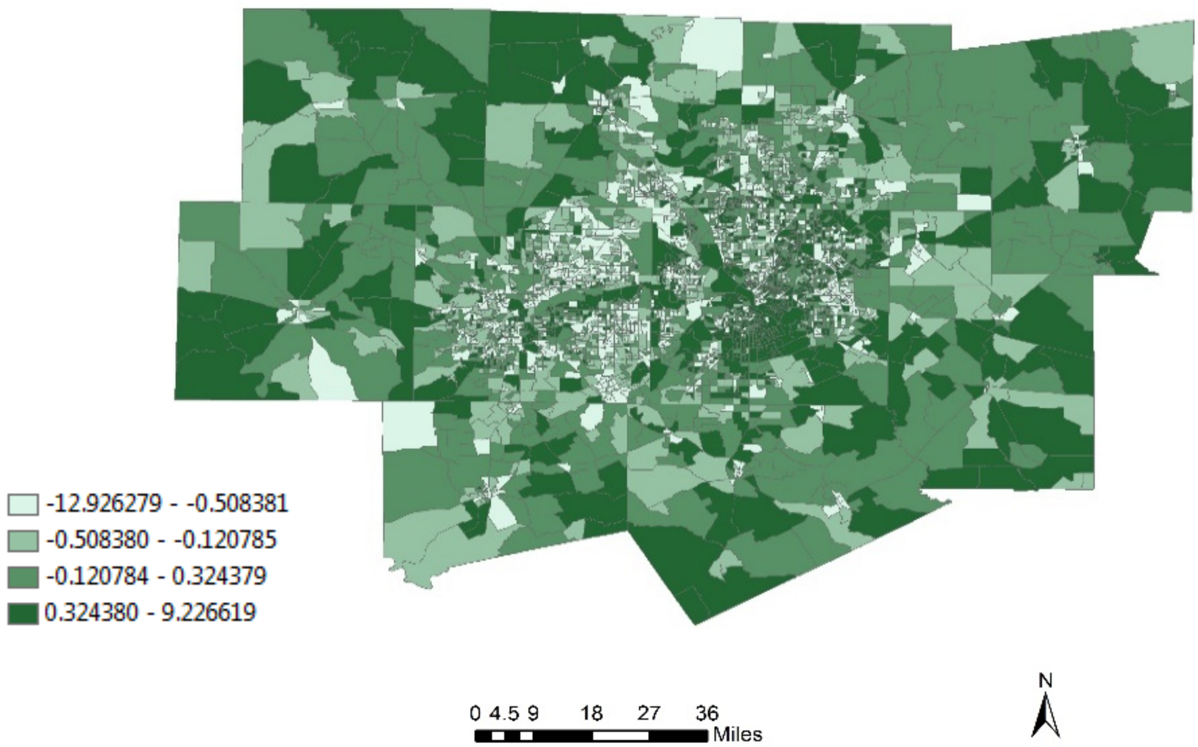

Figure 22. Low income and older population distribution in the DFW MSA

5.1.9 Jobs per household PC

The ninth component has only one indicator, jobs per household. Figure 23 displays the distribution of this factor in DFW MSA; lower values tend to occur in rural areas, but a few exceptions exist.

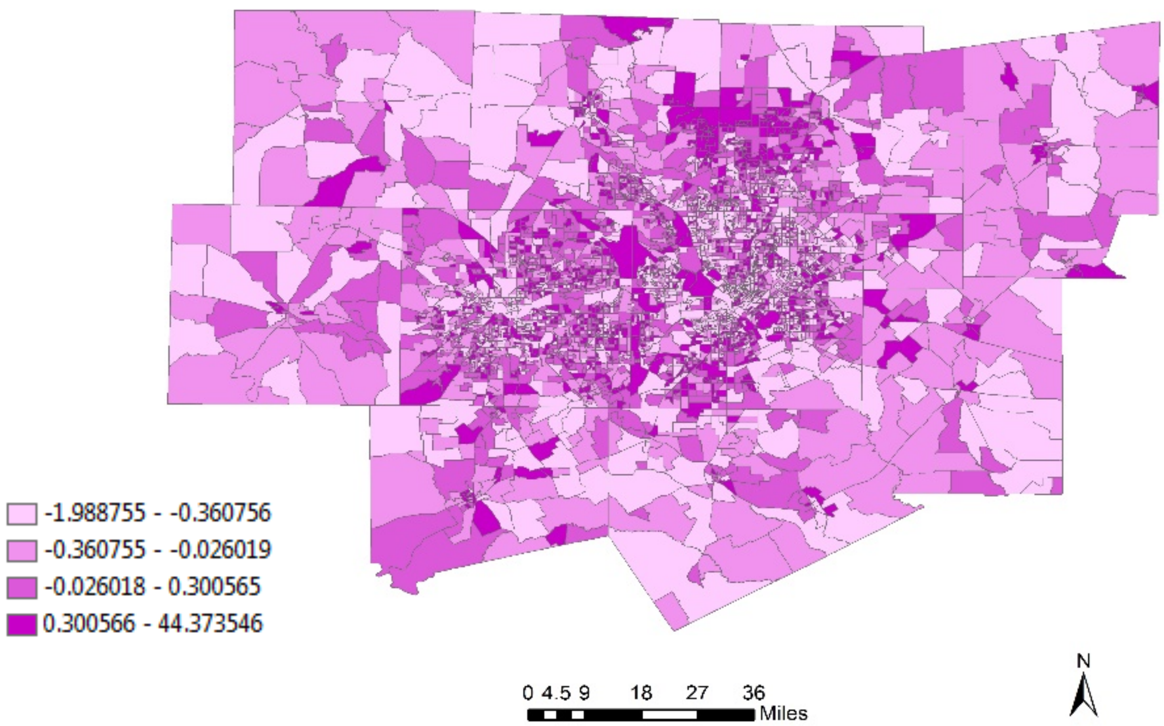

Figure 23. Jobs per household distribution in the DFW MSA

5.2 OLS Regression

The components obtained from PCA are then used as inputs for Ordinary Least Squares (OLS) regression. The R^2 is 0.48, and the fourth component, older adults, is not statistically significant in explaining the risk of respiratory diseases. This study examines autocorrelations among variables with the residuals of the OLS model using Moran's I test and finds positive spatial autocorrelations in the variables. Table 7 provides a summary of the OLS model results.

Table 7. Results of OLS model and Moran's I test-DFW

	OLS model			**Moran's I test**	
Variable	Coefficient	Standard Error	t-Statistic	Moran's I Index	z-Score
1: Demographic	0.002197 *	0.0004	5.3984	0.3694*	211.6744
2: Transit Access to Jobs	0.007717*	0.0004	18.9630	0.5945*	348.0639
3: Automobile Access	0.023060*	0.0004	56.6644	0.5155*	295.3785
4: Older Adults	-0.000009	0.0004	-0.0219	NA[1]	NA[1]
5: Miles Driven	0.001987*	0.0004	4.8820	0.0381*	22.0026
6: Active Population	0.002740*	0.0004	6.7327	0.1369*	78.6101
7: Auto Mode Use	0.003698*	0.0004	9.0868	0.0698*	40.1351
8: Low Income and Older Population	0.001652*	0.0004	4.0586	0.2205*	126.7185
9: Jobs Per Household	-0.002685*	0.0004	-6.5979	0.0241*	16.0875

*indicates a statistically significant p-value at .05 level.
R^2 = 0.48; adjusted **R^2** =0.47; Akaike information criterion = -18360.201; Koenker (BP) statistic = 460.923 (p-value = .0000*).

[1] We did not test Moran's I for component 4, because it was not statistically significant.

5.3 GWR

After reducing the feature dimensions using PCA and confirming spatial autocorrelations in OLS, the researchers estimate a GWR model to account for spatial relationships among variables. This study uses ArcGIS to develop the model. Table 8 shows the estimation results from the GWR model. The descriptive statistics of the estimates (4,128 sets) for the eight factors appear in Table 8. The results show that the GWR model significantly improves the overall fit compared to OLS where the median local R^2 is 0.83 and lower quartile is 0.70.

Table 8. Estimated GWR coefficients-DFW

Variable	Median	Max.	Min	Upper Quartile	Lower Quartile	SD
Constant	0.4743	1.3893	-1.622	0.4975	0.4520	0.0923
Component 1 Demographic	0.0008	0.1882	-0.3120	0.0070	-0.0044	0.0194
Component 2 Transit Access to Jobs	0.0021	1.0761	-2.0348	0.0167	-0.0102	0.0937
Component 3 Automobile Access	0.0026	0.4787	-1.0579	0.0177	-0.0075	0.0503
Component 5 Miles Driven	0.0008	0.1855	-0.2280	0.0059	-0.0035	0.0188
Component 6 Active Population	0.0004	0.1652	-0.1900	0.0051	-0.0040	0.0146
Component 7 Auto Mode Use	0.0002	0.1219	-0.2213	0.0034	-0.0029	0.0116
Component 8 Low Income and Older Population	0.0009	0.2982	-0.5904	0.0072	-0.0032	0.0260
Component 9 Jobs Per Household	0.0002	0.2862	-0.1629	0.0064	-0.0064	0.0207
Local R^2-Value	0.8342	0.9999	0.0976	0.9160	0.7078	0.1599

Diagnostic: R^2 = .98; adjusted R^2 = .92.
Component 4 has been excluded.

The spatial distribution of R-squared ranges from 9.76% to 99.99%. The explanatory power of the model is good for most of the counties in DFW; however, R^2 is particularly inconsistent in Dallas and Tarrant Counties because the demographic characteristics highly vary in these high population density areas. Only 25 percent of the BGs show R^2 values less than 70 percent, which confirms a good fit between the selected independent variables and RHQ used in the GWR model. Figure 24 shows the distribution of the local R^2 in the DFW region.

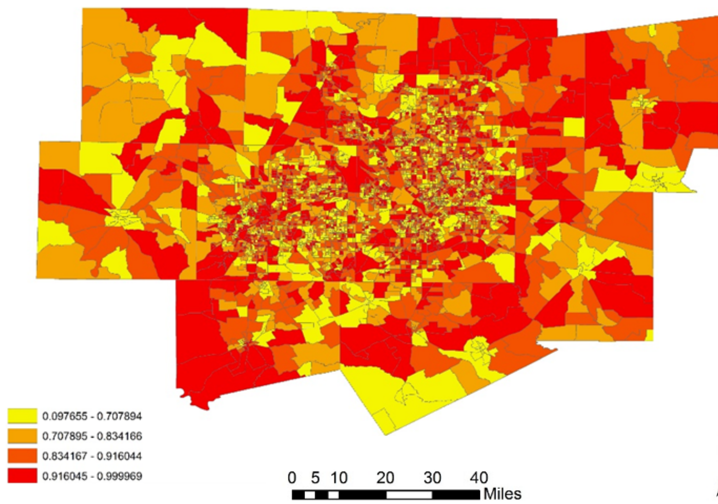

Figure 24. Spatial distribution of the determination coefficient, Local R^2 in the DFW MSA

The GWR model shows that the explanatory variables generally increase the RHQ because they have positive median values; however, some variations by geographic location occur. Figure 25 compares the spatial distribution of estimated coefficients of the eight components. Figure 26 shows the distribution of standard error in the DFW MSA where lower standard error values indicate higher variable significance.

The impact of demographic characteristics and automobile access appears significant in Dallas and Tarrant Counties, which experience the highest population and demographic variations. Southern and western areas including Ellis, Kaufman, Johnson, Parker and Wise Counties, which are characterized by low population density and higher residential land use, show the strong positive relationships between demographic characteristics and automobile access components and respiratory disease risks. However, the Dallas CBD shows an opposite pattern, which indicates that the impacts of demographic characteristics and automobile access remain stronger in the areas with lower population and employment density.

The authors identify a positive relationship between transit access to jobs and respiratory disease risk; a significant cluster appears in north Dallas County and southern Collin County while a smaller cluster appears near the Fort Worth CBD. A strong positive relationship between active population and respiratory disease risk appears outside Dallas and Tarrant Counties. This indicates that young population living in the center of MSA may be at a lower risk of respiratory diseases compared to those living in other areas. However, the impacts of this factor vary significantly in Dallas and Tarrant Counties because the local characteristics of the active population more strongly affect the respiratory disease risk in these areas with higher population and employment density.

This study also shows that a positive relationship between vehicle mile-driven and respiratory disease risk, especially in areas with less population and employment density such as Ellis, Parker and Denton Counties. In addition, the positive relationship between automobile mode use and respiratory disease risk appears significant in the rural counties with lower population density. Low-income and older population has a significant positive correlation with respiratory disease risks in most areas regardless of the local characteristics of the BGs. Lastly, jobs per household and respiratory disease risk show significant positive relationships in suburban areas and negative relationships in urban areas such as Dallas and Fort Worth.

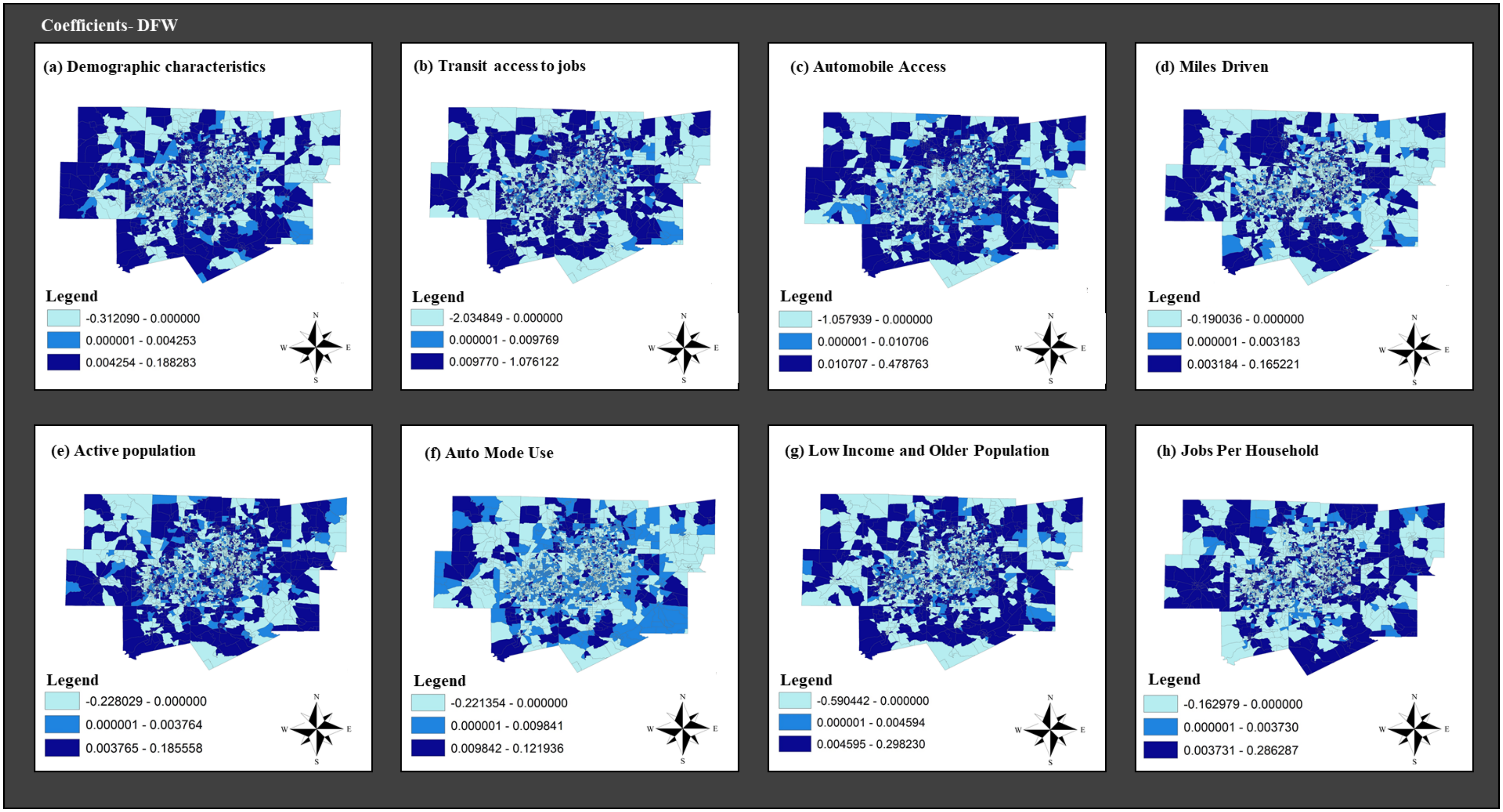

Figure 25. Spatial distribution of estimated coefficients in the DFW

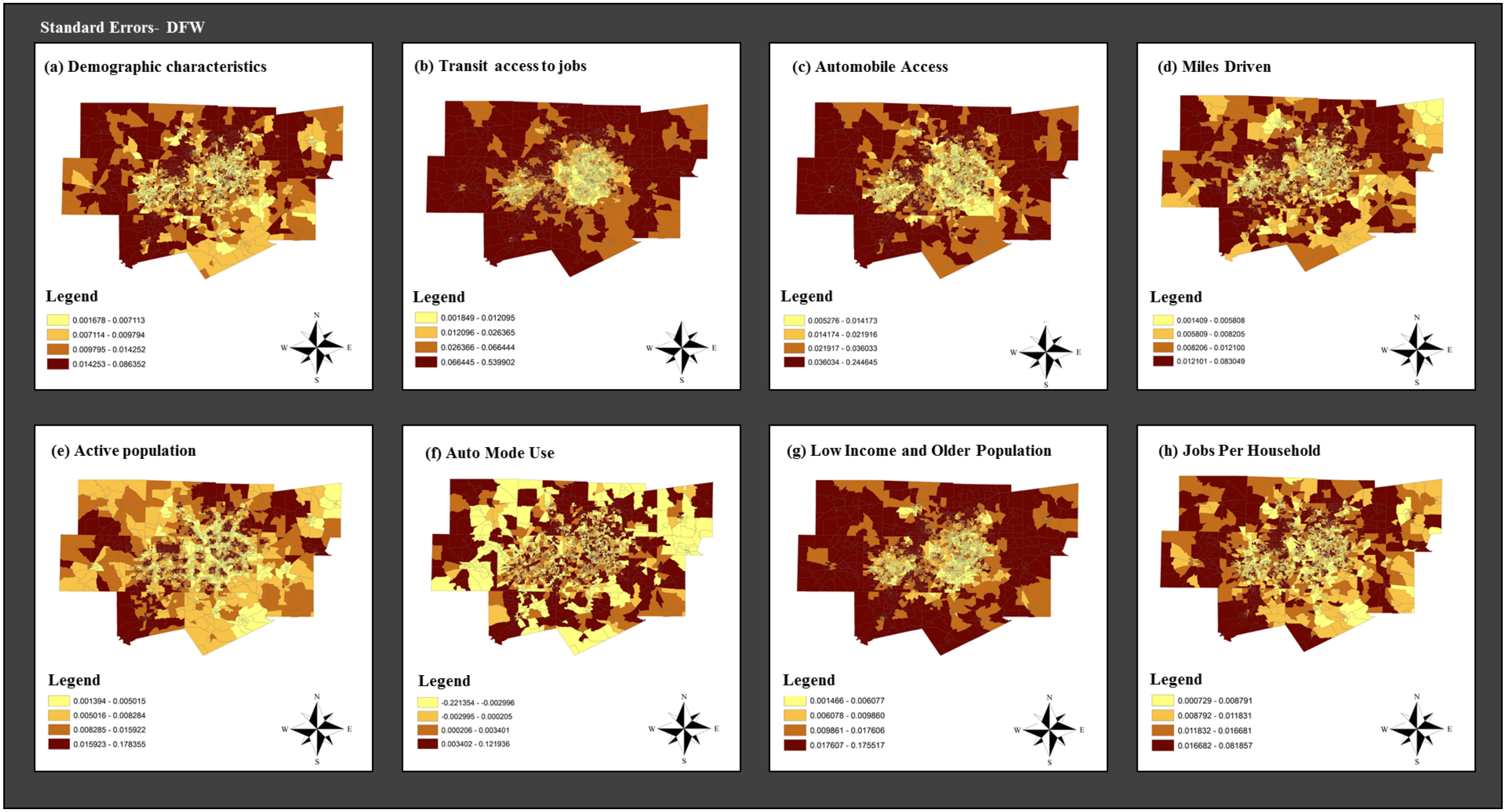

Figure 26. Spatial distribution of Standard Errors in the DFW MSA

6 Chapter 6: Los Angeles MSA Results

6.1 PCA

The PCA eliminates collinearity between the initial twenty-nine independent variables. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy is 0.731 and Bartlett's Test of Sphericity is significant; these results appear acceptable for PCA. The PCA process again produces nine components that explain 72.7% of the variation of the variables.

The first principal component, demographic characteristics, represents six factors including percent of Medicaid insurance coverage, percent of no insurance coverage, percentage of population under age 18, percentage of Hispanic population, percentage of white population and median income of BGs. The percentage of Medicaid insurance, no insurance coverage, percentage of population under age 18 and Hispanic population positively affect the component, while median income and percentage of white population negatively affect the component. The percentage of Medicaid insurance coverage represents the most significant factor in this component and shows a loading value of 0.85. The second component illustrates transit access to jobs using gross population density, workers using public transit, frequency of transit, aggregate frequency of service, and jobs within 45-minute transit commute. The third component describes workplace accessibility and includes jobs within 45-minute auto commute, working age population within a 45-minute travel time transit or auto commute, and distance from jobs to transit stop. The fourth component, older adults, includes percentage of the population over 65 years old, from 18-65 years old, with Medicare coverage, and with Medicare and Medicaid coverage. The fifth component addresses automobile access and includes percentage of the population using private vehicle for work trip and using bike or walk mode for work trip, and average auto ownership. The sixth component explains employment density using gross employment density and employment and household entropy. The seventh component relates to miles driven using AADT, percentage of auto-oriented intersections per square mile, and total road network density. Only one factor, jobs per household represents the eighth component, and the ninth component only contains the percentage of the population carpooling for the work trip. Table 9 shows the nine factors and their total variance explained by the components.

Table 9. Variance and Loadings explained by components obtained from PCA-LA

Component	Factor	loadings	Initial Eigenvalues	Rotation Sums of Squared Loadings
1: Demographic characteristics	EIR-21	0.85	27.07 (27.07 % Cumulative Variance)	15.9 (15.9 % Cumulative Variance)
	EIR-15	0.82		
	EIR-14	-0.80		
	EIR-24	-0.69		
	EIR-23	0.64		
	EIR-16	0.63		
2: Transit access to jobs	EIR-9	0.78	11.44 (38.51 % Cumulative Variance)	11.7 (27.6 % Cumulative Variance)
	EIR-8	0.75		
	EIR-12	0.72		
	EIR-28	0.63		
	EIR-1	0.46		
3: Workplace accessibility	EIR-10	0.84	6.44 (44.95 % Cumulative Variance)	11.0 (38.6 % Cumulative Variance)
	EIR-11	0.83		
	EIR-7	0.66		
	EIR-13	0.60		
4: Older adults	EIR-18	-0.85	6.15 (51.11 % Cumulative Variance)	7.5 (46.1 % Cumulative Variance)
	EIR-17	0.75		
	EIR-20	-0.72		
	EIR-22	-0.41		
5: Automobile access	EIR-29	0.68	5.90 (57.01 % Cumulative Variance)	7.4 (53.5 % Cumulative Variance)
	EIR-19	-0.61		
	EIR-26	-0.59		
6: Employment Density	EIR-4	0.80	4.31 (61.33 % Cumulative Variance)	5.5 (59.0 % Cumulative Variance)
	EIR-2	0.53		
7: Miles Driven	EIR-6	0.88	4.10 (65.42 % Cumulative Variance)	5.5 (64.5 % Cumulative Variance)
	EIR-25	0.64		
	EIR-5	0.56		
8: Jobs per household	EIR-3	-0.86	3.78 (69.20 % Cumulative Variance)	4.2 (68.7 % Cumulative Variance)
9: Carpooling	EIR-27	0.84	3.50 (72.71 % Cumulative Variance)	4.0 (72.7 % Cumulative Variance)

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization a.
a. Rotation converged in 12 iterations.

6.1.1 Demographic characteristics PC

First principal component contains six factors including percent of population with Medicaid insurance coverage, percent of population with no insurance coverage, percentage of Hispanic population, percentage of white population, average median income and percentage of under 18 population. As expected, the percentage of Medicaid insurance and no insurance coverage, percentage of Hispanic population and under 18 population have a positive relationship with the PC because they appear positively correlated with one another. Median income and percentage of the white population have a negative impact on the component because they indicate higher welfare and they have the opposite loading sign. Percent of population with Medicaid insurance coverage represents the most significant factor in this component with a loading value of 0.85. Figure 27 shows the distribution of this component in the Los Angeles MSA; generally, higher scores tend to be observed further from the coast.

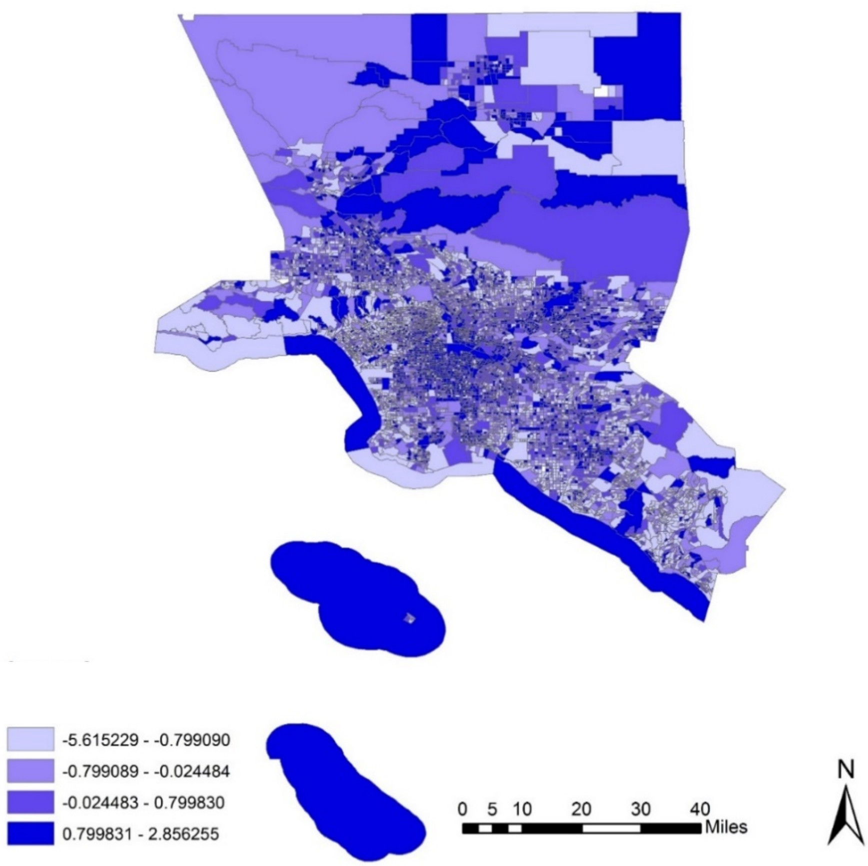

Figure 27. Spatial distribution of the Demographic characteristics in the Los Angeles MSA

6.1.2 Transit access to jobs PC

The second component includes aggregate frequency of transit service within 0.25 miles of block group boundary per hour, aggregate frequency of transit service per square mile, jobs within 45-minute transit commute, percentage of workers using public transit and gross population density. All of these indicators have a positive relationship with the component because they all relate to transit accessibility for employees. The most significant indicator is aggregate frequency of transit per square mile with a loading magnitude of 0.78. Figure 28 shows the distribution of this component in Los Angeles MSA; low scores appear to concentrate in northern Orange County and southern Los Angeles County, especially further from the coast.

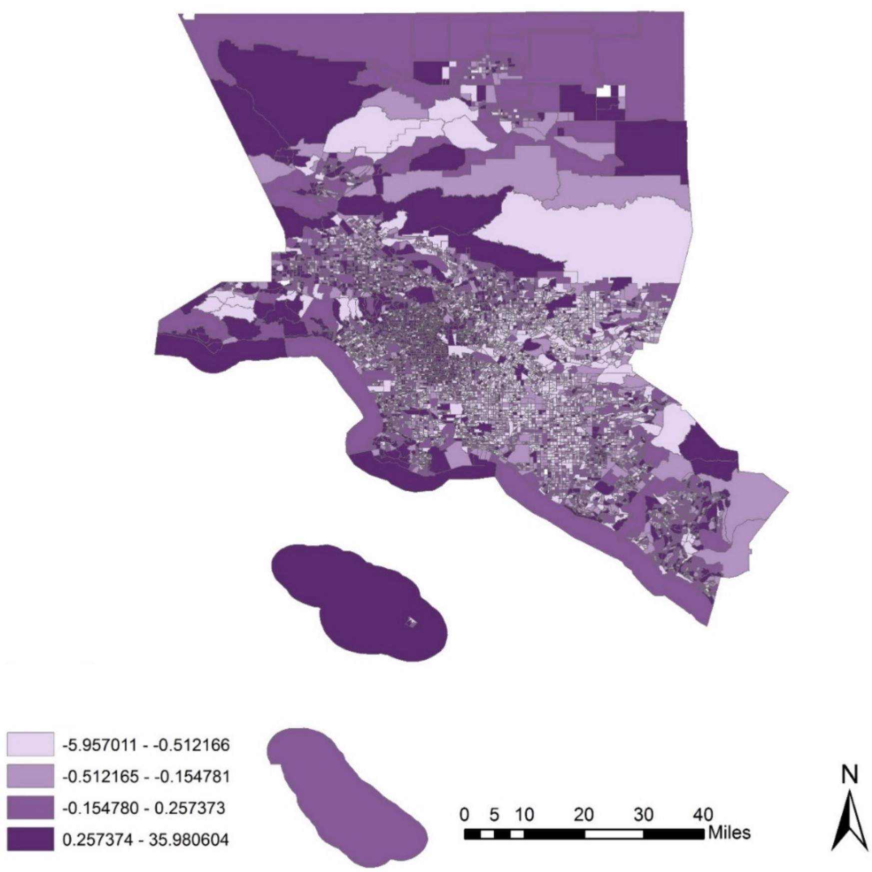

Figure 28. Spatial distribution of transit access to jobs in the Los Angeles MSA

6.1.3 Workplace accessibility PC

The third component contains four indicators (working age population within 45 minutes auto travel time, jobs within 45 minutes auto travel time, population within 45-min transit commute and existence of transit stop within ¾ miles of the population weighted centroid (binary). These indicators describe the ease of access to workplaces, and all of the factors increase accessibility. Jobs within 45 minutes auto travel time has the highest loading in this component (0.84). Figure 29 shows the distribution of this component in the Los Angeles MSA; lower scores appear in southern Orange County and rural areas of both counties.

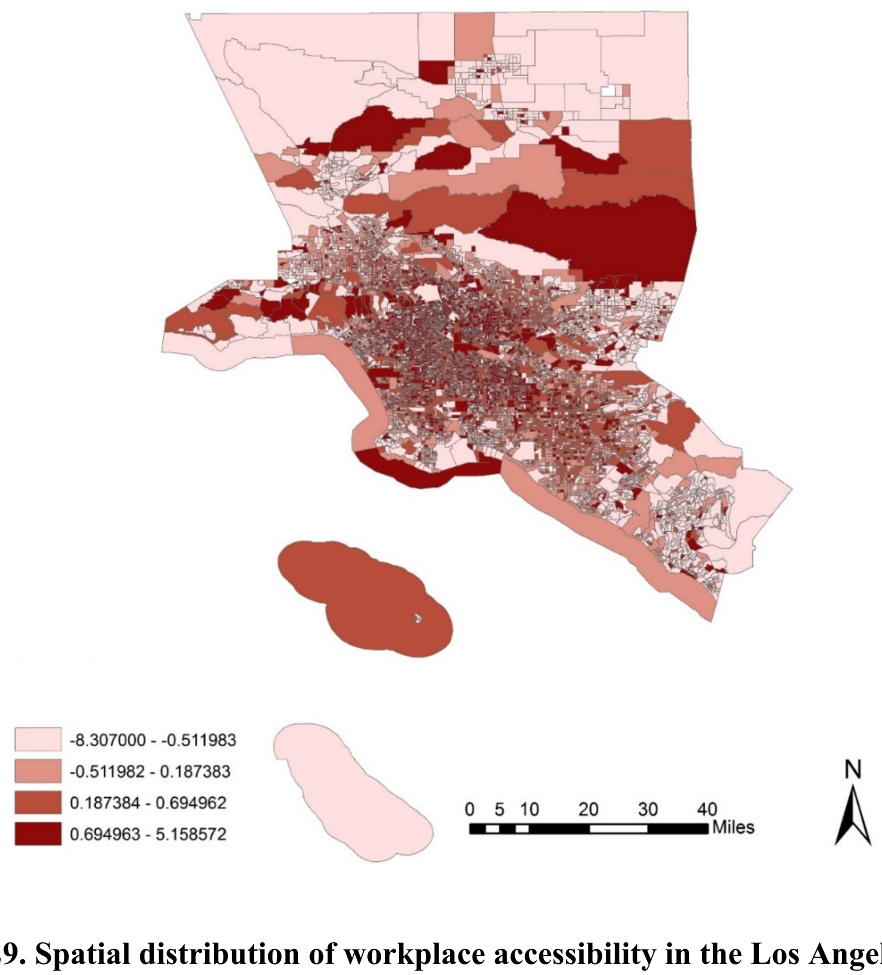

Figure 29. Spatial distribution of workplace accessibility in the Los Angeles MSA

6.1.4 Older adults PC

The fourth component includes the percentage of working age (18 to 65) and older adult (over 65) population, and the percentage of the population with Medicare coverage and both Medicare and Medicaid coverage. Obviously, the percentage of over 65 years old population and percentage of population with Medicare coverage have the same sign; the other factors have a neagive sign for this component because they appear negatively correlated with the first two indicators. The percentage of the population over 65 years old has the highest loading magnitude (-0.85). Figure 30 shows the distribution of this component in Los Angeles MSA; no clear pattern emerges for this component.

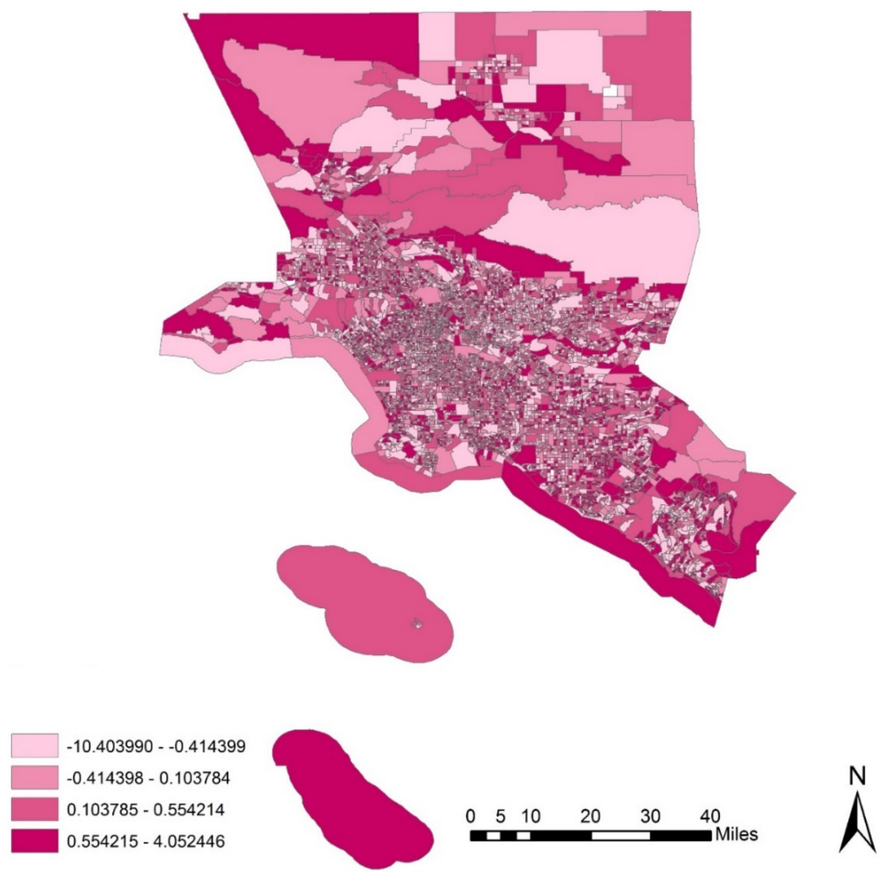

Figure 30. Spatial distribution of older adults in the Los Angeles MSA

6.1.5 Automobile Access PC

The fifth PC includes the percentage of workers using bike or walk for the work trip, average vehicle ownership and percentage of workers using private vehicle for the work trip. The first indicator obviously has a different sign than the other two factors because it describes lower automobile access. Figure 31 shows the distribution of this component in Los Angeles MSA; no clear pattern emerges for this component.

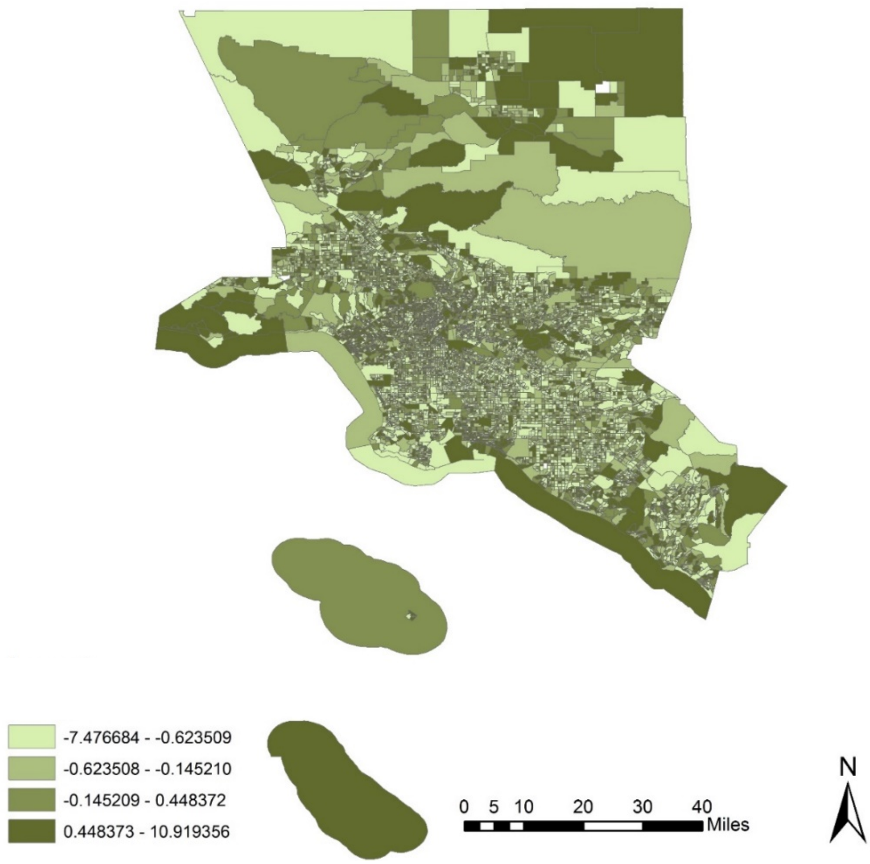

Figure 31. Spatial distribution of car accessibility in the Los Angeles MSA

6.1.6 Employment density PC

The sixth component describes employment density using employment and household entropy and gross employment density. The loading value of employment and household entropy is higher and is 0.8. Figure 32 shows the distribution of this component in Los Angeles MSA; this component appears to be well distributed based on local BG characteristics.

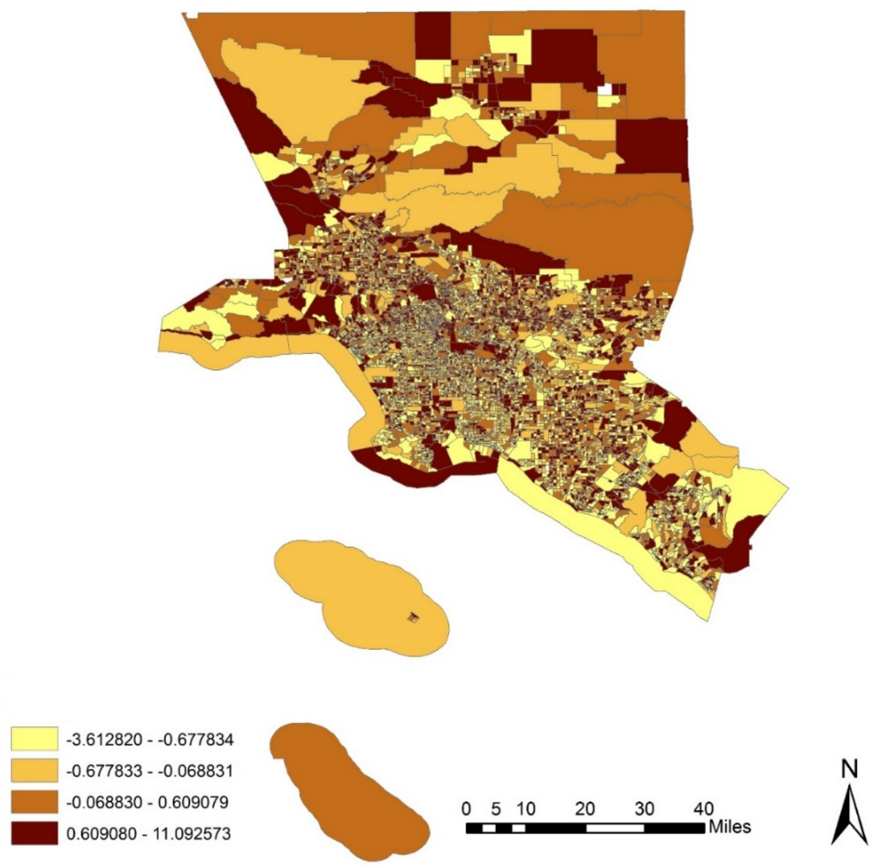

Figure 32. Spatial distribution of employment density in the Los Angeles MSA

6.1.7 Miles driven PC

The seventh component includes intersection density per square mile, total road network density and average annual daily traffic. They all have the same loading sign because they all contribute to higher vehicle use and therefore higher vehicle miles driven. The intersection density per square mile indicator has highest loading value in this component (0.88). Figure 33 shows the distribution of this component in Los Angeles MSA; this component correlates with the Caltrans network.

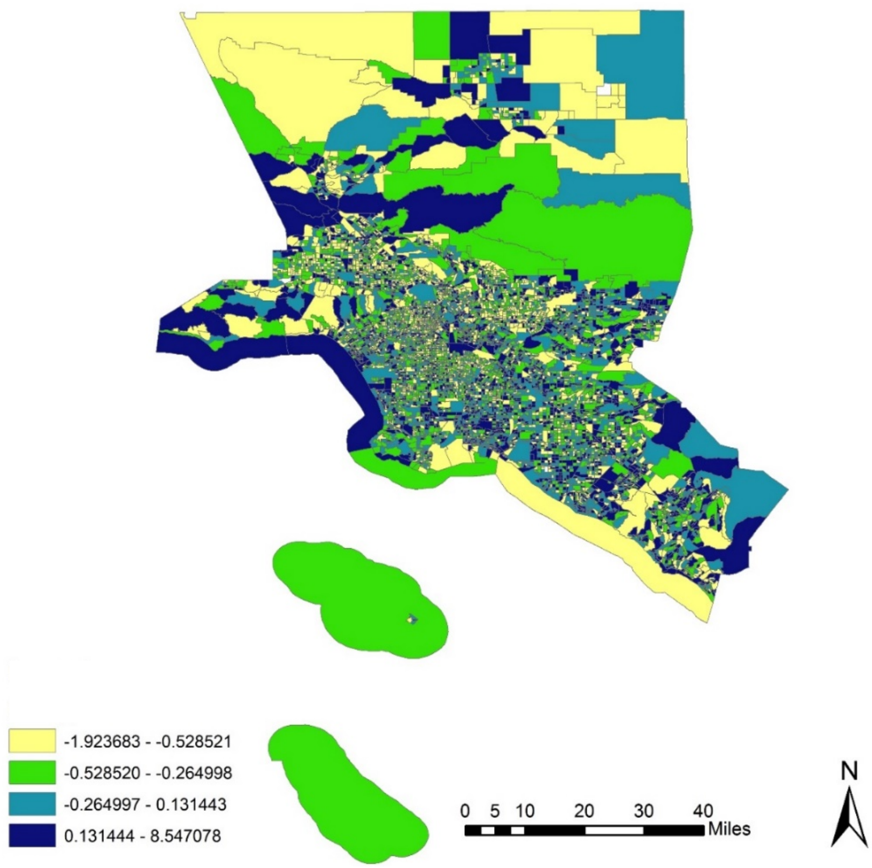

Figure 33. Spatial distribution of miles driven in the Los Angeles MSA

6.1.8 Jobs per household PC

The eighth component includes just one indicator, jobs per household, with a loading value of -0.86. Figure 34 shows the distribution of this component in the Los Angeles MSA, and no clear regional pattern emerges.

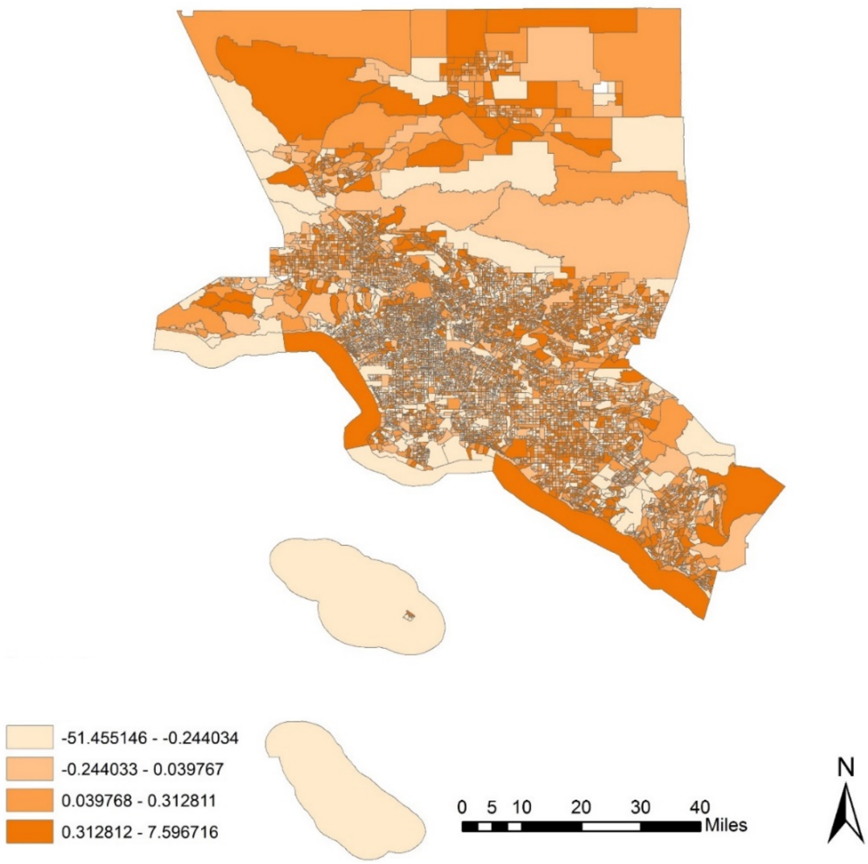

Figure 34. Spatial distribution of jobs per household in the Los Angeles MSA

6.1.9 Carpooling PC

The percentage of workers using carpooling for getting to the work forms the last component with the loading magnitude of 0.84. Figure 35 shows the distribution of this component in the Los Angeles MSA, and areas further from the CBD appear to have higher component values.

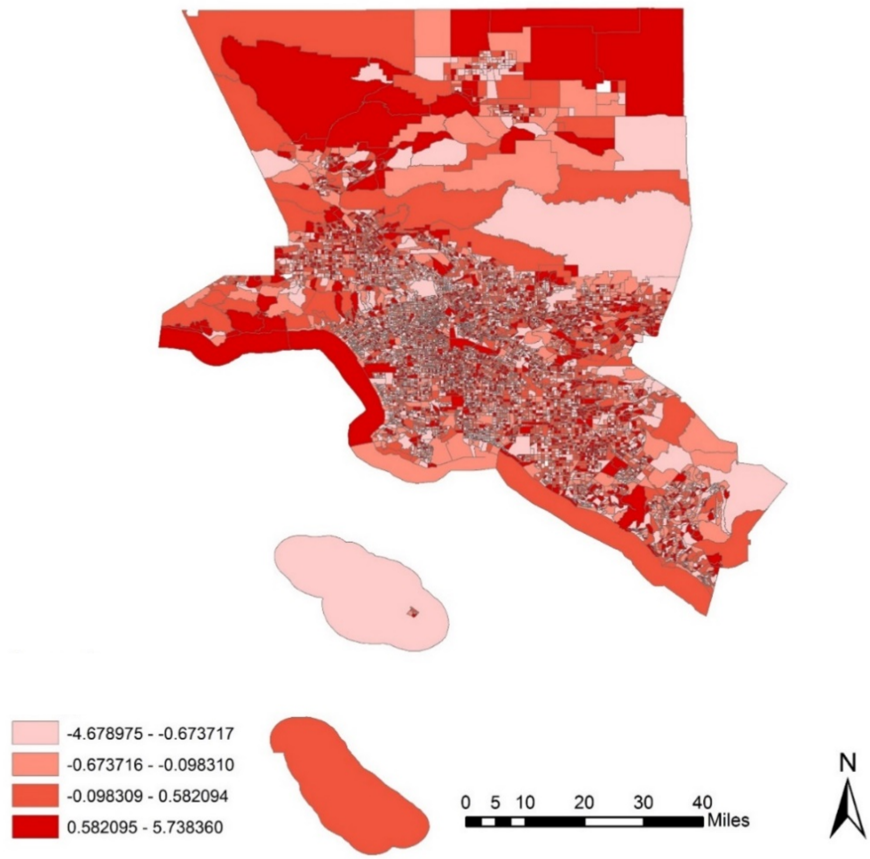

Figure 35. Spatial distribution of carpooling in the Los Angeles MSA

6.2 OLS Regression

Table 10 shows the summary result of the OLS model. The R^2 is 0.48, and the sixth component, which is “employment density”, is not statistically significant in explaining the risk of respiratory diseases in the Los Angeles MSA. Moran’s I test identifies positive spatial autocorrelations in the variables.

Table 10. Results of OLS model and Moran’s I test-LA

	OLS model			**Moran’s I test**	
Variable	Coefficient	Standard Error	t-Statistic	Moran’s I Index	z-Score
1: Demographics	0.0217 *	0.0006	31.7526	0.0699*	230.1568
2: Transit access to jobs	0.0246*	0.0006	36.0319	0.1157 *	385.3916
3: Workplace accessibility	0.0608*	0.0006	88.9070	0.1252*	411.6870
4: Older adults	0.0036*	0.0006	5.2767	0.0062*	20.9759
5: Automobile access	0.0068*	0.0006	10.0518	0.0340*	112.1484
6: Employment Density	0.0011	0.0006	1.7001	NA[1]	NA[1]
7: Miles Driven	0.0072*	0.0006	10.5645	0.0082 *	27.4375
8: Jobs per household	0.0096*	0.0006	14.1234	0.0044*	16.0915
9: Carpooling	0.0019*	0.0006	2.7985	0.0125*	41.5259

*Indicates a statistically significant p-value at .05 level.
R^2 = 0.48; adjusted R^2 =0.56; Akaike information criterion = -22389.03; Koenker (BP) statistic = 1075.005 (p-value = .0000*).

[1] We did not test Moran’s I for component 6, because it was not statistically significant.

6.3 GWR

The results show that the GWR model significantly improves the overall fit compared to the OLS where the median local R^2 is 0.79 and lower quartile is 0.66. The spatial distribution of R-squared ranges from 8.95% to 99.93%. The explanatory power of the model is good for most block groups of both Los Angeles and Orange Counties, but it is particularly good in northern Los Angeles County where the R-squared is mostly above 90%. Despite all the variations in R-squared value in the MSA, only about 30 percent of the BGs have R^2 values lower than 70 percent, which confirms a good fit between the selected independent variables and the RHQ in this area. Figure 36 shows the distribution of local R^2 in the LA MSA.

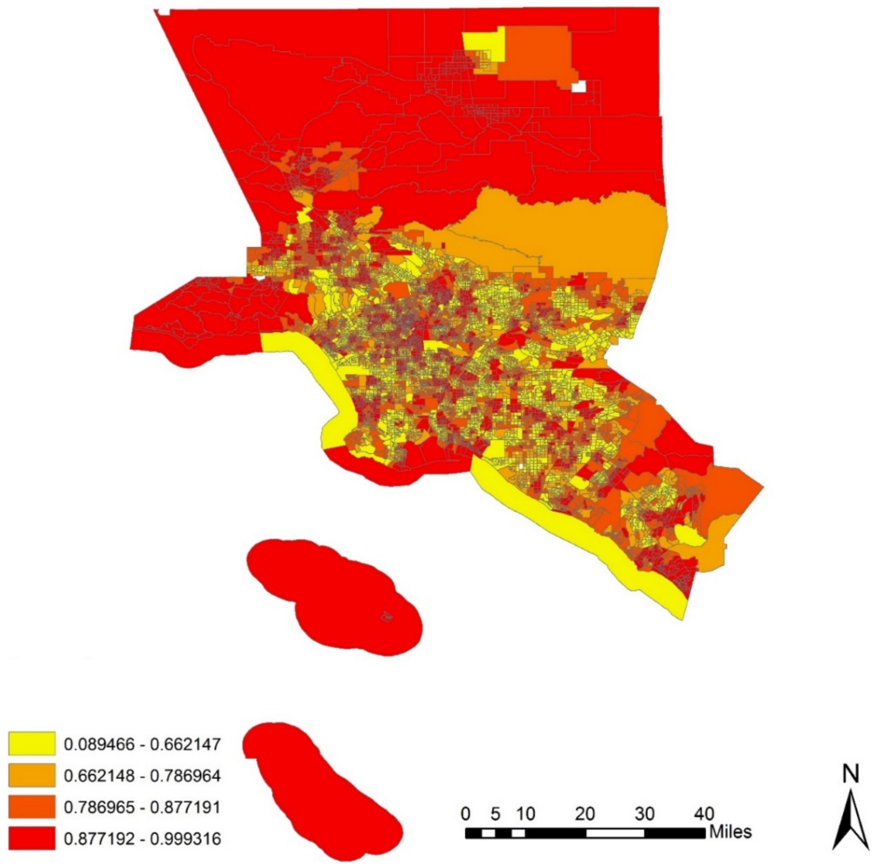

Figure 36. Spatial distribution of the determination coefficient, Local R^2 in the Los Angeles MSA

Table 11 presents the descriptive statistics of the estimates (8,246 sets) for the significant factors obtained from the previous steps. The median value of the coefficients again shows the positive effect of the explanatory variables on the RHQ while variations by geographic locations in Los Angeles and Orange Counties occur.

Table 11. Estimated GWR coefficients-LA

Variable	Median	Max.	Min	Upper Quartile	Lower Quartile	SD
Constant	0.5881	0.8292	0.3843	0.6152	0.5614	0.0456
Component 1 Demographics	0.0143	0.1536	-0.1140	0.0268	0.0020	0.0232
Component 2 Transit Access to Jobs	0.0276	0.4114	-0.3730	0.0409	0.0178	0.0520
Component 3 Workplace Accessibility	0.0181	0.3194	-0.1947	0.0240	0.0143	0.038028
Component 4 Older Adults	0.0142	0.1576	-0.1064	0.0177	0.0114	0.0196
Component 5 Automobile Access	0.0168	0.1934	-0.1954	0.0206	0.0134	0.0244
Component 7 Miles Driven	0.0137	0.2285	-0.1324	0.0198	0.0102	0.0234
Component 8 Jobs Per Household	0.0356	0.3229	-0.3459	0.0437	0.0286	0.0506
Component 9 Carpooling	0.0120	0.1207	-0.0803	0.0141	0.0103	0.0160
Local R^2-Value	0.7868	0.9996	0.0796	0.8771	0.6621	0.1541

Diagnostic: R^2 = .97; adjusted R^2 = .85.
Component 4 has been excluded.

Figure 37 compares the spatial distribution of estimated coefficients of eight components obtained from GWR and Figure 38 shows the distribution of standard error for each variable in Los Angeles and Orange Counties.

The impact of demographic characteristics and automobile access appears significant in central areas of Los Angeles, which experiences high population, employment and demographic variations. Also, the strongest positive relationships between automobile access component and respiratory disease risk typically occur in areas with lower population and employment density.

The positive relationship between transit access to jobs and respiratory disease risk mostly occurs in southern Los Angeles County and does not appear important in Orange County, which has a more limited and more suburban style transit system. According to the standard error distribution in the block groups, the impact of workplace accessibility on respiratory disease risk appears the highest in rural areas of north Los Angeles County and south Orange County.

A strong positive relationship between older adults' component and respiratory disease risk in the northern and southern more rural areas of the MSA exists while the relationship remains mostly negative in CBDs. In more rural areas, the older population experiences greater risks than the rest of the population while in the urbanized areas all of the population faces similarly severe risks. A strong positive relationship between miles driven and respiratory disease risk also appears in rural areas. The relationship between jobs per household and respiratory disease risk is highly varying between block groups but appears more positive in low population density areas. In contrast, the last component, carpooling, appears more important in the urbanized areas of Los Angeles County and it has strong positive relationship with respiratory disease risk in this area and throughout Orange County.

Figure 37. Spatial distribution of estimated coefficients in the LA MSA

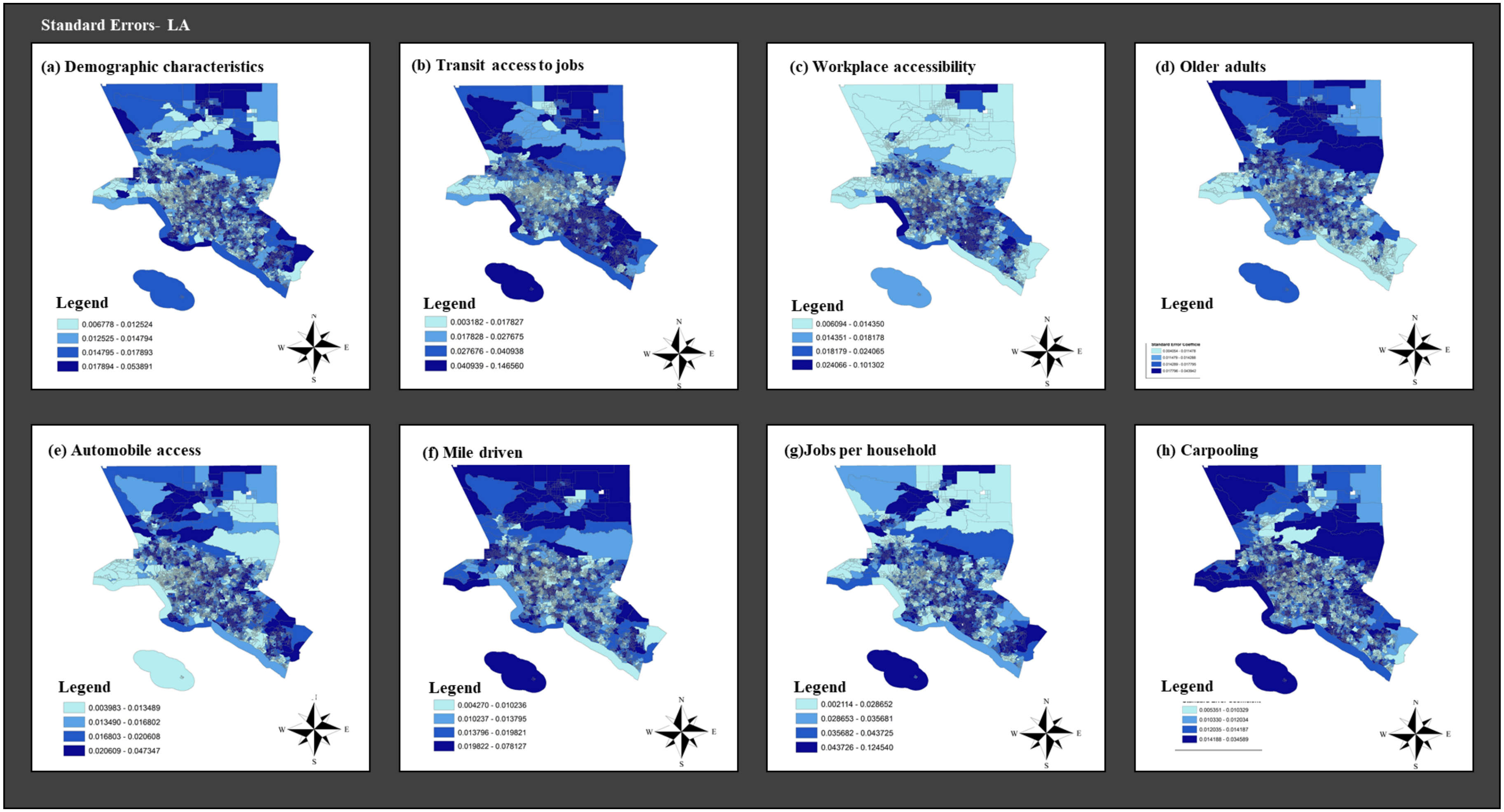

Figure 38. Spatial distribution of Standard Errors in the LA MSA

7 Conclusion

This study investigates the impacts of air pollution on public health along transit routes while controlling for other demographic, transportation and land use factors that could potentially contribute to air pollution and risk of respiratory disease. This study applies PCA, OLS and GWR to investigate the impacts of the selected demographic, land use and transportation factors on the occurrence risk of respiratory diseases in two of the biggest MSAs in the US, Dallas-Fort Worth and Los Angeles, by considering respiratory hazard quotient as the dependent variable. As many of these variables cause multicollinearity problem within the model, the study applies PCA to eliminate multicollinearity and group the initial variables into fewer components, which could be used as OLS and GWR inputs.

The results of the PCA explain about 73 percent of the variation in the dependent variable in both the DFW and Los Angeles MSA using nine components. The OLS model results indicate one of the components appears insignificant for each MSA (old adults in DFW and employment density in Los Angeles), and spatial autocorrelations appear significant. Finally, the researchers use GWR to address the spatial autocorrelations. The GWR results show an overall positive effect of all variables on the independent variable with a median R^2 value of 0.83, compared to 0.48 from OLS in DFW and 0.79 (GWR) and 0.48 (OLS) in Los Angeles.

While demographic characteristics appear the most important determinant of aggregate respiratory disease risk in both MSAs, transit access to jobs represents the second most important component. This indicates that after controlling for demographic effects, higher transit access to jobs clearly indicates a greater risk of respiratory disease, which directly confirms the research question and hypothesis. Those living along transit corridors and likely in transit-oriented development face a greater risk of respiratory disease. While other components experience greater spatial variations in both MSAs, the transit access to jobs displays a clear pattern and significance.

While the specific variables in the components vary slightly between the DFW and Los Angeles MSAs, the components largely measure the same effects as can be noted in their descriptions. The importance of similar effects in both MSAs indicates that large MSAs may experience similar impacts related to transit access to jobs, automobile access, and vehicle miles traveled. The results of the GWR also show the varying effect of chosen variables on the risk of respiratory disease in different area of DFW and Los Angeles. This can be explained by the local characteristics of each factor in different block groups or areas within the MSAs. Analyzing and comparing the results of GWR maps in these two MSAs show that the population living in rural areas of the metropolitan area appear more affected by transportation and land use factors. Demographic and socioeconomic characteristics appear to also play a significant role in risk, especially in urban and suburban BGs.

The respiratory risks in high transit areas may indicate the need for new policies and building codes to provide greater protection to the residents living in those areas. This study also suggests that departments of transportation and local environmental agencies can use the results of a GWR model rather than global models to analyze the key factors and indicators (i.e. land use and transportation) that impact the risk of health issues in different locations. While this study includes two large MSAs further studies in other major MSAs can be useful to achieve a comprehensive and reliable model that confirms transit access to jobs as an indicator of respiratory risk in large urban areas. Future studies should use the same methods to investigate other health outcomes negatively impacted by transportation.

8 Chapter 8: RDC Process

The project included submitting another proposal to the US Centers for Disease Control (CDC) Research Data Center (RDC). This section explains the steps necessary to access restricted National Health Interview Survey (NHIS) data like geocodes. This chapter identifies the current state of the process using process documentation and describes lessons learned.

8.1 Process Scope

This study seeks to investigate the role that individual (e.g., age, income, race/ethnicity, smoking status, diet, physical activity, health status) factors may play in confounding or modifying the health effects of traffic-related air pollution. The study will also explore aggregating these individual level factors to create socio-economic profiles and indicators of health risk due to traffic related air pollution along transit routes in major metropolitan statistical areas such as Dallas- Fort Worth (TX), Los Angeles (CA), Chicago (IL), Miami (FL), and Boston (MA).

8.2 Process Steps

Step 1 (outside of RDC): Emissions levels will be first estimated in a grid system (0.6mi * 0.6 mi) outside of the RDC. The estimated emissions (in a grid level) will be assigned to the center coordinate of each grid system and prepared in a SAS or Stata format.

Step 2 (in RDC): NHIS and HUD restricted LAT (Latitude) and LON (Longitude) data will be first linked to locate individual household in NHIS. Then, the geocoded emissions (prepared outside of the RDC) will be merged with each household based on their coordinates.

For assessing the health impact with restricted variables from NHIS, the research team will bring transportation related air pollution exposure for CO, NO_2 and PM along transit lines for the metropolitan areas of Dallas- Fort Worth (TX), Los Angeles (CA), Chicago (IL), Miami (FL), and Boston (MA). The researchers will apply (i) EPA's Motor Vehicle Emission Simulator (MOVES) to estimate the total emission rate for the various combinations of vehicle fleet and traffic operations, and (ii) R-Line model to identify the exposure level at locations near roads using dispersion modeling. The air pollution exposure outcomes will be estimated in a grid cell and merged with NHIS restricted variables at the RDC. The research team will extract all of the NHIS data that falls within the geographical boundaries where any emissions level (e.g., CO, PM, NO2) was greater than zero. The extracted NHIS data will be screened for the presence of key health and control variables. The researchers will group and create variable profiles for the individual SES and health related factors (i.e., smoking, physical activities). The study also requires the data to be randomly split into two samples and tested for representativeness. If any of the data records must be removed from consideration because it is missing some of the variables, these reduced samples must also be split and tested for representativeness.

Each of the selected health impacts will be modeled separately, but one overall health impact model will be developed as well. The location types (i.e., urban or suburban) may be modeled separately or considered as potential variables in the models. The independent variables selected as control variables will be included with the overall and transportation related emission exposures. The health impact modeling will focus on two approaches:

a. Disaggregate logistic regression models

b. Treed regression models, which combine a logistic regression model with CART.

All models will be validated to avoid overfitting the data to the model structure. The models can be used to characterize the risks related to traffic-related air pollution for different socioeconomic profiles. These profiles will focus on transit dependent populations; however, the socioeconomic profiles of populations targeted by transit-oriented development (TOD) will also be considered. Another set of profiles will be based exclusively on the profiles identified by CART.

8.3 Process Inputs

1. NHIS 2016

Family, Person, Sample Adult, Sample Child, Income Imputation, Cancer, Adult Functioning & Disability, Family Disability, Quality of Life, HUD file

2. Restricted Data: - LAT (from HUD file) = Latitude (in decimal format with up to 6 decimal precision) of residence will be used to examine the effect of air pollution on individuals' health - LON (from HUD file) = Longitude (in decimal format with up to 6 decimal precision) of residence will be used to examine the effect of air pollution on individuals' health

3. Non-NCHS Data: Geocoded Emissions level (CO, PM2.5, PM10, and NO2) will be provided by the team in a SAS or Stata format.

4. Merge Variables: (i) Merge NHIS public data with HUD restricted data Use the variable HHX to link NHIS public data to HUD restricted data.

(ii) Merge geocoded emissions level data with NHIS data

8.4 Process Outputs

This study will focus on assessing health differences for a panel of individuals that participated in the NHIS. The abundance of data within the NHIS will allow the research team to control for individual and household level factors that may also contribute to the adverse health impacts. Given the structure of the NHIS and the statistical analyses considered in this research, the study explores various health outcomes to see if traffic-related air pollution has a significant effect on the rate of associated diseases. This investigation will be exploratory in nature to see if traffic-related air pollution may contribute to unexpected health outcomes not only for respiratory diseases but also low birth weight, and diabetes. The research team will also explore the correlation between negative health outcomes and socioeconomic indicators (e.g. income, race/ethnicity). These relationships will enable public health professionals, urban planners and other policy makers to locate at-risk communities and investigate the potential impacts of remedial activities.

8.5 Lesson Learned

Health information from NHIS is in restricted data category, thus the research team prepared a proposal with the following process:

1. Proposal Format: The RDC proposal has been designed to effectively summarize the required data needs related to the diseases such as asthma, lung cancer, type II diabetes and low birth weight. The completed proposal followed the RDC proposal format and it explained the need for restricted variables, the analytic plan, and the plan for reporting results.

2. Student Advisor Agreement: As part of the required documents, student advisor agreement form has been filled out and signed by the research team and added to the proposal.

3. Creating the Data Dictionary: Since the data dictionary was an essential part of the RDC proposal, the research team spent a significant amount of time to prepare the data dictionary. The

team identified factors from the comprehensive list of all NHIS variables and then organized the selected variables into three different categories of public data, restricted data and non-NCHS (National Center for Health Statistics) data.

After preparing the material we UTA submitted the proposal to the Research Data Center (RDC) on December 8, 2018, but the initial submission required revisions. After two rounds of editing requested by the RDC review committee, UTA finalized and submitted RDC proposal on February 7, 2019 and RDC approved it on April 24, 2019.

Confidentiality and Disclosure

Maintaining confidentiality is the primary objective of the Research Data Center. Therefore, the Disclosure Manual outlines the rules and procedures that are required to protect the data and prevent disclosure of confidential information have been reviewed by the research group.

Fees and Invoicing

RDC charges for data processing, and the fee is disclosed after acceptance of the proposal.

Limitations

The research team was unaware of the special sworn status (SSS) procedure required to gain access to the data center. This remains a stumbling block because all individuals seeking SSS must have resided in the United States for more than three years, and very few graduate students meet this criterion. Furthermore, when the research team was initially made aware of the SSS requirement, the SSS processing had been frozen due to the government shutdown.

9 References

1. Adar, S. D., and J. D. Kaufman. "Cardiovascular disease and air pollutants: evaluating and improving epidemiological data implicating traffic exposure." *Inhalation Toxicology* 19, no. sup1 (2007): 135-149.
2. Armstrong, Jennifer M., and Ata M. Khan. "Modelling urban transportation emissions: role of GIS." *Computers, Environment and Urban Systems* 28, no. 4 (2004): 421-433.
3. Badland, Hannah, Carolyn Whitzman, Melanie Lowe, Melanie Davern, Lu Aye, Iain Butterworth, Dominique Hes, and Billie Giles-Corti. "Urban liveability: emerging lessons from Australia for exploring the potential for indicators to measure the social determinants of health." *Social science & medicine* 111 (2014): 64-73.
4. Badoe, Daniel A., and Eric J. Miller. "Transportation–land-use interaction: empirical findings in North America, and their implications for modeling." *Transportation Research Part D: Transport and Environment* 5, no. 4 (2000): 235-263.
5. Bailey, Linda, Patricia L. Mokhtarian, and Andrew Little. *The broader connection between public transportation, energy conservation and greenhouse gas reduction*. No. TCRP Project J-11/Task 3. Fairfax, VA: ICF International, 2008.
6. Baldauf, Richard, Eben Thoma, Michael Hays, Richard Shores, John Kinsey, Brian Gullett, Sue Kimbrough et al. "Traffic and meteorological impacts on near-road air quality: Summary of methods and trends from the Raleigh near-road study." *Journal of the Air & Waste Management Association*58, no. 7 (2008): 865-878.
7. Barone-Adesi, Francesco, Jennifer E. Dent, David Dajnak, Sean Beevers, H. Ross Anderson, Frank J. Kelly, Derek G. Cook, and Peter H. Whincup. "Long-term exposure to primary traffic pollutants and lung function in children: cross-sectional study and meta-analysis." *PloS one* 10, no. 11 (2015): e0142565.
8. Bartholomew, Keith, and Reid Ewing. "Land use–transportation scenarios and future vehicle travel and land consumption: A meta-analysis." *Journal of the American Planning Association* 75, no. 1 (2008): 13-27.
9. Bullard, Robert Doyle, and Glenn S. Johnson. Just transportation: dismantling and class barriers to mobility. New Society, 1997.
10. Cervero, Robert, and Kara Kockelman. "Travel demand and the 3Ds: Density, diversity, and design." *Transportation Research Part D: Transport and Environment* 2, no. 3 (1997): 199-219.

11. Cesaroni, Giulia, Chiara Badaloni, Valeria Romano, Eugenio Donato, Carlo A. Perucci, and Francesco Forastiere. "Socioeconomic position and health status of people who live near busy roads: the Rome Longitudinal Study (RoLS)." *Environmental health* 9, no. 1 (2010): 41.
12. Chiou, Yu-Chiun, Rong-Chang Jou, and Cheng-Han Yang. "Factors affecting public transportation usage rate: Geographically weighted regression." *Transportation Research Part A: Policy and Practice* 78 (2015): 161-177.
13. Colvile, R. N., E. J. Hutchinson, J. S. Mindell, and R. F. Warren. "The transport sector as a source of air pollution." *Atmospheric environment* 35, no. 9 (2001): 1537-1565.
14. Dockery, Douglas W., C. Arden Pope, Xiping Xu, John D. Spengler, James H. Ware, Martha E. Fay, Benjamin G. Ferris Jr, and Frank E. Speizer. "An association between air pollution and mortality in six US cities." *New England journal of medicine* 329, no. 24 (1993): 1753-1759.
15. Edwards, Jerry Lee. "RELATIONSHIPS BETWEEN TRANSPORTATION ENERGY CONSUMPTION AND URBAN SPATIAL STRUCTURE." (1976): 3505-3505.
16. Ellaway, Anne, Sally Macintyre, Rosemary Hiscock, and Ade Kearns. "In the driving seat: psychosocial benefits from private motor vehicle transport compared to public transport." Transportation Research Part F: Traffic Psychology and Behaviour 6, no. 3 (2003): 217-231.
17. Ewing, Reid H., Rolf Pendall, and Donald DT Chen. Measuring sprawl and its impact. Vol. 1. Washington, DC: Smart Growth America, 2002.
18. Ewing, Reid, and Robert Cervero. "Travel and the built environment: a meta-analysis." *Journal of the American Planning Association* 76, no. 3 (2010): 265-294.
19. Ewing, Reid, Lawrence D. Frank, Jim Chapman, and Richard Kreutzer. "Understanding the relationship between public health and the built environment: A report prepared for the LEED-ND Core Committee." (2006).
20. Ewing, Reid, Michael Greenwald, Ming Zhang, Jerry Walters, Mark Feldman, Robert Cervero, Lawrence Frank, and John Thomas. "Traffic generated by mixed-use developments—six-region study using consistent built environmental measures." *Journal of Urban Planning and Development* 137, no. 3 (2010): 248-261.
21. Favarato, Graziella, H. Ross Anderson, Richard Atkinson, Gary Fuller, Inga Mills, and Heather Walton. "Traffic-related pollution and asthma prevalence in children. Quantification of associations with nitrogen dioxide." *Air Quality, Atmosphere & Health* 7, no. 4 (2014): 459-466.
22. Frank, Lawrence D., James F. Sallis, Terry L. Conway, James E. Chapman, Brian E. Saelens, and William Bachman. "Many pathways from land use to health: associations between neighborhood

walkability and active transportation, body mass index, and air quality." Journal of the American planning Association 72, no. 1 (2006): 75-87.

23. Gehring, Ulrike, Olena Gruzieva, Raymond M. Agius, Rob Beelen, Adnan Custovic, Josef Cyrys, Marloes Eeftens et al. "Air pollution exposure and lung function in children: the ESCAPE project." *Environmental health perspectives* 121, no. 11-12 (2013): 1357-1364.
24. Geurs, Karst T., and Bert Van Wee. "Accessibility evaluation of land-use and transport strategies: review and research directions." *Journal of Transport geography* 12, no. 2 (2004): 127-140.
25. Giles-Corti, Billie, Anne Vernez-Moudon, Rodrigo Reis, Gavin Turrell, Andrew L. Dannenberg, Hannah Badland, Sarah Foster et al., "City planning and population health: a global challenge." *The lancet* 388, no. 10062 (2016): 2912-2924.
26. Health Effects Institute. Panel on the Health Effects of Traffic-Related Air Pollution. *Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects*. No. 17. Health Effects Institute, 2010.
27. Kopits, Elizabeth, and Maureen Cropper. *Traffic fatalities and economic growth*. The World Bank, 2003.
28. Krzyżanowski, Michał, Birgit Kuna-Dibbert, and Jürgen Schneider, eds. *Health effects of transport-related air pollution*. WHO Regional Office Europe, 2005.
29. Litman, Todd. "Developing indicators for comprehensive and sustainable transport planning." *Transportation Research Record* 2017, no. 1 (2007): 10-15.
30. Litman, Todd. "Transportation and public health." *Annual review of public health* 34 (2013): 217-233.
31. Macintyre, Sally, and Anne Ellaway. "Neighborhoods and health: an overview." *Neighborhoods and health* 20 (2003): 42.
32. Marquez, Leorey O., and Nariida C. Smith. "A framework for linking urban form and air quality." *Environmental Modelling & Software* 14, no. 6 (1999): 541-548.
33. Medina, Isabel Canete, G. Schattanek, and F. Nichols Jr. "A framework for integrating information systems in air quality analysis." In *Proceedings of URISA*, p. 339. 1994.
34. Menard, Scott. *Applied logistic regression analysis*. Vol. 106. Sage, 2002.
35. Montgomery County, Maryland. (1989). *A policy vision: Centers and trails.* Silver Spring: Maryland-National Capital Park & Planning Commission. Retrieved July 31, 2007, from http://content.lib.utah.edu/cdm4/
36. NIEHS. (2016, July 21). National Institute of Environmental Health Sciences. Retrieved from https://www.niehs.nih.gov/health/topics/agents/air-pollution/
37. Ramsey, Kevin, and Alexander Bell. "Smart location database." *Washington, DC* (2014).

38. Sallis, James F., Fiona Bull, Ricky Burdett, Lawrence D. Frank, Peter Griffiths, Billie Giles-Corti, and Mark Stevenson. "Use of science to guide city planning policy and practice: how to achieve healthy and sustainable future cities." *The lancet*388, no. 10062 (2016): 2936-2947.
39. Samaranayake, Samitha, Steven Glaser, David Holstius, Julien Monteil, Ken Tracton, Edmund Seto, and Alexandre Bayen. "Real-Time Estimation of Pollution Emissions and Dispersion from Highway Traffic." *Computer-Aided Civil and Infrastructure Engineering* 29, no. 7 (2014): 546-558.
40. Selby, Brent, and Kara M. Kockelman. "Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression." *Journal of Transport Geography* 29 (2013): 24-32.
41. Stevenson, Mark R., Konrad D. Jamrozik, and James Spittle. "A case-control study of traffic risk factors and child pedestrian injury." *International journal of epidemiology* 24, no. 5 (1995): 957-964.
42. Syed, Samina T., Ben S. Gerber, and Lisa K. Sharp. "Traveling towards disease: transportation barriers to healthcare access." *Journal of community health* 38, no. 5 (2013): 976-993.
43. Vallero, Daniel A. *Fundamentals of air pollution*. Academic press, 2014.
44. Venkatram, Akula, Vlad Isakov, Eben Thoma, and Richard Baldauf. "Analysis of air quality data near roadways using a dispersion model." *Atmospheric Environment* 41, no. 40 (2007): 9481-9497.
45. Wang, Bin, Donghua Xu, Zhaohai Jing, Dawei Liu, Shengli Yan, and Yangang Wang. "Effect of long-term exposure to air pollution on type 2 diabetes mellitus risk: a systematic review and meta-analysis of cohort studies." (2014): R173-82.

    Wang, Quan, Jian Ni, and John Tenhunen. "Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems." *Global ecology and biogeography* 14, no. 4 (2005): 379-393.
46. Wegener, Michael. "Overview of land-use transport models." *Handbook of transport geography and spatial systems* 5 (2004): 127-146.
47. Weiner, Edward. *Urban transportation planning in the United States*. Praeger, 1987.
48. World Health Organization. International classification of functioning, disability, and health. Available at: http://www3.who.int/icf/icftemplate.cfm. Accessed April 22, 2003.
49. Zhao, Jie, Wenlei Wang, and Qiuming Cheng. "Application of geographically weighted regression to identify spatially non-stationary relationships between Fe mineralization and its controlling factors in eastern Tianshan, China." *Ore Geology Reviews* 57 (2014): 628-638.
50. Zhu, Yifang, William C. Hinds, Seongheon Kim, and Constantinos Sioutas. "Concentration and size distribution of ultrafine particles near a major highway." *Journal of the air & waste management association* 52, no. 9 (2002): 1032-1042