5-2019

Incorporating The Home Address of Road Users Involved in Traffic Crashes in Road Safety Analysis

Amin Mohamadi Hezaveh
University of Tennessee, amohamad@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Amin Mohamadi Hezaveh entitled "Incorporating The Home Address of Road Users Involved in Traffic Crashes in Road Safety Analysis." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Civil Engineering.

Christopher R. Cherry, Major Professor

We have read this dissertation and recommend its acceptance:

Asad Khattak, Trond Nordfjærn, Candace Brakewood, Lee Han

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Incorporating The Home Address of Road Users Involved in Traffic Crashes in Road Safety Analysis

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Amin Mohamadi Hezaveh
May 2019

DEDICATION

This dissertation work is dedicated to my wife, Haniyeh, who has been a constant source of encouragement and support during my challenging PhD program and every second of my life. I am truly thankful for having you in my life.

ABSTRACT

Traditionally, road safety metrics are measured at the location of the crash and its surrounding area. For example, if a crash occurs at an intersection, depending on the scope of the study, the researchers or practitioners may count crashes at intersection level, corridor level, or at a coarser geographic area such as Traffic Analysis Zone (TAZ), city level, or county level. Attributing crash to the location of the crash helps us learn about the relationship between road, environment, traffic, and weather and road safety. Based on this practice, several countermeasures have been developed to prevent crashes or reduce the severity of traffic crashes. As a result, a large body of road safety literature was allocated to road and geometry design and their effect on traffic crashes. In my dissertation, I set out to take a more epidemiological approach to road safety analysis, looking at factors such as social geography and travel behavior surrounding the home addresses of the road users involved in traffic crashes –i.e., a Home-Based Approach. Knowing more about the role of a human factor origin, and expressly sociodemographic, and travel behavior could help us to understand road safety from a different perspective that enables researchers and road safety practitioners to target individuals with proper countermeasure and intervention with the intention of reducing crash risk or eliminating aberrant behaviors of road users. My dissertation consists of five chapters. I explored different applications of the Home-Based Approach (HBA) methods in economical cost of traffic crashes, seat belt use analysis, and negative externalities of the tourism industry.

Keywords: Home-Based Approach; Road Safety; Traveler; Travel Behavior

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Background and motivation

One of the main negative externalities of the transportation system is traffic crashes, which is among the top ten causes of premature death globally and kills more than 1.25 million annually (World Health Organization 2015). Traffic crashes cost 1-2% of Gross Domestic Product (GDP) of high-income countries and 3% of GDP in low and middle-income countries (Jacobs et al. 2000, WHO 2015, Wijnen and Stipdonk 2016). In the United States, the economic cost and societal harm of traffic crashes were estimated to be over $242 billion and $871 billion in 2010, respectively (Blincoe et al. 2015); these numbers reflect 32,999 fatalities, 3.9 million non-fatal injuries, and 24 million damaged vehicles.

Factors influencing road safety could be classified into three main categories: human, road and environment, and vehicles. The human factor is the main contributing factor in road safety and contributes to 90% of traffic crashes. More importantly, the human factor is the sole contributor to traffic crashes in more than 50% of the crashes (Pakgohar et al. 2011, World Health Organization 2017).

Traditionally, road safety metrics are measured at the location of the crash and its surrounding area. For example, if a crash occurs at an intersection, depending on the scope of the study, the researchers or practitioners may count crashes at intersection level, corridor level, or at a coarser geographic area such as Traffic Analysis Zone (TAZ), city level, or county level. Attributing crash to the location of the crash helps us to understand the relationship between factors such as road geometry, environment, traffic, and weather and crash frequency and severity of a specific transportation network feature. Based on this practice, several countermeasures have been developed to prevent crashes or reduce the severity of traffic crashes. Highway Safety Manual and the Handbook of Road Safety Measures provides several examples of this practice (Elvik et al. 2009, HSM 2010). Therefore, I can conclude that most strategies target engineering solutions and design of the road infrastructure rather than focusing on road users' role which has a substantial role in traffic crashes.

Police crash reports as the main source of road safety analysis only record limited. Information about road users involved in traffic crashes such as age, gender, road user type, seating position, safety equipment use (e.g., helmet, seat belt), driving license status, and road users' violations (e.g., distraction, speeding, DUI) (MMUCC 2012). Although this information provides a valuable contribution to safety science, this information about road users seems trivial compared to the substantial role of road users in traffic crashes.

Unlike engineering approaches in road safety analysis, in epidemiology and public health studies, residential characteristics of the population play an important role, and usually, the issue of interest (i.e., the health problem) is typically attributed to the residential address of population such as neighborhood, city, state or country. This is also the case for travel demand models. In travel demand models such as the four-step model or activity-based models, transportation planners measure and study the travel behavior of the road users based at the traffic analysis zone corresponding to the home address of the road user (Kanafani 1983). Travel demand models and other sources of data such as Census data are valuable sources of data and using them in road safety analysis would provide a complete picture of road safety (Cherry et al. 2018). It is also worthy of mentioning that travel demand models and other sources of data such as Census Bureau measure the travel behavior and demographics of the population at their residential address. Bearing in mind the aforementioned, one way to increase the role of road users in road safety analysis is to use the demographic factors and travel behavior of the road users as a proxy for human factor in the road safety analysis.

Nonetheless, due to privacy concerns, using the home address of the road users existing in police crash reports has very limited applications in road safety analysis (MMUCC 2012). Besides, most studies that used the home address of the road users, they mostly applied an epidemiologic approach and identified groups that are more prone to higher crash rates or fatality rates. In addition, these studies usually used a coarse geographic level such as zip code of the road users (Mayrose and Jehle 2002, Braver 2003, Campos-Outcalt et al. 2003, McAndrews et al. 2013) and mostly focused on fatally injured road users (Schiff and Becker 1996, Baker et al. 1998, Harper et al. 2000). Moreover, these studies mainly focused on the crash outcome and did not examine the role of travel behavior and demographics of road users.

In my dissertation, I set out to take a more epidemiological approach to road safety analysis by looking at factors surrounding the home addresses of the road users involved in traffic crashes –i.e., a Home-Based Approach (HBA). This dissertation focuses on the various applications of the HBA expressly considering sociodemographic and travel behavior as well as attributing traffic crash and its characteristics to the residential address of the road users involved in traffic crashes. HBA enables us to understand road safety from a new perspective. HBA also enables researchers and road safety practitioners to target neighborhoods with proper countermeasures and interventions to reduce the burden of traffic crashes or eliminate aberrant behaviors.

Key purposes

This dissertation consists of five different studies. The primary purposes of this dissertation are as follows. First, many studies explore groups that are more prone to the burden of traffic crashes. In addition to sociodemographic factors, I am exploring the effect of the residential location of the road users on the burden of traffic crashes. The residential address reflects unobserved factors that are hard to capture in the aspatial analysis.

Second, I explore the effect of sociodemographic factors and particularly travel behavior of the road users in one geographic area, and their relationship with the burden of traffic crashes at the zonal level (e.g., Traffic analysis zone, neighborhood). Understanding this relationship would assist us to develop data-driven policies that help safety practitioners and transportation planners to improve road safety and reduce the burden of traffic crashes at the aggregate level for those who are more prone to traffic crashes. In Chapter I and Chapter II of this dissertation, I explored these two hypotheses in more details.

Third, I explore whether road safety of one neighborhood is influenced by its neighbors. I assume that road safety in one neighborhood is not solely determined by their internal factors, it is also influenced by the safety of their neighbors. From this point of view, we can argue that a traffic crash is an unfortunate interaction between two road users. This issue could be explored by modeling the spatial dependency at the zonal level. In chapter I and Chapter II, I examined the spatial dependency of the burden of traffic crashes (e.g., HBA crash rate and cost of a traffic crash at the zonal level). In chapter IV, I explored this issue regarding seat belt use in Tennessee. The presence of spatial dependency, in this case, it could be attributed to the presence of social influence process that shows how one person's decision (to wear a seat belt) affect others decision as well.

Dissertation structure

In this dissertation, due to the limitation of space, I only explored some examples of HBA method application; namely its application on crash rate, economic cost of traffic crashes, seat belt use analysis, and negative externalities of the tourism industry. Table 1 presents the summary of the five studies, including the numbers of observations, methodology, study area, key application and main takeaways of each chapter.

Chapter I and II: Burden of traffic crashes

In chapter I, I applied the HBA method to measure the likelihood of involvement in traffic crashes at the zonal level in the Knoxville Regional Travel Demand model. Furthermore, due to the strong tie of the HBA and transportation planning; I measured the association between the travel behavior, network characteristics and the likelihood of involvement in traffic crashes. In chapter II, I estimated the association between travel behavior, network characteristics and economic cost of the traffic crashes at the zonal level. Analysis indicates that both travel behavior and network characteristics influence the burden of traffic crashes. Statistical methods suggest that road safety level in one TAZ is under the influence of the neighboring units. Moreover, the spatial distribution of the burden of traffic crashes is more tangible for neighborhoods who live near high-speed roads.

Table 1 Comparison of five studies

	Chapter I	Chapter II	Chapter III	Chapter IV	Chapter V
HBA application	Geographic distribution Residential crash rate	Geographic distribution Economic cost of traffic crash per capita (ECCPC)	Geographic distribution of seat belt use	Presence of social influence in seat belt use	Measure Traveler negative externality
Study area	Knoxville Metropolitan area	Knoxville Metropolitan area	Tennessee	Tennessee	Tennessee
Sample Size	• 148K+ individuals • 1100+ TAZ	• 148K+ • 1100+ TAZ	• 500K+ • 4100+ Census tract	• 1.5M+ • 4100+ Census tract	• 2M+ • 97 Counties
Data Source	• Travel Demand model • Police Crash report	• Travel Demand model • Police Crash report	• US Census • Police Crash report	• US Census • Police Crash report	• US Travel Association • Police Crash report
Methodology	• Spatial lag model	• Spatial lag model	• Tobit Model	• Spatial lag model • Spatial Error Model	•
Key takeaways	• Burden of traffic crashes map • Relationship between travel behavior and HBA crash rate	• Burden of traffic crashes map • Relationship between travel behavior and ECCPC	• Generating Geographic map of seat belt use	• Evidence of presence of social influence in seat belt non-use	• Equality and distribution of the benefit and cost of tourism industry

Chapter III and IV: Geographical distribution of seat belt use

In Chapter III and IV focuses on seat belt use. In these chapters, I extracted the home-address of the individuals and their seat belt use at the time of traffic crashes to study the geographical distribution of seat belt use at the fine geographic level. This chapter also explores the association between seat belt use and sociodemographic variables at the zonal level. Findings of this chapter indicate that police crash reports have the potential to be used as a source to examine seat belt use at the neighborhood level and presents valid findings. Using the home-address of the individuals extracted from police crash report could be used to identify areas with lower seat belt use rate, which could be useful in the design of safety campaigns. In chapter IV, I explored the presence of the social influence process in the seat belt use study in Tennessee based on the home address of the vehicle occupants. Presence of highly spatially correlated observations suggests that seat belt use is not produced solely by the internal structural factors represented in the non-spatial models. Seat belt use in Southern metropolitan areas in Tennessee (Memphis and Chattanooga) is also consistent with an influence process– e.g., modification of one person's responses by the actions of another. The observation of spatial effects thus indicates that further inquiry is needed to learn about the underlying mechanism of social influence in future studies.

Chapter V: Negative externality of the Tourism industry

In this chapter, I focused on tourism, and the role of tourist in traffic crashes in Tennessee. It is recognized that the tourism industry contributes to the economic growth of a country. Travelers generate both trips and contribute to generated vehicle miles traveled (VMT) in any geographic area which contribute to an increase in the number of traffic crashes. Less is known about the magnitude of traffic crashes involving travelers and the negative externality of travelers' crashes (NETC) imposed on non-travelers. We find that 19.2% (127,031 out of 694,276 from 2014-2016) of traffic crashes in Tennessee involve a traveler. The injury cost of non-traveler crashes due to a crash with a traveler (i.e., monetized value of NETC) exceeds $7.6 billion, or 12.3% of tourist expenditures between 2014-2016. Analyzing the net impact of travel (tourist expenditures minus NETC) at county level reveals that the NETC exceeds tourist expenditures in 19 of 97 counties (or 20%) in Tennessee. The results of this analysis reveal that an overlooked negative externality of tourism is traffic crashes involving travelers, which warrants further study and potentially policy remediation.

References

Baker, S.P., Braver, E.R., Chen, L.-H., Pantula, J.F., Massie, D., 1998. Motor vehicle occupant deaths among Hispanic and black children and teenagers. Archives of pediatrics & adolescent medicine 152 (12), 1209-1212.

Blincoe, L., Miller, T., Zaloshnja, E., Lawrence, B.A., 2015. The Economic and Societal Impact of Motor Vehicle Crashes, 2010.(Revised)(Report No. DOT HS 812 013). Washington, DC: National Highway Traffic Safety Administration.

Braver, E.R., 2003. Race, Hispanic origin, and socioeconomic status in relation to motor vehicle occupant death rates and risk factors among adults. Accident Analysis & Prevention 35 (3), 295-309.

Campos-Outcalt, D., Bay, C., Dellapena, A., Cota, M.K., 2003. Motor vehicle crash fatalities by race/ethnicity in Arizona, 1990–96. Injury Prevention 9 (3), 251-256.

Cherry, C., Hezaveh, A.M., Noltenius, M., Khattak, A., Merlin, L., Dumbaugh, E., Ragland, D., Sandt, L., 2018. Completing the Picture of Traffic Injuries: Understanding Data Needs and Opportunities for Road Safety.

Elvik, R., Vaa, T., Erke, A., Sorensen, M., 2009. The handbook of road safety measures Emerald Group Publishing.

Harper, J.S., Marine, W.M., Garrett, C.J., Lezotte, D., Lowenstein, S.R., 2000. Motor vehicle crash fatalities: a comparison of Hispanic and non-Hispanic motorists in Colorado. Annals of emergency medicine 36 (6), 589-596.

Hsm, 2010. Highway Safety Manual. American Association of State Highway and Transportation Officials 2, 10-1.

Jacobs, G., Aeron-Thomas, A., Astrop, A., 2000. Estimating global road fatalities.

Kanafani, A., 1983. Transportation demand analysis.

Mayrose, J., Jehle, D.V., 2002. An analysis of race and demographic factors among motor vehicle fatalities. Journal of Trauma and Acute Care Surgery 52 (4), 752-755.

Mcandrews, C., Beyer, K., Guse, C.E., Layde, P., 2013. Revisiting exposure: fatal and non-fatal traffic injury risk across different populations of travelers in Wisconsin, 2001–2009. Accident Analysis & Prevention 60, 103-112.

Mmucc, 2012. Model Minimum Uniform Crash Criteria. DOT HS 811, 631.

Pakgohar, A., Tabrizi, R.S., Khalili, M., Esmaeili, A., 2011. The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. Procedia Computer Science 3, 764-769.

Schiff, M., Becker, T., 1996. Trends in motor vehicle traffic fatalities among Hispanics, non- Hispanic whites and American Indians in New Mexico, 1958–1990. Ethnicity & health 1 (3), 283-291.

Who, 2015. Global status report on road safety 2015 World Health Organization.

Wijnen, W., Stipdonk, H., 2016. Social costs of road crashes: An international analysis. Accident Analysis & Prevention 94, 97-106.

World Health Organization, 2015. Global status report on road safety 2015. World Health Organization. Violence Injury Prevention World Health Organization.

World Health Organization 2017. Road traffic injuries- Fact Sheet.

CHAPTER I: FACTORS INFLUENCING ROAD USERS' LIKELIHOOD OF INVOLVEMENT IN TRAFFIC CRASHES AT THE ZONAL LEVEL

The authors confirm contribution to the paper as follows: study conception and design: Amin Mohamadi Hezaveh, Christopher Cherry; data collection: Not Applicable; analysis and interpretation of results: Amin Mohamadi Hezaveh, Christopher Cherry; draft manuscript preparation: Amin Mohamadi Hezaveh, Christopher Cherry. All authors reviewed the results and approved the final version of the manuscript.

Abstract

In this study, we measured number of traffic crashes which residents of a traffic analysis zone (TAZ) had between 2014-16 by using home addresses of the individual who were involved in Knoxville metropolitan region (i.e., a Home-Based Approach – HBA). By assigning individuals to the TAZ corresponding to their home-address, we obtained socioeconomics and travel behavior data elements surrounding home-address of the individuals. Next, by dividing the HBA crash frequency to the TAZ population, we measured the HBA crash rate at the zonal level (HBA-CR). Moran's I indicates that the HBA-CR is not randomly distributed in space and it exhibits spatial autocorrelation. We also measured average zonal activity based on the travel demand model outputs to measure average distance traveled from one zone to others on a daily base –i.e., exposure. Statistical tests suggest that the spatial lag model (SLM) is more suitable to predict HBA-CR compared to spatial error model. Model's estimate indicates that average zonal activity has a significant positive association with HBA-CR. This is also the case for interstate, and arterial vehicle miles traveled (VMT), population density, intersection density, the percentage of roads with sidewalks and number of workers per household. On the other hand, median household income, VMT on low-speed roads, and percentage of areas near bus stations have significant negative associations with HBA-CR. Findings are discussed in line with road safety countermeasures.

Keywords: Macroscopic Crash Prediction Models; Home-Based Approach

Introduction

Each year approximately 34 thousand people die, and more than two million people are injured in traffic crashes on the United States roadways. The economic and social cost of car and truck crashes in the United States in 2010 was 871 billion dollars (NHTSA 2014). Road safety studies tend to specify the presence of disparities across road user type, income, race, and ethnicities; for instance, crash fatality rate is approximately double in low and middle-income countries compared to high-income countries (21.5, 19.5, and 10.3 per 100,000 populations respectively) (WHO 2015). This trend also holds within-country; for example, several studies in the United States reported that vulnerable road users (i.e., pedestrians and bicyclists) and lower income neighborhoods have higher fatality rates compared to motorized road users and wealthier neighborhoods, respectively (Marshall and Ferenchak 2017). This also holds for the rural areas where the fatality rate is several times higher than the majority of urban areas (Marshall and Ferenchak 2017). Bearing in mind that the burden of road safety injuries and fatalities does not impact the population equally, we may expect the likelihood of involvement in traffic crashes also impacts different populations unequally.

This variation in the burden of traffic crashes echoes the spatial distribution of the burden of traffic crashes and could be used to identify vulnerable neighborhoods where their residents are more prone to traffic crashes burden. Less is known about the factors influencing the likelihood of involvement in traffic crashes particularly the association between the quality of the road infrastructure and travel behavior at a fine geographical level. In this study, we use the home address of the road users extracted from police crash database to measure the likelihood of involvement in traffic crashes at the zonal level (here defined as Home-Based Approach –HBA).

Although the use of the home address of the traffic victims to obtain information regarding their sociodemographic in road safety is not a new effort, one needs to consider that the majority of this studies used fatally injured road users, used course resolution such as zip code, or only focused on a specific group of road users. Blatt and Furman (1998) used information of the fatally injured drivers in the US from the FARS database. Blatt and Furman (1998) reported that residents of rural and small-town are more prone to fatal crashes. Males (2009) also used FARS database to examine the relationship between fatal crashes rate and demographic variables at two level (e.g., driver-level and State-level). Males (2009) concluded that income per capita, population density, motor vehicle trips per capita, college graduates per capita, unemployment rate, and teen population have a significant association with fatality rates. Furthermore, Stamatiadis and Puccini (2000) studied FARS data in the Southeast USA and extracted the driver and census data to obtain the socioeconomic and demographic variables. Their findings indicate that socioeconomic characteristics have an impact on single-vehicle crashes but have no statistically significant impact on multi-vehicle crash rates. Romano et al. (2006) also used FARS database to explore the association between the role of race/ethnicity, language skills, income, and education level on alcohol-related fatal motor vehicle crashes by using zip code level accuracy. Romano et al. (2006) observed a difference in alcohol-related fatality rates across Hispanic subgroups. Furthermore, they concluded high income and education levels have a protective influence on alcohol-related fatal motor vehicle crashes (Romano et al. 2006).

Clark (2003) also used the National Automotive Sampling System (NASS), General Estimates System (GES) data to explore the relationship between population density and mortality rate. Findings indicated that mortality was higher in locations with populations less than 25,000 and was inversely proportional to the driver's county population density (Clark 2003). Lee et al. (2015) also examined the relationship between sociodemographic and crash-involved pedestrians per residence zipcode in Florida. They concluded that pedestrian crashes do not necessarily occur at their zip code residents (Lee et al. 2015). Furthermore, the proportion of children, population working at home, a household without a vehicle, and household income had a significant association with crash-involved pedestrians per residence zip code in Florida (Lee et al. 2015). Girasek and Taylor (2010) used zip code–level income and educational data to measure the safety relationship between socioeconomic status and motor Vehicle Safety Features in Maryland, VA. Girasek and Taylor (2010) concluded that safer motor vehicles appear to be distributed along socioeconomic lines, with lower income groups experiencing more risk. Hezaveh and Cherry (2019) used seat belt use extracted from police crash reports in Tennessee and census tract data and showed that seat belt use varied at a fine geographic level. In addition, Hezaveh and Cherry (2019) explored sociodemographic factors influencing this variation.

Macroscopic Crash Prediction Models (MCPM) is one set of methods that explore the relationship between road safety at macroscopic level with sociodemographic and transportation infrastructure. By using information surrounding the locations of the traffic crashes at the zonal level, researchers identified several factors that associate with crash frequency at the zonal level such as sociodemographic factors, network characteristics, travel behavior, and traffic pattern (e.g., Hadayeghi et al. 2003, Quddus 2008, Hadayeghi et al. 2010, Naderan and Shahi 2010, Pirdavani et al. 2012b, Lee et al. 2015, Gomes et al. 2017).

Traditionally, in road safety analysis, traffic volume was used as the exposure variable, usually in the form of traffic count, VMT (Vehicle Miles Traveled), DVMT (Daily Vehicle Miles Traveled), or VMT by road classification (Aguero-Valverde and Jovanis 2006, Hadayeghi et al. 2010, Pirdavani et al. 2012b, a, Li et al. 2013, Pirdavani et al. 2013b, Rhee et al. 2016, Hosseinpour et al. 2018). In case of absence of traffic information, other proxies such as road lengths with different speed limit (Abdel-Aty et al. 2011, Siddiqui et al. 2012), road length with different functional classification (Quddus 2008, Hadayeghi et al. 2010), or population has been used (Gomes et al. 2017). In regards to measuring the likelihood of involvement in traffic crashes at the zonal level, using VMT may not reflect the exposure properly. One way to deal with this issue is to use population as a proxy for the exposure variable (Lee et al. 2015, Gomes et al. 2017).

However, the population does not reflect the number of trips generated by residents of a geographic area nor their trip length. Other studies also used trip generation models as a vector to measure exposure (Naderan and Shahi 2010, Abdel-Aty et al. 2011, Dong et al. 2014, Dong et al. 2015, Mohammadi et al. 2018). Although this vector provides information regarding exposure of the road users, it fails to capture trip length. A more inclusive exposure variable for estimating the likelihood of involvement in traffic crashes at zonal level needs to consider both trip length and trip frequency simultaneously.

This study aims to explore the association between sociodemographic variable, travel behavior and HBA-CR at the zonal level. Consequently, we used home-address of the road users extracted from police crash database to measure road safety at the zonal level. Added information from the surrounding demographics and travel behavior enable us to explore the association between travel behavior and safety at the zonal level. We also consider the trip length and frequency simultaneously as an exposure variable based on travel demand model outputs.

The next section discusses the methods used in this study. In the methodology section, we discuss the HBA definition, geocoding process, measuring exposure, and spatial models for analyzing the data. In the last section, we present and discuss the findings of the analysis.

Methodology

Home-Based Approach definition

Home-address of the road users who were involved in a traffic crash is one of the data elements that police officer records at the crash scene (MMUCC 2012). Using home-address to collect information of the road users to collect data element regarding sociodemographic and travel behavior is a common practice in urban travel demand analysis (Kanafani 1983). We use the collected home-address of individuals as a basis for further analysis. To tie traffic crashes to the home addresses of the individuals in this study, we define the HBA crash frequency as the expected number of crashes that road users who live in a certain geographic area experience during a specified period. This definition attributes traffic crashes to individuals and their residential addresses.

Data and geocoding process

This study focuses on the Knoxville metropolitan region, which includes 10 counties and a total population of over one million. Figure 1 presents the Knoxville Region study area that includes Knox, Anderson, Roane, Union, Grainger, Jefferson, Sevier, Blount, and Loudon counties. This region is anchored by the city of Knoxville, but also includes several urbanized areas outside the city. The crash data in this study was provided by Tennessee Integrated Traffic Analysis Network (TITAN).

Each crash record includes information about road user type (i.e., driver, motorcyclist, passenger, pedestrian, bicyclist), coordinates of the crashes and addresses of the individual who were involved in traffic crashes. Records of 60,104 crashes and information on 148,666 individuals who were involved in traffic crashes between 2015 and 2016 in the Knoxville region were retrieved from TITAN. After obtaining the address of road users, we used the Bing application program interface services to geocode the addresses. The quality of the geocoding was checked by controlling for the locality of the addresses. Only those records that had an accuracy level of premises (e.g., property name, building name), address level accuracy, or intersection level accuracy was used for the analysis (Hezaveh and Cherry 2019).

We were able to successfully match 141,514 (95%) of the individuals with a home-location. By dividing HBA crash frequency to TAZ's population (1,000 population), we measured HBA-Crash Rate (HBA-CR). Figure 2 presents the histogram of HBA-CR at the TAZ level. Figure 3 also presents the HBA-CR at the TAZ level. Distribution of the negative externalities of the traffic crashes exhibits that burden of traffic crashes are more tangible in the vicinities of the interstates and multilane highways where TAZs' residents are more prone to high-speed traffic and higher road classification.

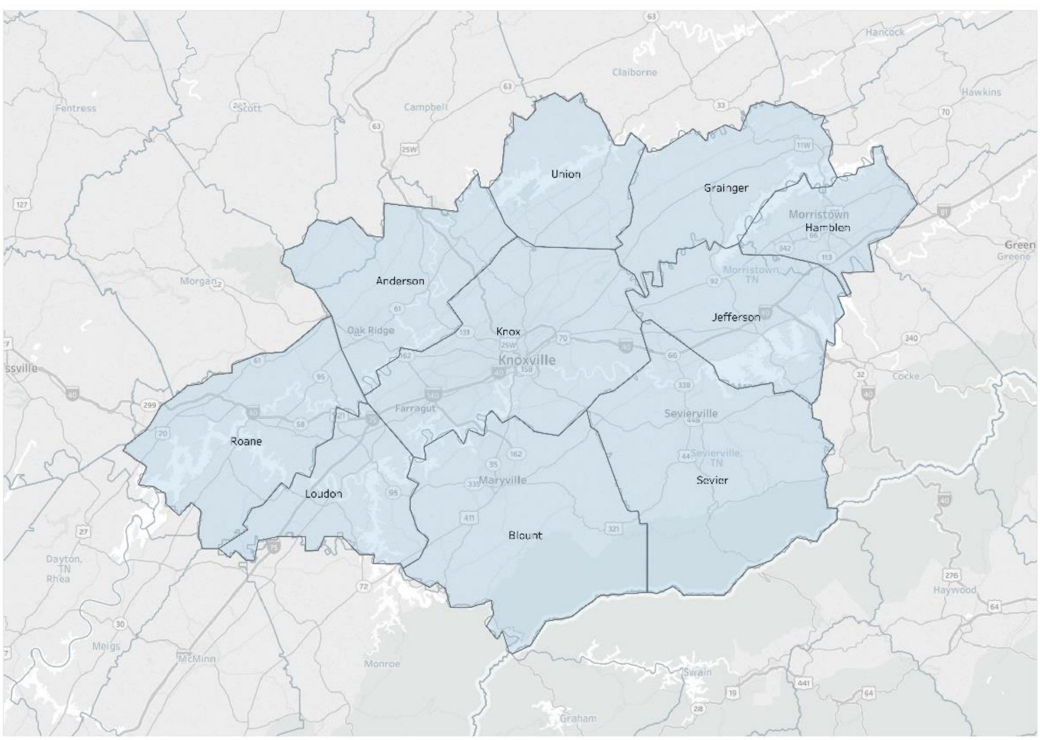

Figure 1 Knoxville Regional Travel Demand

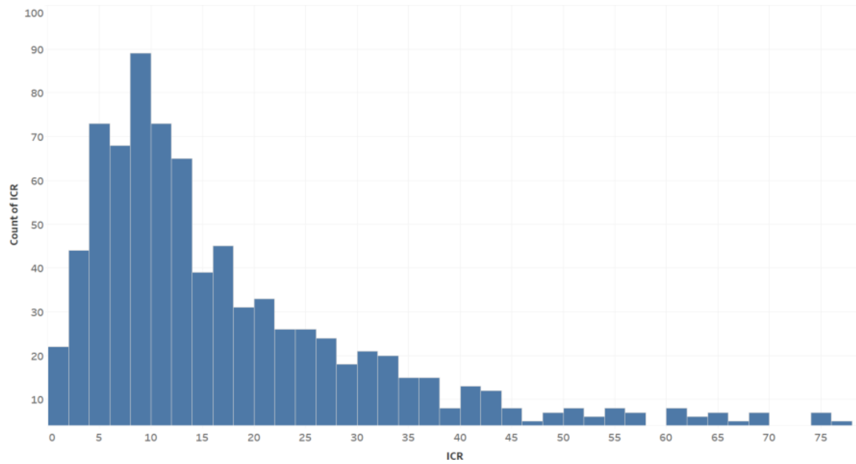

Figure 2 Histogram of HBA-CR at the TAZ level

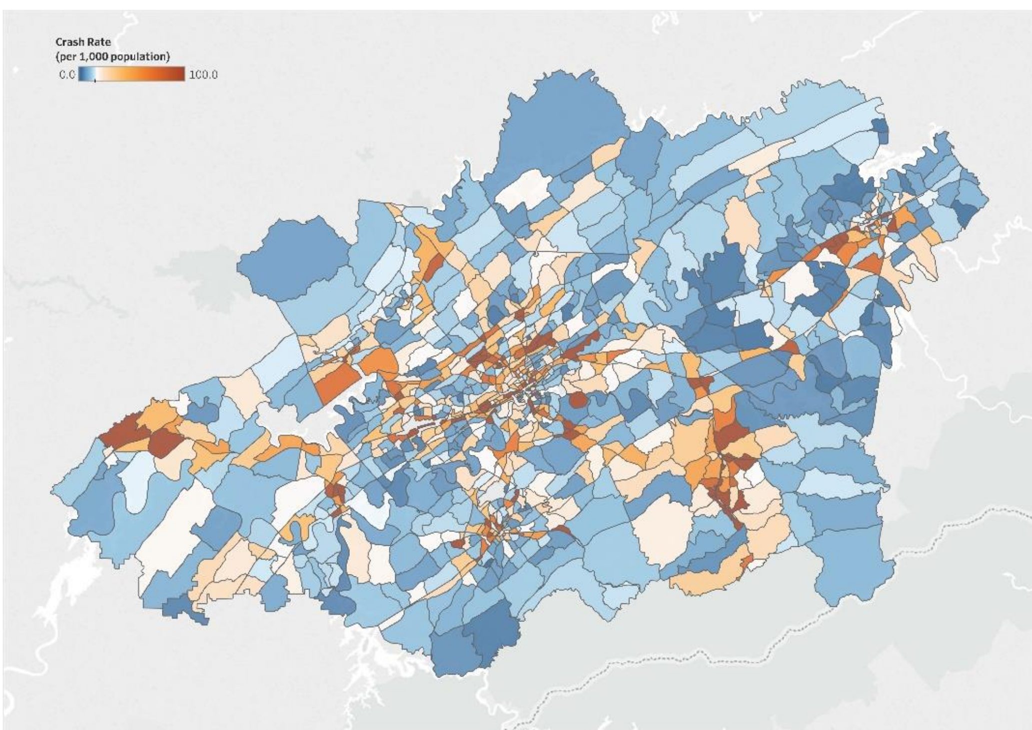

Figure 3 HBA-CR distribution in KRTM

Measuring exposure and travel activity

In this study, one goal was to investigate the relationship between travel behavior and quality of transportation infrastructure with crash cost. To this end, we used the 2014 Knoxville Regional Travel Demand Model. The Knoxville Regional Travel Model (KRTM) has a hybrid design using elements of activity-based model architecture. The model creates a disaggregate synthetic population of households in the region based on the demographic information associated with the traffic analysis zones (TAZs). For more information about Knoxville Regional Travel Demand Model, please see KRTM (2012). The study area includes 1,186 TAZs and includes sociodemographic, economic, and travel information of the residents. Table 2 presents the descriptive statistics of the sociodemographic variables obtained from TAZs. It is worthwhile to mention that 63 zones had no population (e.g., Smoky Mountain National Park, Oak Ridge National Lab) and 135 zones had a population of fewer than 100 individuals. To exclude outliers, we excluded these TAZs from the analysis. Table 2 presents the descriptive statistics of the data elements obtained from the KRTM model.

To evaluate the exposure at the zonal level, we will use zonal activity –i.e., person miles traveled at zonal level (PMT). PMT_i combines trip rate and trip length and is an index for measuring the zonal activity of the trips originated from TAZ_i. To measure PMT_i we will use trip production, distribution, and assignment outputs of the travel demand model. PMT_i is calculated by equation 1:

$$PMT_i = \sum_{j=1}^{n} \frac{P_{ij} L_{ij}}{Pop_i} \qquad \text{Equation 1}$$

where n is the index of TAZ, P_{ij} is the number of trip produced from TAZ i to TAZ j in one day, L_{ij} is the shortest network path between TAZ i to TAZ j, and Pop_i presents the population of the zone i. KRTM was used as a source to extract the number of trips for each pair. Shortest path between each pair was also extracted form traffic assignment at the peak-hour. It is also worthy to mention that PMT reflects all trip purposes and modes in the study area. Figure 4 presents the zonal activity per capita distribution in Knoxville Regional Travel Demand Model at TAZ level. TAZs in the urban and suburban population centers tend to have lower PMT per capita (warmer colors) than outlying rural areas. Visual screening of Figure 4 indicates that the rural areas have higher PMT compared to the urban areas. HBA-CR tended to have more distributed impacts, with higher crash rate along major roads in the study area (e.g., interstate).

Table 2 TAZ descriptive statistics

Variable	Mean	Standard Deviations	Min	Max
Household Income ($)	46655	21075	2349	168227
Workers Per Household	1.21	0.24	0.00	2.10
Students Per Household	0.39	0.18	0.00	1.11
Intersection Density (per square miles)	153	198	3	1657
Percent Road with Sidewalk	0.21	0.32	0.00	1.00
Percent Near Bus Station	0.18	0.36	0.00	1.00
Population Density (Per Square Mile)	1377	2736	3	44072
VMT on Interstate from TAZ (miles)	9625	32673	0	287762
VMT on Arterial from TAZ (miles)	11398	17657	0	163821
VMT on Others from TAZ (miles)	7146	8294	0	76596

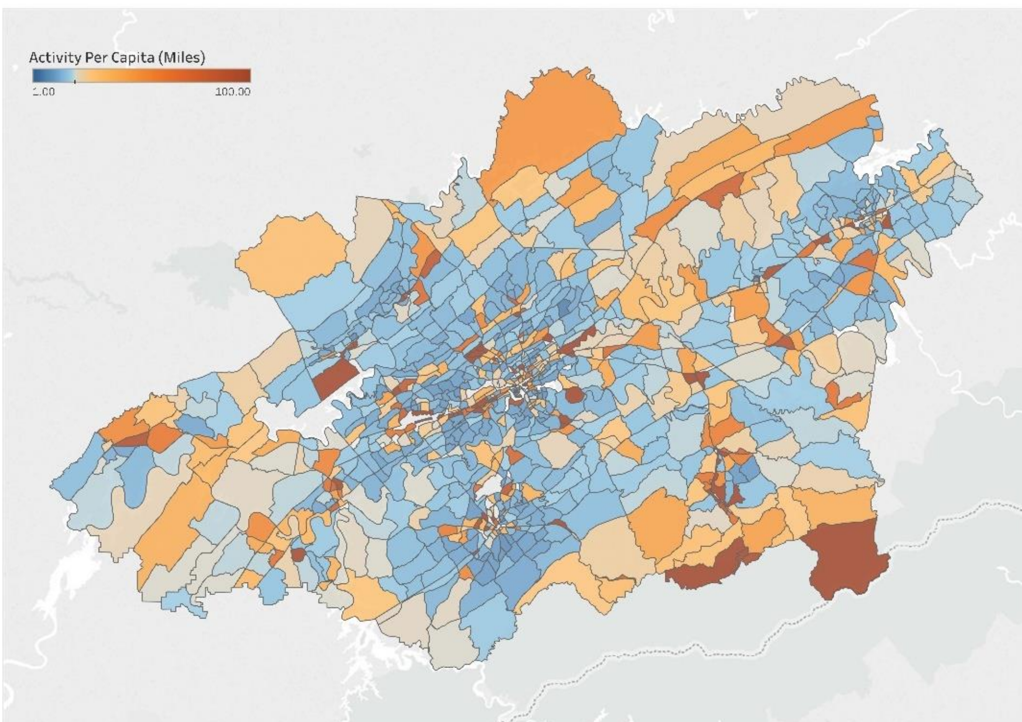

Figure 4 Daily activity per capita (person miles traveled)

Modeling approach

One concern in MCPM modeling is the spatial autocorrelation. Spatial autocorrelation exists when a variable displays interdependence over space (Legendre 1993). Presence of spatial autocorrelation in MCPM was reported in several studies (Quddus 2008, Lee et al. 2015, Rhee et al. 2016). If spatial autocorrelation exists, then the dependent variable is not produced solely by the internal structural factors represented in the non-spatial model. Therefore, disregarding spatial autocorrelation may lead to drawing incorrect inferences.

Testing spatial dependency

Visual inspection of Figure 3 indicates that neighborhoods with better safety records (i.e., blue colors) are surrounded by other TAZs with blue colors. This is also the case for the TAZs with red colors. This may be an indicator of the presence of significant spatial autocorrelation. To diagnose spatial autocorrelation, Global Moran's I (Moran 1950) was used to test whether the model residuals are spatially correlated. Moran's I values range from -1 to +1. Moran's I can be written as:

$$I = \frac{\sum_i \sum_j w_{ij}(y_i - \mu)(y_j - \mu)}{\sum_i (y_i - \mu)^2} \qquad \text{Equation 2}$$

where w_{ij} is an element of a row-standardized spatial weights matrix, y_i is the HBA-CR, and μ is the average HBA-CR in the sample. The statistical significance of the Moran's I is based on the z-score. For more details about the calculation of the Moran's I's Z-score please see Andrew and Ord (1981). The extreme values are indicators of significant spatial autocorrelation where value close to 0 indicates a random pattern between residuals. A significant Moran's I indicates clustering in space of similar HBA-CR.

By assuming the presence of significant spatial autocorrelation, we will use model specifications that consider the spatial dependency in their structure. Spatial error model (SEM)[1] and spatial lag model (SLM) are two common models that are used by researchers to consider spatial autocorrelation in the road safety analysis (Quddus 2008, Lee et al. 2015, Rhee et al. 2016). The distinction between the two models in the method that they incorporate spatial dependency (Doreian 1980, 1982). The SLM model considers the direct effect of one element's response on another's. On the other hand, in the SEM model, the source of the interdependence of the error term is not known.

[1] Not to be mistaken by Structural Equation Modeling

Spatial error model

In the SEM, the models' constant variable is treated as a spatially structured random effect vector. The core assumption in the SEM is that the observational units in close proximity should exhibit effects levels that are similar to those from neighboring units (LeSage and Pace 2009). The SEM is similar to the linear regression models with an additional term for the spatial dependency of errors in neighboring units. The SEM model can be written as:

$$y = X\beta + \varepsilon \qquad \text{Equation 3}$$

$$\varepsilon = \lambda W_\varepsilon + u = (I - \lambda W)^{-1} u \qquad \text{Equation 4}$$

$$y = \lambda W_y + X\beta + \lambda \mathrm{W} \mathrm{X} \beta + \mathrm{u} \qquad \text{Equation 5}$$

where y is a vector of HBA-CR, X is a vector of independent variables presented in Table 2, β is the corresponding vector of estimated coefficients (X). In this model, ε is the error term, which consists of two parts: W_ε and u. W_ε presents the spatially lagged error term corresponding to a weigh matrix W and u refers to the spatial uncorrelated error term that satisfies the normal regression assumption ($u \sim \mathrm{N}(0, \sigma^2 I)$). Last, λ presents the spatial error term parameters, if the value of the spatial error parameters equals zero, the SEM is similar to the standard linear regression model.

Spatial lag model

The spatial lag model, in contrast, incorporates the spatial influence of unmeasured independent variables, but also stipulates an additional effect of neighbors' HBA-CR, via the lagged dependent variable. The SLM model can be represented as:

$$y = \rho W_y + X\beta + \varepsilon \qquad \text{Equation 6}$$

where ρ presents the spatial autoregressive parameter, W_y is a spatially lagged variable corresponding to W matrix, X is a vector of independent variables, β is the vector of estimated coefficients. Last, ε is assumed to be a vector of independent and identically distributed (IID) error terms.

Due to the endogeneity in the W_ε (spatial lag) term, ordinary least-squares (OLS) estimators are biased and inconsistent for the spatial-lag model, and instead, maximum-likelihood estimation (Ord 1975) is used to obtain consistent estimators. (Kim et al. 2003). In order to estimate the SEM and SLM models, we used GeoDa Software (Anselin 2003).

Weight matrix

Choosing a proper weight matrix is crucial for the analysis since it incorporates the prior structure of dependence between spatial units (Baller et al. 2001). Rook and Queen contiguity matrix was used in this analysis to establish the weight matrix. The queen weights matrix define neighbors as TAZs that share a boundary or corner, whereas, rook only considers those TAZ that shares a boundary (Anselin 2003). The selection of optimal weighting matrix could be based on the AICc (Hurvich and Tsai 1989); the weight matrix with the lowest AICc is preferred (Fotheringham and Brunsdon, Nakaya et al. 2005, Hadayeghi et al. 2010, Nakaya 2014). For more information about the weighting matrix, please see Anselin (2003).

Model comparison and assessment

We use the Lagrange Multiplier (LM) principle to choose the proper model specification. These tests are based on the regression residuals obtained from the OLS model. Each of SLM and SEM models has their specific LM statistics, which offers the opportunity to exploit the values of these statistics to suggest the likely alternative. The LM statistic against SEM (LM_{SEM}) and SAR (LM_{SLM}) models take the following forms:

$$LM_{SEM} = \frac{\left(\frac{e'W_e}{s^2}\right)^2}{T} \quad \text{Equation 7}$$

$$LM_{SLM} = \frac{\left(\frac{e'W_e}{s^2}\right)^2}{\frac{(WXb)'M(WXb)}{s^2} + T} \quad \text{Equation 8}$$

where e is a vector of OLS residuals, s^2 its estimated standard error, $T = tr[(W + W')W]$, tr as the matrix trace operator, and $M = I - X(X'X)^{-1}X'$. Both LM_{SEM} and LM_{SAR} are asymptotically distributed as $\chi^2(1)$ under the null. Several researchers illustrate the relative power of these tests by using extensive simulation studies (Anselin and Rey 1991, Anselin and Florax 1995, Anselin et al. 1996).

It is possible that in some cases both LM_{SEM} and LM_{SLM} statistics turn out to be highly significant, which makes it challenging to choose the proper alternative. To deal with this issue, Anselin et al. (1996) developed a robust form of the LM statistics in the sense that each test is robust to the presence of local deviations from the null hypothesis in the form of the other alternative. In other words, the robust Lagrange Multiplier is robust to the presence of spatial lag, and vice versa. The robust tests perform well in a wide range of simulations and form the basis of a practical specification search, as illustrated in (Anselin and Florax 1995, Anselin et al. 1996). In this study, we used GeoDa software to perform the LM tests (Anselin 2003). In addition to LM, to further evaluate the overall model fit and predictive performance, we also used the Akaike Information Criterion ($AICc$) as a measure of the relative goodness.

Results

After assigning the individuals' home-addresses to corresponding TAZs, we calculated the crash frequency at the TAZ level. The average of HBA crash frequency at the TAZ level for the two years period was 95 (*SD* = 107) for the years 2015-16. Average HBA-CR for the study period is 76 per 1,000 populations (*SD* = 141). Results of the global Moran's I indicate that there is a significant spatial autocorrelation exists (*Moran's I* = 0.10 $p < 0.001$). The significant positive value of the Moran's I demonstrates the presence of the spatial pattern, which is an indicator of the clustering in the space of HBA-CR.

Next step we estimated spatial models with consideration of different weight matrices. Considering the non-zero values of ρ and λ, we conclude that both SLM and SAE models are significantly different from linear regression models. By controlling for AICc as well as lag coefficient values for the estimated SLM, and SEM models we learned that the queen contiguity matrix for both SLM and SEM has significantly better performance (significantly lower AICc) compared to the other alternatives.

A Lagrange multiplier test was conducted to select the suitable spatial model. Lagrange multiplier tests (Table 3) revealed that both LM_{SEM} and LM_{SLM} are significant. Therefore, in the next step, we used robust Lagrange multiplier statistics. Only Robust-LM_{SLM} has significant values, which indicates that the SLM model is more suitable. Comparison of the AIC values of estimated models in the Table 3 also indicates that the SLM model has a better performance compared to the OLS and SEM.

Estimated parameters

Table 4 presents the results of the estimated models. In this study, we used the zonal activity as the exposure variable for each TAZ. Therefore, we expected a positive sign for the estimated coefficients. Average zonal activity in all models has a significant positive association with HBA-CR, meaning that as average miles traveled of trips originated from each TAZ increases, the HBA-CR increases. Average Zonal activity implies that those who travel longer distances on daily bases have a higher crash rate.

Number of workers per household and students per household reflect the demographics of a TAZ. The significant positive association of the worker per household variable indicates that as a proportion of workers per household increases HBA-CR also increases. This finding agrees with Naderan and Shahi (2010) study where they reported that the number of work-trips produced at a zonal level has a positive impact with the number of injury crashes, property damage only crashes, and total crash in a TAZ. Similarly, students per household also could be interpreted as a proxy for the number of educational trips produced at each TAZ. The estimated variables in the estimated models are not significant. Nevertheless, the negative sign of students per household indicates that the number of students in each TAZ has a negative correlation with HBA-CR; the negative sign of this variable also agrees with Naderan and Shahi (2010).

The median household income variable also has a negative correlation with HBA-CR, which is consistent with previous studies (Pirdavani et al. 2012b, Pirdavani et al. 2013b, Cai et al. 2017a, Cai et al. 2017b, Gomes et al. 2017, Cheng et al. 2018). Individuals with higher household incomes tend to have lower crash rates. This negative sign also is in agreement with road safety literature (WHO 2015, Marshall and Ferenchak 2017).

Table 3 Results of lagrane multiplier statistics

TEST	VALUE	PROB
Moran's I (error)	5.304	0.000
Lagrange Multiplier (lag)	39.998	0.000
Robust LM (lag)	15.321	0.000
Lagrange Multiplier (error)	25.067	0.000
Robust LM (error)	0.390	0.532

Table 4 Estimated models

	OLS				SLM				SEM			
Variable	**Coef.**	**S. E.**	**T-test**	**P-value**	**Coef.**	**S. E.**	**T-test**	**P-value**	**Coef.**	**S. E.**	**T-test**	**P-value**
Sociodemographics												
Income ($10,000)	-4.794	1.968	-2.437	0.015	-3.232	1.914	-1.689	0.091	-3.623	2.192	-1.653	0.098
Worker Per Household	55.423	17.698	3.132	0.002	47.926	17.170	2.791	0.005	43.076	18.158	2.372	0.018
Student Per Household	-7.747	21.608	-0.359	0.720	-1.856	20.979	-0.088	0.930	-7.179	22.286	-0.322	0.747
Activity Per Capita (Miles Traveled)	1.390	0.069	20.224	0.000	1.347	0.067	20.062	0.000	1.362	0.068	19.916	0.000
Network												
Population Density (per Square miles)	-0.007	0.002	-4.587	0.000	-0.007	0.002	-4.617	0.000	-0.007	0.002	-3.990	0.000
Intersection Density	0.075	0.027	2.801	0.005	0.059	0.026	2.259	0.024	0.067	0.028	2.412	0.016
% Road with Sidewalk	86.125	16.927	5.088	0.000	79.027	16.464	4.800	0.000	86.042	17.427	4.937	0.000
% Near Bus Stop	24.546	14.287	1.718	0.086	18.232	13.875	1.314	0.189	21.932	15.894	1.380	0.168
VMT Interestate	9.767	1.687	5.791	0.000	9.025	1.639	5.505	0.000	9.499	1.714	5.541	0.000
VMT Arterial	12.457	2.058	6.054	0.000	11.181	2.004	5.578	0.000	11.564	2.041	5.665	0.000
VMT Other Roads	-9.411	2.334	-4.032	0.000	-8.455	2.266	-3.731	0.000	-8.779	2.363	-3.716	0.000
Constant	-38.818	20.856	-1.861	0.063	-52.070	20.407	-2.552	0.011	-27.301	22.032	-1.239	0.215
Lag coeff. (Rho)					0.249	0.040	6.256	0.000	0.238	0.047	5.047	0.000
R-squared	0.426				0.453				0.445			
Log likelihood (Full)	-5838.1				-5820.7				-5826.9			
AIC	11700.1				11667.5				11677.8			

As expected, road network characteristics have a significant association with safety level. It is worthy to mention that the network characteristics of a TAZ may reflect the traffic flows and infrastructures that transportation system imposes to residents of a TAZ. Population density also has a negative association with HBA-CR. The negative sign indicates that as density increases the crash frequency of the road users decreases.

Consistent with previous studies VMT also have a significant association with safety outcomes. Comparison of the coefficients indicates that VMT on arterial roads (i.e., major and minor arterials) has a greater impact on HBA-CR compared to the interstate. This differences in the magnitudes could reflect the high access of the arterial roads with more conflicts compared to interstates, which could increase the likelihood of crash occurrence. On the other hand, other road classifications with the lower posted speed limit (e.g., collector, local) has a negative association with HBA-CR. Many studies explored the association between of functional classes and crash frequency at zonal level (e.g., Hadayeghi et al. 2003, Quddus 2008, Xu and Huang 2015), only a few considered the effect of exposure (i.e., VMT) in different road classes. There is also a need to consider that the definition of the functional classes may vary across areas. In a series of studies in Flanders, Belgium, Pirdavani et al. (2013a) and Pirdavani et al. (2012b) reported that VMT on a motorway had a smaller effect on total crash frequency compared to non-motorway VMT. In Florida, Xu and Huang (2015), reported that proportions of the road with speed limits 25 mph or lower had a negative association with crash frequency at a zonal level, whereas, percent of roads at 45 mph and above had positive association on zone crash frequencies. Hadayeghi et al. (2003) also reported that total local road length in a TAZ had a negative association with all crashes and severe crashes, whereas, arterials, expressways, collectors, and ramps had a positive and significant association with crash frequency at the zonal level in a study in Canada.

Percent of roads with sidewalk and number of bus stations also have a significant association with HBA-CR. The positive sign of these two variables may be an indicator of the presence of vulnerable road users. It is likely that due to the less developed network of the pedestrian in the KRTM, vulnerable road users are more prone to traffic crashes and therefore their HBA-CR increases. Cai et al. (2017b) also reported that sidewalk length has a positive association with crash frequency, severe crash, and non-motorized crash frequency. Intersection density in the TAZ also has a significant positive association with HBA-CR. This is in agreement with previous researches that reported the number of intersection could be correlated with higher numbers of conflict and accordingly the higher number of traffic crashes (Hadayeghi et al. 2003, Ladron de Guevara et al. 2004, Lovegrove and Sayed 2006, Abdel-Aty et al. 2011, Pirdavani et al. 2012a, Gomes et al. 2017).

Summary and conclusion

In this study, we measured the likelihood of involvement in traffic crashes based on the home address of individuals (i.e., home-based approach) who were directly involved in traffic crashes at the zonal level. Geographic distribution of the HBA-CR indicates residents' activity and their safety culture.

Analysis indicated that HBA-CR is not randomly distributed in space and it exhibits positive spatial autocorrelation. Highly spatially correlated HBA-CR at zonal level suggest that HBA-CR is not produced solely by the internal structural factors that are captured in the OLS specification. Results of Lagrange multiplier statistics also indicate that the spatial lag model is more suitable compared to the spatial error model. Considering the underlying assumptions of the SLM model, we may conclude that HBA-CR in one TAZ is influenced by HBA-CR in neighboring TAZs. Therefore, we may conclude that a neighborhood with poor traffic safety may pose negative externality to its neighbors and vice versa.

HBA-CR was higher in the vicinities of the high-speed traffic roads and roads with a higher classification. Also, both VMT and average zonal activity have a significant association with HBA-CR. To reduce the burden of traffic crashes, first, designing a transportation network with the aim of diverging high-speed traffic from residential areas or managing the accessibility of the residents near the high-speed, high volume roads could eliminate or discount exposure to high-speed traffics. The second strategy may target average zonal activity by eliminating a portion of trips by managing travel demand and providing strategies and policies that reduce travel demand (Gärling et al. 2002). In addition, the increase in density and mixed land-use design would also degenerate both trips rate, VMT (Cervero and Kockelman 1997), and trip length (Cervero and Kockelman 1997) and eventually zonal activity. Reduction in zonal activity and VMT has a direct impact on the crash rate and eventually burden of traffic crashes. Regarding the significant association of both exposures, we can conclude that policymakers, planners, and safety practitioners may decrease the crash rate by controlling for exposure.

The spatial distribution of the HBA-CR and its association with sociodemographic variables demonstrated potentials of the HBA as a means for identifying the hotspots in which residents have a higher likelihood of involvement in traffic crashes. Proper safety campaigns could be used to address the safety concerns in the TAZs with high HBA crash rate, particularly focusing on behavioral interventions that contribute to higher crash risk and injury burden (e.g., speeding, driving under the influence, seatbelts). Furthermore, road safety culture and driving behavior may also correlate with a crash rate; this issue could be investigated in the future studies. This issue could be explored in future studies.

In addition to the spatial models, we estimated count data models such as negative binomial and Poisson models (both random and fixed coefficients). Comparison of the models suggests that the association between a dependent variable and independent variables were stable. To maintain concision, we did not present the estimated models. Furthermore, the majority of road users in this study was motorized users. In this study, we ran separate models for predicting HBA-CR for all road users and motorized road users. Comparison of the models indicates that the models are similar and findings are broadly in agreement. Therefore, to maintain concision, we did not present the model for the predicting motorized road user crash rate.

It is also worth mentioning that there are difficulties in accessing the crash data with identifiers and it is not possible to obtain this data in some cases. One possible direction for future research could be to partnering with data owners to assist in matching crashes with spatial datasets to preserve confidentiality.

Acknowledgment

The authors would like to thank the Tennessee Department of Safety and Homeland Security for providing the data for this study. This project was supported by the Collaborative Sciences Center for Road Safety, www.roadsafety.unc.edu, a U.S. Department of Transportation National University Transportation Center promoting safety. The study design was reviewed and approved by the University of Tennessee Institutional Review Board. The authors thank Louis Merlin, Eric Dumbaugh, David Ragland, and Laura Sandt for valuable insights.

References

Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating Trip and Roadway Characteristics to Manage Safety in Traffic Analysis Zones. Transportation Research Record: Journal of the Transportation Research Board (2213), 20-28.

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial Analysis of Fatal and Injury Crashes in Pennsylvania. Accident Analysis & Prevention 38 (3), 618-625.

Andrew, C., Ord, J.K., 1981. Spatial Processes: Models and Applications. London: Pion.

Anselin, L., 2003. Geoda 0.9 User's Guide. Urbana 51, 61801.

Anselin, L., Bera, A.K., Florax, R., Yoon, M.J., 1996. Simple Diagnostic Tests for Spatial Dependence. Regional science and urban economics 26 (1), 77-104.

Anselin, L., Florax, R.J., 1995. Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results. New Directions in Spatial Econometrics. Springer, pp. 21-74.

Anselin, L., Rey, S., 1991. Properties of Tests for Spatial Dependence in Linear Regression Models. Geographical analysis 23 (2), 112-131.

Baller, R.D., Anselin, L., Messner, S.F., Deane, G., Hawkins, D.F., 2001. Structural Covariates of Us County Homicide Rates: Incorporating Spatial Effects. Criminology 39 (3), 561-588.

Blatt, J., Furman, S.M., 1998. Residence Location of Drivers Involved in Fatal Crashes. Accident Analysis & Prevention 30 (6), 705-711.

Cai, Q., Abdel-Aty, M., Lee, J., 2017a. Macro-Level Vulnerable Road Users Crash Analysis: A Bayesian Joint Modeling Approach of Frequency and Proportion. Accident Analysis & Prevention 107, 11-19.

Cai, Q., Abdel-Aty, M., Lee, J., Eluru, N., 2017b. Comparative Analysis of Zonal Systems for Macro-Level Crash Modeling. Journal of safety research 61, 157-166.

Cervero, R., Kockelman, K., 1997. Travel Demand and the 3ds: Density, Diversity, and Design. Transportation Research Part D: Transport and Environment 2 (3), 199-219.

Cheng, W., Gill, G.S., Ensch, J.L., Kwong, J., Jia, X., 2018. Multimodal Crash Frequency Modeling: Multivariate Space-Time Models with Alternate Spatiotemporal Interactions. Accident Analysis & Prevention 113, 159-170.

Clark, D.E., 2003. Effect of Population Density on Mortality after Motor Vehicle Collisions. Accident Analysis & Prevention 35 (6), 965-971.

Dong, N., Huang, H., Xu, P., Ding, Z., Wang, D., 2014. Evaluating Spatial-Proximity Structures in Crash Prediction Models at the Level of Traffic Analysis Zones. Transportation Research Record: Journal of the Transportation Research Board (2432), 46-52.

Dong, N., Huang, H., Zheng, L., 2015. Support Vector Machine in Crash Prediction at the Level of Traffic Analysis Zones: Assessing the Spatial Proximity Effects. Accident Analysis & Prevention 82, 192-198.

Doreian, P., 1980. Linear Models with Spatially Distributed Data: Spatial Disturbances or Spatial Effects? Sociological Methods & Research 9 (1), 29-60.

Doreian, P., 1982. Maximum Likelihood Methods for Linear Models: Spatial Effect and Spatial Disturbance Terms. Sociological Methods & Research 10 (3), 243-269.

Fotheringham, A., Brunsdon, C., M. Charlton. 2002. Geographically Weighted Regression-the Analysis of Spatially Varying Relationships. Chichester, UK: John Wiley & Sons.

Gärling, T., Eek, D., Loukopoulos, P., Fujii, S., Johansson-Stenman, O., Kitamura, R., Pendyala, R., Vilhelmson, B., 2002. A Conceptual Analysis of the Impact of Travel Demand Management on Private Car Use. Transport Policy 9 (1), 59-70.

Girasek, D.C., Taylor, B., 2010. An Exploratory Study of the Relationship between Socioeconomic Status and Motor Vehicle Safety Features. Traffic injury prevention 11 (2), 151-155.

Gomes, M.J.T.L., Cunto, F., da Silva, A.R., 2017. Geographically Weighted Negative Binomial Regression Applied to Zonal Level Safety Performance Models. Accident Analysis & Prevention 106, 254-261.

Hadayeghi, A., Shalaby, A., Persaud, B., 2003. Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. Transportation Research Record: Journal of the Transportation Research Board (1840), 87-95.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of Planning Level Transportation Safety Tools Using Geographically Weighted Poisson Regression. Accident Analysis & Prevention 42 (2), 676-688.

Hezaveh, A.M., Cherry, C.R., 2019. Neighborhood-Level Factors Affecting Seat Belt Use. Accident Analysis and Prevention 122, 153-161.

Hosseinpour, M., Sahebi, S., Zamzuri, Z.H., Yahaya, A.S., Ismail, N., 2018. Predicting Crash Frequency for Multi-Vehicle Collision Types Using Multivariate Poisson-Lognormal Spatial Model: A Comparative Analysis. Accident Analysis & Prevention.

Hurvich, C.M., Tsai, C.-L., 1989. Regression and Time Series Model Selection in Small Samples. Biometrika 76 (2), 297-307.

Kanafani, A., 1983. Transportation Demand Analysis.

Kim, C.W., Phipps, T.T., Anselin, L., 2003. Measuring the Benefits of Air Quality Improvement: A Spatial Hedonic Approach. Journal of environmental economics and management 45 (1), 24-39.

KRTM, 2012. Knoxville Regional Travel Model Update 2012 Model Development and Validation Report. Bernardin, Lochmueller & Associates, Inc., Evansville, IN, USA.

Ladron de Guevara, F., Washington, S., Oh, J., 2004. Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. Transportation Research Record: Journal of the Transportation Research Board (1897), 191-199.

Lee, J., Abdel-Aty, M., Choi, K., Huang, H., 2015. Multi-Level Hot Zone Identification for Pedestrian Safety. Accident Analysis & Prevention 76, 64-73.

Legendre, P., 1993. Spatial Autocorrelation: Trouble or New Paradigm? Ecology 74 (6), 1659-1673.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics Chapman and Hall/CRC.

Li, Z., Wang, W., Liu, P., Bigham, J.M., Ragland, D.R., 2013. Using Geographically Weighted Poisson Regression for County-Level Crash Modeling in California. Safety science 58, 89-97.

Lovegrove, G.R., Sayed, T., 2006. Macro-Level Collision Prediction Models for Evaluating Neighbourhood Traffic Safety. Canadian Journal of Civil Engineering 33 (5), 609-621.

Males, M.A., 2009. Poverty as a Determinant of Young Drivers' Fatal Crash Risks. Journal of safety research 40 (6), 443-448.

Marshall, W.E., Ferenchak, N.N., 2017. Assessing Equity and Urban/Rural Road Safety Disparities in the Us. Journal of Urbanism: International Research on Placemaking and Urban Sustainability 10 (4), 422-441.

MMUCC, 2012. Model Minimum Uniform Crash Criteria. DOT HS 811, 631.

Mohammadi, M., Shafabakhsh, G., Naderan, A., 2018. Effects of Modal Shares on Crash Frequencies at Aggregate Level. Accident Analysis & Prevention 120, 295-303.

Moran, P.A., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37 (1/2), 17-23.

Naderan, A., Shahi, J., 2010. Aggregate Crash Prediction Models: Introducing Crash Generation Concept. Accident Analysis & Prevention 42 (1), 339-346.

Nakaya, T., 2014. Gwr4 User Manual. WWW Document. Available online: http://www. st-andrews. ac. uk/geoinformatics/wp-content/uploads/GWR4manual_201311. pdf (accessed on 4 November 2013).

Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically Weighted Poisson Regression for Disease Association Mapping. Statistics in medicine 24 (17), 2695-2717.

NHTSA, 2014. The Economic and Societal Impact of Motor Vehicle Crashes, 2010. Report DOT HS 812, 013.

Ord, K., 1975. Estimation Methods for Models of Spatial Interaction. Journal of the American Statistical Association 70 (349), 120-126.

Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., 2012a. Application of Different Exposure Measures in Development of Planning-Level Zonal Crash Prediction Models. Transportation Research Record: Journal of the Transportation Research Board (2280), 145-153.

Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., Year. Developing Zonal Crash Prediction Models with a Focus on Application of Different Exposure Measures.

Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., 2013a. Evaluating the Road Safety Effects of a Fuel Cost Increase Measure by Means of Zonal Crash Prediction Modeling. Accident Analysis & Prevention 50, 186-195.

Pirdavani, A., Brijs, T., Bellemans, T., Wets, G., Year. Spatial Analysis of Fatal and Injury Crashes in Flanders, Belgium: Application of Geographically Weighted Regression Technique.

Quddus, M.A., 2008. Modelling Area-Wide Count Outcomes with Spatial Correlation and Heterogeneity: An Analysis of London Crash Data. Accident Analysis & Prevention 40 (4), 1486-1497.

Rhee, K.-A., Kim, J.-K., Lee, Y.-I., Ulfarsson, G.F., 2016. Spatial Regression Analysis of Traffic Crashes in Seoul. Accident Analysis & Prevention 91, 190-199.

Romano, E.O., Tippetts, A.S., Voas, R.B., 2006. Language, Income, Education, and Alcohol-Related Fatal Motor Vehicle Crashes. Journal of ethnicity in substance abuse 5 (2), 119-137.

Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes. Accident Analysis & Prevention 45, 382-391.

Stamatiadis, N., Puccini, G., 2000. Socioeconomic Descriptors of Fatal Crash Rates in the Southeast USA. Injury control and safety promotion 7 (3), 165-173.

WHO, 2015. Global Status Report on Road Safety 2015. World Health Organization. Violence Injury Prevention World Health Organization.

Xu, P., Huang, H., 2015. Modeling Crash Spatial Heterogeneity: Random Parameter Versus Geographically Weighting. Accident Analysis & Prevention 75, 16-25.

CHAPTER II: ASSOCIATION BETWEEN TRAVEL BEHAVIOR AND ECONOMIC COST OF TRAFFIC CRASH AT THE ZONAL LEVEL

The authors confirm contribution to the paper as follows: study conception and design: Amin Mohamadi Hezaveh, Christopher Cherry; data collection: Not Applicable; analysis and interpretation of results: Amin Mohamadi Hezaveh, Christopher Cherry; draft manuscript preparation: Amin Mohamadi Hezaveh, Christopher Cherry. All authors reviewed the results and approved the final version of the manuscript.

Abstract

Road safety literature provides abundant examples of studies that measure economic cost of traffic crashes at coarse geographic level. The current practice of road safety economic assessment attributes traffic crash costs to the location of traffic crashes. Therefore, it is challenging to estimate the economic cost of traffic crashes of individuals who live in a specific geographic area. To address this limitation, we used home-address of individuals who were involved in traffic crashes in East Tennessee between 2015-2016. After geocoding the home-addresses, we assigned 110,312 individuals to the Traffic Analysis Zone (TAZ) corresponding to their home address and calculated the economic cost of traffic crashes per capita (ECCPC). The average ECCPC in the study area was $1,399. The Knoxville regional Travel demand model output was used for extracting travel behavior data elements for modeling ECCPC at zonal level. We also established an index to measure exposure individuals' activity in the transportation system –i.e., average zonal activity– for residents of each TAZ. The burden of traffic crashes (ECCPC per income) was also more tangible in the TAZs with lower income and higher zonal activities. Gini index coefficient was also 0.59, which is an indicator of the unequal distribution of the burden of traffic crashes. The spatial autoregressive (SAR) model with a queen contiguity weights matrix was more suitable compared to the spatial error model and ordinary least squares regression. SAR model implies that ECCPC in one TAZ is affected by traffic safety of the adjacent TAZs. Findings indicate that average zonal activity has a significant positive association with ECCPC. In addition, posted speed limit, percentage of worker per household, percentage of road with sidewalk, percentage of area near bus stations, and VMT have a significant positive association with ECCPC. On the other hand, median household income has a negative association with ECCPC. Findings are discussed in line with the road safety policy.

Keywords: Economic Cost of Traffic Crashes; Macroscopic Crash Prediction Model; Home-Based Approach; Home-Address

Introduction

One of the main negative externalities of the transportation system is traffic crashes. They are among the top ten causes of death globally, killing more than 1.25 million annually (WHO 2015b). Traffic crashes cost 1-2% of Gross Domestic Product (GDP) of high-income countries and 3% of GDP in low and middle-income countries (Jacobs et al. 2000, WHO 2015a, Wijnen and Stipdonk 2016). The relative magnitude of this externality is larger in low and middle-income countries compared to the developed countries (WHO 2015b). Road safety literature has abundant examples of estimating cost of traffic crashes at coarse geographic level (i.e., country-level) (Mohan 2002, García-Altés and Pérez 2007, Wegman and Oppe 2010, Ahadi and Razi-Ardakani 2015, Blincoe et al. 2015); however, to the best of our knowledge, there are no studies that explored this matter at fine geographic level (e.g., traffic analysis zone) and factors correlated with it.

Traffic crashes cost can be measured in two ways; the economic cost of traffic crashes and societal harm. Economic costs of traffic crashes include lost productivity, medical costs, legal and court costs, emergency service costs (EMS), insurance administration costs, congestion costs, property damage, and workplace losses (Blincoe *et al.* 2015). In addition to the economic cost of traffic crashes, the societal harm includes lost quality-of-life. The economic cost of traffic crashes reflects the tangible part of the traffic crashes; whereas the societal harm of traffic crashes reflects both tangible and intangible cost of traffic crashes (Mohan 2002, García-Altés and Pérez 2007, Ahadi and Razi-Ardakani 2015, Blincoe *et al.* 2015, Harmon *et al.* 2018). In the United States, the economic cost and societal harm of traffic crashes were estimated to be over $242 billion and $871 billion in 2010, respectively (Blincoe *et al.* 2015); these numbers reflect 32,999 fatalities, 3.9 million non-fatal injuries, and 24 million damaged vehicles.

Road safety studies tend to specify the presence of disparities across road user type, income, race, and ethnicities; for instance, the crash fatality rate is approximately double in low- and middle-income countries compared to high-income countries (21.5, 19.5, and 10.3 per 100,000 populations respectively) (WHO 2015b). This trend also holds within-country; several studies in the United States reported that vulnerable road users (i.e., pedestrians and bicyclists) and lower income neighborhoods have higher fatality rates compared to motorized road users and wealthier neighborhoods, respectively (Clark 2003, Romano *et al.* 2006, Marshall and Ferenchak 2017). In rural areas, the fatality rate tends to be several times higher than in urban areas (Blatt and Furman 1998, Marshall and Ferenchak 2017). Additionally, some ethnicities such as Hispanic, African-American, and Native American have higher crash rates (Mayrose and Jehle 2002, Braver 2003, Campos-Outcalt *et al.* 2003, McAndrews *et al.* 2013) and fatality rates (Schiff and Becker 1996, Baker *et al.* 1998, Harper *et al.* 2000).

The current practice of road safety measure safety at the location of the crash. As a result, it is challenging to measure the likelihood of involvement in a traffic crash at the zonal level and subsequently the economic burden of crashes in areas where individuals reside. In order to examine road safety disparities, we measure the crash cost at the zonal level by using the home address of the road users involved in traffic crashes. Although the use of home-address of the traffic victims to obtain information regarding their sociodemographic in road safety is not a new effort, one need to consider that the majority of this studies used fatally injured road users (Blatt and Furman 1998, Stamatiadis and Puccini 2000, Romano *et al.* 2006, Males 2009), used course resolution such as zip code (Romano *et al.* 2006, Lee *et al.* 2015a), census-level (Stamatiadis and Puccini 2000), or focused on a specific group of road users (Lee *et al.* 2015a). Likewise, these studies did not measure the monetize value of road traffic crashes based on injury level.

Macroscopic Crash Prediction Models are a set of methods that provide information regarding the association between road safety at zonal level and data elements at aggregate level such as sociodemographic factors, network characteristics, travel behavior, and traffic pattern (e.g., Hadayeghi *et al.* 2003, Quddus 2008, Hadayeghi *et al.* 2010, Naderan and Shahi 2010, Pirdavani *et al.* 2012b, Lee *et al.* 2015a, Gomes *et al.* 2017). By using a wide range of safety outcomes, researchers explored the association between geographic unit characteristic and number of all traffic crashes (Miaou *et al.* 2003, Naderan and Shahi 2010, Pirdavani *et al.* 2012b, Pirdavani *et al.* 2013b, Huang *et al.* 2016, Cai *et al.* 2017b, Hezaveh and Cherry 2018), number of property damage only crashes (Naderan and Shahi 2010, Aguero-Valverde 2013), frequency of injury/severe crashes (Aguero-Valverde 2013, Xu and Huang 2015, Cai *et al.* 2017b), or crash frequency of specific road users (e.g., non-motorized, bicyclists) (Lee *et al.* 2015b, Cai *et al.* 2017b, Cheng *et al.* 2018, Saha *et al.* 2018) at zonal level. Although many studies used different forms of the road safety, to the best of our knowledge no studies used monetary value of the traffic crashes based on the home address of the road users and factors associating with it.

Traditionally, in road safety analysis, traffic volume was used as the exposure variable, usually in the form of traffic count, VMT (Vehicle Miles Traveled), DVMT (Daily Vehicle Miles Traveled), or VMT by road classification (Aguero-Valverde and Jovanis 2006, Hadayeghi *et al.* 2010, Pirdavani *et al.* 2012b, a, Li *et al.* 2013, Pirdavani *et al.* 2013b, Rhee *et al.* 2016, Hosseinpour *et al.* 2018). In case of absence of traffic information, other proxies such as road lengths with different speed limit (Abdel-Aty *et al.* 2011, Siddiqui *et al.* 2012), road length with different functional classification (Quddus 2008, Hadayeghi *et al.* 2010), or population has been used (Gomes *et al.* 2017). In case of measuring the likelihood of involvement in traffic crashes at the zonal level, using VMT may not reflect the exposure properly. One way to deal with this issue is to use population as a proxy for the exposure variable (Lee *et al.* 2015a, Gomes *et al.* 2017). However, the population does not reflect the number of trips generated by residents of a geographic area nor their trip length. Other studies also used trip generation models as a vector to measure exposure (Naderan and Shahi 2010, Abdel-Aty *et al.* 2011, Dong *et al.* 2014, Dong *et al.* 2015, Mohammadi *et al.* 2018). Although this vector provides information regarding exposure of the road users, it fails to capture trip length. A more inclusive exposure variable for estimating the likelihood of involvement in traffic crashes at zonal level needs to consider both trip length and trip frequency simultaneously.

This study has several aims. First, we will use the home address of the road users who were involved in traffic crashes to measure road safety (i.e., a Home-Based Approach – HBA). We will apply HBA to measure the economic cost of traffic crashes at fine geographic areas and explore the relationship between travel behavior and economic burden of traffic crashes at the zonal level, focusing on whether there is an equitable distribution of crash burden within an urban area. Last, we measure the distribution of the burden of traffic crashes at the traffic analysis zone (TAZ) level to identifies the groups that are more prone to the burden of traffic crashes. Learning about the relationship between exogenous variables, exposure, and traffic crashes cost of residents of a specific geographic area may enable safety practitioners and researchers to allocate resources to the neighborhoods where the burden of traffic crashes is higher than average, or address inequities in the system where groups are bearing a higher proportional economic burden.

In this study, we will use the data from the Knoxville Regional Travel Demand model in Tennessee. Tennessee has a worse crash record compared to US national level (fatality rate: TN = 1.66 vs. US = 1.34 per 100 MVMT). In the next section, we discuss the methodology including the HBA definition, data, and modeling approach. The rest of the paper presents and discusses the findings of this study.

Methodology

Travel activity

In this study, one goal was to investigate the relationship between travel behavior and quality of transportation infrastructure with crash cost. To this end, we used the 2014 Knoxville Regional Travel Demand Model. This Knoxville region is anchored by the city of Knoxville, but also includes several urbanized areas outside the city. The Knoxville Regional Travel Model (KRTM) has a hybrid design using elements of activity-based models. For more information about Knoxville Regional Travel Demand Model, please see KRTM (2012). Figure 1 presents the Knoxville Region study area that includes Knox, Anderson, Roane, Union, Grainger, Jefferson, Sevier, Blount, and Loudon counties. The study area includes 1,186 TAZs and includes sociodemographic, economic, and travel information of the residents. Table 4 presents the descriptive statistics of the sociodemographic variables obtained from TAZs. It is worthwhile to mention that 63 zones had no population (e.g., Smoky Mountain National Park, Oak Ridge National Lab) and 135 zones had a population of fewer than 100 individuals. To exclude outliers, we excluded these TAZs from the analysis. Table 5 presents the descriptive statistics of the data elements obtained from the KRTM model.

Table 5 TAZ descriptive statistics

Variable	Mean	Standard Deviations	Min	Max
Household Income ($)	46655	21075	2349	168227
Workers Per Household	1.21	0.24	0.00	2.10
Students Per Household	0.39	0.18	0.00	1.11
Intersection Density (per square miles)	153	198	3	1657
Percent Road with Sidewalk	0.21	0.32	0.00	1.00
Percent Near Bus Station	0.18	0.36	0.00	1.00
Population Density (Per Square Mile)	1377	2736	3	44072
Average Speed (MPH)	39.09	8.33	20.00	65.00
VMT on Interstate from TAZ (miles)	9625	32673	0	287762
VMT on Arterials from TAZ (miles)	11398	17657	0	163821
VMT on Others from TAZ (miles)	7146	8294	0	76596

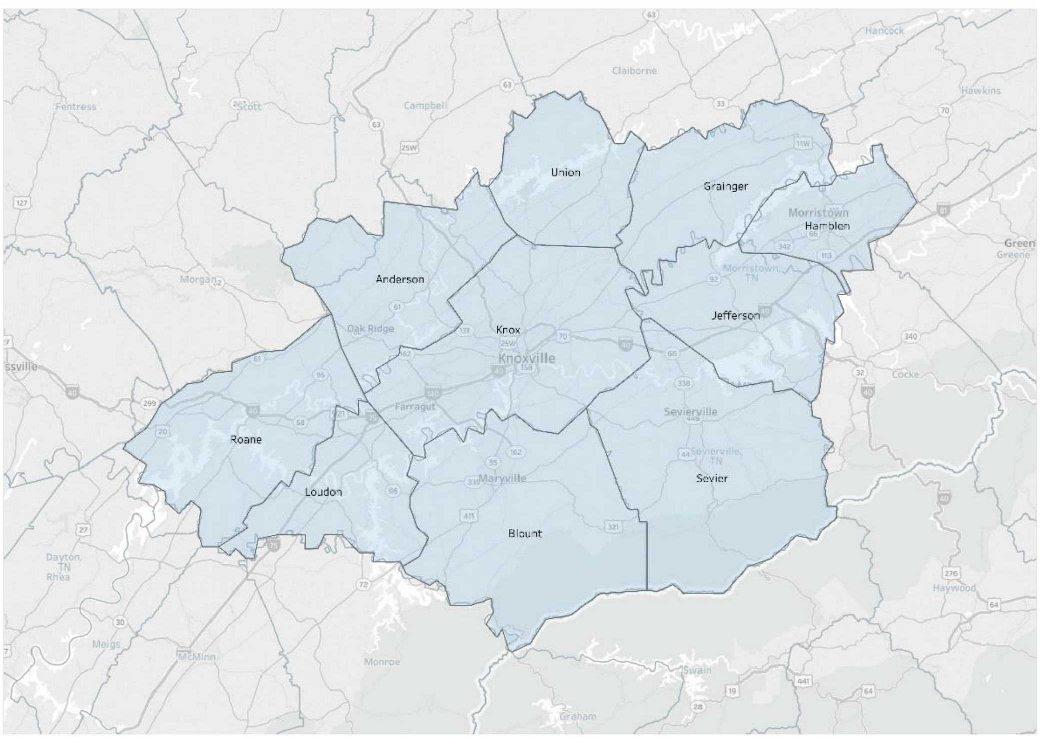

Figure 5 Knoxville Regional Travel Demand Model Extent

Traditionally, in road safety analysis, VMT is used as a variable to measure exposure. However, the VMT alone might not reflect the activity of residents and the amount of travel in the transportation network. To evaluate the activity of road users at the TAZ level (i.e., individual's exposure), we will use the zonal activity as Person Miles Traveled at the zonal level (PMT). PMT_i combines modeled trip rate and trip length for all population in zone i and is an index for measuring the zonal activity in each TAZ . PMT is calculated by equation 1:

$$PMT_i = \sum_{j=1}^{n} \frac{P_{ij} L_{ij}}{Pop_i} \qquad \text{Equation 9}$$

where n is the index of the destination TAZ, P_{ij} is the number of trips produced from TAZ i to TAZ j in one day, L_{ij} is the shortest network path between TAZ i to TAZ j, and Pop_i presents the population of the zone i. KRTM was used as a source to extract the number of trips for each pair. Shortest path between each pair was also extracted from the traffic assignment model at the peak-hour. Figure 4 presents the distribution of daily activity (PMT) per capita in the KRTM Model at the TAZ level. Visual screening of Figure 4 indicates that the rural areas have higher PMT compared to the urban areas.

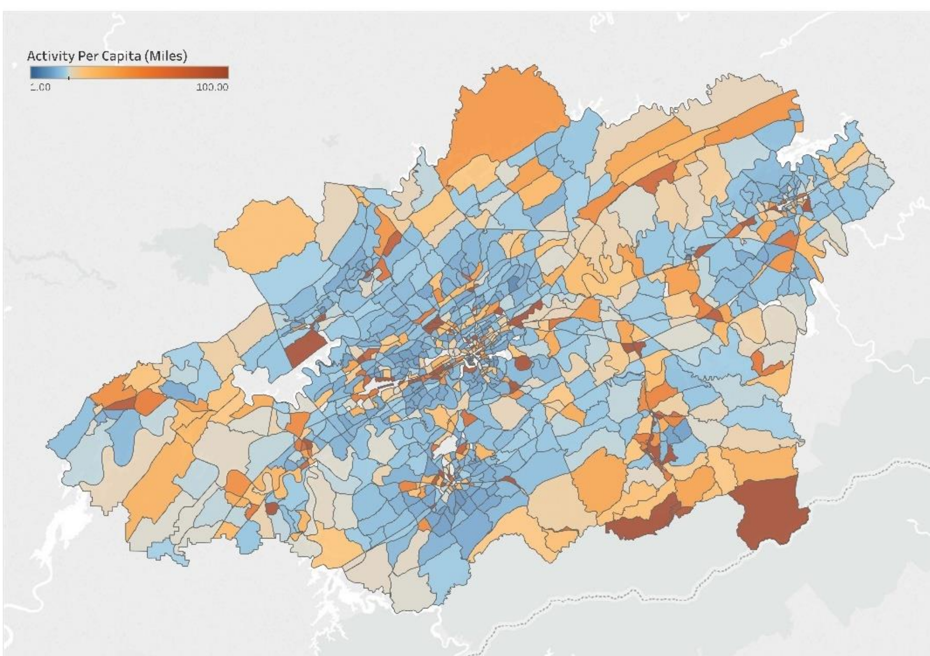

Figure 6 Average zonal activity (person miles traveled)

Data and geocoding process

The crash data in this study was provided by Tennessee Integrated Traffic Analysis Network (TITAN), the statewide crash data administered by the Tennessee Department of Safety and Homeland Security. The records of 60,104 crashes and information on 148,666 individuals who were involved in traffic crashes between 2015 and 2016 in the Knoxville region were retrieved from TITAN. Each record includes information about road user type (i.e., driver, motorcyclist, passenger, pedestrian, bicyclist, persons in the building, vehicle owner, witness, and property owner), coordinates of the crashes and addresses of the individual who were involved in traffic crashes. Since some of the individuals such as property owner, persons in the buildings (in case of a colliding with a building), witnesses, and vehicle owners did not have a direct role in traffic crashes, we excluded them from our analysis[2]. After obtaining the address of road users, we used the Bing Application Program Interface (API) services to geocode the addresses. The quality of the geocoding was checked by controlling for the locality of the addresses. Only those records that had an accuracy level of premises (e.g., property name, building name), address level accuracy, or intersection level accuracy was used for the analysis. We were able to successfully match 141,514 (95%) of the individuals with a home-location.

The economic cost of traffic crashes

The injury severity in TITAN database follows the KABCO scale for Tennessee provided by FHWA (FHWA 2011). In KABCO scale K, A, B, C, and O respectively stand for a crash with fatal, incapacitating, non-incapacitating evident, possible injury, and no-injury (FHWA 2017). In order to convert the injury severities to crash cost, we used the average values presented in Table 6 recommended by FHWA (Harmon *et al.* 2018) for the year 2010 for the person-injury unit. We transformed the injury cost to 2017 dollar by the inflation rate (Harmon *et al.* 2018). Notably, crashes with injury level of no-injury has a non-zero value; the non-zero value reflects the misclassification of the injury by police officers (Harmon *et al.* 2018). By using numbers presented in Table 6 and counting crash frequencies by severity at each census tract, we measured the total economic cost of the traffic crashes at the TAZ level by using the following equation:

$$ECCPC_i = \frac{(N_{v,i} * Cost_{PDO}) + \sum_{\alpha=\{K,A,B,C,O\}} N_{\alpha,i} * Cost_\alpha}{T * Pop_i} \quad \text{Equation 10}$$

where $N_{\alpha,i}$ represents the number of individual who live in zone i with the level of injury α, $Cost_\alpha$ presents the traffic injury cost per injury presented in Table 6 and T presents the period of the study (T = 2 years). $N_{v,i}$ presents the number of vehicles with a registered address in zone i that were involved in traffic crashes, and $Cost_{PDO}$ presents the vehicle unit damage cost. Figure 7 presents the distribution of the ECCPC at zonal

[2] None of them were injured as a result of traffic crash.

level in the study area. ECCPC tended to have distributed impacts, with high economic cost scattered throughout the region.

Lorenz curve and Gini coefficient

Lorenz curves have typically been used in the field of economics to explore the distribution of the inequalities across a population. This method has been used in transportation to explore the inequality in the transportation studies such as public transit and infrastructure investment (Delbosc and Currie 2011, Zofío *et al.* 2014, Xia *et al.* 2016). In an equitable manner, x% of the population pays x% of the economic cost of traffic crashes. for example, 10% of the population bears 10% of the public ECCPC (Straight-line presented in Figure 8). In reality, the distribution of the crash burden would be different from the straight line, and it is presented by the Lorenz curve. The Lorenz curve presents a graphical representation of inequity across a population. The Gini coefficient is a single value based on the area between the line of equality in perfectly equal distribution and the Lorenz curve representing the actual distribution (Atkinson 1970). The closer the Lorenz curve is to the line of equality the more equal the distribution is and the smaller the area enclosed between the two lines. The Gini coefficients range between 0 and 1. The Gini coefficient represents the area that lies between the line of equality and the Lorenz curve over the total area under the line of equality. The value close to 0 corresponds to perfect ECCPC equity and value close to 1 corresponds to perfect ECCPC inequality.

Table 6 National KABCO person-injury unit costs (2017 dollar)

Injury Type	Economic person-Injury Unit Costs
No Injury	$6,426
Possible Injury	$24,448
Non-Incapacitating Injury	$36,089
Incapacitating Injury	$94,994.3
Fatal Injury	$1,572,521.48
PDO Vehicle*	$6,830.03
Unknown	Not Applicable

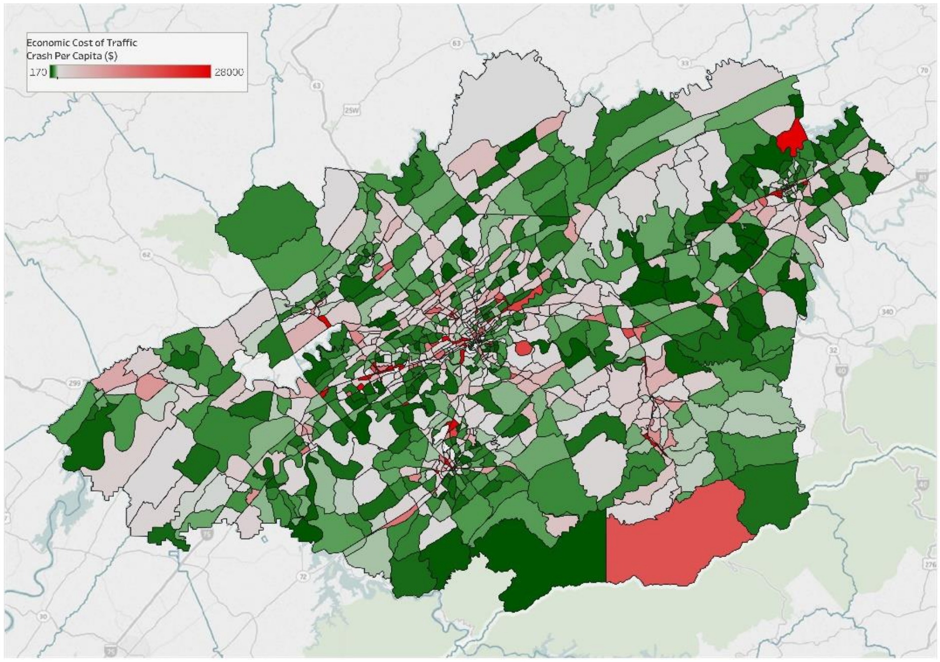

Figure 7 ECCPC distribution in KRTM

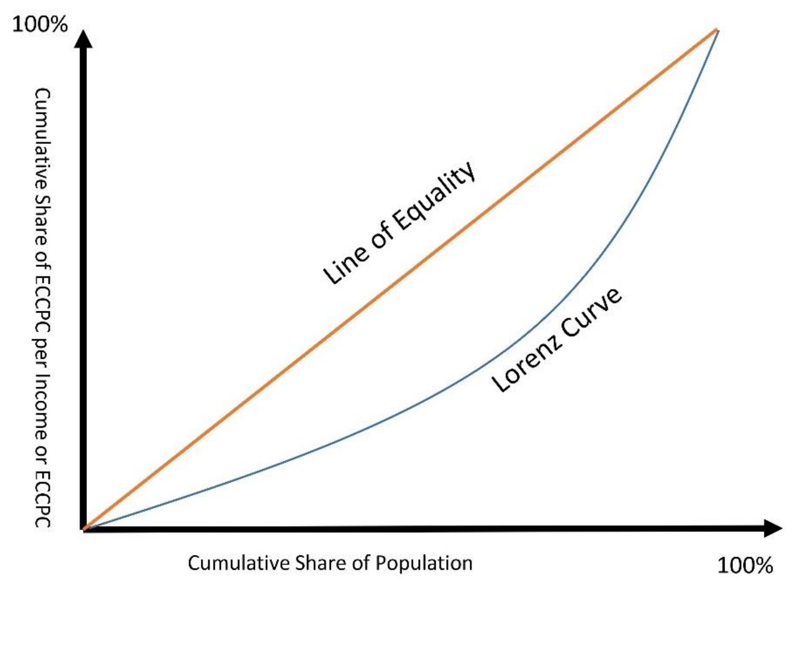

Figure 8 Gini Coefficient and Lorenz curve

Modeling approach

Testing spatial dependency

Visual inspection of Figure 7 indicates that neighborhoods with better safety records (i.e., green colors) are surrounded by other TAZs with blue colors. This is also the case for the TAZs with red colors. This may be an indicator of the presence of significant spatial autocorrelation. Spatial autocorrelation occurs when events occurring at different but nearby locations are correlated. In order to statistically check the presence of spatial autocorrelation, in this study we used global Moran's I statistics. Global Moran's I (Moran 1950) was also used to test whether the model residuals are spatially correlated. Moran's I values range from -1 to +1, where values close to 0 indicate no spatial correlation. Moran's I can be written as:

$$I = \frac{\sum_i \sum_j w_{ij}(y_i - \mu)(y_j - \mu)}{\sum_i (y_i - \mu)^2} \qquad \text{Equation 11}$$

where w_{ij} is an element of a row-standardized spatial weights matrix, y_i is the ECCPC, and μ is the average ECCPC in the sample. The statistical significance of the Moran's I is based on the z-score. For more details about the calculation of the Moran's I's Z-score please see Andrew and Ord (1981). A positive and significant Moran's I score indicates clustering in space of similar ECCPC.

By assuming the presence of significant spatial autocorrelation, we will use model specifications that consider the spatial dependency in their structure. Spatial error model (SEM)[3] and spatial autoregressive model (SAR) are two common models that are used by researchers to consider spatial autocorrelation in the road safety analysis (Quddus 2008, Lee *et al.* 2015a, Rhee *et al.* 2016). The distinction between the two models is the method that they consider spatial dependency (Doreian 1980, 1982). The SAR model considers the direct effect of one element's response on another's. This interdependency is consistent with the presence of an influence process. In the SEM model, the source of the interdependence of the error term is not known and could be due to various unobserved processes that do not involve a direct effect of geographical units on one another (Marsden and Friedkin 1993, Baller *et al.* 2001).

Spatial error model

In the SEM, the models' constant variable is treated as a spatially structured random effect vector. The core assumption in the SEM is that the observational units in close proximity should exhibit effects levels that are similar to those from neighboring units (LeSage and Pace 2009). The SEM is similar to the linear regression models with an additional term for the spatial dependency of errors in neighboring units. The SEM model can be written as:

---

[3] Not to be mistaken by Structural Equation Modeling

$$y = X\beta + \varepsilon \quad \text{Equation 12}$$
$$\varepsilon = \lambda W_{\varepsilon} + u = (I - \lambda W)^{-1}u \quad \text{Equation 13}$$
$$y = \lambda W_y + X\beta + \lambda \mathrm{W X}\beta + \mathrm{u} \quad \text{Equation 14}$$

where y is a vector of ECCPC, X is a vector of independent variables presented in Table 5, β is the corresponding vector of estimated coefficients on X. In this model, ε is the error term, which consists of two parts: W_{ε} and u. W_{ε} presents the spatially lagged error term corresponding to a weigh matrix W and u refers to the spatial uncorrelated error term that satisfies the normal regression assumption ($u \sim N(0, \sigma^2 I)$). Last, λ presents the spatial error term parameters, if the value of the spatial error parameters equals zero, the SEM is similar to the standard linear regression model.

Spatial autoregressive model

A similar approach that accounts for spatial correlation is the SAR model The SAR model can be represented as:

$$y = \rho W_y + X\beta + \varepsilon \quad \text{Equation 15}$$

where ρ presents the spatial autoregressive parameter, W_y is a spatially lagged variable corresponding to W matrix, X is a vector of independent variables, β is the vector of estimated coefficients. Last, ε is assumed to be a vector of independent and identically distributed (IID) error terms. Due to the endogeneity in the W_{ε} (spatial lag) term, ordinary least-squares (OLS) estimators are biased and inconsistent for the spatial-lag model, and instead maximum-likelihood estimation (Ord 1975) is used to obtain consistent estimators. (Kim *et al.* 2003). In order to estimate the SEM and SAR models, we used Geoda Software (Anselin 2003).

Weight matrix

Choosing a proper weight matrix is crucial for the analysis since it incorporates the prior structure of dependence between spatial units (Baller *et al.* 2001). The Rook and Queen contiguity matrix was used in this analysis to establish the weight matrix. The queen weights matrix define neighbors as census tracts that share a boundary or corner, whereas, rook only considers those census tract that shares a boundary (Anselin 2003). The selection of the optimal weighting matrix could be based on the AICc (Hurvich and Tsai 1989); the weight matrix with the lowest AICc is preferred (Fotheringham and Brunsdon, Nakaya *et al.* 2005, Hadayeghi *et al.* 2010, Nakaya 2014). For more information about the weighting matrix, please see Anselin (2003).

Model comparison and assessment

A Lagrange Multiplier (LM) is used to test the specifications against SEM and SAR. These tests are based on the regression residuals obtained from estimated the model under the null hypothesis regression (i.e., OLS). Each of SAR and SEM models has their specific LM statistics, which offers the opportunity to exploit the values of these statistics to suggest the likely alternative. The LM statistic against SEM (LMSEM) and SAR (LMSAR) models take the following forms:

$$LMSEM = \frac{\left(\frac{e'W_e}{s^2}\right)^2}{T} \qquad \text{Equation 16}$$

$$LMSAR = \frac{\left(\frac{e'W_e}{s^2}\right)^2}{\frac{(WXb)'M(WXb)}{s^2} + T} \qquad \text{Equation 17}$$

where e is a vector of OLS residuals, s^2 its estimated standard error, $T = tr[(W + W')W]$, tr as the matrix trace operator, and $M = I - X(X'X)^{-1}X'$. Both LMSEM and LMSAR are asymptotically distributed as $\chi^2(1)$ under the null. Several researchers illustrate the relative power of these tests by using extensive simulation studies (Anselin and Rey 1991, Anselin and Florax 1995, Anselin *et al.* 1996).

It is possible that in some cases both LMSEM and LMSAR statistics turn out to be highly significant, which makes it challenging to choose the proper alternative. To deal with this issue, (Anselin *et al.* 1996) developed a robust form of the LM statistics in the sense that each test is robust to the presence of local deviations from the null hypothesis in the form of the other alternative. In other words, the robust LME is robust to the presence of spatial lag, and vice versa. The robust tests perform well in a wide range of simulations and form the basis of a practical specification search, as illustrated in (Anselin and Florax 1995, Anselin *et al.* 1996). In this study, we used GeoDa software to perform the LM tests. The Queen contiguity matrix was used to generate a spatial weight matrix. In addition to LM, to further evaluate the overall model fit and predictive performance, we also used the Akaike Information Criterion ($AICc$) as a measure of the relative goodness of fit.

Results and discussion

Among those involved in traffic crashes, 308 (residence: 252; non-residence: 56) individuals were fatally injured as a result of traffic crashes in the KRTM Model area. Moreover, another 17,312 (residence: 14,225; non-residence: 3,087) individuals were injured (level A, B, or C). The economic cost of traffic crashes in the region for the two years between 2015-2016 was $2.5 Billion (2017 dollars). Over three quarters (78%) of crash victims were from the KRTM area. The economic costs of residents of the KRTM was $2.08 billion and for non-residents was $503 million. Table 7 presents more details on crash cost based on the driver residential address (KRTM resident vs. non-KRTM resident). For example, KRTM residents bore $263 million out of their pocket due to traffic crashes with a non-KRTM (external) drivers.

The mean and median value of ECCPC for the period between 2015-16 (for selected TAZs) was $1,399 and $702, respectively (max = $28,665), the 90th percentile spans $176 to $3,232. By using average family size at zonal level and normalizing the economic crash cost to median household income per capita; we find that the mean direct cost of traffic crashes consumed 5.6% (median: 3.85%) of annual families' income at zonal level; the 90th percentile spans 0.9 to 20.5%. Figure 9 presents the distribution of ECCPC, ECCPC per income and average zonal activity, as average zonal activity increases, both ECCPC, ECCPC per income increases. For example, TAZs with average zonal activity higher than 40, have a substantially higher ECCPC and ECCPC per income compared to those below 40. This trend also holds for distribution over income. Visual inspection of

Figure 10 indicates that TAZs with median household below $25,000 have substantially higher ECCPC and ECCPC per income compared to wealthier families. For example, TAZs with median household income of less than $15,000, the average ECCPC is equal to $1,500 which is 3 times higher than TAZs with median household income of more than $100,000. Likewise, by normalizing the ECCPC with income, we learned that the value of ECCPC per income for families with income less than $15,000 is 36 times higher (17% v. 0.47) than TAZs with median household income of more than $100,000. Figure 11 also presents the Lorenz curve, and the equity line, the Gini index coefficient for the ECCPC per income is 0.58, which is an indicator of the unequal distribution of the burden of traffic crashes.

Figure 12 presents the spatial distribution of the proportion of the economic cost of traffic crashes to families' income. The gray color in the map exhibits TAZs, where the proportion of the economic cost of traffic crashes to families' income, is less than 6%. The warmer color point out areas, in which direct cost of traffic crashes over families' income level, is more substantial. A visual inspection of traffic crashes in the study area reveals that burden of traffic crashes are larger for TAZs near I-40 (east/west) and multilane highways that connect major cities in the KRTM area (e.g., Knoxville to Maryville, Knoxville to Sevierville). One explanation for more tangible crash burden along the road network is the exposure of the residents to high volume corridors with high traffic speeds. These two factors may increase both crash frequency and severity. Moreover, households who live very close to these corridors could have lower household incomes.

Table 7 Economic cost of traffic crashes by driver and resident types (2017 million dollars)

Person Involved	**Driver Type**		
Residency	**KRTM**	**Non-KRTM**	**Grand Total**
Non-KRTM	19.2	484.1	503.3
KRTM	1,817.8	263.1	2,081.0
Grand Total	**1,837.0**	**747.2**	**2,584.3**

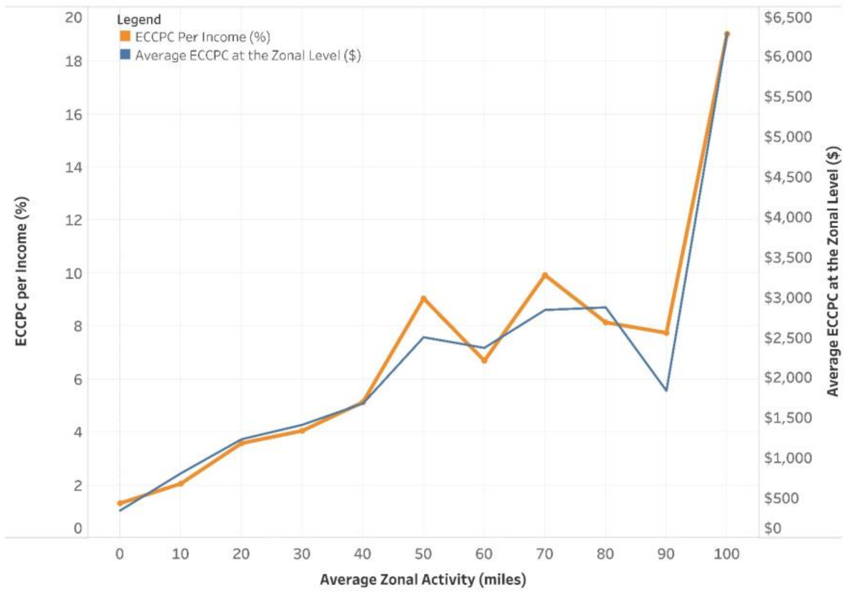

Figure 9 Distribution of the ECCPC & ECCPC per income with regards to average zonal activity

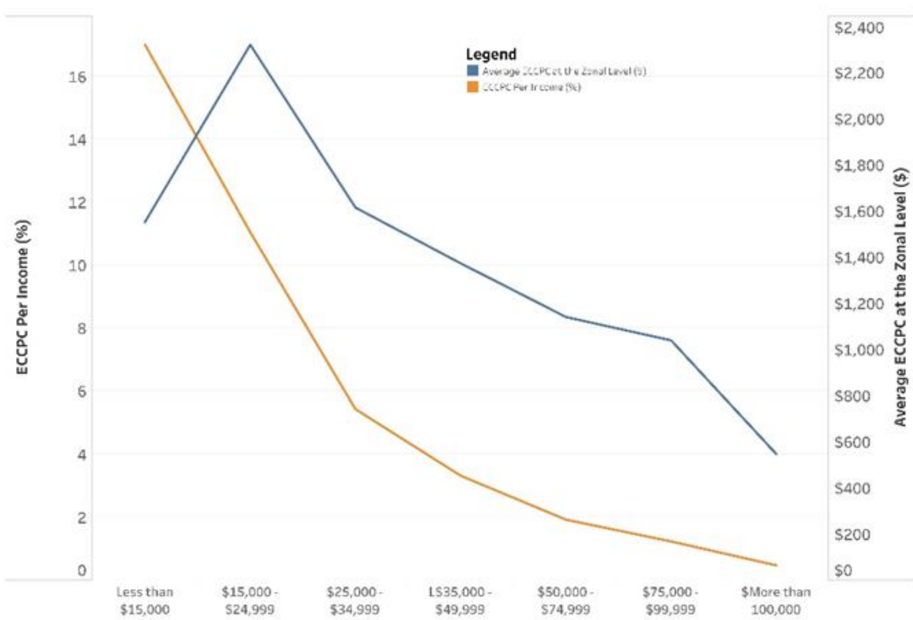

Figure 10 Distribution of the ECCPC & ECCPC per income with regards to median household income

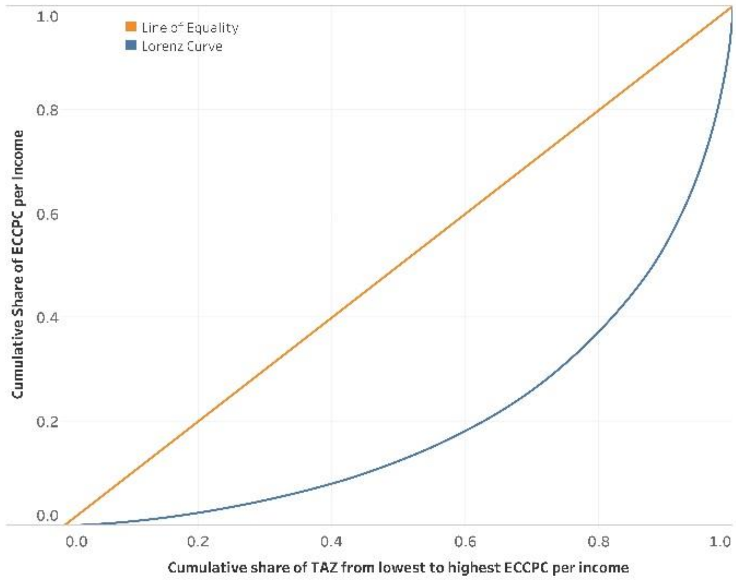

Figure 11 Gini Coefficient

Model evaluation

Results of the global Moran's I, using a Queen contiguity matrix, indicate that there is significant spatial autocorrelation (Moran's I = 0.14, p-value = 0.000). The positive value of the Moran's I indicate the clustering in ECCPC.

By controlling for AICc as well as lag coefficient values for the estimated SAR, and SEM models in different weighting matrices we learned that the queen contiguity matrix for both SAR and SEM has significantly better performance (lower AICc) compared to the other alternatives. Considering the non-zero values of ρ and λ ,we conclude that both SAR and SEM models are significantly different from linear regression models. In addition, Moran's I of residuals in both models (Moran's I: SEM = -0.013; SAR = 0.000) indicate that the residuals are not spatially correlated.

Comparison of the SEM, SAR, model by using LM indicate that both LMSEM and LMSAR are significant. However, using robust-LMSEM and robust-LMSAR tests for comparison indicate that only robust-LMSAR has a significant value. As a result, the SAR model is more suitable compared to the other models. Furthermore, comparison of the AICc and model performance, the SAR model has the lowest value of the AICc; therefore, the SAR model is more suitable compared to OLS and SEM. Table 8 presents the result of the estimated models.

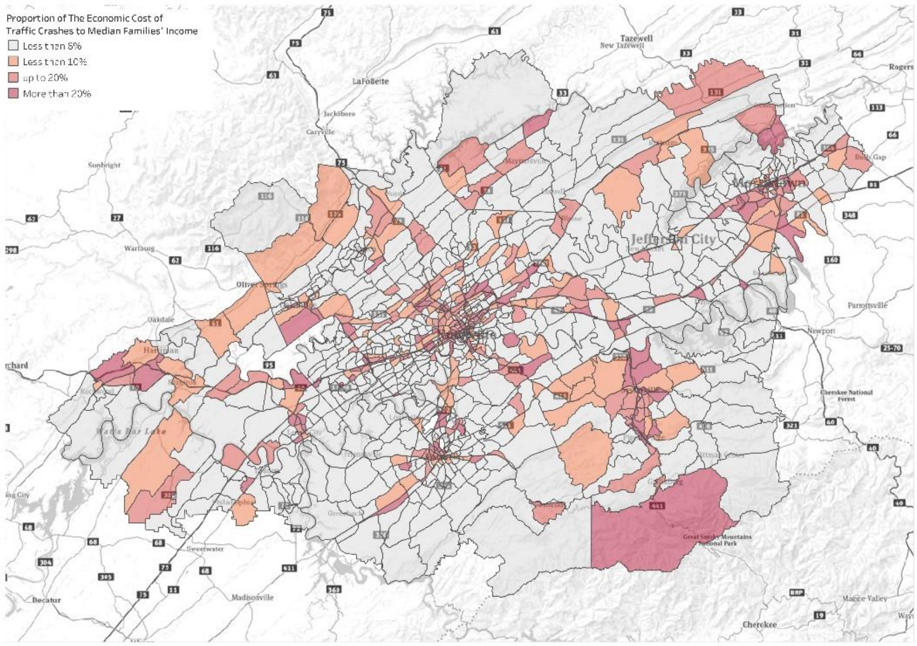

Figure 12 Proportion of The Economic Cost of Traffic Crashes to Median Families' Income

Parameters estimation and discussion

All the variables presented in Table 8 (except student per household and intersection per density) have a significant and intuitive association with ECCPC in three estimated models. In this study, we used the average zonal activity as the individuals' exposure variable for each TAZ. Therefore, we expected a positive sign for the estimated coefficients. Average zonal activity implies that those who travel longer distances are more prone to traffic crashes and traffic crashes have a greater impact on them.

Congruent with previous studies, VMT of roadways in the zone also have a significant association with safety outcomes. (Pirdavani *et al.* 2012b, a, 2013a, Pirdavani *et al.* 2013b, Lee *et al.* 2015b, Cheng *et al.* 2018) Comparison of the coefficients indicates that vehicle miles traveled on arterial roads (i.e., major and minor arterials) has a greater impact on ECCPC compared to the interstate. This differences in the magnitudes could reflect the high access of the arterial roads with more conflicts compared to interstates, which could increase the likelihood of severe crashes; considering the relatively higher speeds on arterials could be another factor contributing to the higher severity of traffic crashes. On the other hand, other road classifications (e.g., collector, local) has a negative association with ECCPC. Although many studies explored the association between of functional classes and crash frequency at zonal level (e.g., Hadayeghi *et al.* 2003, Quddus 2008, Xu and Huang 2015), only a few considered the effect of exposure (i.e., VMT) in different road classes. There is also a need to consider that the definition of the functional classes may vary across areas. In a series of studies in Flanders, Belgium, Pirdavani *et al.* (2013a) and Pirdavani *et al.* (2012b) reported that VMT on a motorway had a smaller effect on total crash frequency compared to non-motorway VMT. In Florida, Xu and Huang (2015), reported that proportions of the road with speed limits 25 mph or lower had a negative association with crash frequency at a zonal level, whereas, percent of roads at 45 mph and above had positive association on zone crash frequencies. Hadayeghi *et al.* (2003) also reported that total local road length in a TAZ had a negative association with all crashes and severe crashes; whereas, arterials, expressways, collectors, and ramps had a positive and significant association with crash frequency at the zonal level in a study in Canada.

The significant positive association of the worker per household variable indicates that as proportion of workers per household increases (i.e., the proposed increase in work trip frequency) ECCPC also increases. This finding agrees with Naderan and Shahi (2010) study where they reported that the number of work-trips produced at zonal level has a positive impact with the number of injury crashes, property damage only crashes, and total crashes in a TAZ.

Population density also has a negative association with the economic cost of traffic crashes; the model predicts that as density increases the ECCPC decreases. The crash frequency in urban areas is higher than rural areas on average; whereas the crash severity is relatively lower (Zwerling *et al.* 2005), as a result, the average economic cost of traffic crashes in the urban areas is lower than rural areas. Furthermore, population density could be used as a surrogate for non-motorized transportation; non-motorized trips are more likely in areas with higher density (Siddiqui *et al.* 2012, Cai *et al.* 2017b); non-motorized road users do not impose a crash risk to other road users.

The household income variable also has a negative association with ECCPC, consistent with previous studies (Pirdavani *et al.* 2012b, Pirdavani *et al.* 2013b, Cai *et al.* 2017a, Cai *et al.* 2017b, Gomes *et al.* 2017, Cheng *et al.* 2018). People with higher household incomes tend to have lower crash rates and, in our model, lower ECCPC. This negative sign also is in agreement with road safety literature (WHO 2015b, Marshall and Ferenchak 2017). In addition, it is possible that individuals with higher income use safer vehicles. As a result, their crash severity and eventually the economic cost of their traffic crashes decreases.

As expected, road network characteristics have a significant association with safety level. Percent of roads with sidewalk and number of bus stations also have a significant positive association with ECCPC. Cai *et al.* (2017b) also reported that sidewalk length has a positive association with crash frequency, severe crash, and non-motorized crash frequency. Considering that sidewalk is utilized by vulnerable road users, we may expect higher injury severity in case of crashes with this road user type and hence, higher ECCPC; this trend also holds on for the number of bus stops, in which more non-motorized road users have access to. Intersection density in the TAZ also has a positive (but non-significant) association with ECCPC. Other literature found that the number of intersection could be correlated with higher numbers of conflict and accordingly the higher number of traffic crashes (Hadayeghi *et al.* 2003, Ladron de Guevara *et al.* 2004, Lovegrove and Sayed 2006, Abdel-Aty *et al.* 2011, Pirdavani *et al.* 2012a, Gomes *et al.* 2017). It is well-established that speed is a contributing factor to both crash frequency and crash severity (Elvik *et al.* 2009, HSM 2010). The average speed of roads in a TAZ has a positive association with ECCPC agreeing with previous research (Hadayeghi *et al.* 2003, Abdel-Aty *et al.* 2011, Pirdavani *et al.* 2012a),

Conclusion

The main aim of this study was to explore the association between travel behavior, and economic cost of traffic crashes at a fine geographic level, aiming to highlight equity challenges associated with disparities in crash cost burden. To explore this problem, we used the home-address of individuals who were involved in traffic crashes in the study area and assigned the economic cost of traffic crashes to their corresponding TAZ. We also determined activity (PMT) per capita for residents of each TAZ to measure their exposure in the transportation network.

By controlling the traffic crash burden by the average zonal activity, we learned that the burden of traffic crashes is higher for those who travel more or have a lower income. The high-value of the Gini index also indicates that ECCPC per income impact residents of the KRTM in an unequable manner. Our analysis indicates that spatial dependency exists in the ECCPC and it is not randomly distributed in space. Our analysis also suggests that that ECCPCs are not generated solely by the internal structural factors represented in the OLS model. Comparison of different spatial models indicates that the SAR model with Queen contiguity matrix is more suitable for interpreting the relationship between ECCPC and travel behavior characteristics at the zonal level. Considering the underlying assumptions of the SAR model, we may conclude that ECCPC in one TAZ is influenced by ECCPC in neighboring TAZs. Therefore, a neighborhood with poor traffic safety outcomes poses negative externality to its neighbors and vice versa.

Table 8 Results of OLS, SAR and SEM models for prediction of ECCPC

	SEM			SAR			OLS		
Variable	**Coefficient**	**Std. Error**	**P-value**	**Coefficient**	**Std. Error**	**P-value**	**Coefficient**	**Std. Error**	**P-value**
Average zonal activity	21.008	1.088	0.000	20.783	1.075	0.000	21.100	1.089	0.000
Average Speed	15.488	8.507	0.069	16.100	8.232	0.050	17.187	8.377	0.040
Income ($10,000)	-82.673	33.438	0.013	-74.930	30.746	0.015	-92.292	31.158	0.003
Worker Per Household	789.818	287.103	0.006	842.896	276.897	0.002	927.897	281.764	0.001
Student per Household	-39.040	349.943	0.911	7.374	336.439	0.983	-45.856	342.180	0.893
Intersection Density (per square miles)	0.663	0.439	0.131	0.631	0.422	0.135	0.765	0.429	0.075
Percent road with Sidewalk	1176.700	273.907	0.000	1132.080	263.859	0.000	1205.690	268.151	0.000
Percent Near Bus Station	485.042	242.838	0.046	433.221	223.289	0.052	503.682	226.838	0.027
Population Density (per Square miles)	-0.112	0.027	0.000	-0.115	0.025	0.000	-0.120	0.025	0.000
VMT Interstate (10,000 miles)	156.811	28.148	0.000	150.221	27.410	0.000	176.145	34.699	0.000
VMT Arterial (10,000 miles)	172.467	34.784	0.000	165.570	34.209	0.000	-155.453	37.024	0.000
VMT Others (10,000 miles)	-147.868	37.433	0.000	-145.169	36.410	0.000	21.100	1.089	0.000
Constant	-788.099	492.471	0.110	16.100	8.232	0.050	-983.910	482.956	0.042
Lag Coef. (Lambda)	0.153	0.049	0.002						
Lag Coef. (Rho)				0.17	0.04	0.00			
Moran's I	-0.013			0.000			0.14		0.000
Log likelihood (Full)	-8473.89			-8470.68			-8437.61		
LMSEM				8.4847			0.004		
Robust LMSEM				0.6037			0.437		
LMSAR	15.0911		0.000						
Robust LMSAR	7.2101		0.007						
Akaike info criterion	16973.8			16969.4			16982.4		
Corrected Akaike info criterion	16894.2			16888.9			16901.8		
R-squared	0.42			0.42			0.41		
Number of Observations	956			956			956		

Geographic distribution of the negative externalities of the traffic crashes shows that the burden of traffic crashes is more tangible in the vicinities of the interstates and multilane highways where TAZs' residents are more prone to high-speed traffic and higher road classification. First, by designing a transportation network with the aim of diverging high-speed traffic from residential areas or managing the accessibility of the residents near the high-speed, high volume roads. The second strategy may target average zonal activity by eliminating a portion of trips by promoting sustainable transport. Moreover, an increase in diversity, mixed land-use design, and non-motorized oriented design would also reduce both trips rate, trip length, modal shift (Cervero and Kockelman 1997) and eventually average zonal activity. Reduction in average zonal activity and VMT has a direct impact on the economic cost of traffic crashes. The economic cost of traffic crashes at the zonal level could also be used as an index for allocating proper countermeasures and interventions to areas where the burden of traffic crashes is more tangible, which can be done by investment in the safer infrastructure and educational interventions.

In summary, in this study, we introduced a method to measure the tangible cost of traffic crashes at the zonal level, which could be straightforwardly integrated to travel demand analysis. The authors recommend using this measure as a criterion to evaluate future scenarios of development of the transportation system in metropolitan areas to identify how those scenarios impact safety costs and distributional impacts of safety externalities.

Acknowledgment

The authors would like to thank the Tennessee Department of Safety and Homeland Security for providing the data for this study. This project was supported by the Collaborative Sciences Center for Road Safety, www.roadsafety.unc.edu, a U.S. Department of Transportation National University Transportation Center promoting safety. The study design was reviewed and approved by the University of Tennessee Institutional Review Board.

References

Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating Trip and Roadway Characteristics to Manage Safety in Traffic Analysis Zones. Transportation Research Record: Journal of the Transportation Research Board (2213), 20-28.

Aguero-Valverde, J., 2013. Multivariate Spatial Models of Excess Crash Frequency at Area Level: Case of Costa Rica. Accident Analysis & Prevention 59, 365-373.

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial Analysis of Fatal and Injury Crashes in Pennsylvania. Accident Analysis & Prevention 38 (3), 618-625.

Ahadi, M.R., Razi-Ardakani, H., 2015. Estimating the Cost of Road Traffic Accidents in Iran Using Human Capital Method. International Journal of Transportation Engineering 2 (3), 163-178.

Andrew, C., Ord, J.K., 1981. Spatial Processes: Models and Applications. London: Pion.

Anselin, L., 2003. Geoda 0.9 User's Guide. Urbana 51, 61801.

Anselin, L., Bera, A.K., Florax, R., Yoon, M.J., 1996. Simple Diagnostic Tests for Spatial Dependence. Regional science and urban economics 26 (1), 77-104.

Anselin, L., Florax, R.J., 1995. Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results. New Directions in Spatial Econometrics. Springer, pp. 21-74.

Anselin, L., Rey, S., 1991. Properties of Tests for Spatial Dependence in Linear Regression Models. Geographical analysis 23 (2), 112-131.

Atkinson, A.B., 1970. On the Measurement of Inequality. Journal of economic theory 2 (3), 244-263.

Baker, S.P., Braver, E.R., Chen, L.-H., Pantula, J.F., Massie, D., 1998. Motor Vehicle Occupant Deaths among Hispanic and Black Children and Teenagers. Archives of pediatrics & adolescent medicine 152 (12), 1209-1212.

Baller, R.D., Anselin, L., Messner, S.F., Deane, G., Hawkins, D.F., 2001. Structural Covariates of Us County Homicide Rates: Incorporating Spatial Effects. Criminology 39 (3), 561-588.

Blatt, J., Furman, S.M., 1998. Residence Location of Drivers Involved in Fatal Crashes. Accident Analysis & Prevention 30 (6), 705-711.

Blincoe, L., Miller, T., Zaloshnja, E., Lawrence, B.A., 2015. The Economic and Societal Impact of Motor Vehicle Crashes, 2010.(Revised)(Report No. Dot Hs 812 013). Washington, DC: National Highway Traffic Safety Administration.

Braver, E.R., 2003. Race, Hispanic Origin, and Socioeconomic Status in Relation to Motor Vehicle Occupant Death Rates and Risk Factors among Adults. Accident Analysis & Prevention 35 (3), 295-309.

Cai, Q., Abdel-Aty, M., Lee, J., 2017a. Macro-Level Vulnerable Road Users Crash Analysis: A Bayesian Joint Modeling Approach of Frequency and Proportion. Accident Analysis & Prevention 107, 11-19.

Cai, Q., Abdel-Aty, M., Lee, J., Eluru, N., 2017b. Comparative Analysis of Zonal Systems for Macro-Level Crash Modeling. Journal of safety research 61, 157-166.

Campos-Outcalt, D., Bay, C., Dellapena, A., Cota, M.K., 2003. Motor Vehicle Crash Fatalities by Race/Ethnicity in Arizona, 1990–96. Injury Prevention 9 (3), 251-256.

Cervero, R., Kockelman, K., 1997. Travel Demand and the 3ds: Density, Diversity, and Design. Transportation Research Part D: Transport and Environment 2 (3), 199-219.

Cheng, W., Gill, G.S., Ensch, J.L., Kwong, J., Jia, X., 2018. Multimodal Crash Frequency Modeling: Multivariate Space-Time Models with Alternate Spatiotemporal Interactions. Accident Analysis & Prevention 113, 159-170.

Clark, D.E., 2003. Effect of Population Density on Mortality after Motor Vehicle Collisions. Accident Analysis & Prevention 35 (6), 965-971.

Delbosc, A., Currie, G., 2011. The Spatial Context of Transport Disadvantage, Social Exclusion and Well-Being. Journal of Transport Geography 19 (6), 1130-1137.

Dong, N., Huang, H., Xu, P., Ding, Z., Wang, D., 2014. Evaluating Spatial-Proximity Structures in Crash Prediction Models at the Level of Traffic Analysis Zones. Transportation Research Record: Journal of the Transportation Research Board (2432), 46-52.

Dong, N., Huang, H., Zheng, L., 2015. Support Vector Machine in Crash Prediction at the Level of Traffic Analysis Zones: Assessing the Spatial Proximity Effects. Accident Analysis & Prevention 82, 192-198.

Doreian, P., 1980. Linear Models with Spatially Distributed Data: Spatial Disturbances or Spatial Effects? Sociological Methods & Research 9 (1), 29-60.

Doreian, P., 1982. Maximum Likelihood Methods for Linear Models: Spatial Effect and Spatial Disturbance Terms. Sociological Methods & Research 10 (3), 243-269.

Elvik, R., Vaa, T., Erke, A., Sorensen, M., 2009. The Handbook of Road Safety Measures Emerald Group Publishing.

FHWA, 2011. Kabco Injury Classification Scale and Definitions.

FHWA, 2017. Kabco Injury Classification Scale and Definitions.

Fotheringham, A., Brunsdon, C., M. Charlton. 2002. Geographically Weighted Regression-the Analysis of Spatially Varying Relationships. Chichester, UK: John Wiley & Sons.

García-Altés, A., Pérez, K., 2007. The Economic Cost of Road Traffic Crashes in an Urban Setting. Injury prevention 13 (1), 65-68.

Gomes, M.J.T.L., Cunto, F., da Silva, A.R., 2017. Geographically Weighted Negative Binomial Regression Applied to Zonal Level Safety Performance Models. Accident Analysis & Prevention 106, 254-261.

Hadayeghi, A., Shalaby, A., Persaud, B., 2003. Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. Transportation Research Record: Journal of the Transportation Research Board (1840), 87-95.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of Planning Level Transportation Safety Tools Using Geographically Weighted Poisson Regression. Accident Analysis & Prevention 42 (2), 676-688.

Harmon, T., Bahar, G., Gross, F., 2018. Crash Costs for Highway Safety Analysis. USDOT FHWA Office, Washington, DC

Harper, J.S., Marine, W.M., Garrett, C.J., Lezotte, D., Lowenstein, S.R., 2000. Motor Vehicle Crash Fatalities: A Comparison of Hispanic and Non-Hispanic Motorists in Colorado. Annals of emergency medicine 36 (6), 589-596.

Hezaveh, A.M., Cherry, C.R., Year. A New Approach to Aggregate Crash Prediction Model: Home-Based Crash Frequency Modeling. In: Proceedings of the Southern District ITE Annual Meeting.

Hosseinpour, M., Sahebi, S., Zamzuri, Z.H., Yahaya, A.S., Ismail, N., 2018. Predicting Crash Frequency for Multi-Vehicle Collision Types Using Multivariate Poisson-Lognormal Spatial Model: A Comparative Analysis. Accident Analysis & Prevention.

HSM, 2010. Highway Safety Manual. American Association of State Highway and Transportation Officials 2, 10-1.

Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., Abdel-Aty, M., 2016. Macro and Micro Models for Zonal Crash Prediction with Application in Hot Zones Identification. Journal of Transport Geography 54, 248-256.

Hurvich, C.M., Tsai, C.-L., 1989. Regression and Time Series Model Selection in Small Samples. Biometrika 76 (2), 297-307.

Jacobs, G., Aeron-Thomas, A., Astrop, A., 2000. Estimating Global Road Fatalities.

Kanafani, A., 1983. Transportation Demand Analysis.

Kim, C.W., Phipps, T.T., Anselin, L., 2003. Measuring the Benefits of Air Quality Improvement: A Spatial Hedonic Approach. Journal of environmental economics and management 45 (1), 24-39.

KRTM, 2012. Knoxville Regional Travel Model Update 2012 Model Development and Validation Report. Bernardin, Lochmueller & Associates, Inc., Evansville, IN, USA.

Ladron de Guevara, F., Washington, S., Oh, J., 2004. Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. Transportation Research Record: Journal of the Transportation Research Board (1897), 191-199.

Lee, J., Abdel-Aty, M., Choi, K., Huang, H., 2015a. Multi-Level Hot Zone Identification for Pedestrian Safety. Accident Analysis & Prevention 76, 64-73.

Lee, J., Abdel-Aty, M., Jiang, X., 2015b. Multivariate Crash Modeling for Motor Vehicle and Non-Motorized Modes at the Macroscopic Level. Accident Analysis & Prevention 78, 146-154.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics Chapman and Hall/CRC.

Li, Z., Wang, W., Liu, P., Bigham, J.M., Ragland, D.R., 2013. Using Geographically Weighted Poisson Regression for County-Level Crash Modeling in California. Safety science 58, 89-97.

Lovegrove, G.R., Sayed, T., 2006. Macro-Level Collision Prediction Models for Evaluating Neighbourhood Traffic Safety. Canadian Journal of Civil Engineering 33 (5), 609-621.

Males, M.A., 2009. Poverty as a Determinant of Young Drivers' Fatal Crash Risks. Journal of safety research 40 (6), 443-448.

Marsden, P.V., Friedkin, N.E., 1993. Network Studies of Social Influence. Sociological Methods & Research 22 (1), 127-151.

Marshall, W.E., Ferenchak, N.N., 2017. Assessing Equity and Urban/Rural Road Safety Disparities in the Us. Journal of Urbanism: International Research on Placemaking and Urban Sustainability 10 (4), 422-441.

Mayrose, J., Jehle, D.V., 2002. An Analysis of Race and Demographic Factors among Motor Vehicle Fatalities. Journal of Trauma and Acute Care Surgery 52 (4), 752-755.

McAndrews, C., Beyer, K., Guse, C.E., Layde, P., 2013. Revisiting Exposure: Fatal and Non-Fatal Traffic Injury Risk across Different Populations of Travelers in Wisconsin, 2001–2009. Accident Analysis & Prevention 60, 103-112.

Miaou, S.-P., Song, J.J., Mallick, B.K., 2003. Roadway Traffic Crash Mapping: A Space-Time Modeling Approach. Journal of Transportation and Statistics 6 (1).

MMUCC, 2012. Model Minimum Uniform Crash Criteria. DOT HS 811, 631.

Mohammadi, M., Shafabakhsh, G., Naderan, A., 2018. Effects of Modal Shares on Crash Frequencies at Aggregate Level. Accident Analysis & Prevention 120, 295-303.

Mohan, D., Year. Social Cost of Road Traffic Crashes in India. In: Proceedings of the Proceedings First Safe Community Conference on Cost of Injury. Viborg, Denmark, pp. 33-38.

Moran, P.A., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37 (1/2), 17-23.

Naderan, A., Shahi, J., 2010. Aggregate Crash Prediction Models: Introducing Crash Generation Concept. Accident Analysis & Prevention 42 (1), 339-346.

Nakaya, T., 2014. Gwr4 User Manual. WWW Document. Available online: http://www. st-andrews. ac. uk/geoinformatics/wp-content/uploads/GWR4manual_201311. pdf (accessed on 4 November 2013).

Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically Weighted Poisson Regression for Disease Association Mapping. Statistics in medicine 24 (17), 2695-2717.

Ord, K., 1975. Estimation Methods for Models of Spatial Interaction. Journal of the American Statistical Association 70 (349), 120-126.

Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., 2012a. Application of Different Exposure Measures in Development of Planning-Level Zonal Crash Prediction Models. Transportation Research Record: Journal of the Transportation Research Board (2280), 145-153.

Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., Year. Developing Zonal Crash Prediction Models with a Focus on Application of Different Exposure Measures.

Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., 2013a. Evaluating the Road Safety Effects of a Fuel Cost Increase Measure by Means of Zonal Crash Prediction Modeling. Accident Analysis & Prevention 50, 186-195.

Pirdavani, A., Brijs, T., Bellemans, T., Wets, G., Year. Spatial Analysis of Fatal and Injury Crashes in Flanders, Belgium: Application of Geographically Weighted Regression Technique.

Quddus, M.A., 2008. Modelling Area-Wide Count Outcomes with Spatial Correlation and Heterogeneity: An Analysis of London Crash Data. Accident Analysis & Prevention 40 (4), 1486-1497.

Rhee, K.-A., Kim, J.-K., Lee, Y.-I., Ulfarsson, G.F., 2016. Spatial Regression Analysis of Traffic Crashes in Seoul. Accident Analysis & Prevention 91, 190-199.

Romano, E.O., Tippetts, A.S., Voas, R.B., 2006. Language, Income, Education, and Alcohol-Related Fatal Motor Vehicle Crashes. Journal of ethnicity in substance abuse 5 (2), 119-137.

Saha, D., Alluri, P., Gan, A., Wu, W., 2018. Spatial Analysis of Macro-Level Bicycle Crashes Using the Class of Conditional Autoregressive Models. Accident Analysis & Prevention.

Schiff, M., Becker, T., 1996. Trends in Motor Vehicle Traffic Fatalities among Hispanics, Non-Hispanic Whites and American Indians in New Mexico, 1958–1990. Ethnicity & health 1 (3), 283-291.

Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes. Accident Analysis & Prevention 45, 382-391.

Stamatiadis, N., Puccini, G., 2000. Socioeconomic Descriptors of Fatal Crash Rates in the Southeast USA. Injury control and safety promotion 7 (3), 165-173.

Wegman, F., Oppe, S., 2010. Benchmarking Road Safety Performances of Countries. Safety science 48 (9), 1203-1211.

WHO, 2015a. Global Status Report on Road Safety 2015 World Health Organization.

WHO, 2015b. Global Status Report on Road Safety 2015. World Health Organization. Violence Injury Prevention World Health Organization.

Wijnen, W., Stipdonk, H., 2016. Social Costs of Road Crashes: An International Analysis. Accident Analysis & Prevention 94, 97-106.

Xia, J.C., Nesbitt, J., Daley, R., Najnin, A., Litman, T., Tiwari, S.P., 2016. A Multi-Dimensional View of Transport-Related Social Exclusion: A Comparative Study of Greater Perth and Sydney. Transportation Research Part A: Policy and Practice 94, 205-221.

Xu, P., Huang, H., 2015. Modeling Crash Spatial Heterogeneity: Random Parameter Versus Geographically Weighting. Accident Analysis & Prevention 75, 16-25.

Zofío, J.L., Condeço-Melhorado, A.M., Maroto-Sánchez, A., Gutiérrez, J., 2014. Generalized Transport Costs and Index Numbers: A Geographical Analysis of Economic and Infrastructure Fundamentals. Transportation Research Part A: Policy and Practice 67, 141-157.

Zwerling, C., Peek-Asa, C., Whitten, P., Choi, S.-W., Sprince, N., Jones, M.P., 2005. Fatal Motor Vehicle Crashes in Rural and Urban Areas: Decomposing Rates into Contributing Factors. Injury Prevention 11 (1), 24-28.

CHAPTER III: NEIGHBORHOOD-LEVEL FACTORS AFFECTING SEAT BELT USE

A version of this chapter was originally published by Amin Mohamadi Hezaveh and Christopher R. Cherry:

Hezaveh, A. M., & Cherry, C. R. (2019). Neighborhood-level factors affecting seat belt use. Accident Analysis & Prevention, 122, 153-161.

The authors confirm contribution to the paper as follows: study conception and design: Amin Mohamadi Hezaveh, Christopher Cherry; data collection: Not Applicable; analysis and interpretation of results: Amin Mohamadi Hezaveh, Christopher Cherry; draft manuscript preparation: Amin Mohamadi Hezaveh, Christopher Cherry. All authors reviewed the results and approved the final version of the manuscript.

Abstract

Despite the well-known safety benefits of seat belt use, some vehicle occupants still do not use their seat belt. This is a challenge in Tennessee, which has a lower seat belt use rate compared to the US national average. Roadside observations and interviews as the two main sources for estimating seat belt use rate are confronted with several limitations (e.g., small sample size, social desirability bias). To address these limitations, we attributed seat belt use of individuals who were involved in traffic crashes (N = 542,776) to their corresponding home-addresses. Home-addresses were retrieved from police crash database and were geocoded and assigned to their corresponding census tract revealing added information about the spatial distribution of seat belt use and socioeconomics of the areas surrounding the crash victim’s home. The average seat belt use rate in the metropolitan area was 88% and for the non-metropolitan area was 87%. A Tobit model was used to evaluate the relationship between the seat belt use rate for both drivers and passengers over 16 years old, with neighborhood sociodemographic variables. Population, age cohorts, race, household vehicles’ ownership, household size, and education were among the predictors of the seat belt use rate. Results of this analysis could be used in safety campaign design to reach geographic areas of groups with a lower seat belt use rate.

Keywords: Seat Belt Use Rate; Tennessee; Census Data; Tobit Model; Home-Address; Seat Belt Hotspots

Introduction

Approximately 1000 individuals die on Tennessee's roads every year, and most of them are vehicle occupants. One known solution that reduces the fatality rate of the vehicles' occupant is a proper use of a seat belt. Several studies have reported the merits of wearing a seat belt in reducing crash fatalities and injury rates. Appropriate use of seat belts increases the chance of vehicle occupants surviving potential fatal crash by 44% - 73% depending on seating position and the type of vehicles involved in a traffic crash (Blincoe *et al.* 2015).

There are mandatory seat belt laws in the United States and its territories (except New Hampshire). In 34 States, the District of Columbia, and Puerto Rico seat belt laws are primary, which enable law enforcement officers to stop vehicles and write citations when they observe a seat belt non-use (IIHS 2018). In 15 States the laws specified secondary enforcement, meaning that law enforcement officers are permitted to issue a seat belt citation only after they stop a vehicle for another primary violation. Notably, only 28 states and two territories enforce rear seat belt use (NHTSA 2017b). In Tennessee, seat belt use is a primary law, and it is mandatory for all the vehicles occupant be restrained by a seat belt (i.e., secured shoulder and lap belts) when riding in the front seat of a vehicle. Licensed passengers 16 years old or older are responsible for their own conduct. Nevertheless, a ten-year trend of traffic crashes shows that 30% of Tennessean who died in traffic crashes failed to wear their seat belt properly at the time of the crash, this rate was 54% and 70% for incapacitating injuries and non-incapacitating injuries, respectively (TITAN 2017).

Based on NHTSA roadside observations, front row passengers in Tennessee had an 88.9% seat belt use rate in 2016, which was 1.2% lower than the National average (NHTSA 2017a). In 2017, roadside observations of 27,000 vehicles' occupants at 190 sites in Tennessee revealed that, on average, 88.5% in Tennessee used their seat belt (CTR 2018), which was still lower than the national average. Females seat belt use rate was 93.8%, and males had an 85.0% seat belt use rate. Furthermore, freeways showed the highest usage rate (91.2%) of all roadway types, while those observed on local roadways had the lowest usage rate (86.1%) (THSO 2016, CTR 2018). In addition, another phone interview in Tennessee in 2017 reported that 90% of respondents always wore their seat belt, females also had higher seat belt use rate than males (Hezaveh *et al.* 2018a).

Seat belt non-use could be attributed to human factors such as forgetfulness, laziness, perceived low risk of injury, and discomfort (Begg and Langley 2001); attitudes, beliefs, and intentions (Fhaner and Hane 1975, Jonah and Dawson 1982, Chliaoutakis *et al.* 2000, Şimşekoğlu and Lajunen 2008); habits (Knapper *et al.* 1976, Chliaoutakis *et al.* 2000, Calisir and Lehto 2002); and lack of enforcement (Jonah *et al.* 1982, Farmer and Williams 2005). Each of these behaviors could be targeted by proper countermeasures through education and enforcement.

Sociodemographic of those who wear their seat belt less frequently is also helpful for identifying and reaching the groups with higher risk. Generally, males have lower seat belt use rates compared to females (Preusser *et al.* 1991, Reinfurt *et al.* 1997, Nelson *et al.* 1998, Calisir and Lehto 2002, Wells *et al.* 2002, Glassbrenner *et al.* 2004, Gkritza and Mannering 2008, Pickrell and Ye 2009). This is also true for younger drivers compared to the older adults (Reinfurt *et al.* 1997, Calisir and Lehto 2002, Glassbrenner *et al.* 2004). Individuals with higher education and/or income tend to have higher seat belt use rates (Preusser *et al.* 1991, Reinfurt *et al.* 1997, Wells *et al.* 2002, Houston and Richardson 2005). Studies in the United States have also shown that African-Americans are less likely to use a seat belt than Whites or Hispanics (Vivoda *et al.* 2004, Gkritza and Mannering 2008, Pickrell and Ye 2009). Several studies have reported that occupants of pickup trucks have the lowest seat belt use rate compared to occupants of other vehicle types (e.g., (Boyle and Vanderwolf 2004, Glassbrenner and Ye 2007, Gkritza and Mannering 2008)).

Nearly all of the studies that investigated seat belt use relied on the direct roadside observation or responses of self-reported surveys. Although these methods are easy to conduct and can provide information at a relatively low cost; they have their limitations that may negatively affect the outcomes of a study. For instance, one issue that could negatively affect results of self-reported questionnaires is social desirability bias (Lajunen and Summala 2003, Nordfjærn *et al.* 2015, Hezaveh *et al.* 2017, Hezaveh *et al.* 2018b). Social desirability bias refers to the tendency of respondents to provide socially desirable answers rather than choosing an answer that that reflects their state of mind (Grimm 2010). Social desirability may bias the respondents answer with regard to questions related to traffic violations (Lajunen *et al.* 1997).

Considering the roadside observations, the amount of data that researcher records are very limited; mainly due to the short amount of the time that the observers have to record the data and conspicuity challenges. In roadside observations, usually observed data elements are limited to the vehicle type, number of front row occupants, gender, age group, and roadside site characteristics (CTR 2018). Also, the number of observations sites are usually a small sample of the transportation network, and they usually take place within daylight or in the nighttime in the areas with sufficient lighting to observe inside of the vehicles.

Police crash reports are the main source for evaluating road safety especially for analyzing crash severity and frequency. However, using police crash reports for studying seat belt use has its own limitations. The main limitation is possible incorrect assignment of seat belt use or crash severity to individuals by a responding officer (Cherry *et al.* 2017). For example, some vehicle occupants who survived a crash may falsely claim that they used a seat belt at the time of the crash in order to avoid a traffic ticket (Cummings 2002). Nevertheless, several studies of police reports show that reported seat belt use is consistent with roadside observations (Li *et al.* 1999) and National Accident Sampling System Crashworthiness Data System (CDS) (Schiff and

Cummings 2004). Using police crash reports have several advantages compared to the roadside observations and self-reported studies. First, it provides a nearly comprehensive dataset of all serious crashes with an objective observer of many reported variables. Second, it covers a vast geographic area with hundreds of thousands of observations.

While the literature on road safety delivers seat belt use rate at coarse geographical level (e.g., county, state, country), it does not provide information about seat belt use rate at fine geographical level such as neighborhood level (e.g., census tract or traffic analysis zone). Knowing about the neighborhood seat belt use rate and seat belt non-use hotspots would benefit safety practitioners by focusing resources on areas where their residents have lower seat belt use rate. This is one of the main challenges in designing an effective and geographically targeted safety campaign. To date, most safety campaigns provide blanket coverage of regions with lower seat belt use rates, rather than precise and targeted messaging. Targeted education could be more cost effective at increasing overall seat belt use rates.

This study aims to propose a new method to measure seat belt use rate at the neighborhood level and evaluate the relationship between seat belt use rate and socio-demographic variables based on the home address of the individual (i.e., home-based approach) who were involved in traffic crashes at zonal level. Although some studies used police crash reports to evaluate seat belt effectiveness and seat belt use rate, to the best of our knowledge no studies used this dataset for investigating the relationship between sociodemographic data elements and seat belt use rate based on home-address of individuals involved in traffic crashes (i.e., drivers, passengers). Using the home-address of the individuals in a large database of the traffic crashes enables researchers to identify the geographic and surrounding socioeconomic factors that affect seat belt use and neighborhoods where their residents have lower seat belt use rate. Additionally, we will compare the seat belt use rate extracted from police crash reports with other sources of the seat belt use rate in Tennessee. Our findings are not only limited to the front row occupants but include all the vehicle occupants in different times of the day, context, weather, light conditions, and road types.

In the next section, we discuss the proposed database, the geocoding process, and the analytical methods. The rest of the paper presents the results and discusses the findings of this study.

Methodology

Database

The data in this study was provided by Tennessee Integrated Traffic Analysis Network (TITAN), a portal provided by Tennessee Highway Patrol (THP) as a repository for traffic crash and surveillance reports completed by Tennessee law enforcement agencies. The traffic crash records from January 1, 2016, through December 31, 2016, were retrieved from TITAN. Each crash record includes information about road user type (e.g., pedestrian), geographic coordinates of the crashes, addresses of the individuals who were involved in a traffic crash, and other variables related to the crash (MMUCC 2012). The Police crash reports database contained 246,777 crashes and information about 580,767 individuals who were involved in traffic crashes in 2016. Data included different road users' classifications; namely driver, motorcyclist, passenger, pedestrian, bicyclist, and other road user types. In order to analyze seat belt use rate, we only considered vehicle occupants. The database included information on 577,131 vehicle occupants (i.e., driver or passenger); 73% of the occupant of the vehicle were drivers, and the rest were passengers.

Geocoding process

Bing API was used in this study for geocoding the residential address of the individuals. Only those addresses with an accuracy level of premise (e.g., property name, building name), address level accuracy, or intersection level accuracy were used for analysis. A sample of addresses was verified by manual inspection. After geocoding the home-addresses, we were able to retrieve home-addresses' coordinates of 542,776 individuals (94% success rate), which met address quality filter criterion. Among geocoded addresses, 62,741 individuals lived out of state. After controlling for age, vehicles' occupants sixteen years old and older were selected for the analysis. Census data from US survey in 2010 was also used for obtaining sociodemographic data elements. Table 9 provides a summary of the sample characteristics of the variables considered as input for model estimation for Tennessee.

Tobit model

In order to model seat belt use rate, first, there is a need to select a suitable model specification. Since the value of seat belt use rate for each zonal level is limited between 0 and 1, it is appropriate to use a regression model with a censored dependent variable. Considering the nature of seat belt use rate at zonal level, we can conclude the dependent variable is left-censored at 0 and right censored at 1. To address censoring in the dependent variables, Tobin (1958) proposed the Tobit model or censored regression model. This model was used by several researchers to model crash rate in various types of road sections (e.g., Anastasopoulos *et al.* 2008, Anastasopoulos *et al.* 2012, Zeng *et al.* 2017).

Table 9 Sample statistic for the state of Tennessee at the census tract

Variable	Mean	Std. Deviation.	[95% Conf. Interval]	
Total Population	1,530.02	788.68	1,505.98	1,554.06
Age Cohort Proportion				
16 Years And Younger	0.23	0.08	0.22	0.23
16-42 Years Old	0.32	0.11	0.32	0.33
43-59 Years Old	0.25	0.08	0.24	0.25
60 Years Old And More	0.20	0.10	0.20	0.20
Age Median	38.96	8.63	38.75	39.27
Race Proportion				
Race White	0.77	0.30	0.76	0.78
Race Black	0.18	0.28	0.18	0.19
Race Indian	0.00	0.01	0.00	0.00
Race Asian	0.01	0.03	0.01	0.01
Race Hawaiian	0.00	0.01	0.00	0.00
Means Of Transportation To Work Proportion				
Personal Vehicle	0.92	0.11	0.92	0.93
Carpool	0.10	0.08	0.10	0.11
Bus	0.01	0.04	0.01	0.01
Motorcycle	0.00	0.01	0.00	0.00
Bicycle	0.00	0.01	0.00	0.00
Walk	0.02	0.05	0.01	0.02
Other Means	0.01	0.03	24.96	25.36
Children (%)	0.20	0.08	0.19	0.20
Household Size	2.72	5.30	2.57	2.89
Education Degree Proportion				
Number Of Educated Over 25 Years	1021.62	514.10	1,005.96	1,037.29
Education Degree Proportion				
High School And Lower	0.52	0.20	0.51	0.53
Some College Degree	0.20	0.08	0.20	0.21
Bachelors' Degree	0.20	0.12	0.19	0.20
Others' Degrees	0.08	0.08	0.07	0.08
Median Household Income ($1,000)	45.9	25.1	45.2	46.7
Occupied Household Proportion	0.87	0.13	0.87	0.88
Vacant Household Proportion	0.12	0.10	0.12	0.12
Household Vehicles' Ownership Proportion				
No-Vehicle	0.07	0.09	0.07	0.07
One Or Two Vehicles	0.70	0.13	0.33	0.33
Three Or More Vehicles	0.22	0.13	0.22	0.23

Data Source: US Census

In the Tobit model, the regression is obtained by making the mean in the preceding correspond to a classical regression model. The general form of the model is usually given in terms of index function as follows:

$$y_i^* = x_i'\beta + \varepsilon_{i,}$$

Where y_i^*defined as:

$$y_i^* = \begin{cases} y_i \; if \; a < y_i < b \\ a \; if \; y_i \leq a \\ b \; if \; y_i \geq b \end{cases}$$

ε_i assumes that the error term is normally distributed with mean 0 and variance equals to σ^2. In this study, the seat belt use rate is the dependent variables, and β is the coefficient corresponding to each independent variable presented in Table 8. The dependent variable is a proportion confined between 0 and 1. In addition to the estimated coefficients, we also measured the elasticities of each coefficient for measuring the sensitivity of the dependent variables to a change in the independent variable. For more information regarding elasticity estimation please see Stata (2015).

Variable selection

A combination of intuition and stepwise regression modeling was used to select the best subset of the predictors with an exclusion criterion of p-values greater than 0.20. Moreover, Variance Inflation Factors (VIF) was used to control for the multicollinearity in each step. Curious readers could refer to O'brien (2007) for more details about the VIF.

Model performance

Veall and Zimmermann (1996) concluded that Maddala pseudo-r-squared is a valid measurement for evaluating the goodness of fit of censored regression. The general form of Maddala pseudo-r-squared displayed below (Maddala 1986):

$$R^2 = 1 - [e^{LL_{Null} - LL_{Full}}]^{2/N}$$

where, LL_{Null} and LL_{Full} are log likelihoods of the null and full model respectively, and *N* is the number of observations. The likelihood function of the Tobit model is:

$$L = \prod_0 \left[1 - \Phi\left(\frac{\beta X}{\sigma}\right)\right] \prod_1 \sigma^{-1} \, \phi[(Y_i - \frac{\beta X}{\sigma})]$$

where, Φ is the standard normal distribution function, and ϕ is the standard normal density function (Anastasopoulos *et al.* 2008).

We also used the Akaike Information Criterion (AIC) as a measure of the relative goodness of fit for identification of the models with a better fit in the sample. AIC is a function of the number of parameters in the model (k) and log-likelihood of the model specification ($\ln(L)$); $AIC = 2k - 2\ln(L)$. As a rule of thumb, a three-point change in an AIC value indicates a significant improvement in the goodness of fit (Bozdogan 1987).

Results

Seat belt use rate

Table 10 presents the average age and gender distribution of the vehicle occupants considering their seat belt use based on the police crash database. The average age of the males' occupants (16 years old and older) was 39.4 (SD = 17.5) and for females was 39.2 (SD = 17.7). In addition, the average age of those who wore a seat belt properly (i.e., lap and shoulder) was 39.4 (SD = 17.6), and those who did not wear a seat belt was 38.8 (SD = 17.1). In general, the average age of those who wore a seat belt was higher than who did not (t = 8.278, P-value = 0.000). Moreover, females (89.1%) had a higher seat belt use rate in comparison to males (87.2%) (t = 23.889, P-value = 0.000). Table 11 also presents the seat belt distributions of the occupants over 16 years old. The highest seat belt use rate was for the front passenger (90.0%) followed by the driver (88.4%). The seat belt use rate dropped as passengers seating position row number increased (Table 11).

Table 12 shows the seat belt use rates under different circumstances. Considering the weather condition, occupants seat belt use rate was higher during the harsh weather, and at its lowest rate during clear weather. Regarding daylight, occupants wore seat belts at higher rates during daylight and less during the night. Seat belt use rates at night were lower when there was no lighting on the road. Regarding the road classification, Interstate and US highways had higher seat belt rate than other route types. In addition, the seat belt use rate was lowest on frontage and urban roads.

Table 10 Age and Gender distribution of the vehicles' occupants over 16 years old

	Female			Male			Total*		
Seat Belt Status	**Mean**	**SD**	**Obs**	**Mean**	**SD**	**Obs**	**Mean**	**SD**	**Obs**
No seat belt	38.70	17.22	25285	38.76	16.97	32178	38.76	17.09	57708
Wear Seat belt	39.24	17.74	205296	39.52	17.54	220700	39.39	17.64	425999
Total	39.18	17.69	230581	39.42	17.47	252878	39.31	17.58	483707

* Including the unknown observations;
Source: Authors' analysis from TITAN data

Table 11 Seat belt use rate (number of observations) among vehicles' occupants over 16 years old regarding seat position

Row	Left	Middle	Right	Other/Unknown
Front	0.88 (395641)	0.55 (912)	0.89 (66464)	0.2 (55)
Second	0.84 (6647)	0.65 (1101)	0.85 (8913)	0.38 (216)
Third	0.74 (424)	0.67 (143)	0.71 (438)	0.12 (54)
Fourth	0.45 (127)	0 (33)	0.50 (166)	0.04 (128)
Other Seats				0.40 (2203)

Source: Authors' analysis from TITAN data

Table 12 Seat belt use rate regarding weather, lighting, and route signage

Variable	Mean	SD	Number of observation
Weather			
Clear	0.868	0.338	395975
Cloudy	0.889	0.314	58743
Fog	0.868	0.339	1377
Smog/Smoke	0.934	0.249	196
Rain	0.884	0.321	54611
Sleet/Hail	0.895	0.307	1181
Snow	0.909	0.287	4749
Blowing Snow	0.912	0.284	272
Severe Cross-Winds	0.902	0.297	123
Blowing Sand/Soil/Dirt	0.922	0.269	51
Other	0.883	0.321	342
Unknown	0.025	0.157	24318
Lighting	Mean	SD	Count2
Daylight	0.879	0.326	389436
Dark-Not Lighted	0.843	0.364	39391
Dark-Lighted	0.860	0.347	69524
Dark-Unknown Lighting	0.787	0.409	1499
Dawn	0.875	0.330	6821
Dusk	0.864	0.343	10632
Other	0.865	0.342	429
Unknown	0.033	0.106	25044
Route Signage	Mean	SD	Count2
Interstate	0.885	0.319	45397
US Route	0.871	0.335	43581
State Route	0.868	0.338	68086
County Route	0.823	0.382	36707
Municipal Route	0.850	0.357	138721
Frontage Road	0.826	0.379	317
Other	0.789	0.408	14054
Unknown	0.796	0.402	195913

Source: Authors' analysis from TITAN data

Seat belt hot spot identification

After assigning respondents home-address to their corresponding census tract, we learned that Tennessee residents had a higher compliance rate (88.2%) than those with out-of-state addresses (86.9%) (t=8.615, P-value = 0.000). By using crash data, 480,035 (7.2% of state population) individual addresses in Tennessee were assigned to their corresponding census tracts. The average seat belt use rate at census tracts for the driver's seat was 0.88 (SD = 0.06) and for the passengers' seat was 0.86 (SD = 0.12); the correlation between driver's seat and passengers' seat was 0.32 (P value < 0.000), which indicated a weak positive linear relationship. Figure 13 and Figure 14 present number of observations and seat belt use rate for driver and passengers over 16 years old at the zonal level. Each point represents a seat belt use rate in a census tract (number of seat belted drivers or passengers divided by total number of participants in a crash), and the number of observations reflects the total number of observations in each census tract. In cases of driver crashes, the range of rates spans 60-100% for tracts with reasonably large crash counts. For passenger seat belt use rate, there are fewer observations in the dataset, and more observations below 60% seat belt use rates.

Figure 15 and Figure 16 present average seat belt use rate for the drivers and vehicle passengers at zonal levels in Tennessee; the red color in the figures indicates census tracts with low seat belt use rate and the green color indicates a high seat belt use rate. The white color also represents the state average. Visual Comparison of Figure 15 and Figure 16 indicate the passengers' seat belt use rate has more variation in Tennessee compared to driver seat belt use rate.

Table 13 presents the average seat belt use rate in metropolitan areas. The average seat belt use rate for vehicle occupants in the metropolitan area was 88% (SD = 0.06), which is slightly higher than the non-metropolitan area (87% SD = 0.06). Moreover, comparing the six metropolitan areas in Tennessee indicates that the driver's seat belt use rate is the highest among residents of Knoxville, followed by Jackson and Tri-cities. This trend also holds for the passenger's seat belt use rate. Chattanooga metropolitan area also has the lowest seat belt use rate for both passenger and driver seat. In addition, Chattanooga is the only metropolitan that passenger seat belt use rate is higher than the driver seat belt use rate.

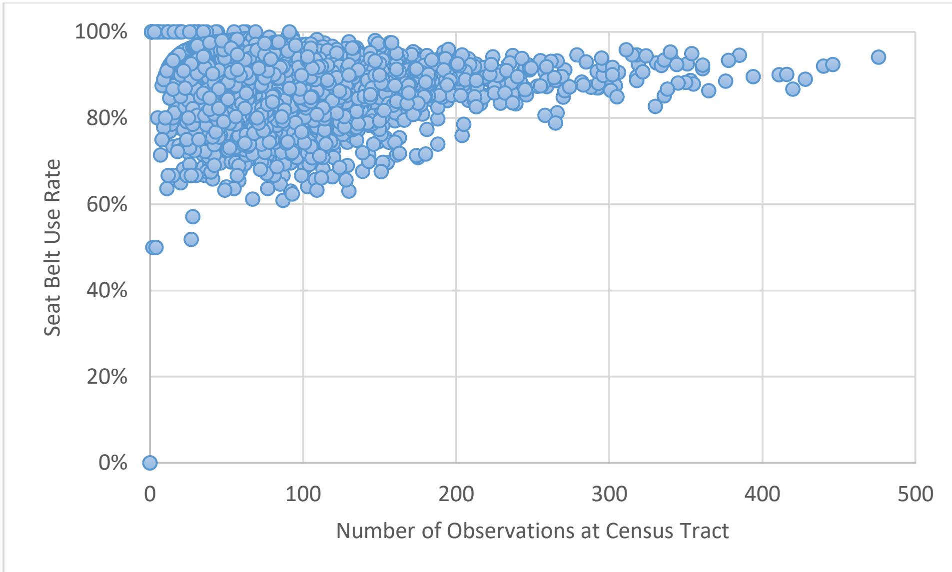

Figure 13 Number of observations and corresponding seat belt use rate at the census tract level for drivers

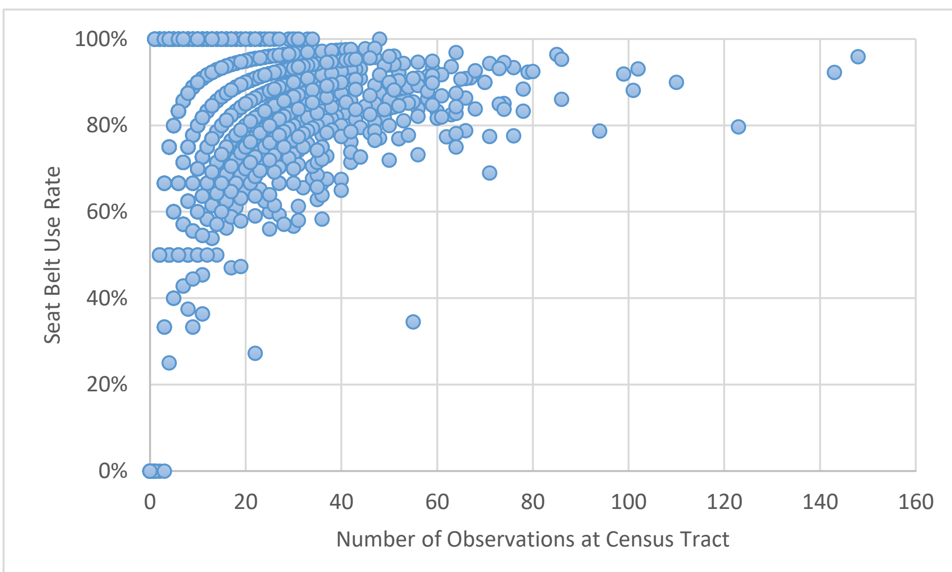

Figure 14 Number of observations and corresponding seat belt use rate at census tract level for passengers (over 16)

Table 13 Mean and Standard Deviation of The Seat Belt Use Rate in Metropolitan Areas

	Driver		**Passenger**		**Overall**	
Metropolitan Area	Mean	SD	Mean	SD	Mean	SD
Knoxville	0.92	0.04	0.90	0.10	0.91	0.04
Nashville	0.89	0.05	0.87	0.11	0.88	0.05
Jackson	0.90	0.04	0.87	0.11	0.90	0.04
Tri-cities	0.89	0.05	0.88	0.13	0.89	0.05
Chattanooga	0.77	0.07	0.81	0.14	0.77	0.06
Memphis	0.87	0.06	0.83	0.12	0.86	0.06
Non-metropolitan area	0.87	0.06	0.86	0.12	0.87	0.06
Grand Total	0.88	0.06	0.86	0.12	0.87	0.06

Source: Authors analysis of TITAN data

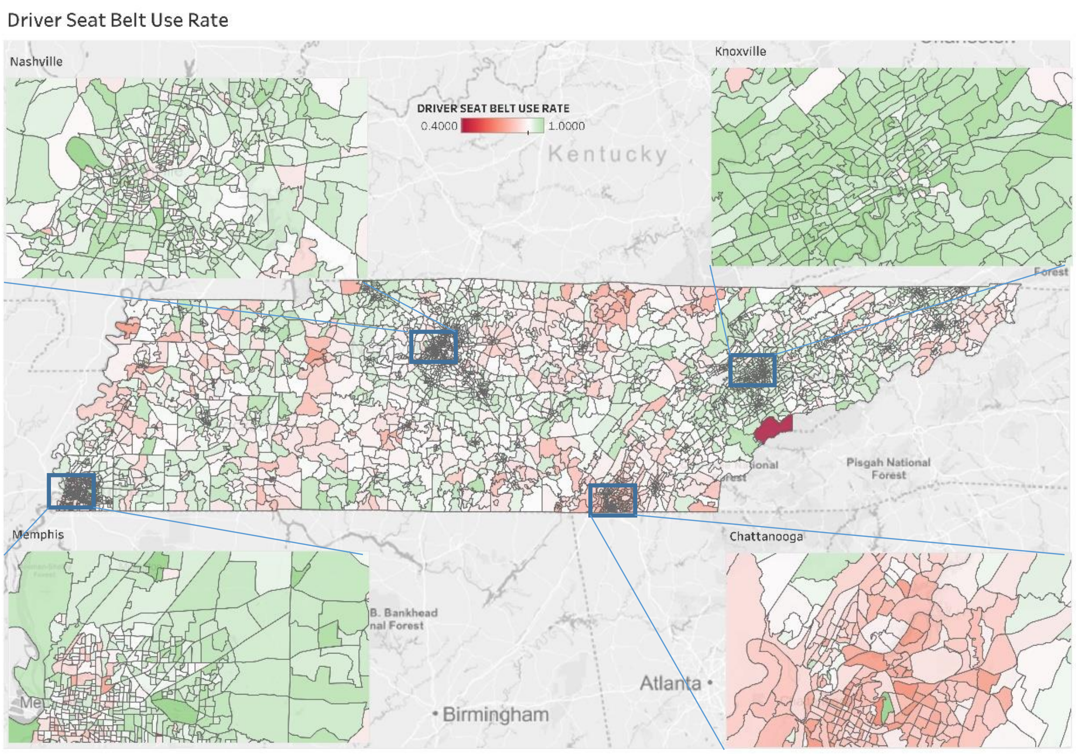

Figure 15 Driver seat belt use rate distribution in Tennessee

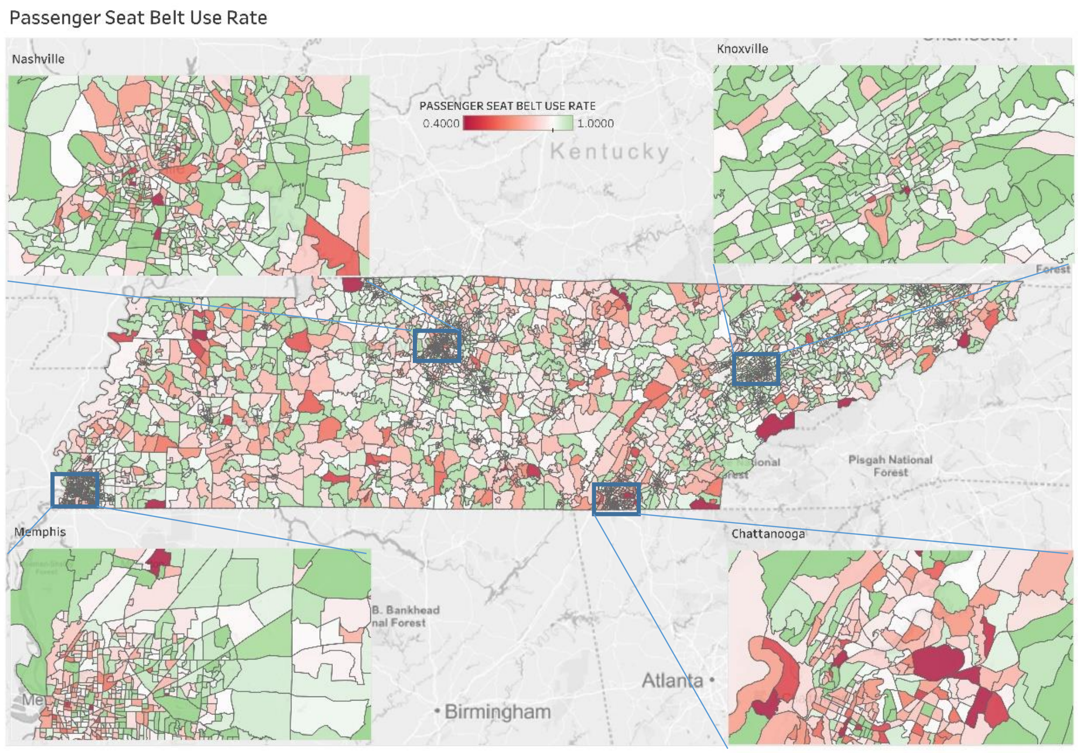

Figure 16 Passengers' seat belt use rate (over 16) distribution in Tennessee

Table 14 Estimated saturated Tobit model for prediction of the seat belt use rate for drivers

	DSBUR		**PSBUR**	
Variable	**Coef.**	**Standard Error**	**Coef.**	**Standard Error**
Population (1000)	0.006***	0.001	0.005*	0.003
Age Cohorts				
Population 16-42	0.001	0.015	-0.059*	0.030
Population 43-59	-0.019	0.027	-0.024	0.053
Population over 60	-0.037	0.029	0.021	0.056
Age Median	2.93E-04	3.41E-04	-8.42E-04	6.71E-04
Race (%)				
Race White	0.012	0.017	0.042	0.035
Race Black	-0.025	0.017	0.007	0.035
Race Indian	-0.031	0.093	0.115	0.182
Race Asian	-0.043	0.037	0.040	0.074
Race Hawaiian	0.004	0.148	0.181	0.291
Travel Mode to Work				
Morning Share Car	-0.029*	0.015	0.031	0.031
Morning Share Carpool	-0.010	0.012	-0.026	0.024
Morning Share Bus	-0.028	0.032	0.007	0.063
Morning Share Motor	-0.060	0.132	0.201	0.259
Morning Share Bicycle	-0.196*	0.103	0.056	0.202
Morning Share Walk	-0.101***	0.027	-0.023	0.057
Education Degree				
% High school degree	-0.057***	0.015	0.027	0.031
% College degree	-0.068***	0.018	0.039	0.036
% Bachelor Degree	-0.036***	0.020	0.105**	0.042
Median Household Income	-1.13E-07	6.39E-07	1.61E-07	1.29E-07
Vehicle Ownership (%)				
Household with no Vehicle	-0.051***	0.015	-0.017	0.029
Household with One or Two Vehicles	-0.014*	0.009	0.041**	0.018
Density (1000 population per square km)	-1.13E-07	1.18E-07	-1.28E-06***	2.34E-07
Constant	0.970***	0.018	0.776***	0.044
Scale Parameter	0.004***	0.000	0.014***	0.000
χ^2	373.78		217.54	
LL_0	5,563.87		2,841.95	
LL_M	5,750.76		2,950.72	
Maddala Pseudo-R^2	0.09		0.09	
N	4,114		4,103	
AIC	-11,453.53		-5,853.44	

* p<.10; ** p<.05; *** p<.01

Source: Authors' analysis of TITAN data and the US Census

Tobit model estimates

Table 14 presents the estimated saturated Tobit model based on the variables in Table 9. After performing stepwise regression and controlling for multicollinearity, insignificant variables were excluded from analysis. Table 15 presents the estimated coefficients for predicting driver seat belt use rate (DSBUR) and passengers seat belt use rate (PSBUR) at the census tract level and their corresponding elasticity values. The chi-square results for all models indicate that both models are significantly different from the null model (DSBUR: $\chi^2 = 328$; PSBUR: $\chi^2 = 233$). The variables that are presented in Table 15 have a significant correlation with both dependent variables. The mean VIF value for DSBUR model and PSBUR are respectively 1.31 (max = 1.59) and 1.34 (max = 1.71).

Findings of estimated models in Table 15 indicate that population size, the percentage of the white race and child percentage at zonal level have a positive association with seat belt use rate in both models. In the DSBUR model, elasticity values indicate that 1% increase in population, child percentage, and portion of white race increase average seat belt use rate by 1.0%, 0.5%, and 0.3% respectively; the corresponding elasticity values for the PSBUR model are higher, 0.8%, 3.7%, and 1.9%, respectively.

Vehicle ownership variables also have a significant association with seat belt use rate; however, the sign of the coefficients are dissimilar in both models. In the DSBUR model, the proportion of household with vehicle (i.e., 0, 1, 2) has a negative association with seat belt use rate. The elasticity values for the proportion of households with one or two vehicles (-2%) is greater than the proportion of households with no-vehicles (-0.6%). In the PSBUR model, the proportion of households with one or two vehicles has a positive correlation with passenger seat belt use, whereas the proportion of households with no-vehicles has a negative association with passenger seat belt use. The elasticity values for the proportion of families with one or two vehicles is 3%, and the corresponding value for households with no-vehicle is -0.3%.

Education-related variable signs are dissimilar in DSBUR model. Percentage of individuals with a college degree has a negative association with drivers' seat belt use rate. On the other hand, the percentage of bachelor degree has a positive association with seat belt use rate in both models. Elasticity values indicate one percent increase in the proportion of the population with bachelor degree increases seat belt use rate by 0.4% and 1.3%, respectively for DSBUR and PSBUR model.

The metropolitan indicator variable in both models has a positive association with seat belt use rate, which indicates that seat belt use in the metropolitan area is higher than non-metropolitan area. The magnitude and elasticity value of the metropolitan coefficient in the DSBUR model is greater than PSBUR model. Alternatively, the population density variable has a negative correlation with PSBUR variable; the elasticity values indicate that one percent change in population density results in 1.1% reduction in seat belt use rate in the PSBUR model. Household size also has a significant negative association with passenger seat belt use; the elasticity value for this variable is -0.4%.

Discussion

Analysis of seat belt use rate for vehicles' occupants over 16 years old indicates that seat belt use rate for drivers and the front passenger in 2016 was approximately 88.2%, which is close to 88.9% observed in roadside observations in Tennessee (THSO 2016, CTR 2018). Comparison of the driver and front row passenger seat belt use rate indicates that front row passenger had higher compliance rate, which is also in line with the roadside observation in Tennessee (THSO 2016, CTR 2018). Generally, the seat belt use rate of passengers (including back row) was lower than the driver, which is influenced by substantially lower seat belt use rate of the passengers in back rows. This lower seat belt use rate for passengers in back rows could be attributed to the current seat belt law in Tennessee, which only covers front row passengers (IIHS 2018).

Table 15 Estimated Tobit model for prediction of the seat belt use rate for drivers

Variable	DSBUR			PSBUR		
	Coef.	**Std. Err**	**Elasticity**	**Coef.**	**Std. Err**	**Elasticity**
Population (1,000)	0.006***	0.001	0.010	0.005*	0.003	0.008
% Children	0.023*	0.012	0.005	0.085***	0.024	0.037
% Race White	0.036***	0.004	0.031	0.042***	0.008	0.019
Vehicle Ownership						
% Household with no Vehicle	-0.078***	0.013	-0.006	-0.041*	0.025	-0.003
% Household with One or Two Vehicles	-0.025***	0.008	-0.020	0.036**	0.016	0.029
Education						
% College degree	-0.032**	0.013	-0.007			
% Bachelor Degree	0.016*	0.009	0.004	0.058***	0.018	0.013
Metropolitan Indicator	0.007***	0.002	0.005	0.015***	0.005	0.012
Household Size				-0.001***	0.000	-0.004
Density (1,000 population per square km)				-1.46E-06***	2.24E-07	-0.011
Constant	0.863***	0.008		0.773***	0.016	
Scale parameter	0.004***	7.97E-05		0.014***	3.03E-04	
χ^2	328.37			233.50		
LL_0	5,563.87			2,841.95		
LL_M	5,728.06			2,958.70		
Maddala Pseudo-R^2	0.077			0.056		
N	4,114			4,103		
AIC	-11,436.1			-5,897.41		

* p<.10; ** p<.05; *** p<.01

Source: Authors' analysis of TITAN data and the US Census

In line with previous studies (Wells *et al.* 2002, Gkritza and Mannering 2008, Ojo 2018), police crash reports analysis indicates that males and younger individuals are more prone to seat belt non-use. Findings also indicate seat belt use rates are higher in daylight and harsh weather, which is consistent with roadside observations (NHTSA 2016). Additionally, seat belt use on interstates is higher than other classes of roads. Variation in seat belt use rate in different circumstances could be attributed to the perception of safety. For instance, those who drive in harsh weather (e.g., rainy or high-speed routes like interstates) may perceive more hazard, and as a result, have a higher seat belt use rate. These abovementioned findings yield the conclusion that police crash reports of seat belt use are broadly in agreement with roadside observations in Tennessee (CTR 2018) and road safety literature (Gkritza and Mannering 2008, Pickrell and Ye 2009, Ogunleye-Adetona *et al.* 2018, Ojo 2018).

Comparison of the driver and passenger seat belt use rate in different metropolitan areas indicates that driver's seat belt use was higher than other passengers, except for Chattanooga metropolitan area. Overall, the Chattanooga region has the lowest seat belt use rate among both metropolitan and non-metropolitan areas. The spatial variation in seat belt use in metropolitan areas reflects different traffic cultures and social and psychological factors within Tennessee. Identifying social and psychological factors (e.g., attitudes, beliefs, and intentions) that affect seat belt use and using them for educational purposes in safety campaigns could increase seat belt use. Moreover, seat belt use rate distribution maps in Tennessee indicate that passenger seat belt use rate has more spatial variation than driver seat belt use. Spatial variation in seat belt use could be attributed to both traffic laws in Tennessee and cultural differences, which should be investigated in the future studies.

A Tobit model was used to investigate the association between seat belt use for the drivers and passengers 16 years or older who were involved in traffic crashes and sociodemographic variables of the occupant's home location, at the aggregate level. This is the first time, to the authors' knowledge, that this type of analysis has been conducted. Results indicate that the percentage of the white race in the neighborhood had a positive impact on seat belt use rate for both models; this finding parallels previous research (e.g., Gkritza and Mannering 2008, Pickrell and Ye 2009, Bhat *et al.* 2015). Using a safety campaign in neighborhoods with a high percentage of non-white populations could be used as an effective method for improving seat belt compliance rates. Consistent with other road safety literature, sociodemographic variables have a significant impact on seat belt use rate (Preusser *et al.* 1991, Reinfurt *et al.* 1997, Wells *et al.* 2002, Houston and Richardson 2005). Different neighborhood education levels also have a different effect on seat belt use rate for both models. Percentage of bachelor's degree have a positive impact on the seat belt use rate for both DSBUR and PSBUR models. On the other hand, in the drivers' model, the percentage of a college degree has a negative association with the driver's seat belt use rate. The motive for not wearing a seat belt for lower education driver could be different; perhaps lower seat belt use of drivers with lower-education could be attributed to their subjective norm and attitude toward wearing a seat belt. On the other hand, for higher education and higher income portion of society lower seat belt use rate could be attributed to perceived behavioral control or over-confidence. This may also explain the negative sign of average neighborhood vehicle ownership for the driver seat belt use. Using social psychological tools to investigate how attitudes, beliefs, and values influence seat belt use for different road users would be beneficial for designing a better safety campaign and targeting human factors that predict seat belt use.

Quite the opposite, passengers' with higher education and higher vehicle ownership have higher seat belt use rate. This behavior may be attributed to the fact that passengers (particularly front row passengers) have little or no control over the driver's behavior (i.e., perceived behavioral control), and as a result, passengers tend to wear their seat belt more frequently when they are not in the driving position. Psychological factor that affects lower seat belt use rates of the back rows passengers needs to be investigated in the future studies. The home-address environment also has a significant effect on seat belt use rate in both models. Results indicate metropolitan indicator has a positive impact on seat belt use rate for both models. Metropolitan indicator could be used as a surrogate for urban areas, which traditionally have a higher seat belt use rate (NHTSA 2017a).

Conclusion and future implications

In sum, results of analysis point out police crash reports have the potential to be used as a source to examine seat belt use at the neighborhood level. Using the home-address of the individuals extracted from police crash report could be used to identify areas with lower seat belt use rate, which could be useful in the design of safety campaigns in programs such as "Click-It or Ticket" to efficiently reach individuals that are more prone to lower seat belt use. This method could be more effective than blanket campaigns that tend to show small population-level effects. There is also a need for developing a methodology that enables researchers and safety practitioners to identify seat belt non-use hot spots. Increase in the enforcement mainly by covering passengers in back rows under the primary seat belt use law in Tennessee could be a practical solution for increasing seat belt use rate of passengers. Besides, findings indicate that there are differences between drivers and passengers in terms of factors correlating with their seat belt use at the zonal level. As a result, the seating position needs to be considered in the design of a road safety campaign.

It is also worth mentioning that there are difficulties in accessing crash data with identifiers and it is not possible to obtain this data in some cases. One possible direction for future researchers could be to develop a methodology to identify seat belt hotspots based on the conventional sources of data (i.e., temporal and spatial transformation of the models). The sample in this study represents an individual who had reported a traffic crash in Tennessee in 2016 and careful consideration needed in order to apply the findings to all residents of the state.

Acknowledgments

The authors would like to express gratitude to the Tennessee Department of Safety & Homeland Security. This project was supported by the Collaborative Sciences Center for Road Safety, www.roadsafety.unc.edu, a U.S. Department of Transportation National University Transportation Center promoting safety. The study design was reviewed and approved by the University of Tennessee Institutional Review Board.

References

Anastasopoulos, P.C., Shankar, V.N., Haddock, J.E., Mannering, F.L., 2012. A multivariate tobit analysis of highway accident-injury-severity rates. Accident Analysis & Prevention 45, 110-119.

Anastasopoulos, P.C., Tarko, A.P., Mannering, F.L., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis & Prevention 40 (2), 768-775.

Begg, D.J., Langley, J.D., 2001. Seat-belt use and related behaviors among young adults. Journal of Safety Research 31 (4), 211-220.

Bhat, G., Beck, L., Bergen, G., Kresnow, M.-j., 2015. Predictors of rear seat belt use among us adults, 2012. Journal of safety research 53, 103-106.

Blincoe, L., Miller, T.R., Zaloshnja, E., Lawrence, B.A., 2015. The economic and societal impact of motor vehicle crashes, 2010. (revised) (report no. Dot hs 812 013). Washington, dc: National highway traffic safety administration.

Boyle, J., Vanderwolf, P., 2004. 2003 motor vehicle occupant safety survey, volume 2: Safety belt report.

Bozdogan, H., 1987. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. Psychometrika 52 (3), 345-370.

Calisir, F., Lehto, M.R., 2002. Young drivers' decision making and safety belt use. Accident Analysis & Prevention 34 (6), 793-805.

Cherry, C.R., Mohamadi Hezaveh, A., Grembek, O., Sandt, L., Merlin, L., Year. Better data integration to create a complete picture for cycling safety. In: Proceedings of the 2017 International Cycling Safety Conference held in Davis, California, USA.

Chliaoutakis, J.E., Gnardellis, C., Drakou, I., Darviri, C., Sboukis, V., 2000. Modelling the factors related to the seatbelt use by the young drivers of athens. Accident Analysis & Prevention 32 (6), 815-825.

CTR, 2018. 2017 survey of safety belt usage in tennessee final report. In: Research, T.U.O.T.C.F.T. ed.

Cummings, P., 2002. Association of seat belt use with death: A comparison of estimates based on data from police and estimates based on data from trained crash investigators. Injury Prevention 8 (4), 338-341.

Farmer, C.M., Williams, A.F., 2005. Effect on fatality risk of changing from secondary to primary seat belt enforcement. Journal of Safety Research 36 (2), 189-194.

Fhaner, G., Hane, M., 1975. Seat belts: Changing usage by changing beliefs. Journal of Applied Psychology 60 (5), 589.

Gkritza, K., Mannering, F.L., 2008. Mixed logit analysis of safety-belt use in single-and multi-occupant vehicles. Accident Analysis & Prevention 40 (2), 443-451.

Glassbrenner, D., Carra, J.S., Nichols, J., 2004. Recent estimates of safety belt use. Journal of safety research 35 (2), 237-244.

Glassbrenner, D., Ye, J., 2007. Seat belt use in 2006—overall results. In: National Highway Traffic Safety Administration, U.S.D.O.T. ed., Washington, DC.

Grimm, P., 2010. Social desirability bias. Wiley international encyclopedia of marketing.

Hezaveh, A.M., Boakye, K., Cherry, C.R., Nambisan, S., 2018a. Factor influencing seat belt use in tennessee: A self-report study. University of Tennessee.

Hezaveh, A.M., Nordfjærn, T., Mamdoohi, A.R., Şimşekoğlu, Ö., 2018b. Predictors of self-reported crashes among iranian drivers: Exploratory analysis of an extended driver behavior questionnaire. PROMET-Traffic&Transportation 30 (1), 35-43.

Hezaveh, A.M., Zavareh, M.F., Cherry, C.R., Nordfjærn, T., 2017. Errors and violations in relation to bicyclists' crash risks: Development of the bicycle rider behavior questionnaire (brbq). Journal of Transport & Health.

Houston, D.J., Richardson, L.E., 2005. Getting americans to buckle up: The efficacy of state seat belt laws. Accident Analysis & Prevention 37 (6), 1114-1120.

IIHS, 2018. Safety belts. Insurance Institute for Highway Safety, Insurance Institute for Highway Safety.

Jonah, B.A., Dawson, N.E., 1982. Predicting reported seat belt use from attitudinal and normative factors. Accident Analysis & Prevention 14 (4), 305-310.

Jonah, B.A., Dawson, N.E., Smith, G.A., 1982. Effects of a selective traffic enforcement program on seat belt usage. Journal of Applied Psychology 67 (1), 89.

Knapper, C., Cropley, A., Moore, R., 1976. Attitudinal factors in the non-use of seat belts. Accident Analysis & Prevention 8 (4), 241-246.

Lajunen, T., Corry, A., Summala, H., Hartley, L., 1997. Impression management and self-deception in traffic behaviour inventories. Personality and individual differences 22 (3), 341-353.

Lajunen, T., Summala, H., 2003. Can we trust self-reports of driving? Effects of impression management on driver behaviour questionnaire responses. Transportation Research Part F: Traffic Psychology and Behaviour 6 (2), 97-107.

Li, L., Kim, K., Nitz, L., 1999. Predictors of safety belt use among crash-involved drivers and front seat passengers: Adjusting for over-reporting. Accident Analysis & Prevention 31 (6), 631-638.

Maddala, G.S., 1986. Limited-dependent and qualitative variables in econometrics Cambridge university press.

MMUCC, 2012. Model minimum uniform crash criteria. DOT HS 811, 631.

Nelson, D.E., Bolen, J., Kresnow, M.-j., 1998. Trends in safety belt use by demographics and by type of state safety belt law, 1987 through 1993. American Journal of Public Health 88 (2), 245-249.

NHTSA, 2016. Traffic safety facts research noteseat belt use in 2015—overall results. In: Analysis, N.S.N.C.F.S.A. ed. NHTSA's National Center for Statistics and Analysis, 1200 New Jersey Avenue SE., Washington, DC 20590.

NHTSA, 2017a. Seat belt use in 2017—use rates in the states and territories. In: Analysis, N.S.N.C.F.S.A. ed.

NHTSA, 2017b. Traffic safety facts 2016 state traffic data. Department of Transportation, National Highway Traffic Safety Administration. DOT-HS-812-581, Washington, DC: US

Nordfjærn, T., Hezaveh, A.M., Mamdoohi, A.R., 2015. An analysis of reported driver behaviour in samples of domestic and expatriate iranians. Journal of Risk Research 18 (5), 566-580.

O'brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Quality & quantity 41 (5), 673-690.

Ogunleye-Adetona, C., Ojo, T., Afukaar, F., 2018. Assessment of seat belt use in university of cape coast campus, ghana. Urban, Planning and Transport Research 6 (1), 22-34.

Ojo, T.K., 2018. Seat belt and child restraint use in a developing country metropolitan city. Accident Analysis & Prevention 113, 325-329.

Pickrell, T., Ye, T., 2009. Traffic safety facts (research note): Seat belt use in 2008–demographic results. National Highway Traffic Safety Administration/Department of Transportation, Washington DC.

Preusser, D.F., Williams, A.F., Lund, A.K., 1991. Characteristics of belted and unbelted drivers. Accident Analysis & Prevention 23 (6), 475-482.

Reinfurt, D., Williams, A., Wells, J., Rodgman, E., 1997. Characteristics of drivers not using seat belts in a high belt use state. Journal of Safety Research 27 (4), 209-215.

Schiff, M.A., Cummings, P., 2004. Comparison of reporting of seat belt use by police and crash investigators: Variation in agreement by injury severity. Accident Analysis & Prevention 36 (6), 961-965.

Şimşekoğlu, Ö., Lajunen, T., 2008. Social psychology of seat belt use: A comparison of theory of planned behavior and health belief model. Transportation Research Part F: Traffic Psychology and Behaviour 11 (3), 181-191.

Stata, A., 2015. Stata base reference manual release 14.

THSO, 2016. Tennessee's average seat belt usage rate soars to 88.95% in 2016. In: Office, T.H.S. ed. Tennessee Highway Safety Office.

TITAN, 2017. Tennessee integrated traffic analysis network. Tennessee Highway Patrol

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica: journal of the Econometric Society, 24-36.

Veall, M.R., Zimmermann, K.F., 1996. Pseudo-r2 measures for some common limited dependent variable models. Journal of Economic surveys 10 (3), 241-259.

Vivoda, J.M., Eby, D.W., Kostyniuk, L.P., 2004. Differences in safety belt use by race. Accident Analysis & Prevention 36 (6), 1105-1109.

Wells, J.K., Williams, A.F., Farmer, C.M., 2002. Seat belt use among african americans, hispanics, and whites. Accident Analysis & Prevention 34 (4), 523-529.

Zeng, Q., Wen, H., Huang, H., Pei, X., Wong, S., 2017. A multivariate random-parameters tobit model for analyzing highway crash rates by injury severity. Accident Analysis & Prevention 99, 184-191.

CHAPTER IV: FACTORS INFLUENCING SEAT BELT NON-USE: INCORPORATING SPATIAL EFFECTS

The authors confirm contribution to the paper as follows: study conception and design: Amin Mohamadi Hezaveh, Trond Nordfjaern, Christopher Cherry; data collection: Not Applicable; analysis and interpretation of results: Amin Mohamadi Hezaveh, Trond Nordfjaern, Christopher Cherry; draft manuscript preparation: Amin Mohamadi Hezaveh, Trond Nordfjaern, Christopher Cherry. All authors reviewed the results and approved the final version of the manuscript.

Abstract

It is well-established that seat belt use in many cases could prevent serious injuries and death of vehicle occupants. However, many car occupants still do not to wear a seat belt. Seatbelt use rates are often spatially correlated with nearby areas. However, very few studies have examined this spatial autocorrelation. In this study, we used exploratory spatial data analysis (ESDA) and spatial regression analysis to model autocorrelation in Tennessee, which has a lower seat belt use than the United States national average. We geocoded home-addresses of vehicle occupants involved in traffic crashes between 2014-16 (n = 1,251,901) and projected them to the census tract corresponding to their home address revealing information about the spatial distribution of seat belt use and socioeconomics of the areas surrounding the crash victim's home. Average seat belt use rate for the three-year period was 89.9%. ESDA analysis reveals a distinctive regional imprint for spatial autocorrelation, in which Southern-metropolitan areas' (Southern-MPOs) census tracts have higher than average seat belt non-use compared to non- Southern-MPOs that form statistically significant clusters. Presence of highly spatially correlated observations suggests that seat belt non-use is not produced solely by the internal structural factors represented in the non-spatial models. Spatial error model and the spatial lag model were suitable for non-Southern-MPOs and Southern-MPOs, respectively. Spatial lag model in the Southern-MPOs is also consistent with an influence process– e.g., modification of one person's responses by the actions of another. The observation of social influence indicates that further inquiry is needed to learn about the underlying mechanism of social influence in future studies. Identifying the underlying mechanism of social influence would be helpful in the design of an effective seat belt campaign, such as communication methods with recipients of the campaign.

Keywords: Seat belt non-use; network influence; diffusion process; ESDA; Spatial Lag Model; Zonal Model

Introduction

It is well-established that seat belt use could reduce serious injuries and fatalities from traffic crashes among car occupants (Blincoe *et al.* 2015). There are mandatory seat belt laws in the United States and not wearing a seat belt violates the state law which could lead to a fine in many jurisdictions. In 34 States, the District of Columbia, and Puerto Rico seat belt laws are primary, which enable law enforcement officers to stop vehicles and write citations when they observe seat belt non-use (IIHS 2018). In 15 States, the laws have specified secondary enforcement, meaning that law enforcement officers are permitted to issue a seat belt citation only after they stop a vehicle for another primary violation. Nevertheless, some occupants do not use their seat belt. In Tennessee, seat belt use is also compulsory and is a primary law (i.e., secured shoulder and lap belts) when riding in the front seat of a vehicle (IIHS 2018). Meanwhile, roadside observations of 190 sites in 2017 revealed that, on average, 88.5% of the vehicle occupants in Tennessee used their seat belt (CTR 2018), which is 1.2% lower than the national average in the United States (US) (NHTSA 2017).

There is compelling evidence suggesting spatial dependency of seat belt use. Spatial dependency may reflect variations in a wide range of factors, including demographic, economic, historical and geographical background, enforcement level, or traffic culture. Roadside observations imply the presence of spatial dependency on seat belt non-use. Figure 17 exhibits the seat belt use rate at the state level based in NHTSA (2017) roadside observations in the United States. Visual screening of this map indicates the presence of spatial clusters of seat belt use (e.g., a state with high seat belt use shares borders with other states with high seat belt use and vice versa). This spatial variation in one observation could be an indicator of the presence of spatial autocorrelation. Spatial autocorrelation exists when a variable displays interdependence over space (Saha *et al.* 2018). Presence of spatial autocorrelation in seat belt use was also reported by Majumdar *et al.* (2004) at the state level.

Wearing a seat belt also is a decision-making problem and as it is expected vehicle occupants' social psychological factors affect seat belt use (Calisir and Lehto 2002, Şimşekoğlu and Lajunen 2008). Several studies used self-reported questionnaires based on the theory of health belief model or the theory of planned behavior to explore factors influencing seat belt use. These studies highlight the role of local effects, such as regulation enforcement, on seat belt use. Subjective norms (i.e., perceived social pressure to perform or not to perform the behavior (Ajzen 1991)), and normative belief (i.e., an individual's perception of social normative pressures, or relevant others' beliefs that he or she should or should not perform such behavior) may represent the effect of social interaction and pressure for seat belt use.

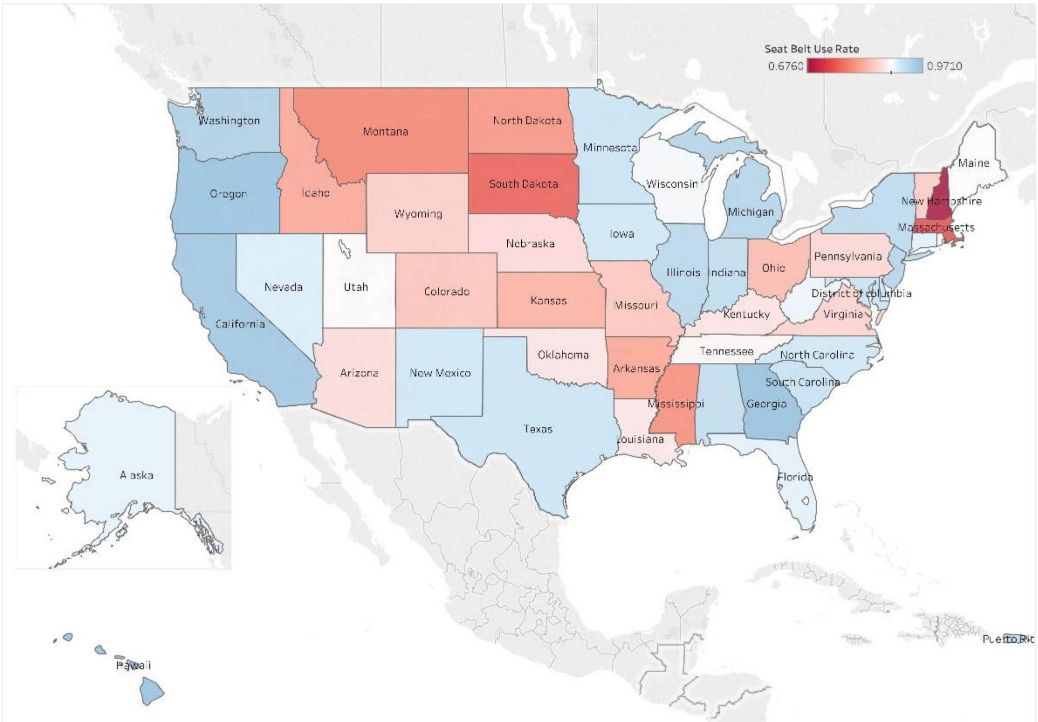

Figure 17 Seat belt use distribution at the state level –2017; adopted from NHTSA (2017)

Furthermore, subjective norms (Şimşekoğlu and Lajunen 2008, Ali *et al.* 2011, Torquato *et al.* 2012), attitude (positive or negative evaluations of seat belt use) (Şimşekoğlu and Lajunen 2008, Ali *et al.* 2011, Torquato *et al.* 2012), or cues to action (e.g., seeing other drivers wearing seat belt) (Şimşekoğlu and Lajunen 2008, Ali *et al.* 2011) also have significant impact on seat belt use. On the other hand, negative attitudes and beliefs about the effectiveness of seat belt use may adversely affect the seat belt use (Fockler and Cooper 1990, Begg and Langley 2000). Many of these psychological factors may reflect the safety culture (Şimşekoğlu *et al.* 2013, Nordfjærn *et al.* 2014) and may drive spatial dependency in seat belt use.

In the social science, the presence of the spatial autocorrelation may be attributed to the social influence phenomenon (Marsden and Friedkin 1993, Leenders 2002, Fujimoto *et al.* 2011, Wang *et al.* 2014, Tranmer *et al.* 2016, Vitale *et al.* 2016, Dittrich *et al.* 2017). This issue could be explored in a social network context by entailing a structural conceptualization of social proximity. A social network displays the relationship between the attitudes and behaviors of the actors who compose a network (Marsden and Friedkin 1993). The general hypothesis is that the proximity (e.g., distance) of two actors (e.g., residences of a geographic unit) in a social network is associated with the occurrence of interpersonal influence between the actors (Marsden and Friedkin 1993). In this context, social influence is the modification of one person's responses by the actions of another (Cartwright 1965, Marsden and Friedkin 1993). Comparison of the attitudes and behavior of the actors happen through processing information about the attitudes and behaviors of other actors and does not require face-to-face interaction between actors. There are several methods for modeling the social influence in the cross-sectional studies. Spatial lag model and spatial error model are two common techniques (Marsden and Friedkin 1993, Baller *et al.* 2001, Leenders 2002) that incorporate the effect of proximity and interaction between actors in the modeling process. This model are discussed in more details in the methodology section. To the best of our knowledge, no study has yet explored the presence of social influence and spatial autocorrelation on seat belt non-use at a fine geographic level within in a city or State.

There are several factors that affect seat belt use. In a recent study, Hezaveh and Cherry (2019) used police crash reports and census tract data and showed that seat belt use varied at a fine geographic level (i.e., census tract) within a state. Moreover, the authors showed that there are several demographic factors besides ethnicity, gender, and age cohorts that influence seat belt use rates at the zonal level, for instance, population density, age, household vehicles' ownership, and household size (Hezaveh and Cherry 2019). Nonetheless, Hezaveh and Cherry (2019) did not consider the effect of spatial structure in their analysis. Considering the demographics of vehicle occupants, males have lower seat belt use rates compared to females (Preusser *et al.* 1991, Reinfurt *et al.* 1997, Nelson *et al.* 1998, Calisir and Lehto 2002, Wells *et al.* 2002, Glassbrenner *et al.* 2004, Gkritza and Mannering 2008, Pickrell and Ye 2009). This also holds for younger drivers compared to older adults (Reinfurt *et al.* 1997, Calisir and Lehto 2002, Glassbrenner *et al.* 2004). Individuals with higher education and/or income tend to have higher seat belt compliance (Preusser *et al.* 1991, Reinfurt *et al.* 1997, Wells *et al.* 2002, Houston and Richardson 2005). Studies in the United States have also shown that African-Americans are less likely to use a seat belt than Whites or Hispanics (Vivoda *et al.* 2004, Gkritza and Mannering 2008, Pickrell and Ye 2009).

In this study, we are exploring the presence of the influence process at fine geographic level by incorporating spatial autocorrelation in the domain of seat belt use in a social network context. This study aims to, first, measure the seat belt non-use rate at the zonal level and evaluate the relationship between seat belt use rate and socio-demographic variables based on the home addresses of the individuals who were involved in traffic crashes at the zonal level. Second, we explore the presence of an influence process regarding seat belt non-use in Tennessee by incorporating the spatial autocorrelation. Evidence of the presence of social influence may provide promising opportunities for understanding the local factors contributing to this phenomenon. Furthermore, it provides insights into designing countermeasures and interventions that increase seat belt use rates in the areas with lower seat belt use rate.

The next section discusses the methods used in this study. In the methodology section, we discuss the geocoding process and measuring seat belt non-use at zonal level by incorporating spatial effects. In the last section, we present and discuss the findings of the analysis.

Methodology

Database and geocoding process

The data in this study were provided by Tennessee Integrated Traffic Analysis Network (TITAN), which is a statewide repository for traffic crashes and surveillance reports completed by Tennessee law enforcement agencies. For the years 2014, 2015, and 2016, the TITAN records include 694,276 crashes and information about 1,607,995 vehicle occupants who were involved in traffic crashes. The Bing API was used in this study for geocoding the residential address of the individuals. Only those addresses with an accuracy level of the premise (e.g., property name, building name), address level accuracy, or intersection level accuracy were used in the analysis. A sample of addresses was verified by manual inspection. After geocoding the home-addresses, we were able to retrieve home-addresses' coordinates of 1,510,506 individuals (94% success rate), which met address quality filter criterion. Among geocoded addresses, 162,447 individuals lived out of Tennessee. After controlling for seat belt use type (i.e., excluding child seat boosters), 1,252,139 observations with a Tennessee address were selected for assignment to the census tract data.

Following the MMUCC (2012), TITAN provides information regarding restraint use by occupants at the time of the crash. For this study, we defined seat belt non-use as vehicle occupants who did not restrain both lap and shoulder seat belt at the time of a traffic crash. Accordingly, we estimated seat belt non-use rates at the zonal level as the percentage of seat belt non-use cases over a total number of observations at a specific geographic area. Census data from the US survey in 2010 were also used for obtaining sociodemographic data elements. Table 16 summarizes the sample characteristics of the variables considered as input for model estimation for Tennessee. To prevent outliers, we only considered the census tracts that had more than 20 observations.

Figure 18 further presents the geographical distribution of seat belt non-use at the census tract level. Red colors indicate a higher level of seat belt non-use, while blue colors show a higher level of compliance. Visual inspection indicates that census tracts are clustered together, meaning that blue colors are usually surrounded by blue neighbors and vice versa. Moreover, seat belt non-use indicates that Chattanooga and Memphis metropolitan areas (here defined as Southern-MPOs) have higher seat belt non-use rate compared to the Knoxville and Nashville metropolitan areas (here defined as non-Southern-MPOs). Seat belt non-use in the rural also follows a similar trend.

Table 16 Sample statistic for the state of Tennessee at the census tract

Variable	Mean	Std. Deviation.	[95% Conf. Interval]	
Total Population	1,530	789	1,506	1,554
Age Cohort %				
<16 Years	0.23	0.08	0.22	0.23
16-42 Years	0.32	0.11	0.32	0.33
43-59 Years	0.25	0.08	0.24	0.25
> 59 Years	0.20	0.10	0.20	0.20
Age Median	38.96	8.63	38.75	39.27
Race %				
White	0.77	0.30	0.76	0.78
Non-White	0.23	0.23	0.22	0.24
Means Of Transportation To Work Proportion				
Personal Vehicle	0.92	0.11	0.92	0.93
Carpool	0.10	0.08	0.10	0.11
Bus	0.01	0.04	0.01	0.01
Motorcycle	0.00	0.01	0.00	0.00
Bicycle	0.00	0.01	0.00	0.00
Walking	0.02	0.05	0.01	0.02
Other Means	0.01	0.03	24.96	25.36
Education Degree %				
High School And Lower	0.52	0.20	0.51	0.53
Some College Degree	0.20	0.08	0.20	0.21
Bachelors' Degree	0.20	0.12	0.19	0.20
Other Degrees	0.08	0.08	0.07	0.08
Median Household Income ($10,000)	45.9	25.1	45.2	46.7
Household Vehicles' Ownership %	0.93	0.01	0.92	0.93
Data Source: US Census				

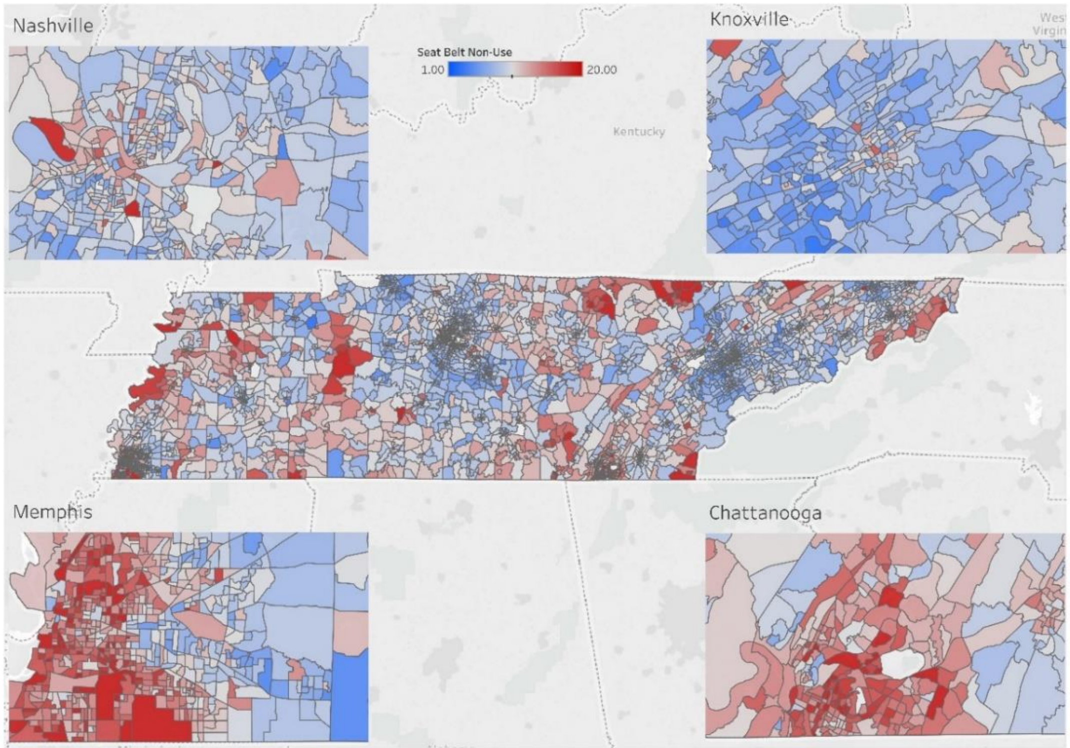

Figure 18 Seat belt non-use map

Spatial analysis

In this study, we are taking an explicitly ecological approach by examining census tract-level seat belt non-use in Tennessee. Our methodology consists of several parts. First, we will examine the spatial clustering of seat belt non-use at the census tract level in search for distinctive spatial regimes in the data. With the assumption of the presence of the spatial regime, we will estimate separate models for each regime to learn whether the models are significantly varying across each regime. Next, with the assumption of substantially different models, we estimate the effects on seat belt non-use of structural variables with adjustments for spatial dependency. Lastly, we will assess the extent to which any observed spatial dependence is best described with reference to the effects of unmeasured predictor variables (the spatial error model) or with reference to the influence of seat belt non-use in neighboring census tracts (the spatial lag model). Evidence consistent with the latter would be suggestive of possible influence processes.

ESDA

Visual inspection of Figure 18 indicates that census tracts with higher seat belt use (i.e., blue colors) are surrounded by other census tracts with blue colors. This is also the case for the census tracts with warmer colors (e.g., lower seat belt use rate). In this study, we apply Exploratory Spatial Data Analysis (ESDA) (Anselin 1999). ESDA discovers patterns of spatial association, or clusters, and suggest spatial regimes or other forms of spatial heterogeneity (Anselin 1990, 1999, Baller *et al.* 2001). We will use the insights gained from ESDA in the spatial structure of the model.

The first stage in the ESDA is to identify spatial autocorrelation. We use Global Moran's I statistics (Moran 1950) to investigate the presence of spatial autocorrelation. Global Moran's I was also used to test whether the model residuals are spatially correlated. Moran's I values range from -1 to +1. The extreme values are indicators of significant spatial autocorrelation where values close to 0 indicate a random pattern between residuals. A positive and significant Moran's I indicate clustering in the space of similar seat belt non-use. Moran's I can be written as:

$$I = \frac{\sum_i \sum_j w_{ij}(y_i - \mu)(y_j - \mu)}{\sum_i (y_i - \mu)^2} \qquad \text{Equation 1}$$

where w_{ij} is an element of a row-standardized spatial weights matrix, y_i is the seat belt non-use, and μ is the average seat belt non-use in the sample. The statistical significance of the Moran's I is based on the Z-score. For more details about the calculation of the Moran's I's z-score please see Andrew and Ord (1981).

Next, we will test Local Indicator of Spatial Association (LISA) statistics. The LISA statistic checks for *local* spatial autocorrelation by applying local Moran's statistics, which indicates to what extent the pattern of the seat belt non-use rate in one geographic unit is compatible with spatial randomness. Rejection of the null hypothesis indicates that local clustering of high-high (high values surrounded by high values), low-

low (low values surrounded by low values), high-low (high values surrounded by low values), and low-high (low values surrounded by high values) exist. The Local Moran's I can be calculated by the following equation:

$$I_i = \left(\frac{z_i}{\sum_i {z_i}^2}\right) \sum_j w_{ij} z_j \qquad \textbf{Equation 2}$$

where z refers to the seat belt in mean-deviation form. For more details about the Moran's I please see Anselin and Florax (1995). Local Moran's I is helpful to identify regimes that could be targeted by separate models.

Stability of the coefficients

The exploratory phase in the analysis is started by an ordinary least squares (OLS) regression. We apply a spatial regime regression by using two separate OLS models, which allows the coefficients to be different in each regime (High-high vs. others). By conducting spatial Chow test on the stability of these coefficients across regimes (Chow 1960, Anselin 1990, Myers *et al.* 2017), which produces a statistic similar to the F-statistic, to detect whether there are differences in selected covariates between census tracts between two regimes. The Chow test is useful for two reasons. First, it allows us to explicitly test the spatial structural variance of regression coefficients, which can reveal different social mechanisms by region or different relative significance of the covariates in the model. Second, if regional stability is rejected, the modeling allows for varying spatial processes to be considered in each region (Baller *et al.* 2001).

Regression models

Assuming the presence of spatial autocorrelation, spatial lag model (SLM) and spatial error model (SEM) are common to address this concern. The methodological distinction between the two models is how they consider spatial dependency (Doreian 1980, 1982). The SLM considers the spatial dependency as a spatial lag, which is a weighted average of values for the dependent variable in neighboring locations. In the SEM, spatial dependency is incorporated into the regression term. Spatial dependence in the SLM model suggests a possible influence process, whereas in the SEM model the source of the interdependence in the error term is not known. (Marsden and Friedkin 1993, Baller *et al.* 2001).

Spatial error

A satisfactory spatial error model implies that it is unnecessary to posit the distinctive effects of the lagged dependent variable (Anselin 1990). In the SEM, the constant variable is treated as a spatially structured random effect vector. The SEM is similar to linear regression models with an additional term for the spatial dependency of errors in neighboring units. The general form of the SEM model is as follows:

$$y = X\beta + \varepsilon \qquad \text{Equation 3}$$

$$\varepsilon = \lambda W_\varepsilon + u = (I - \lambda W)^{-1} u \qquad \text{Equation 4}$$

$$y = \lambda W_y + X\beta + \lambda \mathrm{W} \mathrm{X} \beta + \mathrm{u} \qquad \text{Equation 5}$$

where y is a vector of seat belt non-use, X is a vector of independent variables presented in Table 16, β is the corresponding vector of estimated coefficients (X). In this model, ε is the error term, which contains of two parts: W_ε and u. W_ε presents the spatially lagged error term corresponding to a weigh matrix W and u refers to the spatial uncorrelated error term that satisfies the normal regression assumption ($u \sim \mathrm{N}(0, \sigma^2 I)$). Last, λ presents the spatial error term parameters, if the value of the spatial error parameters equals zero, the SEM is similar to the standard linear regression model.

Spatial lag

The spatial lag model, in contrast, incorporates the spatial influence of unmeasured independent variables, but also stipulates an additional effect of neighbors' seat belt non-use, via the lagged dependent variable. The Spatial lag model can be presented as:

$$y = \rho W_y + X\beta + \varepsilon \qquad \text{Equation 6}$$

where y is a vector of seat belt non-use, where ρ presents the spatial autoregressive parameter, W_y is a spatially lagged variable corresponding to W matrix, X is a vector of independent variables, β is the vector of estimated coefficients. Last, ε is assumed to be a vector of independent and identically distributed (IID) error terms. The model is appealing since it integrates the effect of both independent variables (X) on the outcome y with the network (interdependence) effect of W_y (Marsden and Friedkin 1993) –i.e., a strategic interaction. The corresponding "reduced form" of equation 1 is

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon \qquad \text{Equation 7}$$

This reduced form illustrates two important points. First, the spatial error model (Equation 7) is subsumed by the spatial lag model, although in non-nested form. Second, Equation 7 illustrates how the dependent variable at each location is not only determined by X, but also by the X at all other locations through the "Leontief inverse" $(I - \rho W)^{-1}$. This is the model most compatible with common notions of influence processes because it implies an influence of neighbors' seat belt non-use that is not simply a produce of measured or unmeasured independent variables.

Weight matrix

A crucial concept in these methods is that of a spatial weight matrix (W), which incorporates the prior structure of dependence between spatial units. It is important to keep in mind that all analyses are conditional on the choice of the spatial weights. Different types of weighting matrix were considered in this analysis to obtain the most suitable model; namely rook, queen order 1 and 2, and distance-based weight matrix were used for the analysis. The queen weights matrix define neighbors as census tracts that share a boundary or corner, whereas, rook only considers those census tract that shares a boundary (Anselin 2003). The selection of optimal weighting matrix could be based on the AICc (Hurvich and Tsai 1989); the weight matrix with the lowest AICc is preferred (Fotheringham and Brunsdon, Nakaya *et al.* 2005, Hadayeghi *et al.* 2010, Nakaya 2014).

Model comparison and assessment

A Lagrange Multiplier principle was also used to test the specifications against SEM and SLM. These tests are based on the regression residuals obtained from model estimates under the null hypothesis regression (i.e., OLS). Each of SLM and SEM models has their specific LM statistics, which offers the opportunity to exploit the values of these statistics to suggest the likely alternative. The LM statistic against SEM (LM_{SEM}) and SLM (LM_{SLM}) models take the following forms:

$$LM_{SEM} = \frac{\left(\frac{e'W_e}{s^2}\right)^2}{T} \qquad \text{Equation 18}$$

$$LM_{SLM} = \frac{\left(\frac{e'W_e}{s^2}\right)^2}{\frac{(WXb)'M(WXb)}{s^2} + T} \qquad \text{Equation 19}$$

where e is a vector of OLS residuals, s^2 its estimated standard error, $T = tr[(W + W')W]$, tr as the matrix trace operator, and $M = I - X(X'X)^{-1}X'$. Both LM_{SEM} and LM_{SLM} are asymptotically distributed as $\chi^2(1)$ under the null. Several researchers illustrate the relative power of these tests by using extensive simulation studies (Anselin and Rey 1991, Anselin and Florax 1995, Anselin *et al.* 1996).

It is possible that in some cases both LM_{SEM} and LM_{SLM} statistics turn out to be highly significant, which may make it challenging to choose the proper alternative. To deal with this issue, (Anselin *et al.* 1996) developed a robust form of the LM statistics in the sense that each test is robust to the presence of local deviations from the null hypothesis in the form of the other alternative. In other words, the robust LME is robust to the presence of spatial lag, and vice versa. The robust tests perform well in a wide range of simulations and form the basis of a practical specification search, as illustrated in (Anselin and Florax 1995, Anselin *et al.* 1996). In this study, we used GeoDa software to perform the LM tests (Anselin 2003). Queen contiguity matrix was used to generate a spatial weight matrix for this test. Furthermore, we used the White statistics to check the presence of heteroscedasticity (White 1980). Variance Inflation Factors (VIF) was also used to control for potential multicollinearity in each step (O’brien 2007).

Results

After controlling for the census tracts with less than 20 crash observations and census tract with no populations; we used 1,251,901 observations for measuring seat belt non-use at each census tract, which yielded to an average sample size of 20.4% (SD = 8.3) over a three-year period. Average seat belt non-use in each census tract is 10.1% (SD = 4.1), which indicates that the seat belt use rate for a three-year period in Tennessee is close to 89.9%. The seat belt use rate is close to the roadside observations in

Tennessee (88%) (NHTSA 2017). Figure 19 presents the seat belt non-use histograms at the zonal level.

Spatial diagnosis

Global Moran's I value (I = 0.56) based on the queen contiguity matrix indicates the presence of substantial spatial dependency. Moran's I statistic indicates that there is spatial autocorrelation in the OLS model, the positive sign of the Moran's I shows that the neighborhoods with higher seat belt non-use are clustered together and vice versa. Figure 20 presents the visual map of local Moran's I. The clusters with high rates (i.e., high-high) are located in Chattanooga and Memphis metropolitans' areas as well as some scattered clusters in the rural areas. Alternatively, the clusters with low rates (i.e., low-low) are located in other metropolitan areas in Tennessee, namely the suburban areas surrounding the Nashville metropolitan area (except the urban core of Nashville), Knoxville, Clarksville, and Kingsport. Based on Figure 20, we conclude that there are two regimes in Tennessee; the regime of southern metropolitans and rural areas (i.e., Memphis and Chattanooga) –Southern MPOs – and other metropolitan areas –i.e., non-Southern MPOs. The average seat belt non-use in the Southern metropolitan areas is 16% (90^{th} percentile range between 12-21%). On the other hand, seat belt non-use in the non-Southern metropolitan areas is substantially lower with average seat belt non-use of 9% (90^{th} percentile range between 5-13%).

Regression estimation

Table 17 presents the separate OLS models for seat belt non-use in Tennessee by considering a dummy variable for the regional effect to capture the effect of Southern MPOs. Significant Positive value of the Moran's I (0.169, $p < 0.001$) and White test (409.03, $p < 0.001$), reveal a strong presence of both spatial dependency and heteroscedasticity in the model.

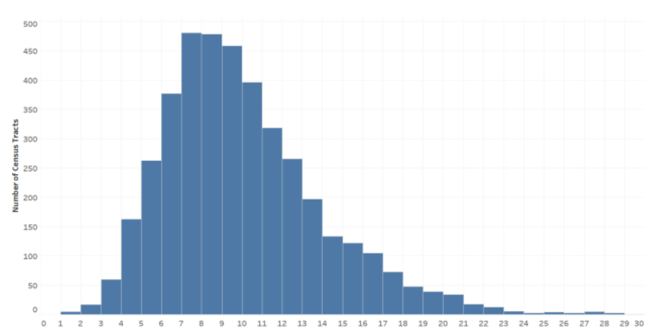

Figure 19 Distribution of seat belt non-use at the zonal level

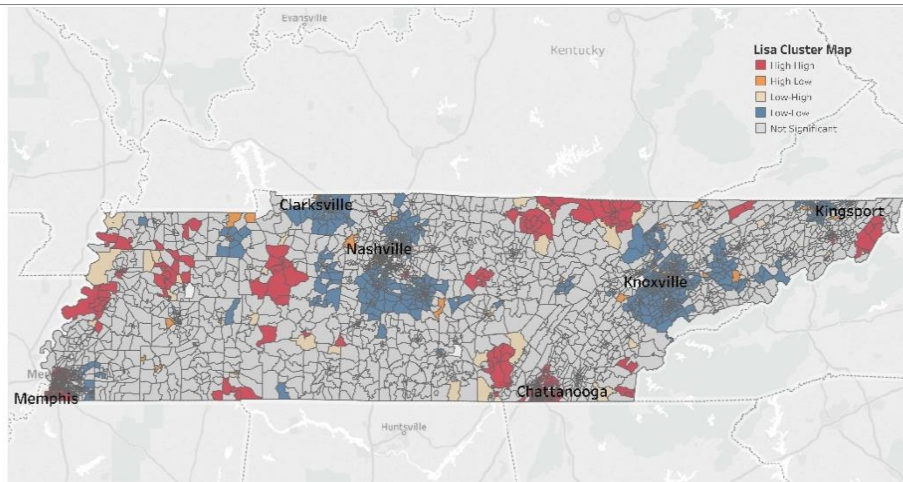

Figure 20 Moran's scatterplot map

Table 17 Ordinary Least Square regression of seat belt non-use and Chow test statistics

					Chow statistics	
Variable	**Coef.**	**S. E.**	**t**	**P-value**	**Value**	**P-value**
Household with Vehicle	-3.038	0.592	-5.130	0.000	1.427	0.232
Age Median	-0.001	0.008	-0.190	0.849	0.003	0.953
Child Percentage	-0.989	0.723	-1.367	0.172	1.845	0.174
Population Density (per Square miles)	0.000	0.000	1.822	0.069	0.798	0.372
% with College Education	-4.784	0.549	-8.719	0.000	6.172	0.013
% with Bachelor Education	-2.255	0.638	-3.537	0.000	0.550	0.459
Income ($10,000)	-0.121	0.029	-4.191	0.000	1.617	0.204
% White	-1.121	0.231	-4.857	0.000	4.621	0.032
Constant	15.041	0.632	23.805	0.000	13.910	0.000
Region (1: Southern MPOs; 0 = otherwise)	5.721	0.153	37.449	0.000		
Global Chow					1134.235	0.000
Sigma-square	9.249					
AIC	20750.47					
Log-likelihood (Full)	-10365.2					
R-squared	0.4412					
Moran's I	0.1688*					
White Test (DF. 44)	409.03*					

* P-value < 0.001

Interpretation of the model without the spatial effect (Table 17) indicates that the sign of proportion of the white population at each census tract indicates that as the proportion of this group increases, seat belt non-use at census tract decreases. As expected, education variables also have a significant association with seat belt use; the negative signs of the estimated coefficients in the models indicate that the proportion of the population with a bachelor degree and college degree negatively impacts seat belt non-use. Income is negatively associated with seat belt non-use. Income, race, and education-related variables are consistent with previous research (Preusser *et al.* 1991, Reinfurt *et al.* 1997, Wells *et al.* 2002, Houston and Richardson 2005) (Vivoda *et al.* 2004, Gkritza and Mannering 2008, Pickrell and Ye 2009).

Vehicle ownership also has a significant association with seat belt non-use. Average vehicle ownership increases reduce seat belt non-use. Population density is correlated with lower seat belt use. This negative impact could be attributed to the shorter distances in the urban areas and a relatively lower speed of travel in general. As a result, vehicle occupants may decide not to use their seat belt in urban areas. Findings regarding the effect of vehicle ownership and population density are in agreement with Hezaveh and Cherry (2019). In this study, the median age did not have a significant association with seat belt non-use, whereas many studies reported that younger populations are more prone to not wearing their seat belt (Reinfurt *et al.* 1997, Calisir and Lehto 2002, Glassbrenner *et al.* 2004).

Table 17 also presents the results of an examination of coefficient stability for the Southern MPO versus non-Southern MPO regimes. The spatial Chow test clearly rejects the null hypothesis of coefficient stability. A closer examination of the individual tests on coefficient stability across regimes supports the conclusion that the proportions of white population and population with a bachelor's degree exert significantly different effects across regions.

Table 18 and Table 19 present the results of Moran's I, White test, and LM test for separate regression models for each region. As presented in Table 18, the significant values of the Moran's I indicates that spatial dependency exists in both regimes. Interestingly, White test statistics indicate that heteroscedasticity exists in the Southern MPOs whereas in the Non-Southern MPOs heterogeneity exists.

Given the strong evidence of distinct spatial regimes in Tennessee (i.e., Chow Test statistics), we estimate separate models for each regime and will scrutinize on the presence of spatial dependence. Applying the LM test (Table 19) suggests that for the Southern MPOs area a spatial lag model is more suitable, whereas in the rest of the study area a spatial error model is more suitable. Table 20 presents the estimated SLM and SEM model for each region.

Table 20 indicates that the effects of the Southern MPO areas lags of seat belt non-use is positive and statistically significant. This finding implies that seat belt non-use in Southern MPOs in Tennessee are influenced by seat belt non-use in nearby census tracts, which is consistent with an influence process. On the other hand, for the non-south metropolitan areas, the SEM model implies that the residual spatial autocorrelation can be adequately accounted for in terms of unmeasured predictor variables. Therefore, we may conclude that an influence process seems unlikely in non-Southern MPOs.

Comparison of the significant association among covariates and seat belt non-use in both models indicate that in the non-Southern MPOs only age median and child percentage do not have a significant association with seat belt non-use. Whereas for the Southern MPOs, only income and percentage of household with the vehicle have a significant association with seat belt non-use. Interestingly, the percentage of the white population at census tract, density, and education do not have a significant association with seat belt non-use in the Southern MPOs regime.

Table 18 Moran's I and White test statistics for each regime

Test	Non-Southern MPOs		Southern MPOs	
	Value	P-value	Value	P-value
Moran's' I	0.2756	0.000	0.139	0.000
White Test	895.69	0.000	21.939	0.997

Table 19 LM test statistics for each regime

Test	Non-Southern MPOs		Southern MPOs	
	Value	P-value	Value	P-value
Lagrange Multiplier (lag)	470.315	0.000	34.246	0.000
Lagrange Multiplier (error)	599.036	0.000	23.309	0.000
Robust LM (lag)	0.863	0.353	12.039	0.001
Robust LM (error)	129.584	0.000	1.102	0.294

Table 20 Results of the spatial models

	SEM (for Non-Southern MPOs)				SLM (for Southern MPOs)			
Variable	**Coef.**	**S. E.**	**Z**	**P-value**	**Coef.**	**S. E.**	**Z**	**P-value**
% White	-1.083	0.331	-3.271	0.001	-0.064	0.479	-0.133	0.894
% with Bachelor Education	-1.079	0.656	-1.645	0.100	-2.617	1.646	-1.590	0.112
% with College Education	-4.131	0.579	-7.138	0.000	-1.007	1.586	-0.635	0.526
Income ($10,000)	-0.114	0.030	-3.756	0.000	-0.220	0.098	-2.240	0.025
Child Percentage	-1.045	0.726	-1.440	0.150	1.385	1.880	0.736	0.462
Population Density (per Square miles)	0.000	0.000	3.268	0.001	0.000	0.000	0.147	0.883
Age Median	-0.006	0.008	-0.855	0.393	0.002	0.021	0.075	0.940
% Household with Vehicle	-2.549	0.642	-3.969	0.000	-3.545	1.250	-2.835	0.005
Constant	14.299	0.690	20.729	0.000	15.304	1.786	8.570	0.000
Lag Coef. (Lambda)	0.493	0.021	23.173	0.000				
Lag Coeff. (Rho)					0.303	0.055	5.479	0.000
AIC	16761.000				3391.440			
Log-likelihood (Full)	-8371.492				-1685.720			
R-squared	0.282				0.159			
Sample Size								

Conclusion

In this study, we used seat belt use reported by police officers at crash sites to explore the spatial dependency of seat belt non-use at the zonal level. We found that seat belt non-use rate is not randomly distributed in space. Southern-MPOs census tracts have higher-than-average seat belt non-use rates that form statistically significant clusters. In addition, ESDA, Chow statistics, and Lagrange multiplier analysis reveal distinctive regional imprint for spatial autocorrelation. Results of LM statistics also indicate that in the Southern-MPOs, the spatial lag model is more suitable, whereas, in the non-Southern MPOs, the spatial error model is more suitable. The LM statistic implies that residents of Southern-MPOs seat belt non-use are under the influence of the other residents in this regions. The positive spatial autocorrelation in Southern-MPO also implies that a census tract seat belt non-use is not produced solely by the internal structural factors and it is influenced by their neighboring units. The presence of the social influence also warrants research for investigating the underlying mechanism of social influence in future studies. Identifying the underlying mechanism would be helpful in the design of an effective road safety campaign, such as communication methods with recipients of the campaign.

The current practice for selecting the seat belt campaigns rely on blanket coverage for areas with lower seat belt use rate. Using the methodology presented in this study provides a spatial analysis of seat belt non-use, which could be used to identify clusters and outliers of seat belt non-use. Moreover, this data visualization could help practitioners to decide on seat belt campaign geographic scopes.

The present study population consists of vehicles occupants with a home address in Tennessee who had a traffic crash in Tennessee during 2014-16. It is likely that the population of this study is skewed towards those who are more prone to unsafe behavior (i.e., they were involved in crashes). Nevertheless, the sample used in this study consists of 1.25M observations or about 19% of the state population. These findings present a sample of Tennessean vehicle occupants, and careful consideration is needed to transfer these findings to other settings. Nevertheless, the method and results that this study present could be generalizable to other contexts as well.

References

Ajzen, I., 1991. The Theory of Planned Behavior. Organizational behavior and human decision processes 50 (2), 179-211.

Ali, M., Haidar, N., Ali, M.M., Maryam, A., 2011. Determinants of Seat Belt Use among Drivers in Sabzevar, Iran: A Comparison of Theory of Planned Behavior and Health Belief Model. Traffic injury prevention 12 (1), 104-109.

Andrew, C., Ord, J.K., 1981. Spatial Processes: Models and Applications. London: Pion.

Anselin, L., 1990. Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis. Journal of Regional Science 30 (2), 185-207.

Anselin, L., 1999. Interactive Techniques and Exploratory Spatial Data Analysis. Geographical Information Systems: principles, techniques, management and applications 1, 251-264.

Anselin, L., 2003. Geoda 0.9 User’s Guide. Urbana 51, 61801.

Anselin, L., Bera, A.K., Florax, R., Yoon, M.J., 1996. Simple Diagnostic Tests for Spatial Dependence. Regional science and urban economics 26 (1), 77-104.

Anselin, L., Florax, R.J., 1995. Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results. New Directions in Spatial Econometrics. Springer, pp. 21-74.

Anselin, L., Rey, S., 1991. Properties of Tests for Spatial Dependence in Linear Regression Models. Geographical analysis 23 (2), 112-131.

Baller, R.D., Anselin, L., Messner, S.F., Deane, G., Hawkins, D.F., 2001. Structural Covariates of Us County Homicide Rates: Incorporating Spatial Effects. Criminology 39 (3), 561-588.

Begg, D.J., Langley, J.D., 2000. Seat-Belt Use and Related Behaviors among Young Adults. Journal of Safety Research 31 (4), 211-220.

Blincoe, L., Miller, T.R., Zaloshnja, E., Lawrence, B.A., 2015. The Economic and Societal Impact of Motor Vehicle Crashes, 2010. (Revised) (Report No. Dot Hs 812 013). Washington, Dc: National Highway Traffic Safety Administration.

Calisir, F., Lehto, M.R., 2002. Young Drivers’ Decision Making and Safety Belt Use. Accident Analysis & Prevention 34 (6), 793-805.

Cartwright, D., 1965. Influence, Leadership, Control.

Chow, G.C., 1960. Tests of Equality between Sets of Coefficients in Two Linear Regressions. Econometrica: Journal of the Econometric Society, 591-605.

Ctr, 2018. 2017 Survey of Safety Belt Usage in Tennessee Final Report. In: Research, T.U.O.T.C.F.T. ed.

Dittrich, D., Leenders, R.T.A., Mulder, J., 2017. Bayesian Estimation of the Network Autocorrelation Model. Social Networks 48, 213-236.

Doreian, P., 1980. Linear Models with Spatially Distributed Data: Spatial Disturbances or Spatial Effects? Sociological Methods & Research 9 (1), 29-60.

Doreian, P., 1982. Maximum Likelihood Methods for Linear Models: Spatial Effect and Spatial Disturbance Terms. Sociological Methods & Research 10 (3), 243-269.

Fockler, S.K., Cooper, P.J., 1990. Situational Characteristics of Safety Belt Use. Accident Analysis & Prevention 22 (2), 109-118.

Fotheringham, A., Brunsdon, C., M. Charlton. 2002. Geographically Weighted Regression-the Analysis of Spatially Varying Relationships. Chichester, UK: John Wiley & Sons.

Fujimoto, K., Chou, C.-P., Valente, T.W., 2011. The Network Autocorrelation Model Using Two-Mode Data: Affiliation Exposure and Potential Bias in the Autocorrelation Parameter. Social networks 33 (3), 231-243.

Gkritza, K., Mannering, F.L., 2008. Mixed Logit Analysis of Safety-Belt Use in Single- and Multi-Occupant Vehicles. Accident Analysis & Prevention 40 (2), 443-451.

Glassbrenner, D., Carra, J.S., Nichols, J., 2004. Recent Estimates of Safety Belt Use. Journal of safety research 35 (2), 237-244.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of Planning Level Transportation Safety Tools Using Geographically Weighted Poisson Regression. Accident Analysis & Prevention 42 (2), 676-688.

Hezaveh, A.M., Cherry, C.R., 2019. Neighborhood-Level Factors Affecting Seat Belt Use. Accident Analysis and Prevention 122, 153-161.

Houston, D.J., Richardson, L.E., 2005. Getting Americans to Buckle Up: The Efficacy of State Seat Belt Laws. Accident Analysis & Prevention 37 (6), 1114-1120.

Hurvich, C.M., Tsai, C.-L., 1989. Regression and Time Series Model Selection in Small Samples. Biometrika 76 (2), 297-307.

Iihs, 2018. Safety Belts. Insurance Institute for Highway Safety, Insurance Institute for Highway Safety.

Leenders, R.T.A., 2002. Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix. Social networks 24 (1), 21-47.

Majumdar, A., Noland, R.B., Ochieng, W.Y., 2004. A Spatial and Temporal Analysis of Safety-Belt Usage and Safety-Belt Laws. Accident Analysis & Prevention 36 (4), 551-560.

Marsden, P.V., Friedkin, N.E., 1993. Network Studies of Social Influence. Sociological Methods & Research 22 (1), 127-151.

Mmucc, 2012. Model Minimum Uniform Crash Criteria. DOT HS 811, 631.

Moran, P.A., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37 (1/2), 17-23.

Myers, C.A., Slack, T., Broyles, S.T., Heymsfield, S.B., Church, T.S., Martin, C.K., 2017. Diabetes Prevalence Is Associated with Different Community Factors in the Diabetes Belt Versus the Rest of the United States. Obesity 25 (2), 452-459.

Nakaya, T., 2014. Gwr4 User Manual. WWW Document. Available online: http://www. st-andrews. ac. uk/geoinformatics/wp-content/uploads/GWR4manual_201311. pdf (accessed on 4 November 2013).

Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically Weighted Poisson Regression for Disease Association Mapping. Statistics in medicine 24 (17), 2695-2717.

Nelson, D.E., Bolen, J., Kresnow, M.-J., 1998. Trends in Safety Belt Use by Demographics and by Type of State Safety Belt Law, 1987 through 1993. American Journal of Public Health 88 (2), 245-249.

Nhtsa, 2017. Seat Belt Use in 2017—Use Rates in the States and Territories. In: Analysis, N.S.N.C.F.S.A. ed.

Nordfjærn, T., Şimşekoğlu, Ö., Zavareh, M.F., Hezaveh, A.M., Mamdoohi, A.R., Rundmo, T., 2014. Road Traffic Culture and Personality Traits Related to Traffic Safety in Turkish and Iranian Samples. Safety science 66, 36-46.

O'brien, R.M., 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & quantity 41 (5), 673-690.

Pickrell, T., Ye, T., 2009. Traffic Safety Facts (Research Note): Seat Belt Use in 2008–Demographic Results. National Highway Traffic Safety Administration/Department of Transportation, Washington DC.

Preusser, D.F., Williams, A.F., Lund, A.K., 1991. Characteristics of Belted and Unbelted Drivers. Accident Analysis & Prevention 23 (6), 475-482.

Reinfurt, D., Williams, A., Wells, J., Rodgman, E., 1997. Characteristics of Drivers Not Using Seat Belts in a High Belt Use State. Journal of Safety Research 27 (4), 209-215.

Saha, D., Alluri, P., Gan, A., Wu, W., 2018. Spatial Analysis of Macro-Level Bicycle Crashes Using the Class of Conditional Autoregressive Models. Accident Analysis & Prevention.

Şimşekoğlu, Ö., Lajunen, T., 2008. Social Psychology of Seat Belt Use: A Comparison of Theory of Planned Behavior and Health Belief Model. Transportation Research Part F: Traffic Psychology and Behaviour 11 (3), 181-191.

Şimşekoğlu, Ö., Nordfjærn, T., Zavareh, M.F., Hezaveh, A.M., Mamdoohi, A.R., Rundmo, T., 2013. Risk Perceptions, Fatalism and Driver Behaviors in Turkey and Iran. Safety science 59, 187-192.

Torquato, R., Franco, C., Bianchi, A., 2012. Seat Belt Use Intention among Brazilian Undergraduate Students. Revista Colombiana de Psicología 21 (2), 253-263.

Tranmer, M., Pallotti, F., Lomi, A., 2016. The Embeddedness of Organizational Performance: Multiple Membership Multiple Classification Models for the Analysis of Multilevel Networks. Social Networks 44, 269-280.

Vitale, M.P., Porzio, G.C., Doreian, P., 2016. Examining the Effect of Social Influence on Student Performance through Network Autocorrelation Models. Journal of Applied Statistics 43 (1), 115-127.

Vivoda, J.M., Eby, D.W., Kostyniuk, L.P., 2004. Differences in Safety Belt Use by Race. Accident Analysis & Prevention 36 (6), 1105-1109.

Wang, W., Neuman, E.J., Newman, D.A., 2014. Statistical Power of the Social Network Autocorrelation Model. Social Networks 38, 88-99.

Wells, J.K., Williams, A.F., Farmer, C.M., 2002. Seat Belt Use among African Americans, Hispanics, and Whites. Accident Analysis & Prevention 34 (4), 523-529.

White, H., 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica: Journal of the Econometric Society, 817-838.

CHAPTER V: TRAVELER-INVOLVED TRAFFIC CRASHES AS A NEGATIVE EXTERNALITY OF TOURISM INDUSTRY

The authors confirm contribution to the paper as follows: study conception and design: Amin Mohamadi Hezaveh, Candace Brakewood, Christopher Cherry; data collection: Not Applicable; analysis and interpretation of results: Amin Mohamadi Hezaveh, Candace Brakewood, Christopher Cherry; draft manuscript preparation: Amin Mohamadi Hezaveh, Candace Brakewood, Christopher Cherry. All authors reviewed the results and approved the final version of the manuscript.

Abstract

Although it is well established that travelers have a higher risk of injury in traffic crashes compared to non-travelers, less is known about the magnitude of traffic crashes involving travelers and the negative externality of travelers' crashes (NETC) imposed on non-travelers. In this note, we rely on the U.S. Travel Association's definition of a traveler to conduct an empirical analysis focusing on the state of Tennessee; we define travelers as those who travel more than 50 miles from home or have a home-address outside of Tennessee state. We find that 19.2% (127,031 out of 694,276 from 2014-2016) of traffic crashes in Tennessee involve a traveler. The injury cost of non-traveler crashes due to a crash with a traveler (i.e., monetized value of NETC) exceeds $7.6 billion, or 12.3% of tourist expenditures between 2014-2016. Analyzing the net impact of travel (tourist expenditures minus NETC) at county level reveals that the NETC exceeds tourist expenditures in 19 of 97 counties (or 20%) in Tennessee. The results of this analysis reveal that an overlooked negative externality of tourism is traffic crashes involving travelers, which warrants further study and potentially policy remediation.

Introduction

In the United States, the direct contribution of travel and tourism to the Gross Domestic Product (GDP) in 2017 was $509.4 billion (2.6% of GDP) (World Travel & Tourims Council 2018). In Tennessee, tourism is a major industry along with agriculture and manufacturing, and the state is a Top 10 travel destination in the United States. Based on U.S. Travel Association (2017) estimates, travelers spent $56.7 billion in Tennessee between 2014 and 2016 (adjusted for 2017 inflation). Furthermore, in the last three years, more than 100 million people visited Tennessee annually, which resulted in $11.8 billion in direct payroll impacts and employed 176 thousand people in 2016.

Travelers increase roadway traffic en-route and at their final destination, which can lead to increased traffic crashes at these locations. This study aims to explore the association between road safety and travelers. Several studies showed that travelers - particularly foreign travelers and out of state drivers - have a higher injury risk compared to domestic drivers (Petridou et al. 1997, Leviäkangs 1998, Summala 1998, Petridou et al. 1999, Wilks et al. 1999, Claret et al. 2002, Yannis et al. 2007) or in-state drivers (Harootunian et al. 2014a, Harootunian et al. 2014b). Studies in the United States indicate that the odds of out-of-state drivers being at-fault for a crash are higher than in-state drivers Harootunian et al. (2014a), and a driver's general unfamiliarity with one's surroundings may cause more trouble compared to other factors such as language barriers, culture, or infrastructure quality Harootunian et al. (2014b).

It is apparent that travelers spend money en-route and at their final destination, resulting in a direct local economic benefit. However, in case of a crash that involves a traveler and a non-traveler (i.e., a local road user), we argue that a negative externality has been imposed on the non-traveler. To the best of our knowledge, no study has explored the magnitude of this negative externality. Therefore, this study aims to provide empirical evidence that sheds light on this important topic. First, we identify traffic crashes in which at least one of the drivers was a traveler. Second, we measure the monetary value of negative externality of travelers' crash (NETC) imposed on non-travelers both at the state level and county level in Tennessee. Last, we measure the net impact of travelers' crashes (NITC) at the county level to shed light on potential geographic inequalities.

Methodology

Definition of a traveler and traveler-involved traffic crash

In order to identify travelers, we classify road users based on their distance between the location of traffic crashes and their home addresses. The U.S. Travel Association (2017) defines travel as activities associated with all overnight trips away from home in paid accommodations and day or overnight trips to places 50 miles or more (one way) from traveler's origin. In line with this definition, we defined a ***traveler*** as any person whose distance between the location of the traffic crash and his/her home address is beyond 50 miles or whose home address is outside of Tennessee State. Those who do not fit this definition are referred to as ***non-travelers*** (i.e., local residents) and are road users whose distance between their home address and the location of traffic crashes is less than 50 miles. It is also worthy to mention that we do not have access to overnight stay data, which could underestimate travel, and that we did include out-of-state trips with home addresses in close proximity to the Tennessee state border, potentially overestimating travel in border counties.

Next, we define a ***Traveler-involved Traffic Crash (TTC)*** as a crash in which at least one of the drivers, pedestrians, or bicyclists who were involved in the traffic crashes was a traveler. Otherwise, we labeled the crash as a ***non-Traveler-involved Traffic Crash (non-TTC)***. In this case, all the pedestrian, bicyclists, and drivers are local residents.

Data and geocoding

The crash data in this study was provided by Tennessee Integrated Traffic Analysis Network (TITAN), the statewide crash data administered by the Tennessee Department of Safety and Homeland Security. We retrieved records of 694,276 crashes involving 1,501,044 individuals between 2014-2016. Each record includes information about road user type, coordinates of the crashes, and reported road users' home addresses. We used the Bing application program interface services to geocode the addresses. Only those records that had an accuracy level of premises, address level accuracy, or intersection level accuracy was used for the analysis (Hezaveh and Cherry 2019).

Part 1: Identifying traveler-involved traffic crashes and estimating their economic costs

The injury severity in the TITAN database followed the KABCO scale for Tennessee provided by FHWA (FHWA 2011). In KABCO scale, K, A, B, C, and O respectively stand for a crash with a fatality, Incapacitating, Non-Incapacitating Evident, Possible Injury, and No Injury (FHWA 2017). In order to convert the injury severities to crash cost, we used the monetary figures presented in Table 21 recommended by FHWA (Harmon et al. 2018) for the year 2010 for the person-injury unit. To account for inflation, we converted 2010 USD to 2017 USD. By using the numbers presented in Table 21, we estimated the monetary value imposed on local residents in a crash involving travelers using Equation 1, which we refer to as the ***negative externality of travelers' crashes (NETC):***

$$NETC_i = \left(N_{v,i} * Cost_{PDO}\right) + \sum_{\alpha=\{K,A,B,C,O\}} N_{\alpha,i} * Cost_\alpha \qquad \text{Equation 1}$$

where $N_{v,i}$ and $N_{\alpha,i}$ respectively present the number of vehicles registered in county i and individuals with level of injury α who were involved in traffic crashes with a traveler-driver. $Cost_{PDO}$ and $Cost_\alpha$ represent the injury unit costs presented in Table 21.

Table 21 National KABCO person-injury unit costs

Injury Type	Crash Cost Per Injury			
	Economic person-Injury Unit Costs*	QALY Person -Injury Unit Costs*	Comprehensive Crash Cost (2010 Dollars) *	Comprehensive Crash Cost (2017 Dollars) **
No Injury*	\$5,717	\$2,563	\$8,280	\$9,308
Possible Injury	\$21,749	\$49,926	\$71,675	\$80,570
Non-Incapacitating Injury	\$32,105	\$97,974	\$130,079	\$146,222
Incapacitating Injury	\$84,507	\$363,324	\$447,832	\$503,408
Fatal Injury	\$1,398,916	\$7,747,082	\$9,145,998	\$10,281,016
Unknown	\$0	\$0	\$0	\$0
Vehicle unit cost	\$6,076	\$0	\$6,076	\$6,830

* The cost reflects the cases where injury severity was falsely assigned. Source: Harmon et al. (2018)
** Adjusted person-injury unit cost. Source: Authors

Part 2: Estimating the net impact of traveler-involved traffic crashes

The Travel Economic Impact Model (TEIM) was developed by the U.S. Travel Association to provide annual estimates of the impact of the travel activity of U.S. residents on national, state and county economies in the United States. TEIM is a disaggregated model that has the capability to estimate the various types of travel such as business and vacation trips, by various transportation modes, and type of accommodations used. For more details about the TEIM model, please see U.S. Travel Association (2017). The TEIM model measures the travelers' expenditures (TE) at the county level. TE includes spending by travelers on goods and services during their trips. For this analysis, we assume that TE represents the economic benefit of travelers to a region. We then used equation 2 to calculate the difference between this benefit and the negative externality of traffic crashes involving travelers, which we refer to as the ***net impact of travelers' crashes (NITC)***, as follows:

$$NITC_i = TE_i - NETC_i \qquad \text{(Equation 2)}$$

where $NETC_i$ represents the negative externality of travelers' crashes (NETC) for county i and TE_i represents the tourist expenditure (TE) reported by the U.S. Travel Association (2017) for county i. A negative value would indicate that travelers' crash negative externalities exceeds the tourist expenditures in that county.

Results

Part 1: Traveler crash frequency and economic costs

Between 2014 and 2016 in Tennessee, we were able to extract the home addresses of 1,501,044 (92.9%) of traffic crash victims. Table 22 presents the road user type involved in traffic crashes and their corresponding traveler status. 1,196,353 drivers were involved in traffic crashes. Out of those, we were able to retrieve the state of 112,0251 drivers, pedestrians, and bicyclists as well as the exact home address of 992,306 individuals that met the minimum address quality. Further, we were able to identify the home address of the drivers, pedestrians, and bicyclists in 660,919 traffic crashes (95.2%) of reported crashes in TITAN. We also flagged 146,285 (13.1%) of the drivers, pedestrians, and bicyclists as travelers. The percentage of travelers was greater than what Harootunian *et al.* (2014b) reported in the Florida, Maine, Minnesota, and Nevada. From this, we also labeled 127,031 (19.2%) of traffic crashes that occurred in Tennessee between 2014 and 2016 as TTC.

We also estimated the number of travel-involved traffic crashes at the county level. On average, 15.7% of the crashes at the county level involved a traveler. Tourism hubs Shelby (24,248), Davidson (18,970), Hamilton (16,311), and Knox (7,151) counties had the highest frequency of TTC crashes; these counties are the main hubs of travelers in Tennessee and respectively include Memphis, Nashville, Chattanooga, and Knoxville. However, Marion (46%; 1,064), Sevier (38%; 5,063), and Polk (36%; 873) counties had the highest percentage of TTC crashes. Figure 21 presents the spatial distribution of the TTC percentage (top) and TTC frequency (bottom).

Table 22 Road user type and traveler frequency

Road User Type	Frequency			Percentage	
	Local Resident	Traveler	Total	Local Resident	Traveler
Driver	973,966	146,285	1,120,251	86.9%	13.1%
Passenger	321,600	51,826	373,426	86.1%	13.9%
Pedestrian	4,627	421	5,048	91.7%	8.3%
Bicyclist	1,229	82	1,311	93.8%	6.3%
Other or Unknown	748	260	1,008	74.2%	25.8%
Total	1,302,170	198,874	1,501,044	86.8%	13.2%

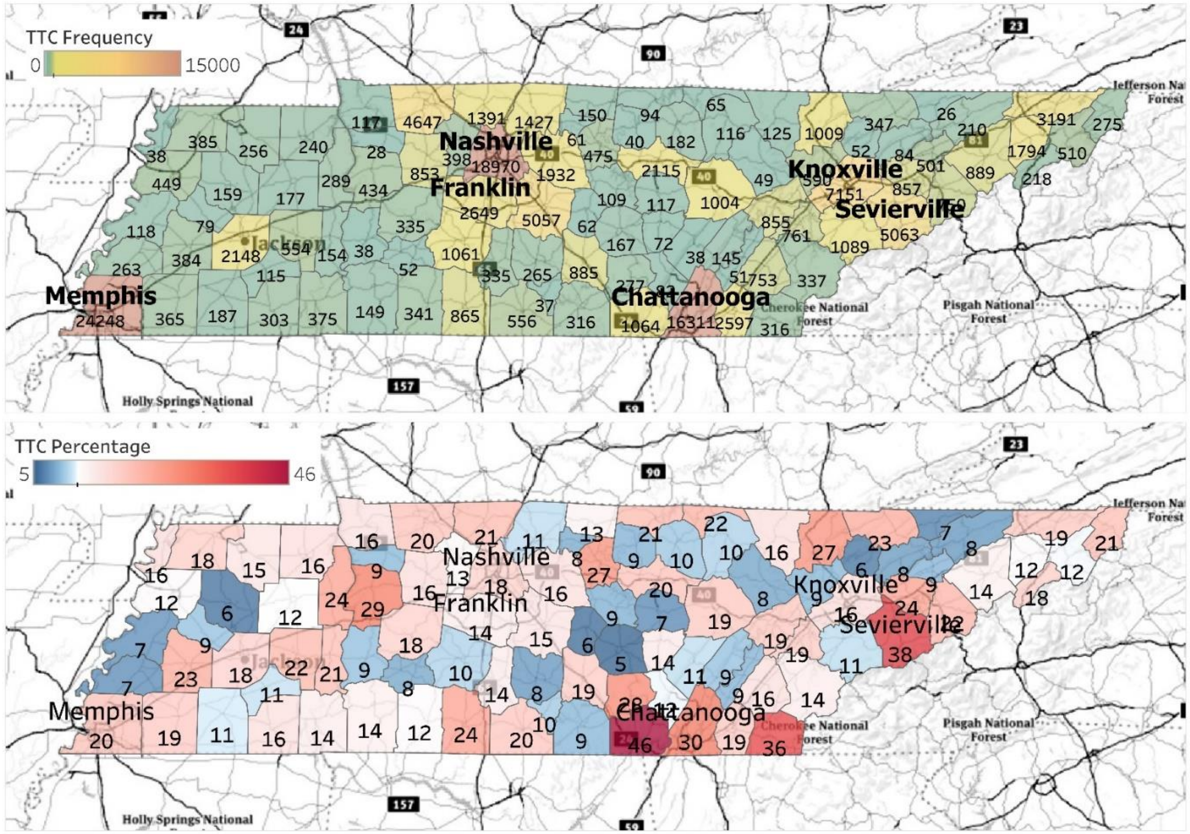

Figure 21 TTC percentage (top) and frequency (bottom) at the county level

Part 2: Net impact of traveler-involved traffic crashes

The comprehensive cost of traffic crashes in Tennessee between 2014 and 2016 was $79.9 billion. We estimate that traveler-involved traffic crashes imposed approximately $7 billion of externalities on local residents. The monetized value of the NETC is equivalent to 12.3% of all traveler expenditures (adjusted for inflation).

The negative externalities of travelers' crashes at the county level was equal to 16% (Median = 14%; 90% interval 7%~28%) of traveler expenditures in Tennessee. Figure 22 presents the spatial distribution of the NITC. Visual examination of Figure 22 shows the disparities between counties. Overall, the sum of travelers' expenditure in tourist hubs Davidson, Shelby, Sevier, Knox, Hamilton counties were $49 billion, and the NETC was $3.4 billion (49% of all NETC). The average TTC percentage in these counties was 24% (ranges 16 to 38%). Although the average TTC crash frequency and relative TTC crash frequency in these counties are higher than the state average, due to the substantial travel expenditures (61% of statewide expenditures), these counties generally benefit from travelers visits. Overall, 19 counties have lower traveler benefits than crash costs; the sum of travelers' expenditure in these counties was $586 million, and NETC equals $908 million. The average TTC crash percentage in these counties was 14% (ranges 6 to 27%). These counties are located adjacent tourist hubs or in some cases are clustered in rural areas.

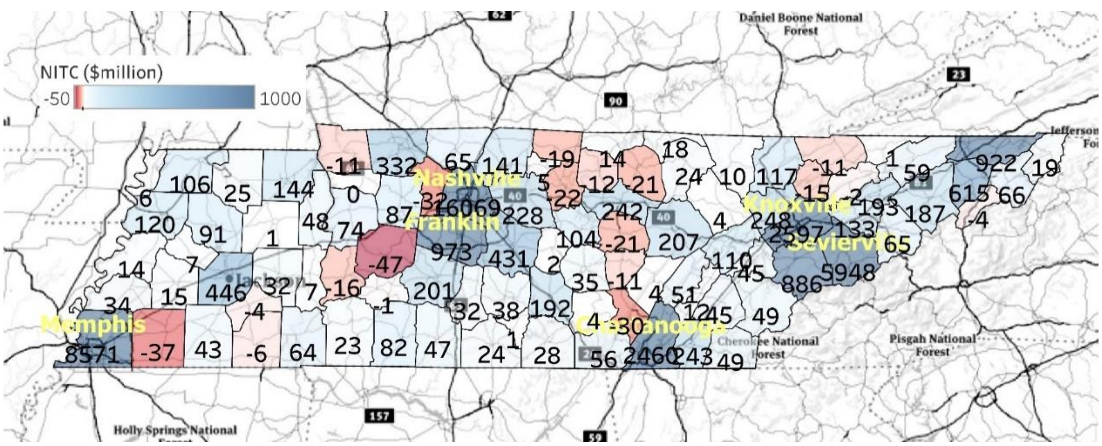

Figure 22 Spatial distribution of the NITC

Discussion

Equitable transportation policy aims to generate fair distribution of economic resources and costs to individuals. In transportation finance, this often manifests as distributing transportation revenue from one group (donors) to another (donees) often to reflect higher marginal cost burdens or externalities in the transportation system. One of the negative externalities of the tourism industry is the traffic crashes due to travelers. Travelers negatively impact local residents, yet bring economic opportunity that tends to outweigh their negative safety local impact. However, the spillover effects tend to burden other counties that do not reap local benefits.

This study contributes to the road safety and tourism industry literature by quantifying the negative externality of travelers' crashes at the county level and identifying the "winning" (done) and "losing" (donor) counties in the state of Tennessee. In total, 19 counties were identified as having negative externalities from traffic crashes that outweigh the economic benefits received from these travelers. Based on this finding, we hope to stimulate discussion and future research on approaches to redistribute and/or remedy the negative crash costs imposed by travelers in these "losing" counties. For example, these pass-through counties could benefit from improved infrastructure or education programs, paid for by traveler-targeted tax revenue streams (e.g., occupancy tax, car rental fees) that aim to reduce the burden of crashes on rural counties. Moreover, we hope that researchers in other regions will utilize the proposed method to examine the geographic distribution of traveler-involved traffic crashes and assess if other localities experience similar externalities from tourism and travel.

Acknowledgment

Thanks to Patrick Dolan from the Tennessee Department of Safety & Homeland Security. This project was partially supported by the Collaborative Sciences Center for Road Safety, a U.S. Department of Transportation National University Transportation Center promoting safety.

References

Claret, P.L., Del Castillo, J.L., Moleón, J.J., Cavanillas, A.B., Martín, M.G., Vargas, R.G., 2002. Influence of driver nationality on the risk of causing vehicle collisions in spain. Journal of Epidemiology & Community Health 56 (5), 394-398.

Fhwa, 2017. Kabco injury classification scale and definitions.

Harmon, T., Bahar, G., Gross, F., 2018. Crash costs for highway safety analysis. USDOT FHWA Office, Washington, DC

Harootunian, K., Aultman-Hall, L., Lee, B.H., 2014a. Crash fault analysis of out-of-state drivers in vermont. Journal of Transport Safety and Security 6 (3), 207-219.

Harootunian, K., Lee, B.H., Aultman-Hall, L., 2014b. Odds of fault and factors for out-of-state drivers in crashes in four states of the USA. Accident Analysis & Prevention 72, 32-43.

Hezaveh, A.M., Cherry, C.R., 2019. Neighborhood-level factors affecting seat belt use. Accident Analysis and Prevention 122, 153-161.

Leviäkangs, P., 1998. Accident risk of foreign drivers—the case of russian drivers in south-eastern finland. Accident Analysis & Prevention 30 (2), 245-254.

Petridou, E., Askitopoulou, H., Vourvahakis, D., Skalkidis, Y., Trichopoulos, D., 1997. Epidemiology of road traffic accidents during pleasure travelling: The evidence from the island of crete. Accident Analysis & Prevention 29 (5), 687-693.

Petridou, E., Dessypris, N., Skalkidou, A., Trichopoulos, D., 1999. Are traffic injuries disproportionally more common among tourists in greece? Struggling with incomplete data. Accident Analysis & Prevention 31 (6), 611-615.

Summala, H., 1998. American drivers in europe: Different signing policy may cause safety problems at uncontrolled intersections. Accident Analysis & Prevention 30 (2), 285-289.

U.S. Travel Association, 2017. The economic impact of travel on tennessee counties 2016. Research Department of the U.S. Travel Association, Washington, D.C.

Wilks, J., Watson, B.C., Johnston, K.L., Hansen, J.A., 1999. International drivers in unfamiliar surroundings: The problem of disorientation. Travel Medicine International 17 (6), 162-167.

World Travel & Tourims Council, 2018. Travel & tourism economic impact 2018. The Harlequin Building, 65 Southwark Street, London SE1 0HR, United Kingdom.

Yannis, G., Golias, J., Papadimitriou, E., 2007. Accident risk of foreign drivers in various road environments. Journal of safety research 38 (4), 471-480.

CONCLUSION

The Home-Based Approach (HBA) offers a new perspective and enables us to develop new data-driven policy recommendation that focuses on the role of road users involved in traffic crashes. HBA offers many advantages and new research ideas. In this dissertation, I provided several examples of the application of HBA in the road safety analysis in five chapters. Key contributions of each chapter of this dissertation are listed as follow:

Chapter I and Chapter II:

- Developing a methodology for identifying the areas that burden of road safety is more tangible.
- Developing a methodology for integrating HBA and travel demand model to analyze factor influencing the burden of traffic crashes based on the residential address of road users.
- Developing a methodology for measuring the economic cost of traffic crashes at a very fine geographic level.

Chapter III, Chapter IV:

- Developing a methodology for measuring seat belt use rate at a very fine geographic level.
- Exploring the spatial dependencies in seat belt use rates at the zonal level.
- Developing a methodology for exploring the spatial patterns in seat belt use and identifying spatial regimes in Tennessee.

Chapter V:

- Developing a methodology for identifying traveler involved in traffic crashes
- Developing a methodology for measuring the travelers' negative externality of the traffic crashes imposed on non-traveler.

Key takeaways

In Chapter I and Chapter II, HBA incorporated sociodemographic and travel behavior of the road users. Incorporating of the travel behavior of the road users based on their home addresses indicates sociodemographic factors, population density, trip rate, and trip length has key roles in analyzing the burden of traffic crashes (i.e., HBA crash rate, the economic cost of traffic crash per capita). Moreover, considering the spatial effect in the model indicate that road safety in one neighborhood is not solely produced by internal factors, yet it is affected by road safety level in neighboring TAZs. Analyzing the spatial distribution of the burden of traffic crashes also indicates that road users who live in certain geographic areas (e.g., close to high-speed roads) are more prone to the burden of traffic crashes.

Findings based on chapter III, IV indicates that police reports of seat belt use of the vehicle occupants are consistent with roadside observations and phone interviews in

Tennessee. Accordingly, by applying the HBA method and assigning the seat belt use of the vehicle occupants to their residential neighborhood, we measured seat belt use rates at a fine geographical area. Moreover, the analysis indicates that seat belt use is not produced solely by the internal structural factors and their neighboring units influence it. Furthermore, HBA provides valuable information regarding seat belt use distribution at fine geographic level with hundreds of thousands of observations that have the potential to replace the traditional sources of seat belt use data and help practitioners in the design of an effective road safety campaign.

Findings of Chapter V point out that approximately 14% of drivers involved in traffic crashes in Tennessee are travelers and almost in one on five crashes in Tennessee a traveler is involved. These crashes imposed a $7.7 billion cost to local residents in Tennessee. Moreover, analyzing the net impact of travel (tourist expenditures minus negative externality of travelers' crashes –NETC) at county level reveals that the NETC exceeds tourist expenditures in 19 of 97 counties (or 20%) in Tennessee. The results of this analysis reveal that an overlooked negative externality of tourism is traffic crashes involving travelers, which warrants further study and potentially policy remediation.

Future directions

HBA offers a new perspective to road safety analysis and has the potential to be applied to several aspects of road safety. Regarding the case studies presented in this dissertation, HBA has the potential to integrate to travel demand models and provide an evaluation of transportation planning alternatives and their effect on the distribution of the burden of traffic crashes. Moreover, the travel demand model that I used in this study is not sensitive to mode choice. One possible direction for future studies is to develop models that are sensitive to modal split particularly transit ridership and evaluate the effect of building massive transit structure and reduction in road users exposure to traffic.

Another direction for future studies could be to measure the aberrant behavior of the road users at the zonal level by applying the HBA method. HBA method could be used to identify areas that their residents are more prone to a particular type of violations such as driving under the influence, speeding, and distracted driving. Similar to chapter III and Chapter IV, application of the HBA would enable researchers and practitioners for designing educational campaigns that targets groups and geographic areas that their residents are more prone to specific aberrant behaviors.

Moreover, HBA could be used to identify the crash characteristics of the traveler in Tennessee. One direction for future research could be identified as travelers' crash hotspots and exploring their crash characteristics. Traveler crashes could have their own characteristics and that distinguish them from non-travelers. This distinction could be attributed to the non-familiarity of drivers with the transportation infrastructure. Knowledge of contributing factors such as built environment and engineering design would help researchers and safety practitioners to provide a safe structure that adapts to the complexities of travelers' behavior such as their unfamiliarity.

VITA

Amin Mohamadi Hezaveh was born in Arak, Iran. Amin has a Master degree in transportation planning and engineering from Tarbiat Modares University in Tehran, Iran. Amin's research focuses on transportation planning and road safety particularly safety of vulnerable road users as well as promoting the use of safety equipment such as helmet and seat belt. Amin received his Ph.D. in Civil and Environmental Engineering at the University of Tennessee, Knoxville in 2019.