Collaborative Sciences Center for ROAD SAFETY

FINAL REPORT

Appendices: Applications to integrating spatial safety data into transportation planning processes

December, 2019

Christopher R. Cherry
Amin M. Hezaveh
University of Tennessee, Knoxville

Louis Merlin
Florida Atlantic University

Contents

HBA Application 1: Factors influencing road users' likelihood of involvement in traffic crashes at the zonal level

Amin Mohamadi Hezaveh
University of Tennessee, Knoxville
311 JD Tickle Building, TN 37996
Phone: (385) 259-5148; Email: amohamad@vols.utk.edu

Christopher R. Cherry
University of Tennessee, Knoxville
321 JD Tickle Building, TN 37996
Phone: (865) 974-7710
Fax: (865) 974-2669; Email: cherry@utk.edu

Abstract

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, less is known about the relationship between travel behavior and the likelihood of involvement in traffic crashes at the zonal level. *By using the home address of the road users who were involved in a traffic crash, we measured the number of traffic crashes which residents of a traffic analysis zone had (i.e., a Home-Based Approach –HBA) in* Knoxville metropolitan region *between 2014-16. Next, by dividing HBA crash frequency to the traffic analysis zone population, we measured the HBA crash rate at the zonal level (HBA-CR). Furthermore, we obtained socioeconomics and travel behavior data elements surrounding home-address of the individuals from Knoxville regional travel demand model.* We also measured average zonal activity based on the travel demand model outputs by using average distance traveled from one zone to others on a daily basis –i.e., individuals' exposure. Moran's I indicates that the HBA-CR is not randomly distributed in space, and it exhibits spatial autocorrelation. Analysis indicates that HBA-CR varies substantially over income, average zonal activity, and traffic 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 also holds for interstate, and arterial vehicle miles traveled (VMT), intersection density, the percentage of roads with sidewalks, percentage of areas near bus stations and number of workers per household. On the other hand, median household income, population density, and VMT on low-speed roads have significant negative associations with HBA-CR. *Findings are discussed in line with road safety countermeasures.*

Keywords: Macroscopic Crash Prediction Models; Home-Based Approach; Spatial Lag Model; Residence crash rate

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 population respectively) (World Health Organization 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). 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.

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 unevenly. Less is known about the factors influencing the likelihood of involvement in traffic crashes based on the residential address of the road users 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).

Several studies used the home address of the road users involved in traffic crashes to explore factors affecting road safety. For example, Lee, Abdel-Aty, and Choi (2014) investigated the characteristics of the at-fault drivers in traffic crashes in Florida by using the zip code of the drivers. Lee, Abdel-Aty, and Choi (2014) reported that population, age, commute mode, and income were associated with the number of at-fault drivers. Moreover, Lee et al. (2015) also examined the relationship between sociodemographic and crash-involved pedestrians per residence zip code in Florida. They concluded that pedestrian crashes do not necessarily occur at their zip code residents (Lee et al. 2015). Likewise, 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). Blatt and Furman (1998) used information of the fatally injured drivers in the US from the Fatality Analysis Reporting System (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 and 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, in a study in the Southeast USA, Stamatiadis and Puccini (2000) extracted the driver address 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, Tippetts, and Voas (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, Tippetts, and Voas (2006) observed a difference in alcohol-related fatality rates across Hispanic subgroups. Furthermore, Romano, Tippetts, and Voas (2006) concluded high-income and education levels have a protective influence on alcohol-related fatal motor vehicle crashes. 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 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. In a recent study, Hezaveh and Cherry (2019b) 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 (2019b) explored sociodemographic factors influencing seat belt use rates variation.

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 studies that relied on home addresses of traffic victims used fatally injured road users. These studies used course geographic units such as zip code, or only focused on a specific group of road users. Although the relationship between sociodemographic factors and road safety is well explored, less is known about the relationship between travel behavior and the likelihood of involvement in traffic crashes at the zonal level.

To explore the spatial variation of the **likelihood of involvement in traffic crashes and its relationship with travel behavior, we will use macroscopic crash prediction models** (MCPM). MCPM is one set of methods that explores 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, and travel behavior (e.g., Gomes, Cunto, and da Silva 2017; Hadayeghi, Shalaby, and Persaud 2003; Hadayeghi, Shalaby, and Persaud 2010b; Lee et al. 2015; Naderan and Shahi 2010; Pirdavani et al. 2012b; Quddus 2008).

Traditionally, in road safety analysis as well as MCPM, 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, Shalaby, and Persaud 2010b; Li et al. 2013; Rhee et al. 2016; Pirdavani et al. 2012b, 2012a; Pirdavani, Brijs, Bellemans, and Wets 2013; 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, Abdel-Aty, and Choi 2012), road length with different functional classification (Hadayeghi, Shalaby, and Persaud 2010b; Quddus 2008), or population has been used (Gomes, Cunto, and da Silva 2017). In regard to measuring the likelihood of involvement in traffic crashes at the zonal level based on the home address of the road users, 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; Lee, Abdel-Aty, and Choi 2014). 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 (Dong et al. 2014; Dong, Huang, and Zheng 2015; Abdel-Aty et al. 2011; Naderan and Shahi 2010; Mohammadi, Shafabakhsh, and Naderan 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 travel behavior, sociodemographic variables, and the likelihood of involvement in traffic crashes at the zonal level. Instead of relying on the zip code of the road users, we used home-address of the road users extracted from police crash database to measure road safety at the zonal level. High resolution of the home address enables us to explore the association between travel behavior and safety at the zonal level by linking the data to a travel demand model. Furthermore, we also consider the trip length and frequency simultaneously to measure road users' exposures in the transportation networks based on travel demand model outputs.

The next section discusses the methods used in this study. In the methodology section, we discuss the HBA, 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-addresses of the road users who were involved in a traffic crash is one of the data elements of 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 Home-Based Approach (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 rather than the location of traffic crashes.

Data and Geocoding Process

This study focuses on the Knoxville metropolitan region with a total population of over one million (Figure 1) and includes ten counties namely Knox, Anderson, Roane, Union, Grainger, Hamblen, Jefferson, Sevier, Blount, and Loudon. 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 the 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. We were able to successfully match 141,514 (95%) of the individuals with a home-location and accordingly to a TAZ corresponding to their home address.

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 HBA-CR indicates that the 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.

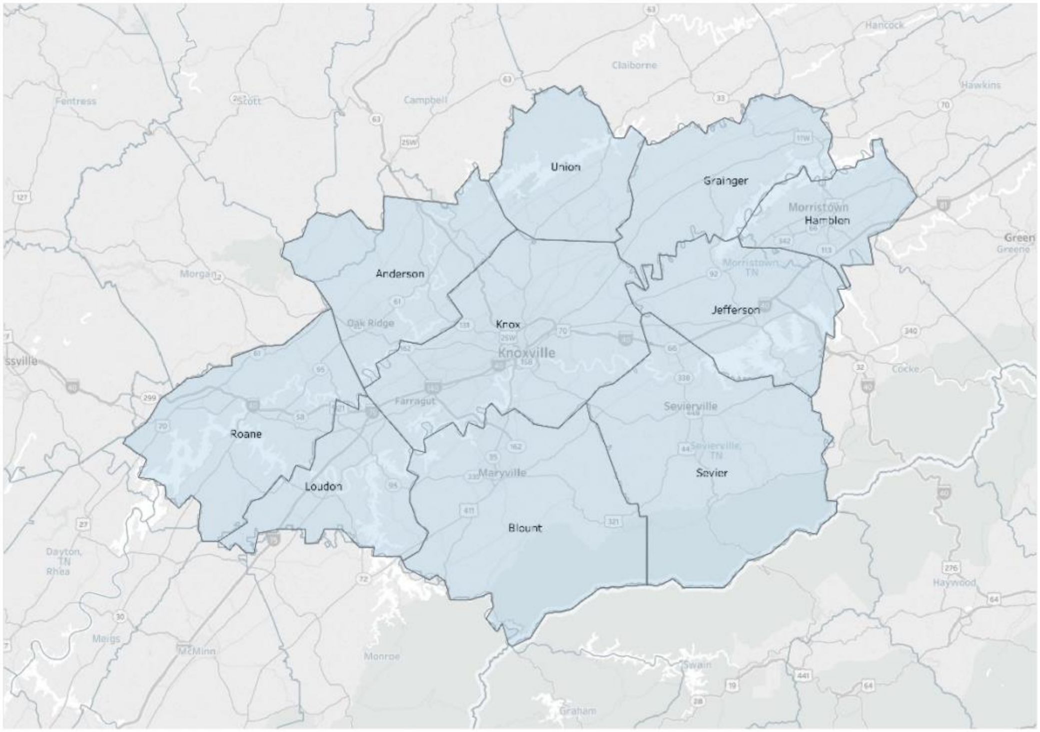

Figure 1 Knoxville Regional Travel Demand

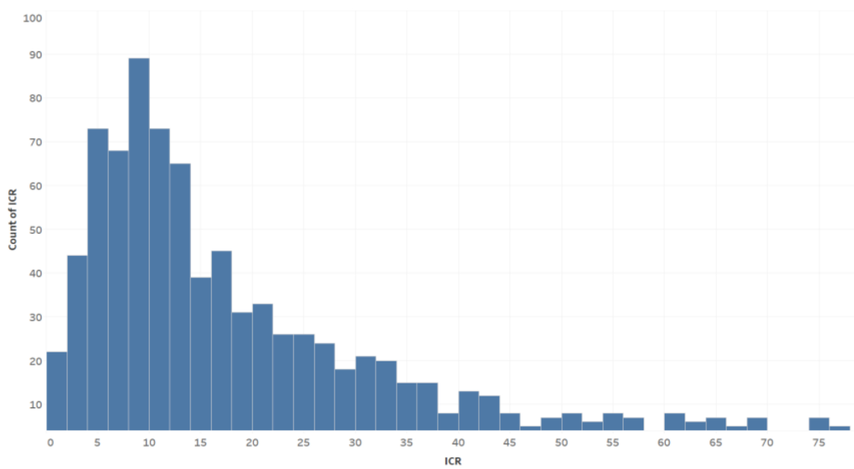

Figure 2 Histogram of HBA-CR at the TAZ level

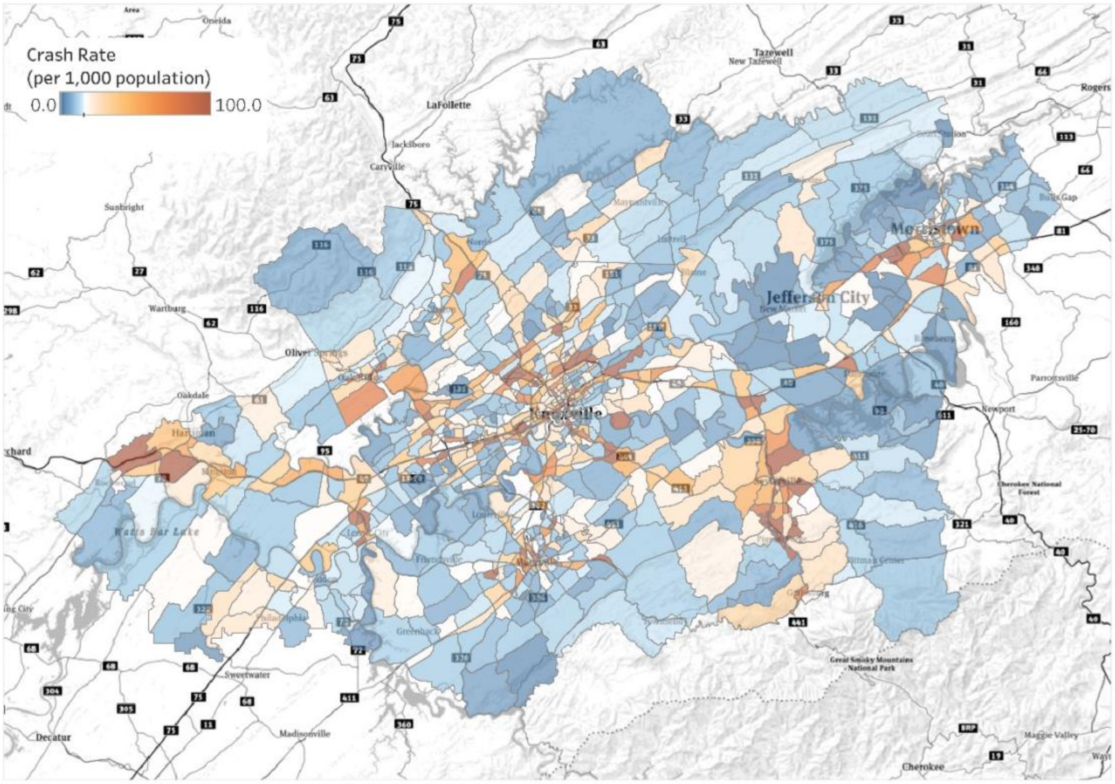

Figure 3 HBA-CR distribution in KRTM

Measuring exposure and travel activity

In this study, one goal was to investigate the relationship between travel behavior, quality of transportation infrastructure, and HBA-CR. 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 1 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 1 presents the descriptive statistics of the data elements obtained from the KRTM model.

Table 1 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

To evaluate the exposure at the zonal level, we use average person miles traveled at zonal level (PMT). PMT_i combines trip rate and trip length, and is an index for measuring the average 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 average zonal activity 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 (blue 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).

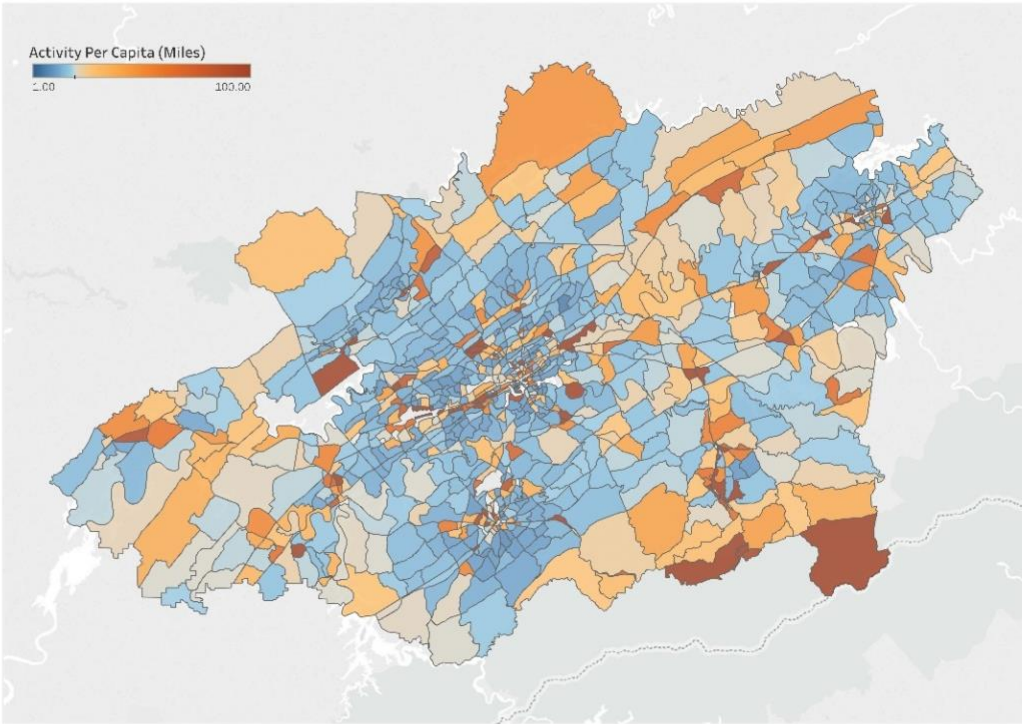

Figure 4 Daily average zonal activity (person miles traveled)

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 (Rhee et al. 2016; Lee et al. 2015; Quddus 2008). 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 of Moran's I indicate a significant spatial autocorrelation where value close to 0 indicates a random pattern between residuals. A significant and positive Moran's I indicates clustering in space of similar HBA-CR.

By hypothesizing the presence of significant spatial autocorrelation, we will use model specifications that consider the spatial dependency in their structure. Spatial error model (SEM) and spatial lag model (SLM) are two common models that are used by researchers to consider spatial autocorrelation in the road safety analysis (Lee et al. 2015; Rhee et al. 2016; Quddus 2008). The distinction between the two models in the method is 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.

Spatial error model

SEM model is similar to the ordinary least squares (OLS) model. However, 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 effect levels that are similar to those from neighboring units (LeSage and Pace 2009). Compared to the OLS, the SEM has 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 1, 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, Phipps, and Anselin 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 corrected Akaike information criterion –AICc (Hurvich and Tsai 1989); the weight matrix with the lowest AICc is preferred (A. Fotheringham and Brunsdon ; Nakaya 2014; Nakaya et al. 2005; Hadayeghi, Shalaby, and Persaud 2010b). 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:

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

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

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 Florax 1995; Anselin and Rey 1991; 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 Lagrange Multiplier (LM) 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 was 95 (SD = 107). Average HBA-CR for the study period was 76 per 1,000 populations (SD = 141). Figure 5 and Figure 6 present the visual distribution of the HBA-CR over the dependent variables. Visualizing HBA-CR over different functional classes of VMT indicates that as VMT increases, the HBA-CR also increases. For example, the HBA-CR for TAZs with high VMT is 2-3 times more than areas with no interstate roads. This is also the case for the arterial roads and other road classification. Notably, the HBA-CR distribution also varied substantially over average zonal activity. For example, in TAZs with very low average zonal activity (<10 PMT), the average HBA-CR is 17, whereas for TAZs with very high average zonal activity (PMT > 50) the corresponding value is greater than 100. HBA-CR also has substantial variation over income categories. For instance, the groups with income below $25K the HBA-CR is 2.4 times more than the income group between $50-75K and four times more than the group with a median household income over $100K.

In a nutshell, HBA-CR has a linear relationship with the dependent variables except for the population density as population density increases (smaller than 800 people per square miles) the HBA-CR increases, and then again HBA-CR decreases after the 800 people per square miles point. Furthermore, both workers per household and student per household have a negative relationship with HBA-CR. As worker per household and student per household increases, the HBA-CR decreases.

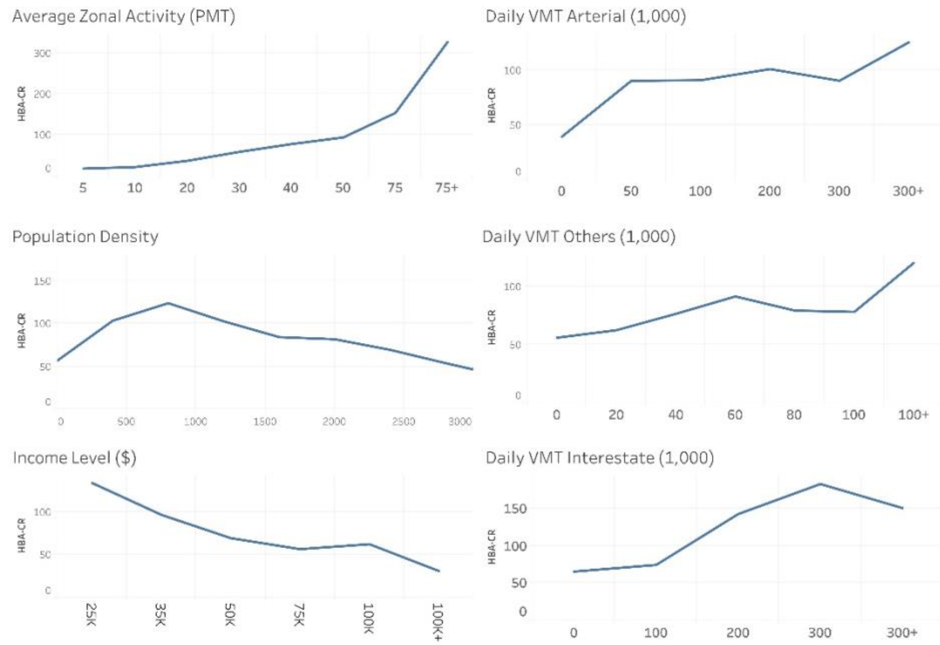

Figure 5 Relationship between the HBA-CR and dependent variables

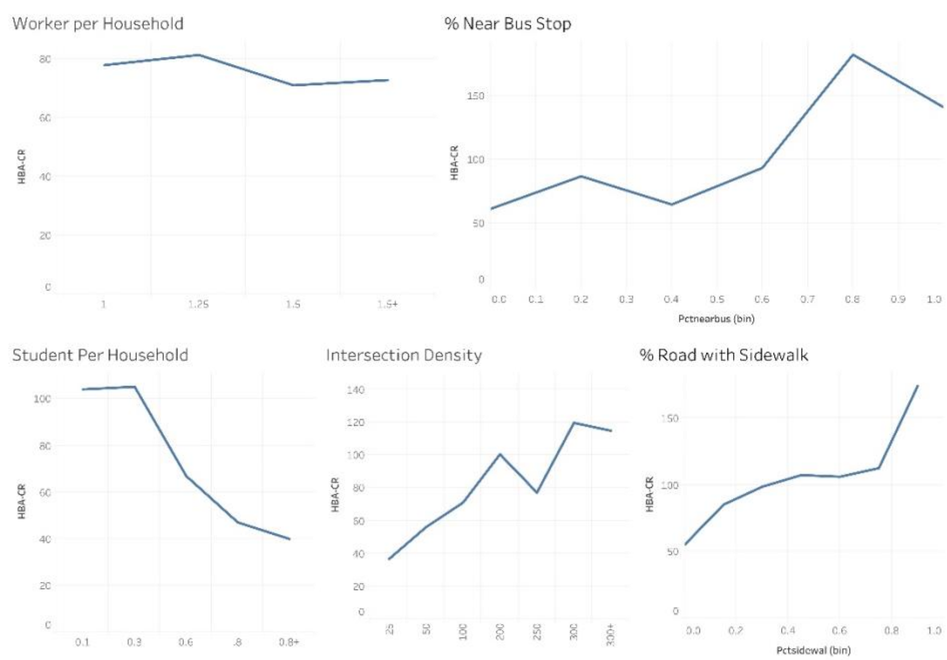

Figure 6 Relationship between the HBA-CR and dependent variables

Results of the global Moran's I indicate that 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. In the 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, 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 LM test was conducted to select the suitable spatial model. LM tests (Table 2) revealed that both LM_{SEM} and LM_{SLM} are significant. Therefore, in the next step we used robust LM statistics. Only Robust-LM_{SLM} has significant values, which indicates that the SLM model is more suitable. Comparison of the AICc values of estimated models in the Table 3 also indicates that the SLM model has a better performance compared to the OLS and SEM.

Table 2 Results of Lagrange multiplier statistics

TEST	VALUE	PROB
Moran's I (error)	5.304	0.000
LM_{SLM}: Lagrange Multiplier (lag)	39.998	0.000
$Robust\ LM_{SLM}$: Robust LM (lag)	15.321	0.000
LM_{SEM}: Lagrange Multiplier (error)	25.067	0.000
$Robust\ LM_{SEM}$: Robust LM (error)	0.390	0.532

Estimated Parameters

In this study, we used the average 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 TAZs with longer travel distances on daily bases have a higher crash rate.

The median household income variable also has a negative correlation with HBA-CR which is consistent with previous studies (Cai, Abdel-Aty, and Lee 2017; Cai et al. 2017; Pirdavani et al. 2012b; Pirdavani, Brijs, Bellemans, and Wets 2013; Gomes, Cunto, and da Silva 2017; Cheng et al. 2018; Lee, Abdel-Aty, and Choi 2014). Individuals with higher household incomes tend to have lower crash rates. This negative sign also is in agreement with road safety literature (World Health Organization 2015; Marshall and Ferenchak 2017; Girasek and Taylor 2010).

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 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.

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) have a negative association with HBA-CR. Many studies explored the association between of functional classes and crash frequency at zonal level (e.g., Hadayeghi, Shalaby, and Persaud 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, Brijs, Bellemans, Kochan, et al. (2013) 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, Shalaby, and Persaud (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 HBA-CR increases. Cai et al. (2017) 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 a higher number of traffic crashes (Ladron de Guevara, Washington, and Oh 2004; Pirdavani et al. 2012a; Hadayeghi, Shalaby, and Persaud 2003; Lovegrove and Sayed 2006; Abdel-Aty et al. 2011; Gomes, Cunto, and da Silva 2017).

Table 3 OLS, SLM, and SEM Estimations

	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
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
Network												
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			

Summary and Conclusion

In this study, we measured the likelihood of involvement in traffic crashes based on the on the home address of individuals (i.e., home-based approach) who were directly involved in traffic crashes at the zonal level. Analysis of the HBA-CR over different categories indicates that HBA-CR substantially varies over VMT classification, average zonal activity, and income variables. Spatial analysis showed 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 aspatial specification. Results of Lagrange Multiplier (LM) 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. Regarding the significant and positive association between both exposure variables and HBA-CR, we can conclude that HBA-CR may decrease by controlling for exposure variables. First, by reducing the VMT of the roads with higher classifications, for example, 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. Both trip length and frequency influence average zonal activity. Therefore, by eliminating a portion of trips by managing travel demand and providing strategies and policies that reduce travel demand (Gärling et al. 2002) may impact HBA-CR. Besides, it is well-established that an increase in density and mixed land-use design would degenerate both trip rate (Cervero and Kockelman 1997), and trip length (Cervero and Kockelman 1997). Hence, an increase in both density and mixed land-use would eventually reduce average zonal activity, VMT and improve the road safety of the road users.

The spatial distribution of the HBA-CR and its association with sociodemographic variables demonstrated potentials of the HBA as a means for identifying the TAZ's 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, mainly 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 crash rate; this issue could be investigated in the 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 the dependent variable and the 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. Moreover, we ran separate models for predicting HBA-CR for all road users and drivers crash rate. Comparison of the models indicates the models are similar, and findings are broadly in agreement. This is due to the fact that pedestrian and bicyclists consist a small portion of road users in this study. Alternatively, average zonal activity reflects trip rates of all road users. Therefore, to maintain concision, we did not present the model for 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 the future could be in 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* (pp. 21-74): Springer.

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. (2017). 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. (2017). 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., . . . 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*. doi:https://doi.org/10.1016/j.aap.2018.05.003

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. Retrieved from 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. (2014). Analysis of residence characteristics of at-fault drivers in traffic crashes. *Safety science, 68*, 6-13.

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. (2012b). Developing zonal crash prediction models with a focus on application of different exposure measures.

Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., & Wets, G. (2013). 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. (2013). 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.

World Health Organization. (2015). *Global status report on road safety 2015*. Retrieved from

Xu, P., & Huang, H. (2015). Modeling crash spatial heterogeneity: random parameter versus geographically weighting. *Accident Analysis & Prevention, 75*, 16-25.

HBA Application 2: Exploring the Cost of Traffic Crash at the Traffic Analysis Zone Level

Amin Mohamadi Hezaveh
University of Tennessee, Knoxville
311 JD Tickle Building, TN 37996
Phone: (385) 259-5148; Email: amohamad@vols.utk.edu

Christopher R. Cherry
University of Tennessee, Knoxville
321 JD Tickle Building, TN 37996
Phone: (865) 974-7710
Fax: (865) 974-2669; Email: cherry@utk.edu

Abstract

Road safety literature provides abundant examples of studies that measure the economic cost of traffic crashes at the coarse geographic level. The current practice of road safety economic assessment attributes traffic crashes 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. After geocoding the home address of individuals who were involved in traffic crashes in Knoxville metropolitan area 2014-16 (n = 183,833) and assigning them to the Traffic Analysis Zone (TAZ) corresponding to their home address; we measured the Economic Cost of traffic Crashes (ECC) and Comprehensive Cost of traffic Crashes Economic (CCC) at the zonal level by using monetary value of the person-injury cost. The average ECC and CCC at the TAZ level were respectively, $920 K and $ 2.7 M. The adjusted Gini index coefficient for the ECC per capita (ECCPC) and CCC per capita (CCCPC) was respectively 0.43 and 0.54 which is an indicator of the unequal distribution of the burden of traffic crashes. Travel demand model output was used as an input for a negative binomial model for exploring the factor correlating with CCC and ECC at the zonal level. Overall, both models were largely consistent. Findings indicate that person miles traveled in network, transportation network characteristics, and demographic information significantly correlates with burden of traffic crashes. Burden of traffic crashes at the zonal level could be used as an index for allocating proper countermeasures and interventions to groups and areas where the burden of traffic crashes is more tangible.

Keywords: Economic Cost of Traffic Crashes; Home-Based Approach; Home-Address; Equity

Introduction

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, Aeron-Thomas, and Astrop 2000; Wijnen and Stipdonk 2016; World Health Organization 2015). Road safety literature has abundant examples of estimating cost of traffic crashes at coarse geographic level (i.e., country-level) (Ahadi and Razi-Ardakani 2015; Blincoe et al. 2015; García-Altés and Pérez 2007; Wegman and Oppe 2010; Mohan 2002); however, to the best of our knowledge, there are no studies that explored this matter at fine geographic level (i.e., traffic analysis zone) and factors correlated with it.

Traffic crash cost has several components and based on the selected components, researchers measure traffic crashes 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 such as the value of pain, suffering, and quality of life loss to victims and their families (Mohan 2002). In other words, 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 (Ahadi and Razi-Ardakani 2015; Blincoe et al. 2015; García-Altés and Pérez 2007; Wegman and Oppe 2010; Mohan 2002). 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 population respectively (World Health Organization 2015). 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; Marshall and Ferenchak 2017; Romano, Tippetts, and Voas 2006). 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 both higher crash rates (Mayrose et al. 2005) (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 attributes safety to the location of the traffic crash. As a result, it is challenging to measure and attribute 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 instead of the location of traffic crashes.

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 studies used fatally injured road users (Blatt and Furman 1998; Males 2009; Romano, Tippetts, and Voas 2006; Stamatiadis and Puccini 2000), course geographic units such as zip code (Lee et al. 2015; Romano, Tippetts, and Voas 2006), census-level (Stamatiadis and Puccini 2000), or only focused on a specific group of road users (Lee et al. 2015). Likewise, these studies did not measure the monetized value of road traffic crashes based on person-injury cost. Monetized value of the traffic crashes consider the effect of both crash frequency and severity simultaneously.

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 an aggregate level such as sociodemographic factors, network characteristics, and travel behavior. By using a wide range of safety outcomes, researchers explored the association between geographic unit characteristic and number of all traffic crashes (Naderan and Shahi 2010; Pirdavani et al. 2012b; Pirdavani, Brijs, Bellemans, and Wets 2013; Huang et al. 2016; Miaou, Song, and Mallick 2003; Cai et al. 2017; Hezaveh and Cherry 2018), number of property damage only crashes (Naderan and Shahi 2010; Aguero-Valverde 2013), frequency of injury/severe crashes (Xu and Huang 2015; Aguero-Valverde 2013; Cai et al. 2017), or crash frequency of specific road users (e.g., non-motorized, bicyclists) (Cai et al. 2017; Cheng et al. 2018; Saha et al. 2018; Lee, Abdel-Aty, and Jiang 2015) at zonal level mostly based on the location of traffic crashes. Although many studies used different dependent variables to measure 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 travel-related factors associating with it.

The population of a TAZ to some extent represents its residents' miles traveled in the transportation system. However, the population variable does not capture the number of trips generated by residents of a geographic area nor their trip length (e.g., activity). Some studies that focused on the modeling the crash frequency of a TAZ based on the location of traffic crashes used trip generation models as a vector to reflect the activity of one TAZ (Naderan and Shahi 2010, Abdel-Aty et al. 2011, Dong et al. 2014, Dong et al. 2015, Mohammadi et al. 2018). Although trip generation vector provides information regarding the activity of the road users, it fails to capture trip length. A more inclusive variable for estimating the economic cost of traffic crashes at a 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). Accordingly, we convert the HBA crash frequency (based on injury severity) to measure the economic cost of traffic crashes at fine geographic areas and subsequently explore the relationship between travel behavior, and economic burden of traffic crashes at the zonal level. We also explore the equitable distribution of crash burden within an urban area based on TAZs characteristics. We measure the distribution of the burden of traffic crashes at the traffic analysis zone (TAZ) level to identify the groups that are more prone to the burden of traffic crashes. Learning about the relationship between exogenous variables, travel activity, and traffic crash 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 transportation system where specific groups are bearing a higher proportional economic burden.

In the next section, we discuss the methodology, including the HBA definition, data, and modeling approach. The rest of the paper presents results and discusses the findings of this study.

Methodology

Person Miles Traveled

In this study, one goal was to investigate the relationship between travel behavior and quality of transportation infrastructure with the crash cost. We used the data from the Knoxville Regional Travel Demand Model (KRTM) in Tennessee. This region is anchored by the city of Knoxville but also includes several urbanized areas outside the city. The 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 7** presents the Knoxville Region study area that includes Knox, Anderson, Roane, Union, Grainger, Jefferson, Sevier, Blount, and Loudon counties. The study area also includes 1,186 TAZs and includes sociodemographic, economic, and travel information of the residents. Table 1 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 we excluded these zones from our analysis.

Table 1 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

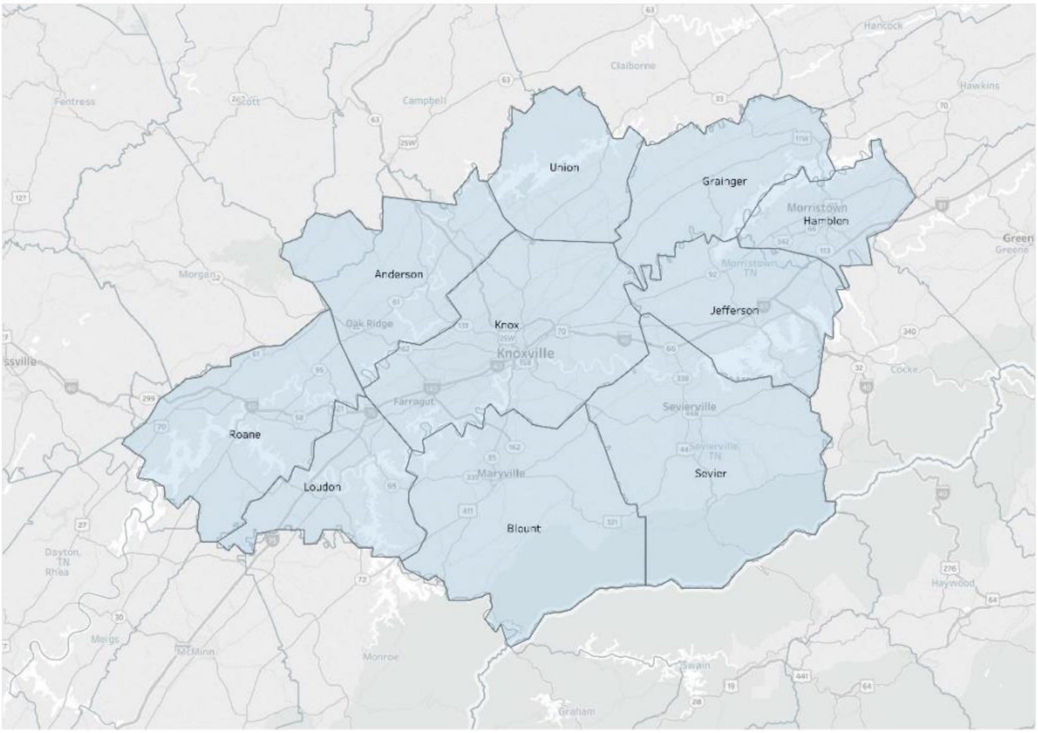

Figure 7 Counties in the Knoxville Regional Travel Demand Model

To evaluate the activity of road users at the TAZ level (i.e., individual's exposure to transportation system), we will use Person Miles Traveled (PMT) at the zonal level. 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:

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

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 output was used as a source to extract the number of trips for each pair. The shortest path between each pair was also extracted from the traffic assignment model at the peak-hour. **Figure 8** presents the distribution of daily PMT in the KRTM Model at the TAZ level. Visual screening of **Figure 8** indicates that the rural areas have higher PMT compared to the urban areas.

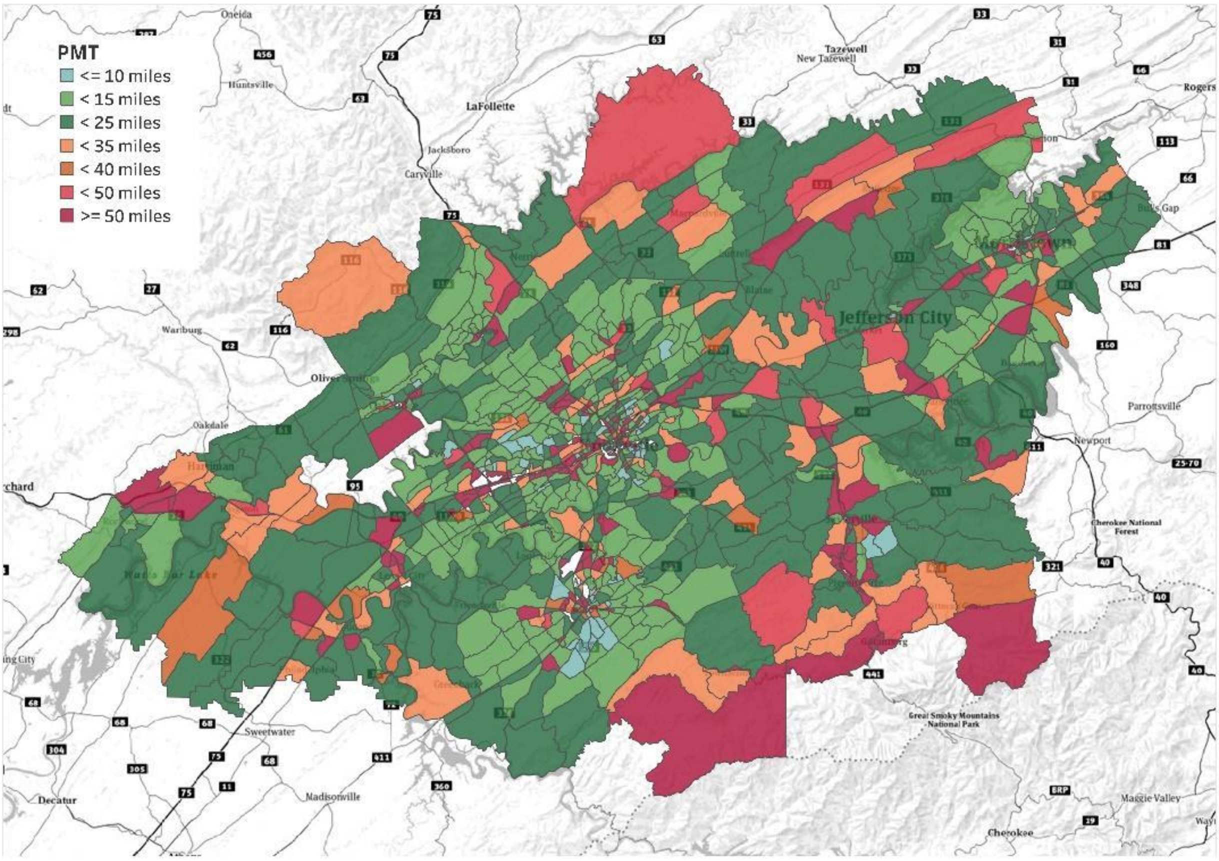

Figure 8 PMT (person miles traveled) distribution across the study area

Home-Based Approach definition

The home address of the road users who were involved in a traffic crash is one of the data elements that police officers record 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), but is not often used in the road safety analysis due to privacy concerns and geocoding challenges.

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, by severity, 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. Next, we use crash frequency and crash severity to calculate the economic cost of traffic crashes in each zone.

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 89,380 crashes that occurred in KRTM area were retrieved from the TITAN. After geocoding the home addresses of the individuals involved in the crashes with Bing Application Program Interface (API) services and quality control of addresses; we were able to assign 183,833 (95% success rate) of the road users to a TAZ corresponding to their home addresses. For more details about the geocoding process, please see Hezaveh and Cherry (2019b) and Hezaveh, Arvin, and Cherry (2019).

Cost of traffic crashes

The injury severity in TITAN database follows the KABCO scale provided by the Federal Highway Administration (FHWA 2017). In the KABCO scale, K, A, B, C, and O respectively stand for an injury with fatal, incapacitating, non-incapacitating evident, possible injury, and no-injury (FHWA 2017). FHWA offers two units for crash cost analysis; person-injury and crash-unit. Person-injury should be applied to the number of involved-persons in crashes whereas the crash-unit cost should be applied to the number of crashes (Harmon, Bahar, and Gross 2018). In order to monetize the value of injury severities, we used the person-injury unit costs presented in Table 2 recommended by FHWA (Harmon, Bahar, and Gross 2018) based on the year 2010. We also converted the person-injury cost to 2018 dollar by adjusting for inflation and income (Harmon, Bahar, and Gross 2018).

Although it is probable that an individual may sustain multiple injuries, person-injuries only considers the most severe injury, and each individual is only counted once (Harmon, Bahar, and Gross 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, Bahar, and Gross 2018).

Next, we measured the total economic cost of the traffic crashes (ECC) and total comprehensive crash cost (CCC) for each TAZ. Furthermore, we measured the economic cost of traffic crashes per capita (ECCPC), and comprehensive cost of traffic crashes per capita (CCCPC) at the TAZ level by dividing the total cost to the population of each TAZ:

$$ECCPC_i = \frac{ECC_i}{T * Pop_i} = PCI * \frac{\left(N_{v,i} * Cost_{PDO}\right) + \sum_{\alpha=\{K,A,B,C,O\}} N_{\alpha,i} * Cost_{E\alpha}}{T * Pop_i} \quad \text{Equation 2}$$

$$CCCPC_i = \frac{CCC_i}{T * Pop_i} = PCI * \frac{\left(N_{v,i} * Cost_{PDO}\right) + \sum_{\alpha=\{K,A,B,C,O\}} N_{\alpha,i} * Cost_{C\alpha}}{T * Pop_i} \quad \text{Equation 3}$$

$N_{\alpha,i}$ is the number of individuals who live in zone i with the level of injury α, $Cost_{E\alpha}$ and $Cost_{C\alpha}$ respectively presents the economic and comprehensive the traffic injury cost per injury in Table 2. T also presents the period of the study (T = 3 years). $N_{v,i}$ is the number of vehicles with a registered address in zone i that were involved in traffic crashes, and $Cost_{PDO}$ is the vehicle unit damage cost. PCI is Tennessee per capita income ratio adjustment factor which is equal to 0.855 for year 2018 (Bureau of Economic Analysis 2018).

Table 2 National KABCO person-injury unit costs (2018 dollar)

Injury Type	Crash Cost Per Injury		
	Economic person-Injury Unit Costs	QALY Person-Injury Unit Costs	Comprehensive Crash Cost (2018 Dollars)
No Injury†	6,553 (5,717*)	2,938 (2,563*)	9,491 (8,280*)
Possible Injury	24,930 (21,749*)	57,227 (49,926*)	82,157 (71,675*)
Non-Incapacitating Injury	36,800 (32,105*)	112,302 (97,974*)	149,102 (130,079*)
Incapacitating Injury	96,866 (84,507*)	41,6459 (363,324*)	513,325 (447,832*)
Fatal Injury	1,603,502 (1,398,916*)	8,880,060 (7,747,082*)	10,483,562 (9,145,998*)
Unknown			
Vehicle unit cost	6,965 (6,076*)		6,965 (6,076*)

† The cost reflects the cases where injury severity was falsely assigned.
* Source: adjusted person-injury cost based on 2010 US Dollar based on Harmon, Bahar, and Gross (2018)

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 (Xia et al. 2016; Zofío et al. 2014; Delbosc and Currie 2011). In an equitable manner, x% of the population pays x% of the economic cost of traffic (Straight-line presented in Figure 9). 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 inequality across a population.

The Gini coefficient is a single value based on the area between the line of equality and the Lorenz curve (Atkinson 1970) and ranges between 0 and 1. 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 value close to 0 corresponds to perfect equality and value close to 1 corresponds to perfect ECCPC inequality.

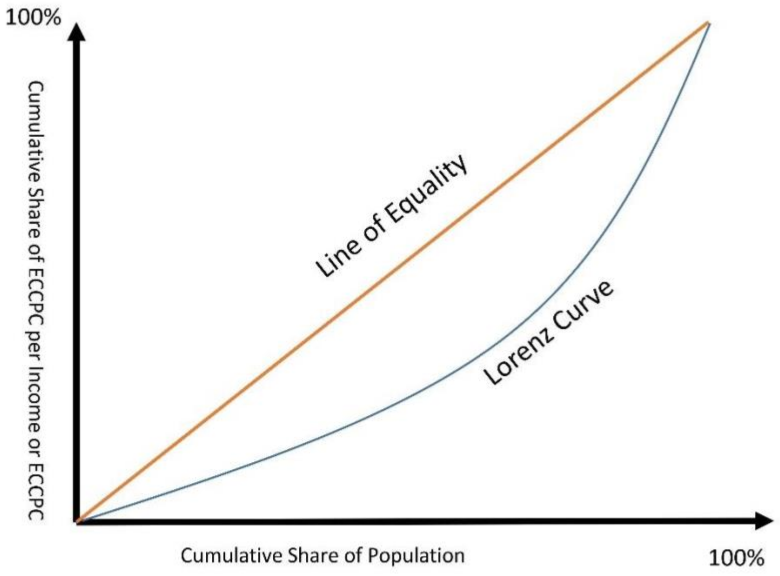

Figure 9 Gini Coefficient and Lorenz curve

Modeling approach

To evaluate safety at zonal level, traditionally, count data models are commonly utilized owing to the nature of traffic crashes that are measured as non-negative integers in a specific period of time (Anastasopoulos and Mannering 2009). Likewise, the cost of the traffic crash is a non-negative integer. Hence the models that would be used to evaluate cost of traffic crashes must follow the nature of counts model (Hezaveh, Arvin, and Cherry 2019).

The Poisson model and negative binomial models are two common model specifications for count data. The main difference between these two specifications is the restriction of equality of the mean and variance $(E[n_i] = Var[n_i])$ of the observations (here ECC and CCC). In the case of traffic crashes, this assumption usually is not met. To take account for the inequality of mean and variance, the more generalized negative binomial model is proposed:

$$\lambda_i = exp\ (\beta X_i + \varepsilon_i) \qquad \text{Equation 4}$$

where $Exp(\varepsilon_i)$ is a gamma-distributed error term with mean 1 and variance α. The addition of this term allows the variance to differ from the mean $(Var[n_i] = E[n_i] + \alpha E[n_i]^2)$. The negative binomial probability density function has the form:

$$P(n_i) = \left(\frac{\frac{1}{\alpha}}{\left(\frac{1}{\alpha}\right) + \lambda_i}\right)^{\frac{1}{\alpha}} \frac{\Gamma\left[\left(\frac{1}{\alpha}\right) + n_i\right]}{\Gamma\left(\frac{1}{\alpha}\right) n_i!} \left(\frac{\lambda_i}{\left(\frac{1}{\alpha}\right) + \lambda_i}\right)^{n_i} \qquad \text{Equation 5}$$

where, $\Gamma(.)$ is a gamma function. In the negative binomial model, if the value of α approaches zero, the negative binomial model yield to the Poisson model. Therefore, the negative binomial model is appropriate when the value of the dispersion parameter (α) significantly differ from zero (S.P. Washington, Karlaftis, and Mannering 2010).

Furthermore, we measured elasticity effects for each variable. Elasticity can be interpreted as the percent effect of a 1% change in a variable has on the severity outcome probability (Khorashadi et al. 2005).

Results and discussion

Among MPO residents, 382 individuals were fatally injured as a result of traffic crashes in the study area. Moreover, 20,705 individuals were injured (level A, B, or C). The economic and comprehensive cost of traffic crashes in the region for 2014-16 were $3.1 and $9.2 Billion (2018 dollars). The average ECC and CCC at the zonal level were respectively $920 K, 95th percent interval $ 866-975K) and $ 2.74 M (95th percent interval $2.54- 2.95 M). Figure 10 and Figure 11 present the spatial distribution of the ECC and CCC at zonal level in the study area. These variables tended to have distributed impacts with high economic cost scattered throughout the region.

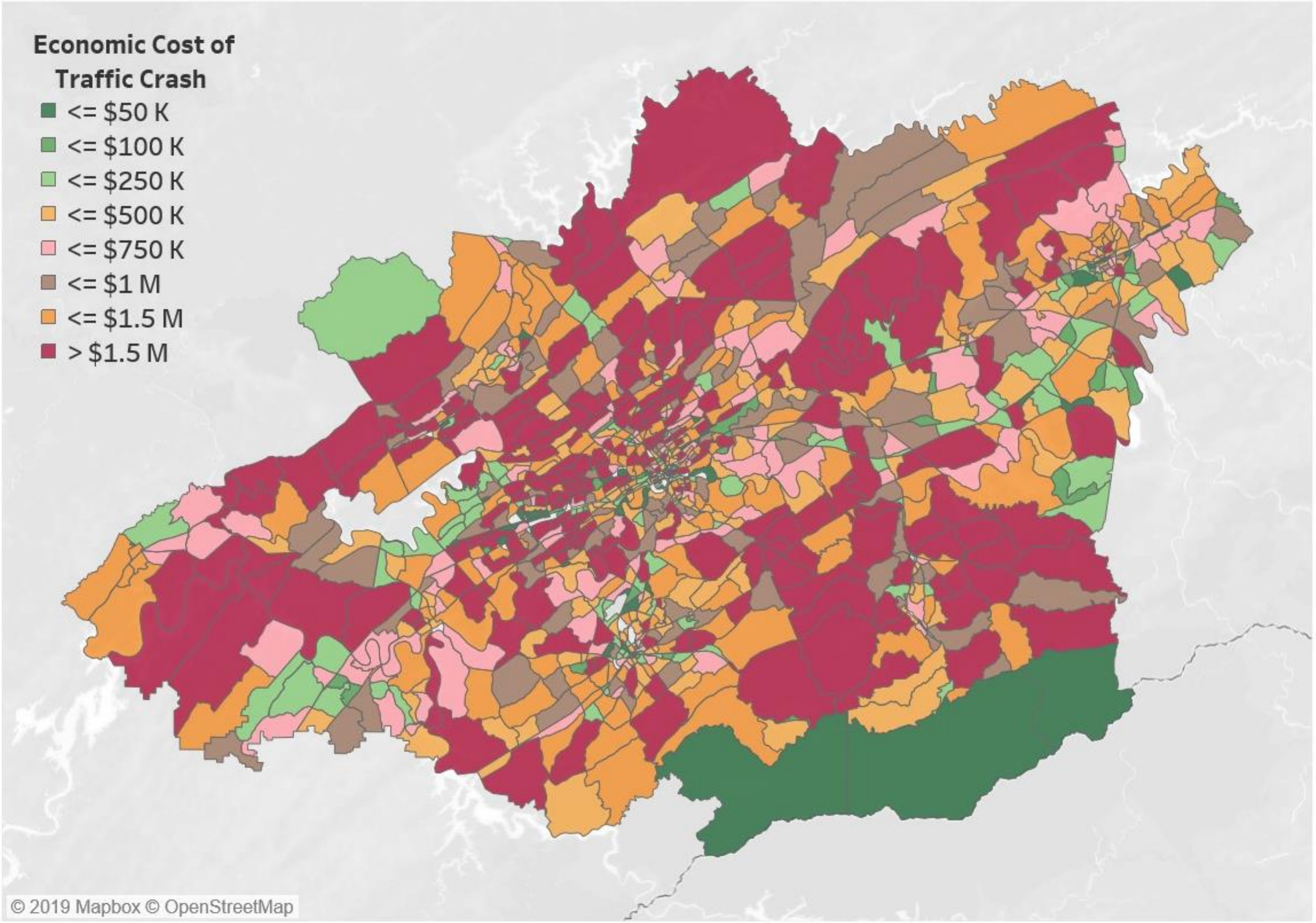

Figure 10 ECC distribution in KRTM; average of 2014-16

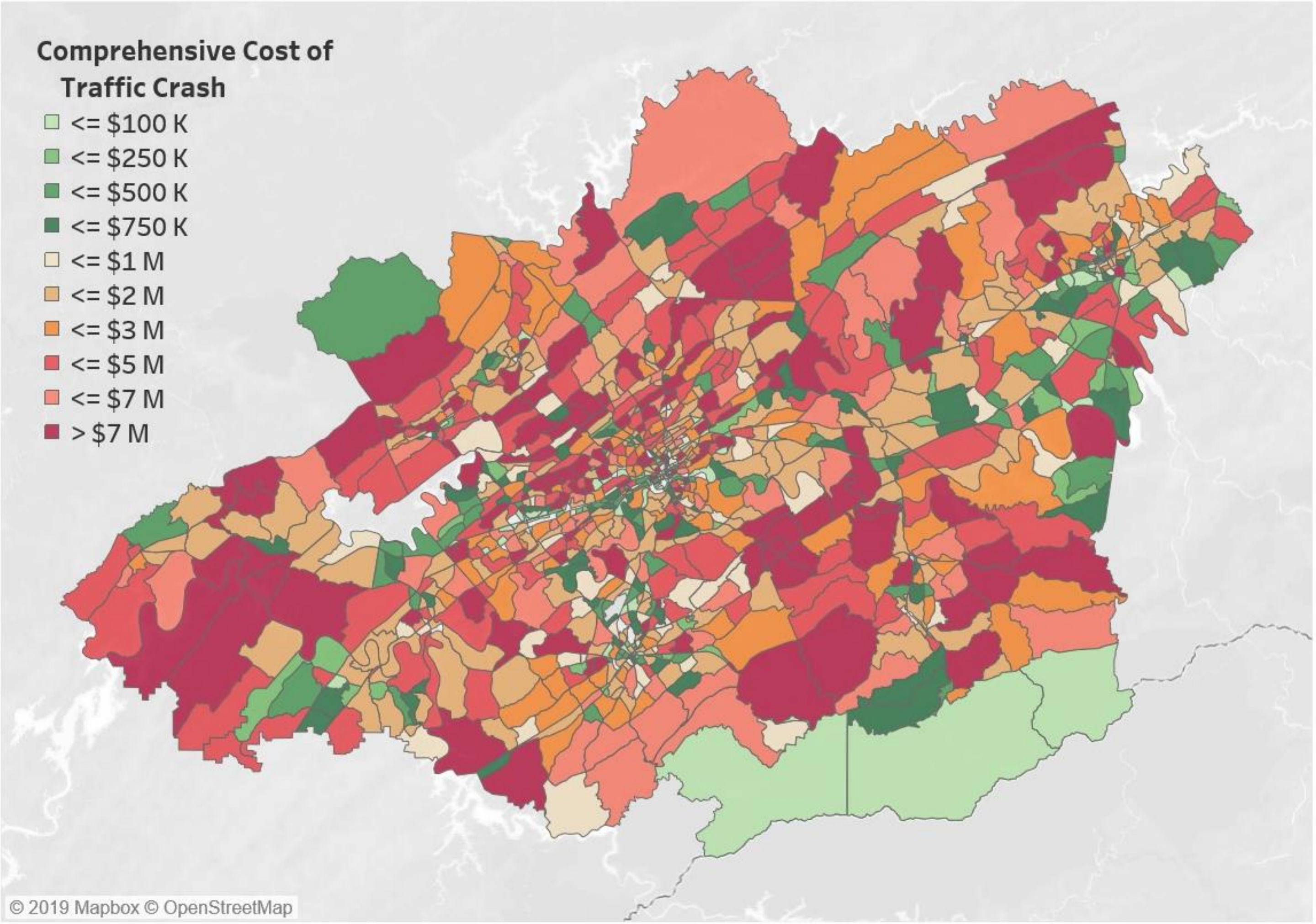

Figure 11 CCC distribution in KRTM; average of 2014-16

Figure 12 and Figure 13 presents the spatial distribution of the proportion of the economic and comprehensive cost of traffic crashes per capita to individual's income. Average ECCPC per income and CCCPC per income are respectively 6.5% (Q1: 2.7%, Median: 6.5%, Q3: 13.1%) and 14.9% (Q1: 5.7%, Median: 14.8%, Q3: 39.1%). The gray color in the maps exhibits TAZs, in which the proportion of the economic cost of traffic crashes to families' income is less than average. The warmer colors point out areas in which 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 in the northern part of the Knoxville and near multilane highways that connect major cities in the KRTM area (e.g., Knoxville to Maryville, Knoxville to Sevierville); on the other hand, in areas near I-40 (major corridor of the study area), the burden of traffic crashes are lower. One explanation for more tangible crash burden along the road network is the exposure of the residents to high volume corridors with high accessibility and high traffic speeds. These factors are known to contribute to both crash frequency and severity. Moreover, households who live very close to these corridors have lower household incomes. This is also supported by literature that individuals with lower socioeconomic status (e.g., income) experience higher residential exposure to traffic and traffic exposure and traffic-related pollution than non-minorities and persons of higher socioeconomic status.(Pollution 2010; Apelberg, Buckley, and White 2005; Gunier et al. 2003; Parker et al. 2012; Woghiren-Akinnifesi 2013)

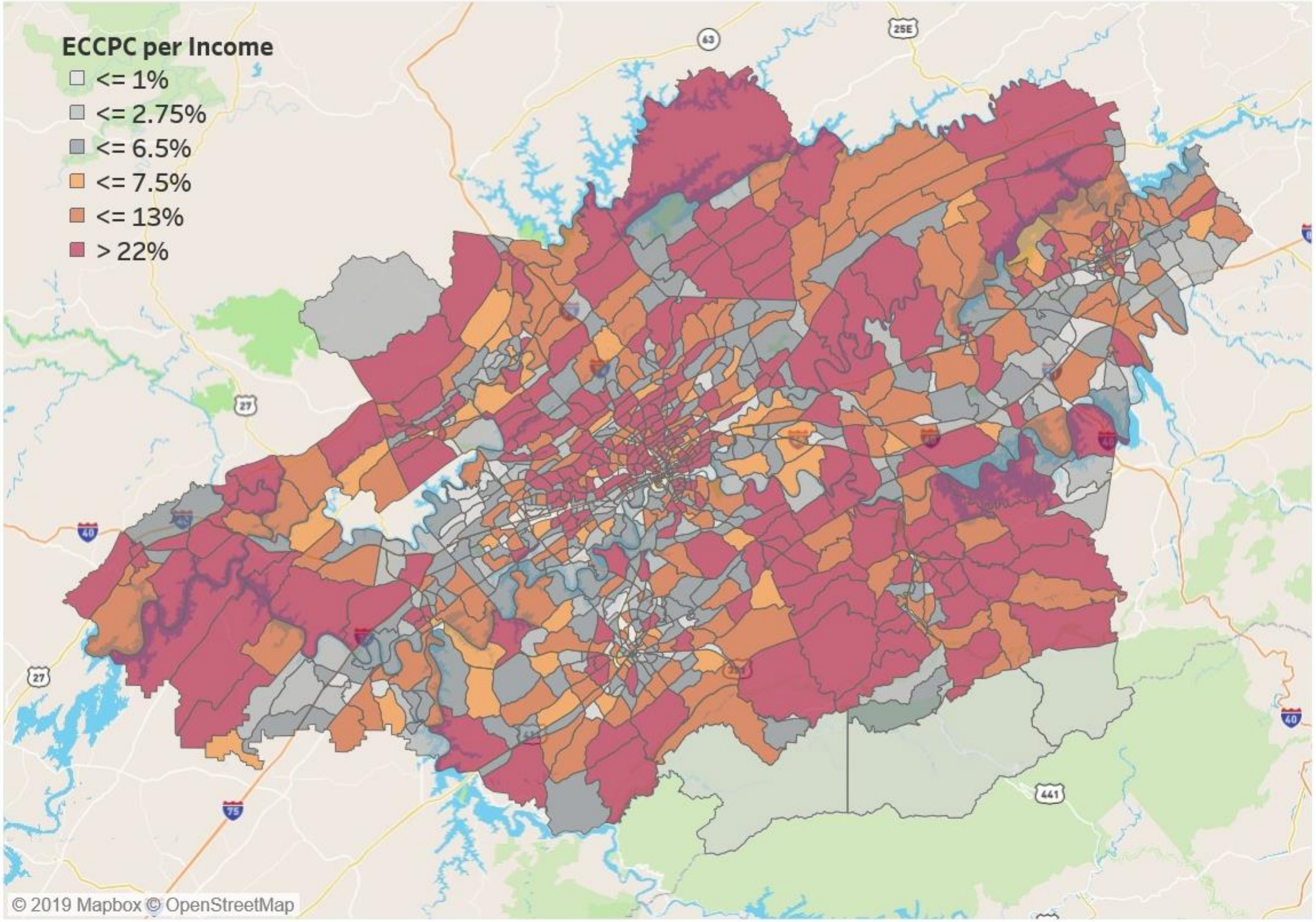

Figure 12 Proportion of The Economic Cost of Traffic Crashes per Capita to Median Families' Income

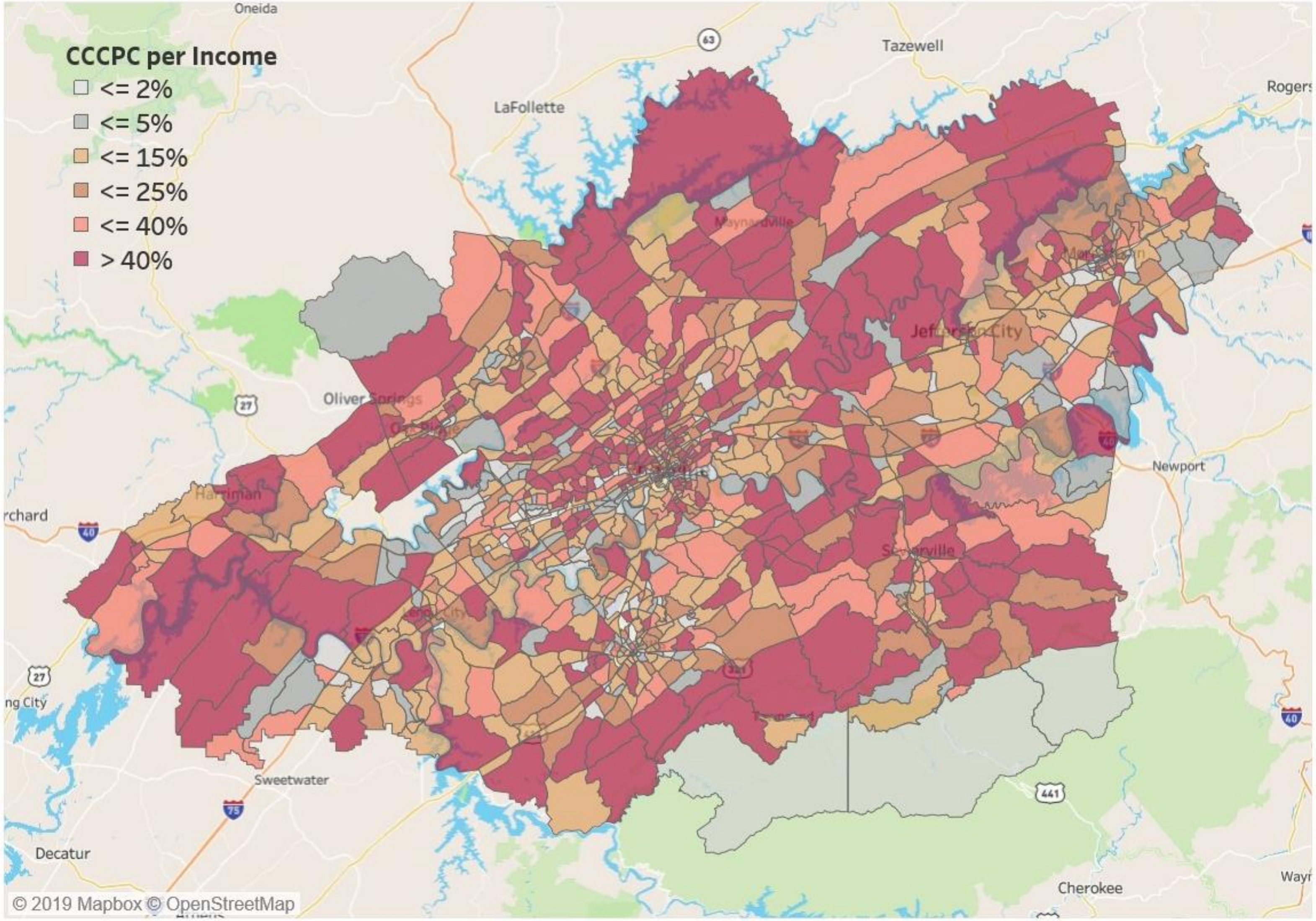

Figure 13 Proportion of The Comprehensive Cost of Traffic Crashes per Capita to Median Families' Income

Figure 14 presents the distribution of crash cost per capita and crash cost per capita per income over PMT. As PMT increases, the burden of traffic crashes increases. For example, TAZs with PMT higher than 50 have a substantially higher ECCPC/CCCPC and ECCPC/CCCPC per income compared to those below 50. This trend also holds for distribution over income.

Visual inspection of **Figure 15** indicates that TAZs with median household below $25,000 have a substantially higher burden of traffic crashes compared to wealthier families. For example, TAZs with a median household income of less than $25,000, the average ECCPC is equal to $2,200 which is 3 times higher than TAZs with a 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 $25,000 is 8 times higher (9% vs. 1%) than TAZs with a median household income of more than $100,000.

The value of the Gini coefficient for the ECC and CCC are respectively 0.51 and 0.58. Furthermore, the value of the Gini coefficient for the ECCPC and CCCPC are respectively 0.43 and 0.54.

Figure 14 Distribution of the ECCPC & ECCPC per income with regards to PMT

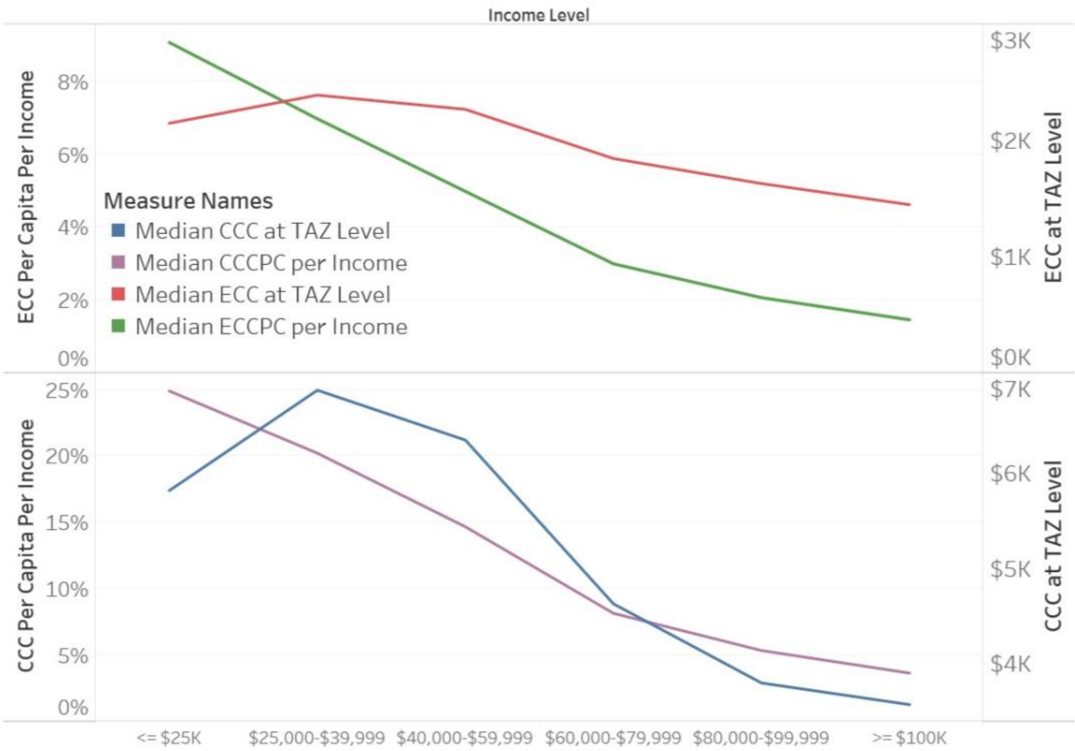

Figure 15 Distribution of the ECCPC & ECCPC per income with regards to median household income

Parameters estimation and discussion

In both estimated ECC and CCC models (presented in Table 4), we used the TAZ population as the offset variable to predict the Economic/comprehensive cost of traffic crashes at the zonal level. The significant value of dispersion parameters justified the use of negative binomial mode over the Poisson model. Significant variables in Table 4 are presented in bold. Comparison of the estimated models and their significance level indicate estimated coefficients in both models are largely consistent. Rest of the section discusses the study findings based on the CCC model output.

Analysis of elasticity indicates that none of the dependent variables have elastic effect. The average speed at the TAZ level has the highest negative impact on the burden of traffic crashes followed by intersection density. Alternatively, median household income, percentage of households with senior population, and number of workers per household had a positive impact on burden of traffic crashes. These findings are discussed in more details.

Table 4 Results of negative binomial model for predicting ECC and ECCPC at the zonal level

	ECC				CCC			
ec18average	**Coef.**	**Std. Err.**	**P>|z|**	**Elasticity**	**Coef.**	**Std. Err.**	**P>|z|**	**Elasticity**
Average zonal activity	3.41E-04	4.26E-05	0.000	0.046	2.59E-04	4.66E-05	0.000	0.035
Average Speed (MPH)	0.009	0.003	0.005	0.345	0.010	0.004	0.015	0.376
Income ($10,000)	-0.059	0.011	0.000	-0.272	-0.068	0.015	0.000	-0.314
Percentage of households with senior population	-1.183	0.236	0.000	-0.333	-1.023	0.304	0.001	-0.288
Worker Per Household	-0.299	0.113	0.008	-0.358	-0.316	0.146	0.031	-0.379
Student per Household	0.118	0.135	0.382	0.045	0.041	0.170	0.808	0.016
Intersection Density (per square miles)	4.15E-04	1.38E-04	0.003	0.072	6.63E-04	1.82E-04	0.000	0.115
Percent road with Sidewalk	-0.132	0.086	0.124	-0.031	-0.346	0.112	0.002	-0.082
Percent Near Bus Station	-0.032	0.070	0.648	-0.006	-0.117	0.091	0.196	-0.023
Population Density (per Square miles)	-8.82E-05	7.74E-06	0.000	-0.116	-9.47E-05	1.01E-05	0.000	-0.125
VMT Interstate	-2.63E-06	7.06E-07	0.000	-0.025	-2.83E-06	8.95E-07	0.002	-0.027
Daily VMT Arterial	6.98E-06	1.40E-06	0.000	0.078	4.24E-06	1.72E-06	0.014	0.047
Daily VMT Other road classes	-3.07E-06	2.65E-06	0.246	-0.022	2.58E-07	3.38E-06	0.939	0.002
Constant	7.620	0.218	0.000		8.711	0.280	0.000	
ln(Population)	1	(exposure)			1	(exposure)		
alpha	0.424	0.017			0.681	0.027		
Log likelihood (null)	-15060.41				-16159.56			
Log likelihood (model)	-14842.35				-16039			
LR $\chi^2(13)$	241.11				241.11			
Df	15				15			
AIC	29714.7				32108			

Positive sign of the PMT suggests that residents of those zones that originated trips are made more frequently or for longer instances are more prone to traffic crashes, and traffic crashes have a greater monetary impact on them. This is also supported by visual inspection of Figure 8. Population density has a negative effect on cost of traffic crashes. It is well established that the crash frequency in urban areas is higher than rural areas on average; whereas the crash severity is relatively lower (Marshall and Ferenchak 2017; Zwerling et al. 2005); hence, the average cost of traffic crashes in the urban areas is lower than rural areas. Moreover, it is expected in the urban areas with higher population density, road users travel shorter distances which may yield to lower trip length (Pucher and Renne 2005). The negative sign of the income variable and crash cost is in agreement with road safety literature (Marshall and Ferenchak 2017; World Health Organization 2015). Furthermore, it is more likely that individuals with lower income use less safe vehicles with fewer safety features (Girasek and Taylor 2010); therefore, their crash severity and eventually the cost of their traffic crashes increases.

Unlike the number of students per household, number of workers per household have a significant association with the dependent variables. This difference may reflect the difference in travel behavior of these groups. It should be noted that the majority of the students use alternative modes of transportation such as school bus, carpooling, or usually depending on adults to commute to school or other places. The proportion of individuals over 60-years-old has a significant negative association with economic cost of traffic crashes. Although, one may expect the senior population due to their vulnerability will suffer from higher injury severity (Yee, Cameron, and Bailey 2006); conversely, the senior population has lower trip rates compared to other groups (Williams and Carsten 1989, Massie et al. 1995, KRTPO 2008). Thus, in this study, percentage of household with seniors reduces the economic impact of traffic crashes compared to other age cohorts.

As expected, road network characteristics have a significant association with safety level. Percent of roads with sidewalk has a significant negative association with the crash cost. This was unlike Cai, Abdel-Aty, and Lee (2017) that reported sidewalk length has a positive association with crash frequency, severe crash, and non-motorized crash frequency.

It is well established that modal shift from private vehicle to other transportation modes (in particular public transit) has a positive effect on road safety (Tiwari, Jain, and Rao 2016; Schepers and Heinen 2013; Elvik et al. 2009). However, in this study, percent of roads near bus stop does not have a significant association with cost of traffic crashes. This could be explained by very low demand for public transit in the study area and higher automobile dependency in Knoxville Metropolitan area.

Speed is known as a contributing factor to both crash frequency and crash severity (Elvik et al. 2009; Highway Safety Manual 2010). As it was expected, the average speed of roads in a TAZ has a positive association with the crash cost. VMT on interstate road has a negative association with both comprehensive crash cost and economic crash cost models. Although this finding may sound counterintuitive at first glance, it should be noted that 85% (900 TAZs) of the TAZs do not have access to this road classification function. Moreover, TAZs with interstate access has a lower population compared to the others; therefore, the overall cost of traffic crashes was lower. Furthermore, due to the design nature of the interstates, interstates have lower access compared to arterial and local streets, and consequently local traffic is not mixed with higher speed traffics. VMT of the arterial roads, on the other hand, has a positive association with crash cost, which could be explained by relatively higher speed as well as their higher accessibility which increase number of traffic conflicts, likelihood of traffic crashes, and severity of injuries. Unlike two previous classes of the roads, other road classification does not have a significant association with crash cost.

Conclusion

The main aim of this study was to explore the factor influencing the burden of traffic crashes at a fine geographic level as well as highlighting the equality challenges associated with disparities in burden of traffic crashes. 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 measured Person Miles Traveled (PMT) to measure the average trip rate weighted by trip length originated at each TAZ. By controlling the traffic crash burden over the PMT, we learned that the burden of traffic crashes is higher for those TAZs which have a higher PMT or their residents have a lower income.

Additionally, the value of the Gini index indicated that the burden of traffic crashes is not equally distributed across the population. Establishing progressive policies concerning educational outreach and additional funding resources to ease the burden of traffic crashes for neighborhoods that are more prone to traffic crashes particularly neighborhoods consisting of disadvantaged groups (i.e., lower-income families) may alleviate the burden of traffic crashes.

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 multilane highways where TAZs' residents are more prone to high-speed traffic and higher road classification. A transportation network designs that diverge high-speed traffic from residential areas or managing the accessibility of the residents near the high-speed, high volume roads could be used to improve the safety of residents of these areas. Moreover, eliminating a portion of trips by promoting sustainable transport and targeting PMT could reduce the residents' exposure to traffic. Likewise, an increase in diversity and mixed land-use design would also reduce both trips rate, and trip length (Cervero and Kockelman 1997) and eventually PMT. Reduction in PMT and subsequently, VMT has a direct impact on both economic and comprehensive cost of traffic crashes.

In summary, in this study we introduced a method to measure the cost of traffic crashes at the zonal level, which could be straightforwardly integrated to travel demand analysis. 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. 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.

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis & Prevention 41 (1), 153-159.

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.

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.

Bureau of Economic Analysis, 2018. Per capita real gdp by state (chained 2009 dollars).

Cai, Q., Abdel-Aty, M., Lee, J., Eluru, N., 2017. 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.

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.

García-Altés, A., Pérez, K., 2007. The economic cost of road traffic crashes in an urban setting. Injury prevention 13 (1), 65-68.

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.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., 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.

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.

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

Kanafani, A., 1983. Transportation demand analysis.

Krtm, 2012. Knoxville regional travel model update 2012 model development and validation report. Bernardin, Lochmueller & Associates, Inc., Evansville, IN, USA.

Krtpo, 2008. 2008 east tennessee household travel survey final report. Knoxville Regional Transportation Planning Organization. Knoxville Regional Transportation Planning Organization, 206 Wild Basin Rd., Suite A-300 Austin, Texas 78746.

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.

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.

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.

Massie, D.L., Campbell, K.L., Williams, A.F., 1995. Traffic accident involvement rates by driver age and gender. Accident Analysis & Prevention 27 (1), 73-87.

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.

Naderan, A., Shahi, J., 2010. Aggregate crash prediction models: Introducing crash generation concept. Accident Analysis & Prevention 42 (1), 339-346.

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., 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.

Washington, S.P., Karlaftis, M.G., Mannering, F., 2010. Statistical and econometric methods for transportation data analysis CRC press.

Wegman, F., Oppe, S., 2010. Benchmarking road safety performances of countries. Safety science 48 (9), 1203-1211.

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.

Williams, A.F., Carsten, O., 1989. Driver age and crash involvement. American Journal of Public Health 79 (3), 326-327.

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

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.

Yee, W., Cameron, P.A., Bailey, M., 2006. Road traffic injuries in the elderly. Emergency Medicine Journal 23 (1), 42-46.

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.

HBA Application 3: Factors influencing cost of traffic crash at the traffic analysis zone level: incorporating spatial effects

Amin Mohamadi Hezaveh
University of Tennessee, Knoxville
311 JD Tickle Building, TN 37996
Phone: (385) 259-5148; Email: amohamad@vols.utk.edu

Christopher R. Cherry
University of Tennessee, Knoxville
321 JD Tickle Building, TN 37996
Phone: (865) 974-7710
Fax: (865) 974-2669; Email: cherry@utk.edu

Abstract

Road safety literature provides abundant examples of studies that measure the economic cost of traffic crashes at the 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 spatial autoregressive (SAR) model with a queen contiguity weights matrix was more suitable compared to 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 and traffic exposure have a significant positive association with ECCPC. The ECCPC could be used as an index for allocating proper countermeasures and interventions to groups and areas where the burden of traffic crashes is more tangible. This could be done by investment in the safer infrastructure and educational interventions.

Keywords: Economic Cost of Traffic Crashes; Home-Based Approach; Home-Address; Equity

Introduction

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 (WHO 2015; Jacobs, Aeron-Thomas, and Astrop 2000; Wijnen and Stipdonk 2016). The relative magnitude of this externality is larger in low and middle-income countries compared to the developed countries. Road safety literature has abundant examples of estimating cost of traffic crashes at coarse geographic level (i.e., country-level) (Wegman and Oppe 2010; García-Altés and Pérez 2007; Mohan 2002; Blincoe et al. 2015; Ahadi and Razi-Ardakani 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. 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 (Blincoe et al. 2015; Mohan 2002; García-Altés and Pérez 2007; Ahadi and Razi-Ardakani 2015; Harmon, Bahar, and Gross 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 population respectively) (World Health Organization 2015). 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 (Marshall and Ferenchak 2017; Romano, Tippetts, and Voas 2006; Clark 2003). In rural areas, the fatality rate tends to be several times higher than in urban areas (Marshall and Ferenchak 2017; Blatt and Furman 1998). 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 and attribute 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 the traffic victims' home-addresses to obtain information regarding their sociodemographic in road safety is not a new effort, one should consider that the majority of these studies used fatally injured road users (Blatt and Furman 1998; Males 2009; Romano, Tippetts, and Voas 2006; Stamatiadis and Puccini 2000), used course resolution such as zip code (Romano, Tippetts, and Voas 2006; Lee et al. 2015), census-level (Stamatiadis and Puccini 2000), or focused on a specific group of road users (Lee et al. 2015). 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, and travel behavior (e.g., Gomes, Cunto, and da Silva 2017; Hadayeghi, Shalaby, and Persaud 2003; Hadayeghi, Shalaby, and Persaud 2010b; Lee et al. 2015; Naderan and Shahi 2010; Pirdavani et al. 2012b; Quddus 2008). By using a wide range of safety outcomes, researchers explored the association between geographic unit characteristic and number of all traffic crashes (Naderan and Shahi 2010; Pirdavani et al. 2012b; Pirdavani, Brijs, Bellemans, and Wets 2013; Huang et al. 2016; Miaou, Song, and Mallick 2003; Cai et al. 2017; Hezaveh and Cherry 2018), number of property damage only crashes (Naderan and Shahi 2010; Aguero-Valverde 2013), frequency of injury/severe crashes (Xu and Huang 2015; Aguero-Valverde 2013; Cai et al. 2017), or crash frequency of specific road users (e.g., non-motorized, bicyclists) (Cai et al. 2017; Cheng et al. 2018; Saha et al. 2018; Lee, Abdel-Aty, and Jiang 2015) 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, Shalaby, and Persaud 2010b; Li et al. 2013; Rhee et al. 2016; Pirdavani et al. 2012b, 2012a; Pirdavani, Brijs, Bellemans, and Wets 2013; 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, Abdel-Aty, and Choi 2012), road length with different functional classification (Hadayeghi, Shalaby, and Persaud 2010b; Quddus 2008), or population has been used (Gomes, Cunto, and da Silva 2017). In case of measuring the economic cost of traffic crashes based on the home addresses of the traffic victims at the zonal level, using VMT may not reflect the road users' exposures properly. One way to deal with this issue is to use population as a proxy for the exposure variable (Gomes, Cunto, and da Silva 2017; Lee et al. 2015). 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 (Dong et al. 2014; Dong, Huang, and Zheng 2015; Abdel-Aty et al. 2011; Naderan and Shahi 2010; Mohammadi, Shafabakhsh, and Naderan 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 economic cost of traffic crashes at a zonal level based on the home address of the road users, 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). Accordingly, we convert HBA crash frequency and crash severity to measure the economic cost of traffic crashes at five 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 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. We used 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). 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 also includes 1,186 TAZs and includes sociodemographic, economic, and travel information of the residents. Table 1 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 4 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

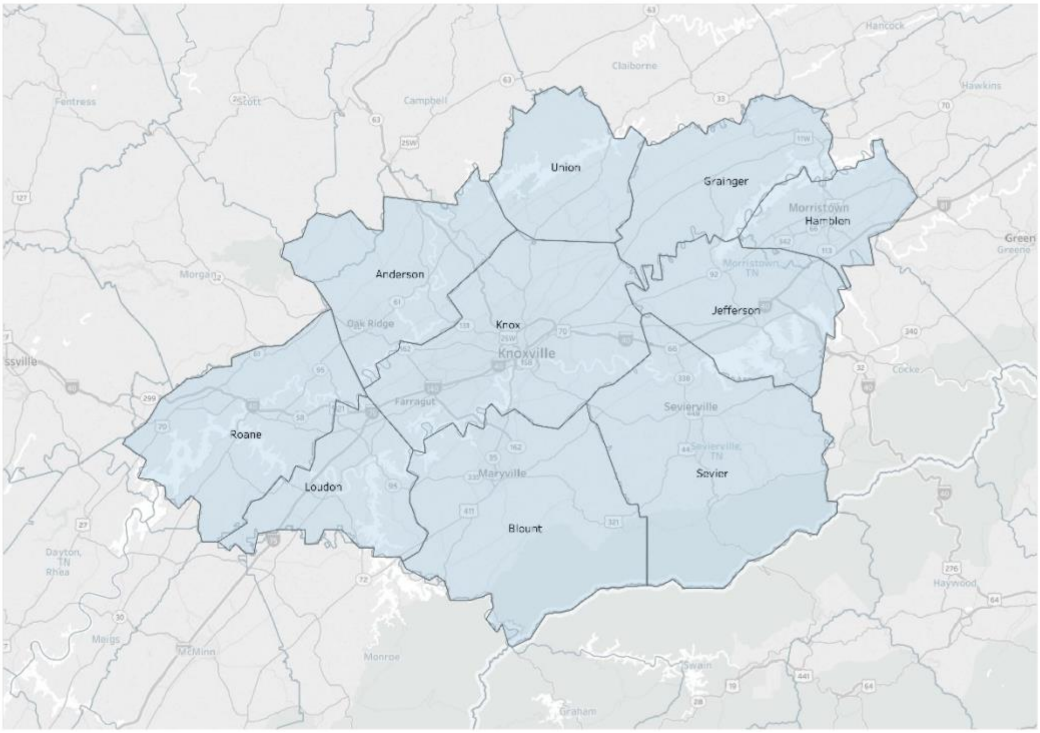

Figure 16 Knoxville Regional Travel Demand Model Extent

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:

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

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.

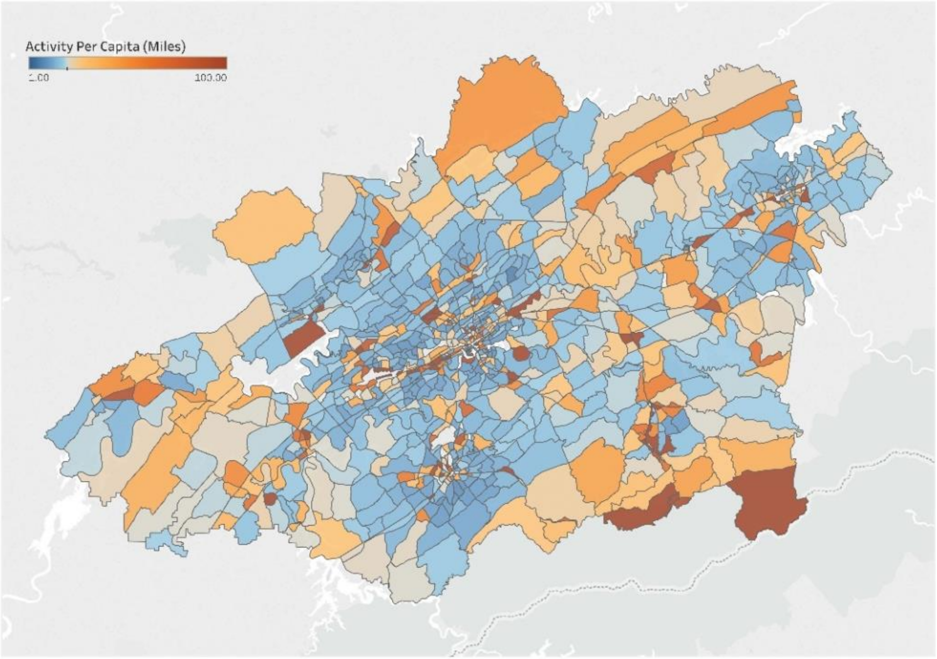

Figure 17 Average zonal activity (person miles traveled)

Home-Based Approach definition

Home address of the road users who were involved in a traffic crash is one of the data elements that a 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) but is not often used in the road safety analysis. We use the home-address of individuals involved in crashes, and reported in the crash database, 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 by severity 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. We use crash frequency and crash severity to calculate the economic cost of traffic crashes at each zone.

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), coordinates of the crashes, and addresses of the individual who were involved in traffic crashes. 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 (Hezaveh and Cherry 2019b). 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 5** recommended by FHWA (Harmon, Bahar, and Gross 2018) for the year 2010 for the person-injury unit. We converted the injury cost to 2017 dollar by the inflation rate (Harmon, Bahar, and Gross 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, Bahar, and Gross 2018). By using numbers presented in **Table 5** 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:

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

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 5** 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 18 presents the distribution of the ECCPC at zonal level in the study area. ECCPC tended to have distributed impacts with high economic cost scattered throughout the region.

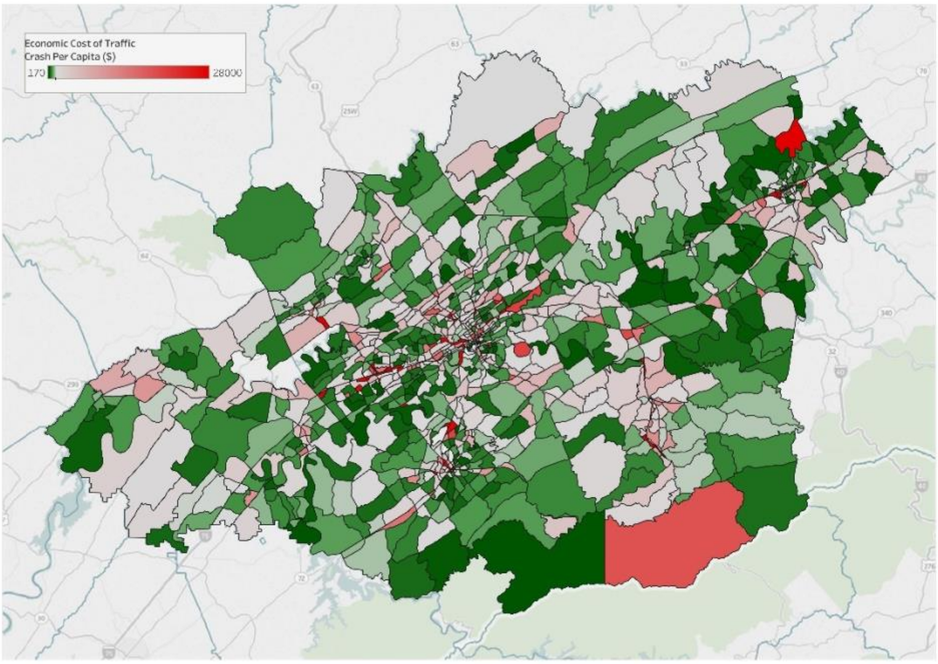

Figure 18 ECCPC distribution in KRTM

Table 5 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

Modeling approach

Testing spatial dependency

Visual inspection of Figure 18 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)[1] and spatial autoregressive model (SAR) are two common models that are used by researchers to consider spatial autocorrelation in the road safety analysis (Lee et al. 2015; Rhee et al. 2016; Quddus 2008). 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:

$$y = X\beta + \varepsilon \qquad \text{Equation 12}$$

$$\varepsilon = \lambda W_\varepsilon + u = (I - \lambda W)^{-1}u \qquad \text{Equation 13}$$

$$y = \lambda W_y + X\beta + \lambda WX\beta + u \qquad \text{Equation 14}$$

[1] Not to be mistaken by Structural Equation Modeling

where y is a vector of ECCPC, X is a vector of independent variables presented in Table 1, 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:

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

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, Phipps, and Anselin 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 were 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 corrected Akaike information criterion –AICc (Hurvich and Tsai 1989); the weight matrix with the lowest AICc is preferred (A. Fotheringham and Brunsdon ; Nakaya 2014; Nakaya et al. 2005; Hadayeghi, Shalaby, and Persaud 2010b). 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 Florax 1995; Anselin and Rey 1991; 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 6** 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 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, and the 90th percentile spans 0.9 to 20.5%. **Figure 19** presents the distribution of ECCPC, ECCPC per income over 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 20** 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 a median household income of less than $15,000, the average ECCPC is equal to $1,500 which is 3 times higher than TAZs with a 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 a median household income of more than $100,000.

Table 6 Economic cost of traffic crashes by driver and resident types (2017 million dollars)

Person Involved Residency	Driver Type 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**

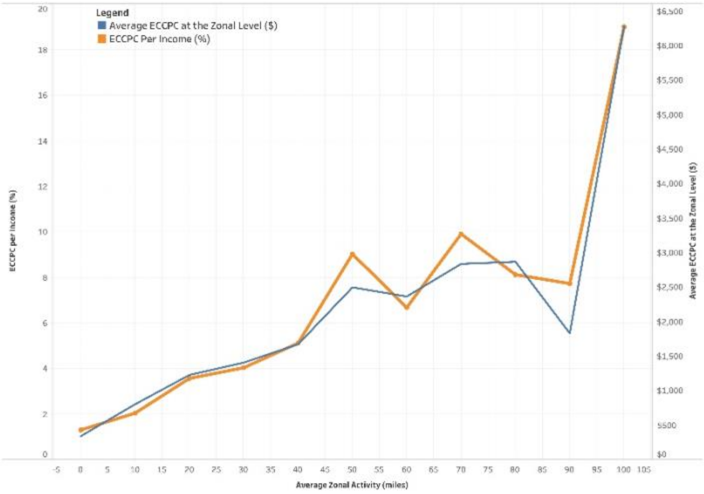

Figure 19 Distribution of the ECCPC & ECCPC per income with regards to average zonal activity

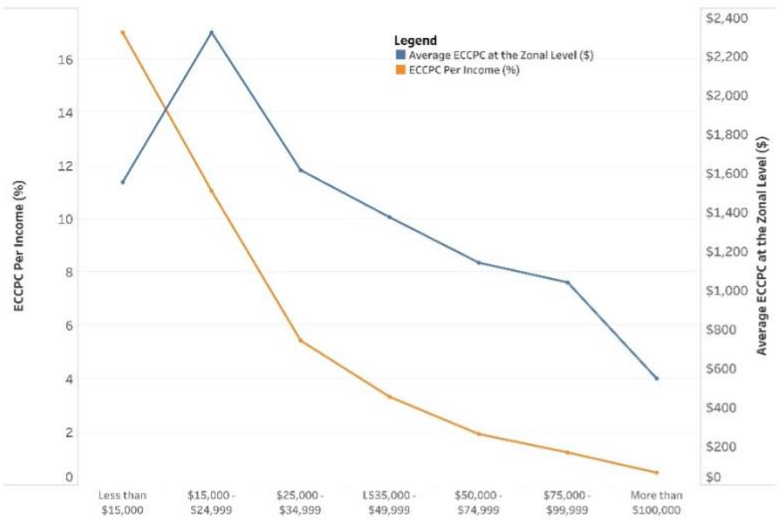

Figure 20 Distribution of the ECCPC & ECCPC per income with regards to median household income

Figure 21 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 points 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.

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 indicates 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.

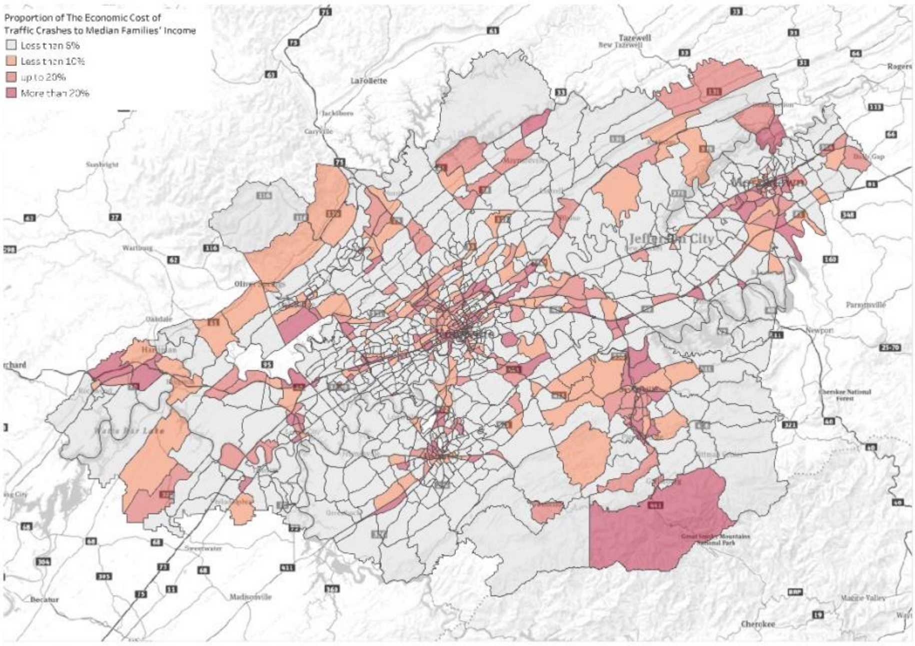

Figure 21 Proportion of The Economic Cost of Traffic Crashes to Median Families' Income

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 7** presents the result of the estimated models.

Table 7 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			0.39		
Number of Observations	956			956			956		

Parameters estimation and discussion

All the variables presented in Table 7 (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. (Cheng et al. 2018; Lee, Abdel-Aty, and Jiang 2015; Pirdavani et al. 2012b, 2012a; Pirdavani, Brijs, Bellemans, Kochan, et al. 2013; Pirdavani, Brijs, Bellemans, and Wets 2013) 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, Shalaby, and Persaud 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, Brijs, Bellemans, Kochan, et al. (2013) 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, Shalaby, and Persaud (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 the Naderan and Shahi (2010) study where they reported 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 (Cai et al. 2017; Siddiqui, Abdel-Aty, and Choi 2012); 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 (Cai, Abdel-Aty, and Lee 2017; Cai et al. 2017; Pirdavani et al. 2012b; Pirdavani, Brijs, Bellemans, and Wets 2013; Gomes, Cunto, and da Silva 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 (World Health Organization 2015; 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. (2017) 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 (Ladron de Guevara, Washington, and Oh 2004; Pirdavani et al. 2012a; Hadayeghi, Shalaby, and Persaud 2003; Lovegrove and Sayed 2006; Abdel-Aty et al. 2011; Gomes, Cunto, and da Silva 2017). It is well-established that speed is a contributing factor to both crash frequency and crash severity (Highway Safety Manual 2010; Elvik et al. 2009). The average speed of roads in a TAZ has a positive association with ECCPC agreeing with previous research (Pirdavani et al. 2012a; Abdel-Aty et al. 2011; Hadayeghi, Shalaby, and Persaud 2003),

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 over the average zonal activity, we learned that the burden of traffic crashes is higher for those who travel more or have a lower income. As a result, these groups require further attention in the transportation design process or in case of allocating funding to ease the burden of traffic crashes.

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 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.

Geographic distribution of the negative externalities of the traffic crashes shows 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. doi:https://doi.org/10.1016/j.aap.2013.06.014

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* (pp. 21-74): Springer.

Anselin, L., & Rey, S. (1991). Properties of tests for spatial dependence in linear regression models. *Geographical analysis, 23*(2), 112-131.

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. (2017). 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. (2017). 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.

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. Retrieved from

FHWA. (2017). KABCO Injury Classification Scale and Definitions. Retrieved from https://safety.fhwa.dot.gov/hsip/spm/conversion_tbl/pdfs/kabco_ctable_by_state.pdf

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*. Retrieved from 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. (2018). *A New Approach to Aggregate Crash Prediction Model: Home-Based Crash Frequency Modeling.* Paper presented at the Southern District ITE Annual Meeting.

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

Highway Safety Manual. (2010). Highway Safety Manual. American Association of State Highway and Transportation Officials, 2, 10-11.

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*. doi:https://doi.org/10.1016/j.aap.2018.05.003

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. Retrieved from 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.

Lee, J., Abdel-Aty, M., & Jiang, X. (2015). 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. (2002). *Social cost of road traffic crashes in India.* Paper presented at the Proceedings First Safe Community Conference on Cost of Injury. Viborg, Denmark.

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. (2012b). Developing zonal crash prediction models with a focus on application of different exposure measures.

Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., & Wets, G. (2013). 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. (2013). 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. (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*. Retrieved from

Xu, P., & Huang, H. (2015). Modeling crash spatial heterogeneity: random parameter versus geographically weighting. *Accident Analysis & Prevention, 75*, 16-25.

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.

HBA Application 4: A Statewide Geographically Weighted Regression to Estimate the Comprehensive Cost of Traffic Crashes at a Zonal Level

Amin Mohamadi Hezaveh
University of Tennessee, Knoxville
311 JD Tickle Building, TN 37996
Phone: (385) 259-5148; Email: amohamad@vols.utk.edu

Christopher R. Cherry
University of Tennessee, Knoxville
321 JD Tickle Building, TN 37996
Phone: (865) 974-7710
Fax: (865) 974-2669; Email: cherry@utk.edu

Abstract

Global road safety records demonstrate spatial variation of comprehensive cost of traffic crashes between countries. To the best of our knowledge, no study has explored the variation of this matter at a local geographical level. This study proposes a method to estimate the comprehensive crash cost at the zonal level by using person-injury cost. The current metric of road safety attributes safety to the location of the crash which makes it challenging to assign the crash cost to home-location of the individuals who were involved in traffic crashes. To overcome this limitation, we defined Home-Based Approach crash frequency as the expected number of crashes by severity that road users who live in a certain geographic area have during a specified period. Using crash data from Tennessee, we assign those involved in traffic crashes to the census tract corresponding to their home address. The average Comprehensive Crash Cost at the Zonal Level (CCCAZ) for the period of the study was $18.2 million (2018 dollars). Poisson and Geographically Weighted Poisson Regression (GWPR) models were used to analyzing the data. The GWPR model was more suitable compared to the global model to capture the spatial heterogeneity. Findings indicate population of people over 60-years-old, the proportion of residents that use non-motorized transportation, household income, population density, household size, and metropolitan indicator have a negative association with CCCAZ. Alternatively, VMT, vehicle per capita, percent educated over 25-year-old, population under 16-year-old, and proportion of non-white races and individuals who use a motorcycle as their commute mode have a positive association with CCCAZ. Findings are discussed in line with road safety literature.

Keywords: Comprehensive Crash Cost; Home-Based Approach; Geographically Weighted Regression; Road Safety.

Introduction

Annually more than 1.2 million individuals lose their lives on roads globally. Likewise, between 20 to 50 million individuals are impacted by serious and sometimes permanent injuries in traffic crashes. World Health Organization reports on road safety indicate road fatality rate (death per 100,000 population) varies across countries (World Health Organization 2015). The traffic fatality rate is approximately two times higher in low and middle-income countries compared to high-income countries (21.5, 19.5, and 10.3 per 100,000 respectively). Spatial variation in road safety performance indicators and social costs of traffic crashes on roads could be attributed to several external factors: namely safety measures and programs, traffic structure, and culture of a country (Koornstra, Lynam, and Nilsson 2002). These factors also reflect study area characteristics such as demographics, weather, reporting practices, and the economy (Koornstra, Lynam, and Nilsson 2002).

Social cost or Comprehensive crash cost of traffic crashes consists of several components including medical cost, loss of production capacity, costs of property damage, administrative costs, and economic valuation of lost quality of life (Elvik et al. 2009; Harmon, Bahar, and Gross 2018). These components could be categorized into two main categories; tangible and intangible cost. Tangible costs reports are the economic costs of traffic crashes that can be directly measured such as medical bills and lost wages. The intangible costs comprise the other impacts of crashes and can be monetized as quality-adjusted life years (QALY) (Harmon, Bahar, and Gross 2018).

Road safety literature provides several instances of studies regarding the comprehensive crash cost at country level (e.g., Wegman and Oppe 2010; García-Altés and Pérez 2007; Mohan 2002; Blincoe et al. 2015; Ahadi and Razi-Ardakani 2015). Findings indicate that traffic crashes cost 1-2% of Gross Domestic Product (GDP) of high-income countries and 3% of GDP in low and middle-income countries. (WHO 2015; Jacobs, Aeron-Thomas, and Astrop 2000). In the United States, societal harm from traffic crashes in 2010 was estimated to be over $836 billion. The economic cost of traffic crashes in the US was estimated to be over $242 billion which is equal to 1.6% of the US GDP (Blincoe et al. 2015). This trend also holds within a country; for example, several studies in the United States showed that in rural areas the fatality rate is several times higher than the majority of urban areas (Marshall and Ferenchak 2017). In addition, some ethnicities such as Hispanic, African-American, and Native American are more prone to traffic crashes –i.e., they have higher crash rates (Mayrose and Jehle 2002; Braver 2003; Campos-Outcalt et al. 2003; McAndrews et al. 2013); this is also the case for the fatality rate (Schiff and Becker 1996; Baker et al. 1998; Harper et al. 2000). Furthermore, vulnerable road users (i.e., pedestrians and bicyclists) and lower-income neighborhoods have higher fatality rates compared to motorized road users and wealthier neighborhoods (Marshall and Ferenchak 2017).

To the best of our knowledge, the road safety literature has abundant examples of estimating the societal outcome of traffic crashes at aggregate level (i.e., country); however, there are no studies that investigated comprehensive crash cost at a fine geographical level (e.g., traffic analysis zone, census tract) and explores the factors correlating with it by using a Macroscopic Crash Prediction Models (MCPM). Furthermore, we may expect that traffic crashes do not impact geographic areas in equitable ways. Therefore, we expect the comprehensive cost of traffic crashes has spatial variation within a country, city, or finer geographic unit. In addition, learning about the areas where their residents are more prone to the burden of traffic crashes would help safety practitioners and researchers to allocate proper countermeasures to reduce the burden of traffic crashes or providing resources to reduce the burden of traffic crashes.

Macroscopic Crash Prediction Models

In order to study the societal cost of traffic crashes, we will measure societal cost at the zonal level by using MCPM. MCPM is a set of methods that are used to investigate the relationship between safety at zonal level and socioeconomic factors (Huang and Abdel-Aty 2010; Aguero-Valverde and Jovanis 2006;

Hadayeghi, Shalaby, and Persaud 2010b; Pulugurtha, Duddu, and Kotagiri 2013), travel behavior (Naderan and Shahi 2010), road infrastructure and traffic flow (Huang and Abdel-Aty 2010; Abdel-Aty et al. 2011; Quddus 2008; Xu and Huang 2015; Pirdavani et al. 2012b), and environment condition (Aguero-Valverde and Jovanis 2006). Various type of spatial units have been used by researchers; from fine level such as traffic analysis zones (Hadayeghi, Shalaby, and Persaud 2010b; Pirdavani, Brijs, Bellemans, and Wets 2013; Dong et al. 2016; Pirdavani et al. 2012b; Xu and Huang 2015; Gomes, Cunto, and da Silva 2017; Pulugurtha, Duddu, and Kotagiri 2013), census tracts (Ukkusuri, Hasan, and Aziz 2011; Wang and Kockelman 2013; Hezaveh and Cherry 2019a), and block groups (Levine, Kim, and Nitz 1995) to coarser levels such as zip codes (Girasek and Taylor 2010), districts (Haynes et al. 2007), counties (Miaou, Song, and Mallick 2003; Aguero-Valverde and Jovanis 2006; Huang and Abdel-Aty 2010), and regions (S. Washington et al. 1999).

In MCPM, safety usually is measured with different indices, namely number of all traffic crashes (Naderan and Shahi 2010; Pirdavani et al. 2012b; Pirdavani, Brijs, Bellemans, and Wets 2013; Huang et al. 2016; Miaou, Song, and Mallick 2003; Cai et al. 2017; Hezaveh and Cherry 2019a), number of property damage only crashes (Naderan and Shahi 2010; Aguero-Valverde 2013), frequency of injury/severe crashes (Xu and Huang 2015; Aguero-Valverde 2013; Cai et al. 2017), and crashes of specific road users (e.g., non-motorized, bicyclists) (Cai et al. 2017; Cheng et al. 2018; Saha et al. 2018; Lee, Abdel-Aty, and Jiang 2015). Although using different dependent variables enable researchers to investigate the correlation of exogenous variables and traffic crash outcomes, it does not provide information about the association of exogenous variable and ultimate outcome – societal cost (i.e., comprehensive cost of traffic crashes) - of road safety. For example, the crash rates in an urban area are higher than rural area; but the fatal injury rate and crash injury rate in the urban area is lower than the rural area (Zwerling et al. 2005; NHTSA 2013).

Learning about the relationship between exogenous variables and comprehensive cost of traffic crashes would help safety practitioners and researchers prioritize their countermeasures based on the monetary values of traffic crashes regardless of road user type, the location of the crash (e.g., rural vs. urban), and crash severity. The current practices of assessing road safety relies on MCPM based on the location of the crash. This metric was best described by Hauer (1997, 24) "*the number of accidents (crashes) by kind and severity, expected to occur on the entity during a specified period.*" This definition attributes road safety to the location of the crashes rather than individuals who were involved in traffic crashes (i.e., pedestrians, bicyclists, motorcyclists, vehicle occupants, and drivers). As a result, it is challenging to attribute crash burden to the location where individuals reside by using this definition.

Another concern in MCPM models is spatial heterogeneity or spatial non-stationary (LeSage and Pace 2009; A. Fotheringham and Brunsdon). Spatial heterogeneity exists when exogenous variables do not vary identically across space (Xu et al. 2017). One reason for the presence of unobserved heterogeneity in the data is the presence of unknown or known factors that are unlikely to be available for the analysis (Mannering, Shankar, and Bhat 2016). This phenomenon influences the association among exogenous variables and dependent variables; as a result, this relationship may not be constant across the observation. Failing to consider the unobserved heterogeneity in count data analysis would lead to overdispersion; hence, the variance of the exogenous variable is larger than the mean (Gourieroux and Visser 1997; Cameron and Trivedi 1986). Likewise, if unobserved factors correlate with known exogenous factors, the estimates would yield biased parameters which eventually lead to drawing incorrect inferences (Mannering, Shankar, and Bhat 2016).

There are different methods to address the heterogeneity in count models. Random parameters count data, and geographically weighted poison regression (GWPR) are two common methods to address this issue (Xu and Huang 2015; Arvin, Kamrani, and J. 2019). Random parameters models are drawn from some random distribution and are assumed to vary randomly over observations (Xu and Huang 2015). One of the shortcomings of the random parameter model is that it usually fails to reflect the location of observation. Alternatively, spatial models consider the location of the observations to capture spatially structured variability in the effect of contributing factors (Xu et al. 2017; Xu and Huang 2015). Several studies showed the advantage of GWPR models with regards to improvement in model goodness of fit and capability to explore the spatially varying association among dependent variables and contributing factors (Hadayeghi, Shalaby, and Persaud 2010b; Xu and Huang 2015; Xu et al. 2017; Pirdavani et al. 2014). In addition, estimated parameters of the GWPR reflects local characteristics by enabling coefficients to vary across the study area; therefore, GWPR results could be used as a reference for transportation agencies focusing on geographic differences (Chiou, Jou, and Yang 2015).

Regarding the current state of practice in MCPM, this study has three aims. First, it aims to measure comprehensive crash cost at the zonal level by using the home address of the road users (i.e., Home-Based Approach). Secondly, we aim to display the geographical distribution of the comprehensive cost of traffic crash across the study area. Third, we will use a geographically weighted regression to account for spatial heterogeneity and explore the relationship between sociodemographic factors and comprehensive crash cost at the zonal level. The result of this study could help safety practitioners and researchers to identify the neighborhoods where their residents are more prone to the burden of traffic crashes and target them with proper countermeasures or interventions to reduce their risk.

Methodology

Data and Geocoding Process

To achieve the aims of this study, we need to attribute the safety outcome of traffic crashes to the origin of the individuals who were involved in traffic crashes. Therefore, we defined the HBA to measure *the expected number of crashes by severity that road users who live in a certain geographic area have during a specified period*. This definition attributes traffic crashes to individuals and their residential addresses rather than the location of traffic crashes. We use crash frequency by severity to calculate the comprehensive crash cost at each zone.

The data in this study was provided by the Tennessee Integrated Traffic Analysis Network (TITAN), a portal provided by Tennessee Highway Patrol (THP). The records include 694,276 crashes and information on 2,026,666 individuals who were involved in traffic crashes between 2014-2016. Each record includes information about road user type (e.g., driver, motorcyclist, passenger, pedestrian, bicyclist), coordinates of the crashes and addresses of all individuals who were involved in traffic crashes. It is also worthy to mention that the TITAN database does not identify at-fault road users in traffic crashes so we could not conduct any analysis on the role of each person in the traffic crash.

After obtaining the address of the pedestrians, bicyclists, motorcyclists, drivers, and vehicle occupants (n = 1,615,374), we used the Bing and Google 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 (Merlin et al. 2019; Hezaveh and Cherry 2019a, 2019b). After controlling for the address quality, 1,521,583 (94.1%) of the records met the minimum address quality filter. Of those, 1,358,117 had a Tennessee home-address (89.3% of geocoded addresses); the number out of state individuals was 163,466 (10.7% of geocoded addresses).

In this study, one goal was to investigate the relationship between sociodemographic variables and crash frequency at the zonal level. For that reason, we used the census tract as the geographic unit. Census data from the US survey in 2010 was also used to obtain sociodemographic data elements in each census tract in Tennessee.

Table **8** presents the descriptive statistics of the sociodemographic variables obtained from the US census in 2010. Figure 22 presents the histogram of HBA crash frequency by severity at the zonal level in this study.

Furthermore, we used highway performance monitoring system data for Tennessee in 2015 to obtain Average Annual Daily Traffic for each road segment and calculate total Vehicle Miles Travelled (VMT) at the census tract level.

Comprehensive Crash Cost at the Zonal Level

The injury severity in this 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 fatal, Incapacitating, Non-Incapacitating Evident, Possible Injury, and No Injury (FHWA 2017). The comprehensive crash cost consisted of two elements: economic person-injury unit costs and quality-adjusted life years (QALY), which respectively take account for tangible and intangible consequences of traffic crashes. In order to convert the injury severities to crash cost, we used number presented in **Table 9**, which are based on FHWA recommendation for the person-injury unit cost (Harmon, Bahar, and Gross 2018). It is also worthy to mention that in their report, Harmon, Bahar, and Gross (2018) considered injury misclassification that controlled for more accurate injury accounting at emergency departments. For more details about this issue please see Gomes, Cunto, and da Silva (2017). Furthermore, based on the Harmon, Bahar, and Gross (2018) study recommendation, we updated the person injury unit cost to reflect 2018 US Dollars. For more details, please see Harmon, Bahar, and Gross (2018).

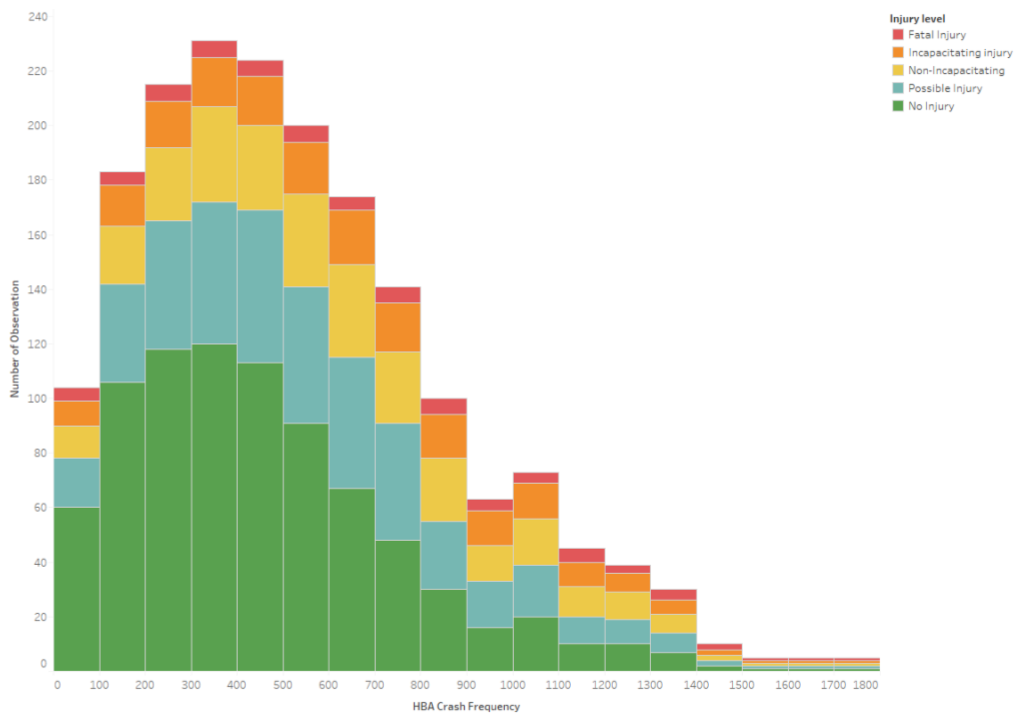

Figure 22 Histogram of HBA Crash frequency by severity type at the census tract level

In this study, comprehensive cost of traffic crashes at the zonal level consists of two parts. The first part reflects the injury severity cost and the second part reflects the vehicle damage cost. By using numbers presented in **Table 9** and counting crash frequencies by severity at each census tract, Comprehensive Crash Cost at Zonal Level (CCCAZ) at census tract was calculated using the following equations:

$$CCCAZ_i = PCI * \left(\sum_{\alpha=\{K,A,B,C,O\}} N_{\alpha,i} * Cost_\alpha + \left(N_{v,i} * Cost_{PDO}\right) \right) \tag{1}$$

where $\mathrm{N}_{\alpha,\mathrm{i}}$ represents the number of individual who live in zone i with the level of injury α, and Cost_α presents the traffic injury cost per injury presented in Table 9. $\mathrm{N}_{\mathrm{v,i}}$ presents the number of vehicles with a registered address in zone i which were involved in traffic crashes, $\mathrm{Cost}_{\mathrm{PDO}}$ also presents the vehicle unit damage cost. PCI also represents the Tennessee crash cost Per Capita Income (PCI) ratio adjustment factor, the PCI of the state of Tennessee for year 2018 was 0.855 (Bureau of Economic Analysis 2018). Figure 23 presents the histogram of the Comprehensive Crash Cost at the zonal level for the period of the study.

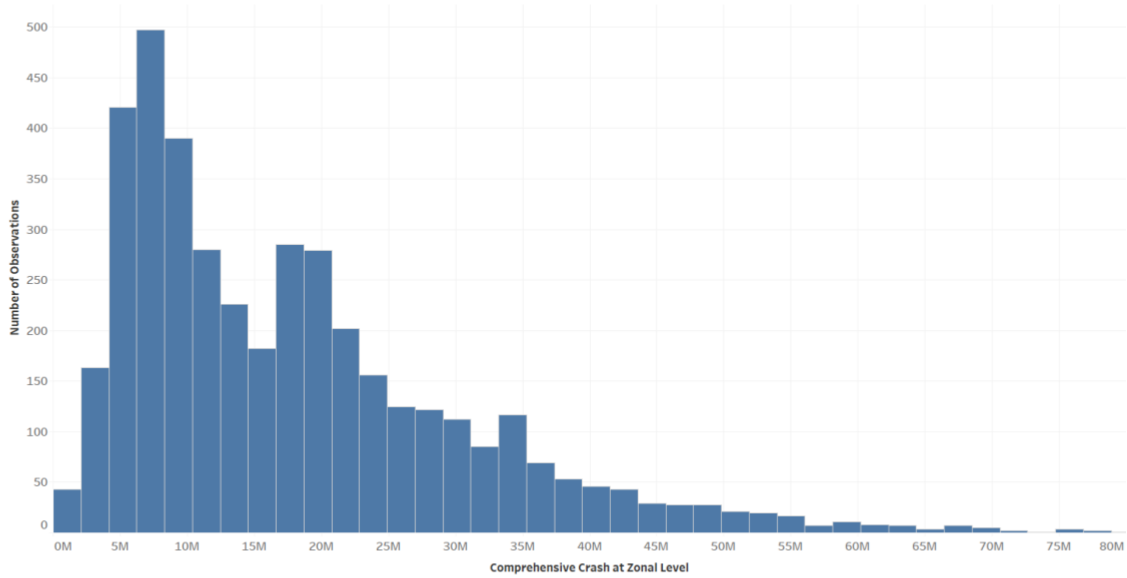

Figure 23 Histogram of CCCAZ (2018 Dollars)

Table 8 Descriptive statistics of the variables

Variable	Mean	SD	Range
Total Population	1526	789	[0, 9281]
Population Density (Person per square km)	625	979	[0, 32989]
Average Household Size	2.72	5.3	[0, 243.18]
Race Proportion			
White	0.77	0.3	[0, 1]
Non-White	0.22	0.28	[0, 1]
Means Of Transportation To Work Proportion			
Personal Vehicle	0.92	0.11	[0, 1]
Carpool	0.1	0.08	[0, 0.82]
Bus	0.01	0.04	[0, 0.62]
Motorcycle	0	0.01	[0, 0.17]
Bicycle	0	0.01	[0, 0.18]
Walk	0.02	0.05	[0, 1]
Other Means	0.01	0.03	[0, 0.6]
Age Cohort Proportion			
16 Years And Younger	0.23	0.08	[0, 0.71]
16-42 Years Old	0.32	0.11	[0, 1]
43-59 Years Old	0.25	0.08	[0, 1]
60 Years Old And More	0.2	0.1	[0, 1]
Vehicles' Ownership Per Capita	0.69	0.16	[0, 1.2]
% Of Educated People Over 25 Years Old	67.67	10.37	[0,99.93]
Housing Unit			
Percent Of Vacant Housing Unit	0.12	0.1	[0, 1]
VMT (1,000,000)	0.57	0.69	[0, .74]
Average Travel Time To Work (Minutes)	25.1	6.6	[0, 65.85]
Median Household Income ($1,000)	45.7	25.09	[0, 249.3]

Source: United States Census and HPMS

Table 9 National KABCO person-injury unit costs in 2018 dollars

Injury Type	Crash Cost Per Injury		
	Economic person-Injury Unit Costs	QALY Person-Injury Unit Costs	Comprehensive Crash Cost (2018 Dollars)
No Injury†	6,553 (5,717*)	2,938 (2,563*)	9,491 (8,280*)
Possible Injury	24,930 (21,749*)	57,227 (49,926*)	82,157 (71,675*)
Non-Incapacitating Injury	36,800 (32,105*)	112,302 (97,974*)	149,102 (130,079*)
Incapacitating Injury	96,866 (84,507*)	41,6459 (363,324*)	513,325 (447,832*)
Fatal Injury	1,603,502 (1,398,916*)	8,880,060 (7,747,082*)	10,483,562 (9,145,998*)
Unknown			
Vehicle unit cost	6,965 (6,076*)		6,965 (6,076*)

† The cost reflects the cases where injury severity was falsely assigned.

* Source: adjusted person-injury cost based on 2010 US Dollar based on Harmon, Bahar, and Gross (2018)

Modeling Approach

To evaluate safety at zonal level, traditionally, count data models such as Poisson, negative binomial and zero-inflated models are commonly utilized owing to the nature of traffic crashes that are usually measured as non-negative integers in a specific period of time (Anastasopoulos and Mannering 2009). Similar to crash frequency, the comprehensive cost of the traffic crash is a non-negative integer. Hence the models that would be used to evaluate CCCAZ must follow the nature of counts model. In addition, variations in the relationship over space also could be present in the data (i.e., spatial heterogeneity) (LeSage and Pace 2009). The stationary relationship may hide some spatial factors affecting the safety at the zonal level, which may eventually affect the accuracy of models only use one constant coefficient for the study area (i.e., global model). Using the analogy between crash frequency and CCCAZ, in this study, we will use the Poisson Regression model and the Geographically Weighted Poisson Regression Model (GWPR). To directly model the CCCAZ as an integer variable, we also used the population of each census tract as the generalized offset variable. We used the GWPR model to reflect the role of spatial heterogeneity in the modeling process by using the coordinates of the center of census tracts.

Poisson Model

In the Poisson regression, the probability that comprehensive cost of the crash at zone i equal to n could be written as (Greene 2003):

$$P(n_i) = \frac{\lambda_i^{n_i} \exp(-\lambda_i)}{n_i!} \tag{2}$$

where λ_i (Poisson parameter) is the expected CCCAZ for zone i in a three year period, $\mathrm{E(n_i)}$. In order to fit the regression model, the Poisson parameter, λ_i, can be written in a logarithm format (Greene 2003):

$$\ln(\lambda_i) = \beta X_i \tag{3}$$

where X_i is the vector of the sociodemographic data element extracted from the census tract and β is a vector of the estimated coefficients. To consider the population variable as an offset variable, we constrained the value of the population's (in logarithm scale) coefficient equal to one (Pérez-Marín and Guillen 2019).

Furthermore, in cases where the mean and the variance are not equal, applying the Poisson regression might lead to inappropriate results. in order to statistically test the existence of over-dispersion in the Poisson model, the Lagrange multiplier method was performed (Greene 2003):

$$LL = \left(\frac{\sum_{i=1}^{N}((y_i - \mu_i)^2 - y_i)}{2\sum_{i=1}^{N} \mu_i^2} \right)^2 \tag{4}$$

where y_i is the observed CCCAZ at zone i, μ_i is the predicted CCCAZ at zone i, and N is the number of zones.

Geographically Weighted Poisson Regression Model

GWPR can be used to examine whether the association between exogenous variables and CCCAZ substantially varies across space (A.S. Fotheringham, Brunsdon, and Charlton 2003). The model can be written as:

$$\ln(\lambda_i) = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)\ln{(E_{vi})} + \sum_{k=1}^{K} \beta_k(u_i, v_i) x_{ij} \tag{5}$$

where (u_i, v_i) denotes the coordinates of zone i. It should be noted that in the GWPR, $\beta_k(u_i, v_i)$ is a function of the coordinates of the center of census tract i. The following equation can be used to estimate $\beta_k(u_i, v_i)$:

$$\hat{\beta}(\mathrm{u_i, v_i}) = (\mathrm{X^T W(u_i, v_i) X})^{-1} \mathrm{X^T W(u_i, v_i) Y} \tag{6}$$

where $\hat{\beta}(\mathrm{u_i, v_i})$ *is the vector of estimated coefficients at zone* i, X *is the matrix of exogenous variables,* Y *is the* $\mathrm{n \times 1}$ *vector of the dependent variable (CCCAZ), and* $\mathrm{W(u_i, v_i)}$ *is* $\mathrm{n \times n}$ *spatial weight matrix:*

$$W(u_i, v_i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & w_{in} \end{bmatrix} \quad (7)$$

where w_{ij} *is the weight of variable* j *at location* i. *In this approach, a regression equation is estimated for each location based on observations at nearby areas. Based on the distance from the regression point each area is weighted (areas that are closer have a higher weight than ones that are farther). The* W *matrix can be estimated using an adaptive bi-square kernel, which can be written as:*

$$w_{ij} \begin{cases} \left(1 - \left({}^{d_{ij}}/_{d_{iN}}\right)^2\right)^2 & \text{if } d_{ij} < d_{iN} \\ 0 \text{ otherwise} \end{cases} \quad (8)$$

where d_{iN} denotes the distance to the Nth nearest zone of zone i. Compared to the fixed bandwidth kernels, the adaptive bi-square bandwidth varies based on the data's sparsity. To determine the bandwidth of the adaptive kernel, the corrected Akaike Information Criteria (AICc) (Hurvich, Simonoff, and Tsai 1998) was used. The best model is the one with the lowest AICc score (A.S. Fotheringham, Brunsdon, and Charlton 2003; Hadayeghi, Shalaby, and Persaud 2010a).

The non-stationarity test was used to evaluate the existence of variation in the estimated coefficients across space (Arvin, Kamrani, and J. 2019; Liu and Khattak 2017). Substantial variations among the estimated coefficients across space exist if the difference between upper and lower quartile ($\delta = \beta_{upper} - \beta_{lower}$) of the estimated coefficients from the GWPR model meets both of the following conditions:

$$\begin{cases} \delta > 1.96 * SE \\ and \\ 1.96 < \max(|z_i|) \end{cases} \quad (9)$$

where SE is the standard error of the coefficient in the global Poisson model and $|z_i|$ is the absolute value of the significance z-score of the GWPR model at census tract i. Otherwise, the coefficient is considered as the global coefficient, which does not have a substantial spatial variation. In order to estimate the GWPR model, GWR4.0 software which is developed by Nakaya et al. (2012) was used.

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.

Measures of Goodness of Fit

To evaluate and compare the performance of traditional Poisson regression, and GWPR, three statistics were utilized to measure estimation accuracy. First, we used AIC, a lower value of AIC (Bozdogan 1987) represents the better goodness of fit. We can measure AIC as following:

$$AIC = D + 2k \ (10)$$

where D denotes the model deviance, and k is the number of parameters. In the GWPR, due to the non-parametric framework of the model, the number of parameters is meaningless. Therefore, an effective number of parameters should be considered, which can be written as (Nakaya et al. 2005)

$$K = trace(S) \ (11)$$

where S is the hat matrix. In addition to the AIC, we will also use Mean Absolute Error (MAE), Root Mean Square Error (RMSE) to compare the model performances. The lower value of MAE and RMSE indicates a better performace.

Results and Discussion

Between 2014-2016, 215,481 individuals were injured in traffic crashes and 3,082 died on Tennessee's road. The total compressive crash cost of Tennessean in a three-year period was $75.0 Billion (2018 dollars). **Table 10** shows the cost of traffic crashes based on road users types. Overall, drivers and passenger have the biggest share of Comprehensive Cost of traffic crashes (94%). Yet, pedestrian and bicyclists crashes average comprehensive cost of traffic crashes was 15.7 and 6.7 times higher than driver crashes, respectively. This value reflects the vulnerability of pedestrians and bicyclist compared to motorized road users.

Table 10 Crash cost by road user type (2018 Dollars)

User Type	Number of users	Total Cost ($ million)			Average Cost ($ 1,000)		
		Compreh. Cost	Economic Costs	QALY Costs	Compreh. Cost	Economic Costs	QALY Costs
Driver	995,670	$41,282	$11,617	$29,815	$41.5	$11.7	$29.9
Passenger	352,389	$12,644	$3,828	$8,850	$35.9	$10.9	$25.1
Pedestrian	5,262	$3,431	$579	$2,836	$652.0	$110.1	$539.0
Cyclists	1,387	$390	$74	$314	$280.9	$53.2	$226.4
Grand Total	**1,354,708**	**$57,746**	**$16,098**	**$41,815**	**$42.6**	**$11.9**	**$30.9**

On average, the comprehensive cost of a traffic crash in Tennessee was $92,374. **Table 11** presents the average number crashes by severity at the zonal level. As crash severity increases, the frequency of individuals who suffered decreases. On average, on each census tract 28, 14, and 5 individuals received possible injury, non-incapacitating injury and incapacitating injury in a three-year period, respectively. On average, 0.7 individuals in each census tract were fatally injured in a traffic crash over the period. The mean comprehensive crash cost at census tract level (CCCAZ) for the study period was $18.2 million (SD = $13.9 million; Median= $15 million). Figure 24 exhibits a geographical distribution of CCCAZ in Tennessee.

Table 11 National KABCO person-injury unit costs and number of crashes in Tennessee 2014-2016 based on 2018 Dollar

Injury Type	Zonal Level Crash Frequency				Total Cost Of Crashes Between 2014-16 (2010 Dollars)**
	Mean	Std. Dev.	Max	Total Number Of Observed Crashes	
No Injury	267.6	179.1	2472	1,099,523	$9,041,888,254
Possible Injury	28.4	20.8	275	116,652	$8,441,878,352
Non-Incapacitating Injury	13.5	9.7	114	55,330	$7,284,213,036
Incapacitating Injury	4.9	3.9	40	20,287	$9,217,372,280
Fatal Injury	0.7	0.9	7	2,725	$25,323,439,334
Unknown	1.7	1.9	27	7,071	$0
PDO Vehicle*	197.1	134.9	19971	810,055	$4,823,938,279
Total					$64,132,729,534

* Unit: Crash cost per vehicle

** Only Reflects Tennesseans' Part

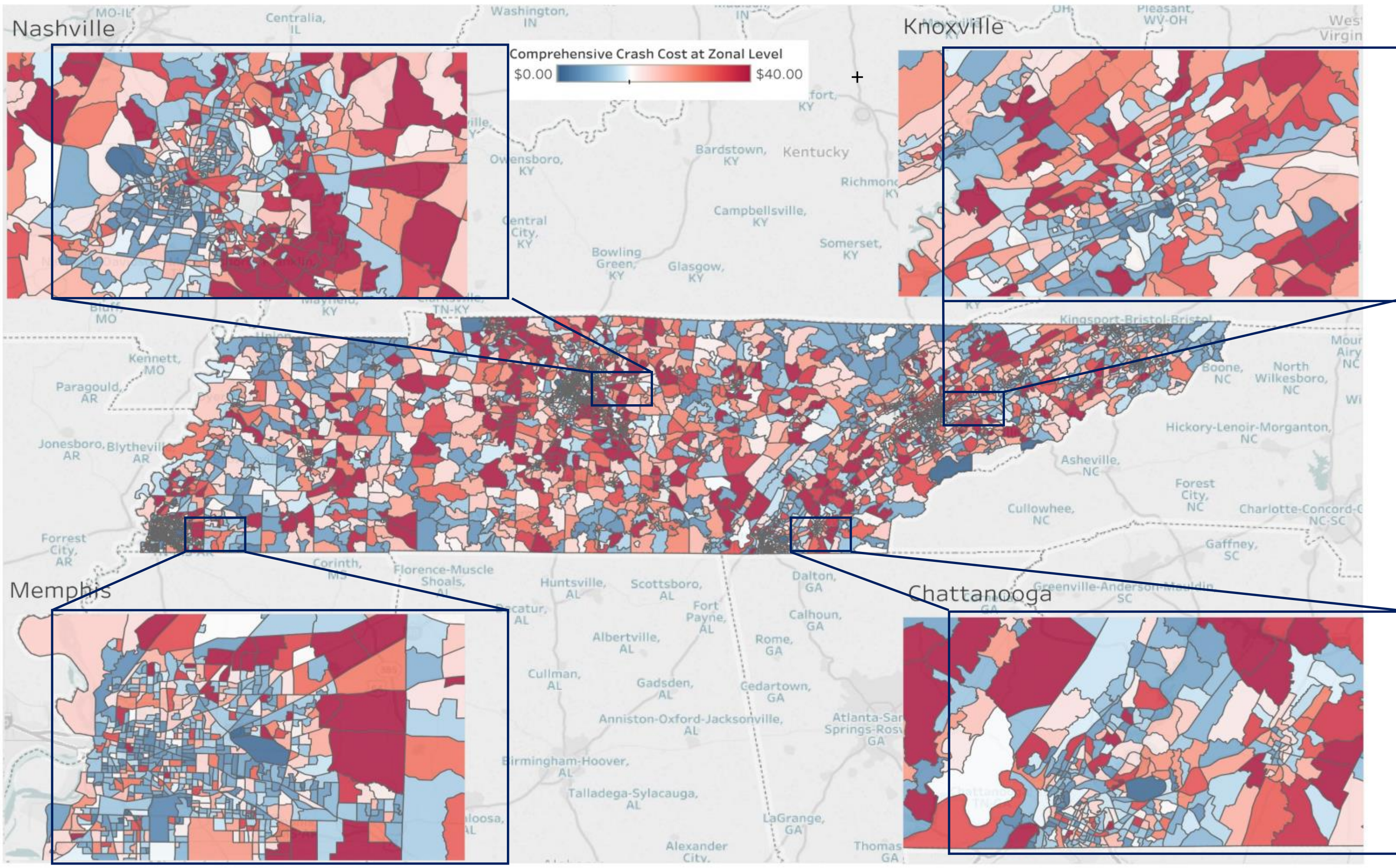

Figure 24 Geographical distribution of the comprehensive crash cost at the zonal level ($ million)

Model Comparison

Table 12 presents the result of the Poisson model and Geographically Weighted Poisson Regression with the adaptive bi-square kernel for predicting CCCAZ. We also tested the sensitivity of the model with fixed Gaussian, fixed bi-square and adaptive Gaussian kernels for model estimation. The adaptive bi-square kernel had the lowest value of the AICc. Therefore, to maintain concision, we only included the adaptive bi-square kernel results. Notably, we found that the results were largely consistent with the adaptive bi-square kernel.

The value of the Lagrange multiplier for the GWPR model and global model are 0.10, 0.13, respectively, which is less than the critical value of Lagrange multiplier ($\chi^2_{(1)}$= 3.84). Therefore, we can conclude the overdispersion is not an issue in this study. Moreover, the comparison of the AIC, AICc, deviance, MAE, and RMSE presented in Table 4 also indicate that the GWPR model is more suitable compared to the global model. The value of the Moran's I of residuals (Moran's I = 0.009) indicate that in the GWPR model the residuals are not spatially correlated. Furthermore, the VIF values (average = 1.6, max = 2.9) also indicate that the multicollinearity in not an issue.

Parameter Estimation

Results of the stationary test indicated in the GWPR model, all the covariates in the GWPR model have a local effect. **Figure 25** presents the spatial effect of the estimated parameter on CCCAZ. Only those variables that have a significant effect are presented in **Figure 25**; the insignificant coefficients are presented with a white color. It is worthy to mention that the estimated coefficients in the traditional fixed models always fall into the range of correspondence counterparts in the spatial models (Xu and Huang 2015), indicating that the estimated parameters in the global models (i.e., fixed models) characterize the average effects of the factors on the dependent variable. The sign of the median of all the variables in the GWPR model (except mean of the proportion of road users who use bus and proportion of road users 16 years old and younger) is consistent with the Poisson model which attest the aforementioned.

Analysis of the local distribution of the estimated coefficients indicates that in most variables (except population density, median family income, and household size) the sign of the variables vary from negative to positive, which is some cases are unexpected. The counterintuitive sign is not an uncommon issue considering the geographically weighted regression models and has been reported in previous studies (Hadayeghi, Shalaby, and Persaud 2010b; Chow et al. 2006; Pirdavani, Brijs, Bellemans, and Wets 2013; Xu and Huang 2015). Some studies attributed this issue to the local multicollinearity (Hadayeghi, Shalaby, and Persaud 2010b); however, this was not the case for this study. Results of the VIF test on areas where signs were counterintuitive did not raise the local multicollinearity issue; VIF values ranged between 1.01 to 3.05. Another issue could be the presence of over-dispersion in the dependent variables (Xu and Huang 2015). As a result, the Poisson model could produce more significant variables compared to the negative binomial model (Lord and Mannering 2010). This issue needs to be investigated in future studies.

Table 12 Results of the Poisson model for predicting CCCAZ ($100,000)

Variable	Estimate	Stand. Error	z(Est/SE)	Mean	STD	Min	lwr Quartile	Median	Upr Quartile	Max	Local
Constant	-3.2939	0.0233	-141.23	-3.1656	0.4972	-4.4814	-3.5978	-3.2531	-2.7313	-1.9068	yes
Age Cohorts											
16 Years And Younger	-0.0004	0.0003	-1.55	0.0017	0.0053	-0.0147	-0.0013	0.0007	0.0062	0.0140	yes
60 Years Old and More	-0.0070	0.0002	-35.42	-0.0044	0.0037	-0.0110	-0.0077	-0.0049	-0.0013	0.0031	yes
Proportion of Minor Race	0.0032	0.0001	47.10	0.0033	0.0020	-0.0056	0.0026	0.0034	0.0045	0.0080	yes
Travel to Work Mode											
Motorcycle	0.0442	0.0017	26.20	0.0351	0.0392	-0.0638	0.0094	0.0364	0.0635	0.1522	yes
Bus	0.0019	0.0004	4.28	-0.0033	0.0133	-0.0365	-0.0143	-0.0034	0.0038	0.0481	yes
Non-Motorized Modes	-0.0043	0.0004	-11.06	-0.0056	0.0094	-0.0366	-0.0117	-0.0079	0.0022	0.0292	yes
Average Travel Time To Work	0.0141	0.0002	63.57	0.0101	0.0045	-0.0031	0.0065	0.0093	0.0137	0.0251	yes
Household Size	-0.0223	0.0014	-15.75	-0.0571	0.0519	-0.2342	-0.0953	-0.0316	-0.0180	0.0544	yes
% Educated Over 25 Years Old	0.0079	0.0003	30.90	0.0092	0.0067	-0.0121	0.0077	0.0096	0.0137	0.0198	yes
Median Family Income ($1,000)	-0.0044	0.0001	-73.89	-0.0048	0.0013	-0.0080	-0.0056	-0.0051	-0.0039	-0.0008	yes
% Vacant Household	0.0057	0.0001	40.06	0.0089	0.0050	-0.0015	0.0059	0.0084	0.0112	0.0216	yes
Vehicle Per Capita	0.6573	0.0152	43.21	0.5253	0.4432	-0.8005	0.1758	0.3955	0.8985	1.9799	yes
Vehicle Miles Traveled (1,000,000)	0.2708	0.0165	16.38	0.0061	0.4124	-1.3157	-0.2697	0.0232	0.3177	0.7178	yes
Population Density (Population Per Square Kilometer)	-9.00E-05	2.0E-06	-39.80	-1.3E-04	6.1E-05	-3.7E-04	-1.6E-04	-1.5E-04	-8.4E-05	-1.9E-05	yes
Metropolitan Indicator	-0.0059	0.0032	-1.85	0.0005	0.1083	-0.2598	-0.0631	0.0285	0.0721	0.2895	yes
AIC	184877			163309							
AICc	184877			163324							
Deviance	184845			162961							
Percent Deviance Explained	0.145			0.246							
Lagrange Multiplier	0.125			0.106							
RMSE	7743.11			6861.31							
Mean Absolute Deviation	66.59			62.63							
R^2 Poisson	0.41			0.49							
Moran's I	0.065			0.009							

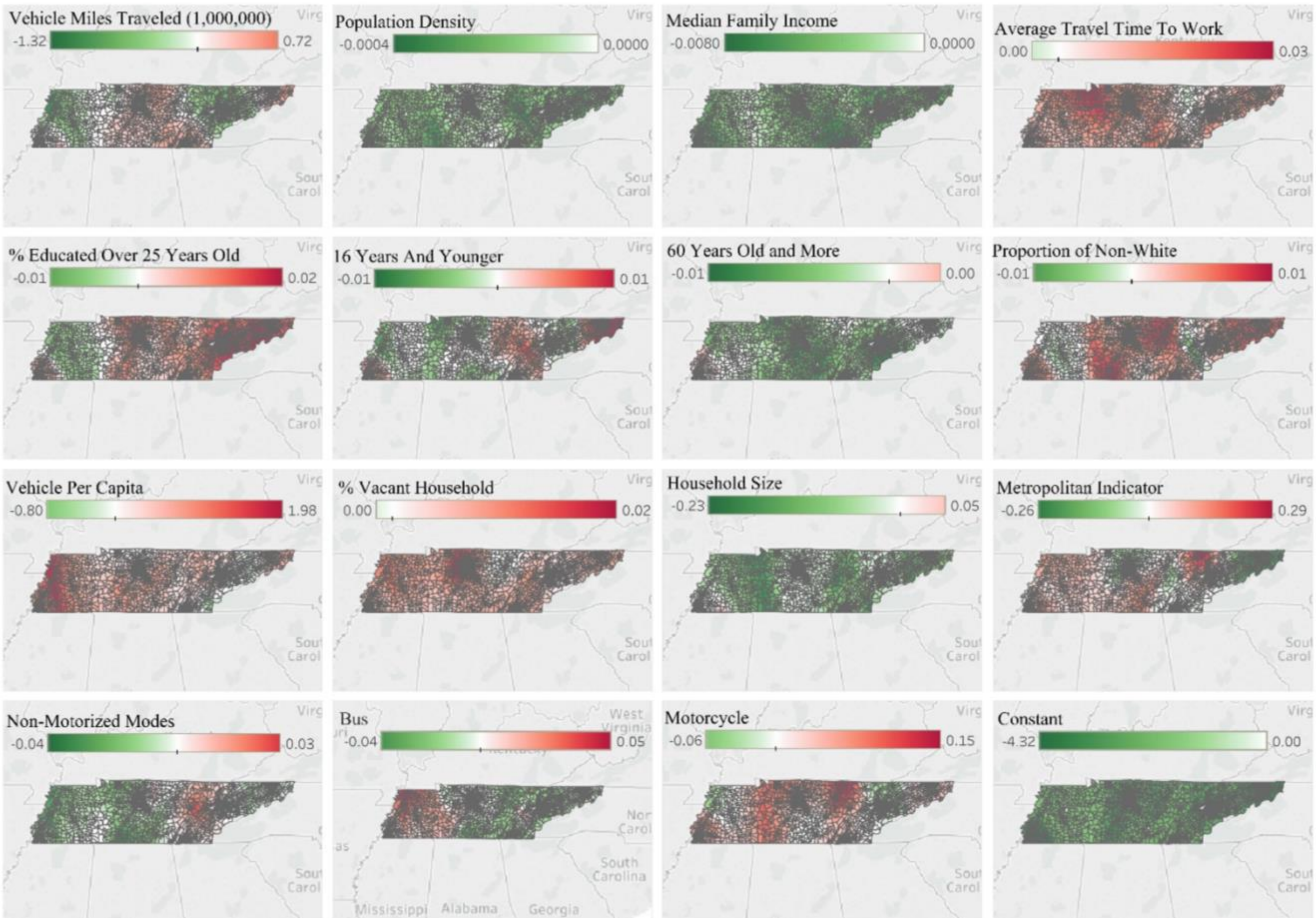

Figure 25 Local effect of the estimated coefficients in the GWPR model

Means of travel to work have a significant impact on CCCAZ. The proportion of road users who use motorcycle also has a positive sign in most of Tennessee; however, the sign of the estimated coefficients has a negative association with CCCAZ in the Knoxville metropolitan area. Moreover, means of estimated coefficients of non-motorized modes (i.e., walk and bicycle) and bus in the local model have a significant negative association with the CCCAZ. Local model estimation indicates that non-motorized modes of transportation has a negative association with CCCAZ in Memphis and Nashville metropolitan areas, whereas, in Knoxville and Chattanooga metropolitan areas this variable has a positive association with CCCAZ. In the Knoxville, Nashville and Chattanooga metropolitan areas, the estimated coefficients for the proportion of road users who use public transit has a negative association with CCCAZ. In the Memphis metropolitan area, this variable has a significant positive association with CCCAZ; however, the magnitude of the estimated coefficients is close to zero. One explanation for the negative sign of both bus and non-motorized road users could be a reduction in motorized vehicle volume in the surrounding of the residential area, which reduces traffic conflicts and eventually exposure to motorized traffic to other residents of the census tract. On the other hand, the poor design of a multimodal network could adversely impact the safety of non-motorized and public transit users. The difference between signs of the estimated coefficients in the different metropolitan areas needs to be investigated in more details in future studies.

Population density also has a negative association CCCAZ; the model predicts that as density increases the CCCAZ decreases. The sign of the population density is intuitive and is in agreement with the previous studies. For example, population density is associated with higher crash frequency for all traffic crashes, and vulnerable road user crashes (Zwerling et al. 2005; Marshall and Ferenchak 2017). Crash frequency in high-density areas such as urban areas or metropolitan areas is usually higher than rural or non-metropolitan areas; but, the crash severity is relatively lower (Zwerling et al. 2005; Clark 2003; Dumbaugh and Rae 2009). As a result, the overall effect of the population density is constructive and reduces the comprehensive cost of traffic crashes. The Metropolitan indicator also is a proxy for urban areas. Interestingly, Knoxville and Memphis Metropolitan areas coefficients have a negative association with CCCAZ, which is different from the corresponding signs in the Nashville Metropolitan.

Along with previous literature, findings indicate that age cohorts have a significant relationship with the crash outcome (e.g., Wier et al. 2009; Gomes, Cunto, and da Silva 2017; Dong et al. 2016). The proportion of population 16 years and younger has a varying sign across the state. While the percentage of individuals over 60-years-old has a significant negative association with CCCAZ (except in the Memphis metropolitan area). One may expect the senior population due to their vulnerability will suffer from higher injury severity (Yee, Cameron, and Bailey 2006); conversely, senior population have a lower trip rate (e.g., exposure to traffic) compared to other groups (KRTPO 2008; Williams and Carsten 1989; Massie, Campbell, and Williams 1995). As a result, in this study percent of the senior population has a negative effect on CCCAZ (with the exception of the Memphis metropolitan area) compared to other age cohorts.

Considering racial distribution, the estimated model indicates that the population of non-white residents has a significant association with increasing CCCAZ. This finding agrees with previous research (Marshall and Ferenchak 2017; McAndrews et al. 2013). The percentage of the population educated over 25 years old (except in some rural areas in West-Tennessee), and the percentage of a vacant houses in a census tract also has a significant positive impact on CCCAZ. Although one may expect safer behavior from educated people, it was surprising that this variable's estimated coefficients' sign is counterintuitive. The negative sign could be attributed to a higher trip rate of this group. This issue needs further analysis. Household size also has a negative association with CCCAZ, which indicates as average household size increases the CCCAZ decreases. One explanation could be the lower per-capita trip rate of individuals in families with bigger household size compared to smaller households in the study area (KRTPO 2008).

Variables that explain the economic status of each census tract are also associated with the CCCAZ. Median family income is a significant predictor of the CCCAZ; a negative sign of the variable suggests that as family income increases the CCCAZ decreases. The sign agrees with previous studies that show road users with lower income are more prone to traffic crashes (Lee, Abdel-Aty, and Choi 2014; Males 2009; World Health Organization 2015; Lee and Abdel-Aty 2018). Furthermore, lower-income families' vehicles usually have fewer safety features which may increase the likelihood of severe injuries (Girasek and Taylor 2010). In contrast, vehicles per capita has a significant impact on the CCCAZ; the positive sign indicates that as this variable increases, the social outcome of traffic crashes gets worse. Vehicles per capita also could be used as a proxy for activity (i.e., amount of vehicle traveled) or lack of multi-modality; we expect a higher trip rate in areas with higher vehicle ownership (e.g., Khattak and Rodriguez 2005; KRTPO 2008). These findings are also in agreement with studies that focused on human factors that show that some groups (e.g., lower income, lower education, young road users) are more prone to aberrant behaviors (Nordfjærn, Hezaveh, and Mamdoohi 2015; Hezaveh et al. 2017; Hezaveh et al. 2018; Davey et al. 2007; Elliott, Baughan, and Sexton 2007; Özkan et al. 2006).

VMT and average travel time to work could be interpreted as proxies to exposure. The expectation was to see higher crash cost as these two variables value increase. The models indicate that VMT in the surrounding area of residents has a positive association with CCCAZ. However, considering the local effect and geographical distribution of this variable in Figure 4, we noticed that in the Knoxville metropolitan area and some rural area, VMT has a negative association with CCCAZ. Analyzing the local coefficient indicate that the multicollinearity was not an issue in the areas with counterintuitive signs. This issue needs to be investigated in future studies.

Average travel time to work, which represent the amount of time that individuals spend in traffic on their work trips also has a significant positive association with CCCAZ. The positive sign in the model indicates that as travel time increases the crash cost at zonal level increases. Travel time could be interpreted as an indicator of accessibility (Merlin et al. 2019; Marshall and Ferenchak 2017). Increase in accessibility would decrease the travel time (by reducing trip length), VMT and eventually would reduce the comprehensive cost of traffic crashes.

Conclusion

In this study, we used the Home-Based Approach crash frequency at the zonal level to calculate the comprehensive crash cost at the zonal level. Unlike traditional road safety analysis that aggregates crashes at the location of the crash, the HBA attributes road safety to the home-address of individuals in a traffic crash. Consequently, we measured the comprehensive cost of traffic crashes at the zonal level by using person-injury crash cost.

Findings indicate that the burden of traffic crashes does not affect the study area in equitable ways. Moreover, over-dispersion is not an issue regarding CCCAZ analysis in this study, hence the Poisson model is suitable for evaluation of the relationship between sociodemographic variables and CCCAZ at the zonal level. Comparison of the performance of the GWPR and Poisson models shows the substantial existence of spatial heterogeneity in the analysis.

This study's findings are broadly in agreement with road safety literature. We find that an increase in population density reduces the societal cost of traffic crashes at the zonal level; increase in residential density, particularly in the urban areas is correlated with the reduction in speeds. On the other hand, an increase in travel time and consequently higher traffic exposure adversely affect the social cost of traffic crashes.

Comprehensive crash cost at the zonal level could be used as a tool for assigning proper countermeasures or interventions to the areas where the disproportionate economic burden of traffic crashes exists or to promote vertical equity in the distribution of road safety countermeasures. Moreover, the HBA could be an advantageous element for developing policies that support groups that are more prone to burden from road traffic crashes.

There are several possible extensions for this study; first we can learn to reduce the injury misclassification error by linking police crash reports to health-oriented databases (Cherry et al. 2018) to get a better understanding of injury outcome and subsequently a more accurate measurement of the injury. Second, the variables that we used in this study was mostly limited to the demographics of residents extracted from the US Census. Adding extra variables regarding transportation network and travel behavior would help us understand the relationship between travel behavior and the comprehensive cost of traffic crashes. Third, based on our findings, we are recommending the use of the home address of the road users to target the areas that are more prone to the burden of traffic crashes by proper education and enforcement countermeasures.

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, Mohamed, Chowdhury Siddiqui, Helai Huang, and Xuesong Wang. 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, Jonathan. 2013. "Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica." *Accident Analysis & Prevention* 59: 365-373. https://doi.org/https://doi.org/10.1016/j.aap.2013.06.014. http://www.sciencedirect.com/science/article/pii/S000145751300242X.

Aguero-Valverde, Jonathan, and Paul P Jovanis. 2006. "Spatial analysis of fatal and injury crashes in Pennsylvania." *Accident Analysis & Prevention* 38 (3): 618-625.

Ahadi, Mohammad Reza, and Hesamoddin Razi-Ardakani. 2015. "Estimating the cost of road traffic accidents in Iran using Human Capital Method." *International Journal of Transportation Engineering* 2 (3): 163-178.

Anastasopoulos, Panagiotis Ch, and Fred L Mannering. 2009. "A note on modeling vehicle accident frequencies with random-parameters count models." *Accident Analysis & Prevention* 41 (1): 153-159.

Andrew, Cliff, and J Keith Ord. 1981. "Spatial Processes: Models and Applications." *London: Pion*.

Anselin, Luc. 2003. "GeoDa 0.9 user's guide." *Urbana* 51: 61801.

Anselin, Luc, Anil K Bera, Raymond Florax, and Mann J Yoon. 1996. "Simple diagnostic tests for spatial dependence." *Regional science and urban economics* 26 (1): 77-104.

Anselin, Luc, and Raymond JGM Florax. 1995. "Small sample properties of tests for spatial dependence in regression models: Some further results." In *New directions in spatial econometrics*, 21-74. Springer.

Anselin, Luc, and Serge Rey. 1991. "Properties of tests for spatial dependence in linear regression models." *Geographical analysis* 23 (2): 112-131.

Apelberg, Benjamin J, Timothy J Buckley, and Ronald H White. 2005. "Socioeconomic and racial disparities in cancer risk from air toxics in Maryland." *Environmental health perspectives* 113 (6): 693-699.

Arvin, Ramin, Mohsen Kamrani, and Khattak. Asad J. 2019. "How instantaneous driving behavior contributes to crashes at intersections: Extracting useful information from connected vehicle message data." *Accident Analysis & Prevention*. https://doi.org/https://doi.org/10.1016/j.aap.2019.01.014. .

Atkinson, Anthony B. 1970. "On the measurement of inequality." *Journal of economic theory* 2 (3): 244-263.

Baker, Susan P, Elisa R Braver, Li-Hui Chen, Janella F Pantula, and Dawn Massie. 1998. "Motor vehicle occupant deaths among Hispanic and black children and teenagers." *Archives of pediatrics & adolescent medicine* 152 (12): 1209-1212.

Baller, Robert D, Luc Anselin, Steven F Messner, Glenn Deane, and Darnell F Hawkins. 2001. "Structural covariates of US county homicide rates: Incorporating spatial effects." *Criminology* 39 (3): 561-588.

Blatt, Jesse, and Susanne M Furman. 1998. "Residence location of drivers involved in fatal crashes." *Accident Analysis & Prevention* 30 (6): 705-711.

Blincoe, LJ, TR Miller, Eduard Zaloshnja, and Bruce A Lawrence. 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*.

Bozdogan, Hamparsum. 1987. "Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions." *Psychometrika* 52 (3): 345-370.

Braver, Elisa 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.

Bureau of Economic Analysis. 2018. "Per capita real GDP by state (chained 2009 dollars)."

Cai, Qing, Mohamed Abdel-Aty, and Jaeyoung Lee. 2017. "Macro-level vulnerable road users crash analysis: a Bayesian joint modeling approach of frequency and proportion." *Accident Analysis & Prevention* 107: 11-19.

Cai, Qing, Mohamed Abdel-Aty, Jaeyoung Lee, and Naveen Eluru. 2017. "Comparative analysis of zonal systems for macro-level crash modeling." *Journal of safety research* 61: 157-166.

Cameron, A Colin, and Pravin K Trivedi. 1986. "Econometric models based on count data. Comparisons and applications of some estimators and tests." *Journal of applied econometrics* 1 (1): 29-53.

Campos-Outcalt, Doug, Curt Bay, A Dellapena, and Marya K Cota. 2003. "Motor vehicle crash fatalities by race/ethnicity in Arizona, 1990–96." *Injury Prevention* 9 (3): 251-256.

Cervero, Robert, and Kara Kockelman. 1997. "Travel demand and the 3Ds: density, diversity, and design." *Transportation Research Part D: Transport and Environment* 2 (3): 199-219.

Cheng, Wen, Gurdiljot Singh Gill, John L Ensch, Jerry Kwong, and Xudong Jia. 2018. "Multimodal crash frequency modeling: Multivariate space-time models with alternate spatiotemporal interactions." *Accident Analysis & Prevention* 113: 159-170.

Cherry, Christopher, Amin Mohamadi Hezaveh, Melanie Noltenius, Asad Khattak, Louis Merlin, Eric Dumbaugh, David Ragland, and Laura Sandt. 2018. "Completing the Picture of Traffic Injuries: Understanding Data Needs and Opportunities for Road Safety."

Chiou, Yu-Chiun, Rong-Chang Jou, and Cheng-Han Yang. 2015. "Factors affecting public transportation usage rate: Geographically weighted regression." *Transportation Research Part A: Policy and Practice* 78: 161-177.

Chow, Lee-Fang, Fang Zhao, Xuemei Liu, Min-Tang Li, and Ike Ubaka. 2006. "Transit ridership model based on geographically weighted regression." *Transportation Research Record: Journal of the Transportation Research Board* (1972): 105-114.

Clark, David E. 2003. "Effect of population density on mortality after motor vehicle collisions." *Accident Analysis & Prevention* 35 (6): 965-971.

Davey, J., D. Wishart, J. Freeman, and B. Watson. 2007. "An application of the driver behaviour questionnaire in an Australian organisational fleet setting." *Transportation Research Part F: Traffic Psychology and Behaviour* 10 (1): 11-21. https://doi.org/http://dx.doi.org/10.1016/j.trf.2006.03.001. http://www.sciencedirect.com/science/article/pii/S1369847806000258.

Delbosc, Alexa, and Graham Currie. 2011. "The spatial context of transport disadvantage, social exclusion and well-being." *Journal of Transport Geography* 19 (6): 1130-1137.

Dong, Ni, Helai Huang, Jaeyoung Lee, Mingyun Gao, and Mohamed Abdel-Aty. 2016. "Macroscopic hotspots identification: a Bayesian spatio-temporal interaction approach." *Accident Analysis & Prevention* 92: 256-264.

Dong, Ni, Helai Huang, Pengpeng Xu, Zhuodi Ding, and Duo Wang. 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, Ni, Helai Huang, and Liang Zheng. 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, Patrick. 1980. "Linear models with spatially distributed data: Spatial disturbances or spatial effects?" *Sociological Methods & Research* 9 (1): 29-60.

---. 1982. "Maximum likelihood methods for linear models: spatial effect and spatial disturbance terms." *Sociological Methods & Research* 10 (3): 243-269.

Dumbaugh, Eric, and Robert Rae. 2009. "Safe urban form: revisiting the relationship between community design and traffic safety." *Journal of the American Planning Association* 75 (3): 309-329.

Elliott, Mark A., Christopher J. Baughan, and Barry F. Sexton. 2007. "Errors and violations in relation to motorcyclists' crash risk." *Accident Analysis & Prevention* 39 (3): 491-499. https://doi.org/http://dx.doi.org/10.1016/j.aap.2006.08.012. http://www.sciencedirect.com/science/article/pii/S0001457506001588.

Elvik, Rune, Truls Vaa, Alena Erke, and Michael Sorensen. 2009. *The handbook of road safety measures*. Emerald Group Publishing.

FHWA. 2011. *KABCO Injury Classification Scale and Definitions.*

---. 2017. "KABCO Injury Classification Scale and Definitions." Accessed June 21. https://safety.fhwa.dot.gov/hsip/spm/conversion_tbl/pdfs/kabco_ctable_by_state.pdf.

Fotheringham, A Stewart, Chris Brunsdon, and Martin Charlton. 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

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

García-Altés, A, and Katherine Pérez. 2007. "The economic cost of road traffic crashes in an urban setting." *Injury prevention* 13 (1): 65-68.

Gärling, Tommy, Daniel Eek, Peter Loukopoulos, Satoshi Fujii, Olof Johansson-Stenman, Ryuichi Kitamura, Ram Pendyala, and Bertil Vilhelmson. 2002. "A conceptual analysis of the impact of travel demand management on private car use." *Transport Policy* 9 (1): 59-70.

Girasek, Deborah C, and Brett Taylor. 2010. "An exploratory study of the relationship between socioeconomic status and motor vehicle safety features." *Traffic injury prevention* 11 (2): 151-155.

Gomes, Marcos José Timbó Lima, Flavio Cunto, and Alan Ricardo da Silva. 2017. "Geographically weighted negative binomial regression applied to zonal level safety performance models." *Accident Analysis & Prevention* 106: 254-261.

Gourieroux, Christian, and Michael Visser. 1997. "A count data model with unobserved heterogeneity." *Journal of Econometrics* 79 (2): 247-268.

Greene, William H. 2003. *Econometric analysis*. Pearson Education India.

Gunier, Robert B, Andrew Hertz, Julie Von Behren, and Peggy Reynolds. 2003. "Traffic density in California: socioeconomic and ethnic differences among potentially exposed children." *Journal of Exposure Science and Environmental Epidemiology* 13 (3): 240.

Hadayeghi, Alireza, Amer Shalaby, and Bhagwant Persaud. 2003. "Macrolevel accident prediction models for evaluating safety of urban transportation systems." *Transportation Research Record: Journal of the Transportation Research Board* (1840): 87-95.

---. 2010a. "Development of planning-level transportation safety models using full Bayesian semiparametric additive techniques." *Journal of Transportation Safety & Security* 2 (1): 45-68.

Hadayeghi, Alireza, Amer S Shalaby, and Bhagwant N Persaud. 2010b. "Development of planning level transportation safety tools using Geographically Weighted Poisson Regression." *Accident Analysis & Prevention* 42 (2): 676-688.

Harmon, Tim, Geni Bahar, and Frank Gross. 2018. *Crash Costs for Highway Safety Analysis.* USDOT FHWA Office (Washington, DC).

Harper, Jennifer S, William M Marine, Carol J Garrett, Dennis Lezotte, and Steven R Lowenstein. 2000. "Motor vehicle crash fatalities: a comparison of Hispanic and non-Hispanic motorists in Colorado." *Annals of emergency medicine* 36 (6): 589-596.

Hauer, Ezra. 1997. *Observational Before-After Studies in Road Safety--Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*.

Haynes, Robin, Andrew Jones, Victoria Kennedy, Ian Harvey, and Tony Jewell. 2007. "District variations in road curvature in England and Wales and their association with road-traffic crashes." *Environment and Planning A* 39 (5): 1222-1237.

Hezaveh, Amin Mohamadi, Ramin Arvin, and Christopher R Cherry. 2019. "A geographically weighted regression to estimate the comprehensive cost of traffic crashes at a zonal level." *Accident Analysis & Prevention* 131: 15-24.

Hezaveh, Amin Mohamadi, and Christopher R. Cherry. 2018. "A New Approach to Aggregate Crash Prediction Model: Home-Based Crash Frequency Modeling." Southern District ITE Annual Meeting.

---. 2019a. "Individuals' Likelihood of Involvement in Traffic Crashes: Introduction of Home-Based Approach – A Complementary Definition of Road Safety." Transportation Research Board 98th Annual Meeting.

---. 2019b. "Neighborhood-level factors affecting seat belt use." *Accident Analysis and Prevention* 122: 153-161.

Hezaveh, Amin Mohamadi, Trond Nordfjærn, Amir Reza Mamdoohi, and Özlem Şimşekoğlu. 2018. "Predictors of Self-reported Crashes among Iranian Drivers: Exploratory Analysis of an Extended Driver Behavior Questionnaire." *PROMET-Traffic&Transportation* 30 (1): 35-43.

Hezaveh, Amin Mohamadi, Mohsen Fallah Zavareh, Christopher R Cherry, and Trond Nordfjærn. 2017. "Errors and violations in relation to bicyclists' crash risks: Development of the Bicycle Rider Behavior Questionnaire (BRBQ)." *Journal of Transport & Health*.

Highway Safety Manual. 2010. "Highway Safety Manual." *American Association of State Highway and Transportation Officials* 2: 10-1.

Hosseinpour, Mehdi, Sina Sahebi, Zamira Hasanah Zamzuri, Ahmad Shukri Yahaya, and Noriszura Ismail. 2018. "Predicting crash frequency for multi-vehicle collision types using multivariate Poisson-lognormal spatial model: A comparative analysis." *Accident Analysis & Prevention*. https://doi.org/https://doi.org/10.1016/j.aap.2018.05.003. http://www.sciencedirect.com/science/article/pii/S0001457518301878.

Huang, Helai, and Mohamed Abdel-Aty. 2010. "Multilevel data and Bayesian analysis in traffic safety." *Accident Analysis & Prevention* 42 (6): 1556-1565.

Huang, Helai, Bo Song, Pengpeng Xu, Qiang Zeng, Jaeyoung Lee, and Mohamed Abdel-Aty. 2016. "Macro and micro models for zonal crash prediction with application in hot zones identification." *Journal of Transport Geography* 54: 248-256.

Hurvich, Clifford M, Jeffrey S Simonoff, and Chih-Ling Tsai. 1998. "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60 (2): 271-293.

Hurvich, Clifford M, and Chih-Ling Tsai. 1989. "Regression and time series model selection in small samples." *Biometrika* 76 (2): 297-307.

Jacobs, Goff, Amy Aeron-Thomas, and Angela Astrop. 2000. "Estimating global road fatalities."

Kanafani, Adib. 1983. "Transportation demand analysis."

Khattak, Asad J, and Daniel Rodriguez. 2005. "Travel behavior in neo-traditional neighborhood developments: A case study in USA." *Transportation Research Part A: Policy and Practice* 39 (6): 481-500.

Khorashadi, Ahmad, Debbie Niemeier, Venky Shankar, and Fred Mannering. 2005. "Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis." *Accident Analysis & Prevention* 37 (5): 910-921.

Kim, Chong Won, Tim T Phipps, and Luc Anselin. 2003. "Measuring the benefits of air quality improvement: a spatial hedonic approach." *Journal of environmental economics and management* 45 (1): 24-39.

Koornstra, Matthijs, David Lynam, and Göran Nilsson. 2002. "SUNflower: a comparative study of the development of road." *Leidschendam: SWOV Institute for Road Safety Research*.

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

KRTPO. 2008. *2008 East Tennessee Household Travel Survey Final Report.* Knoxville Regional Transportation Planning Organization (206 Wild Basin Rd., Suite A-300 Austin, Texas 78746).

Ladron de Guevara, Felipe, Simon Washington, and Jutaek Oh. 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, Jaeyoung, and Mohamed Abdel-Aty. 2018. "Macro-level analysis of bicycle safety: Focusing on the characteristics of both crash location and residence." *International journal of sustainable transportation* 12 (8): 553-560.

Lee, Jaeyoung, Mohamed Abdel-Aty, and Keechoo Choi. 2014. "Analysis of residence characteristics of at-fault drivers in traffic crashes." *Safety science* 68: 6-13.

Lee, Jaeyoung, Mohamed Abdel-Aty, Keechoo Choi, and Helai Huang. 2015. "Multi-level hot zone identification for pedestrian safety." *Accident Analysis & Prevention* 76: 64-73.

Lee, Jaeyoung, Mohamed Abdel-Aty, and Ximiao Jiang. 2015. "Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level." *Accident Analysis & Prevention* 78: 146-154.

Legendre, Pierre. 1993. "Spatial autocorrelation: trouble or new paradigm?" *Ecology* 74 (6): 1659-1673.

LeSage, James, and Robert Kelley Pace. 2009. *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Levine, Ned, Karl E Kim, and Lawrence H Nitz. 1995. "Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators." *Accident Analysis & Prevention* 27 (5): 675-685.

Li, Zhibin, Wei Wang, Pan Liu, John M Bigham, and David R Ragland. 2013. "Using geographically weighted Poisson regression for county-level crash modeling in California." *Safety science* 58: 89-97.

Liu, Jun, and Asad J Khattak. 2017. "Gate-violation behavior at highway-rail grade crossings and the consequences: using geo-spatial modeling integrated with path analysis." *Accident Analysis & Prevention* 109: 99-112.

Lord, Dominique, and Fred Mannering. 2010. "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives." *Transportation Research Part A: Policy and Practice* 44 (5): 291-305.

Lovegrove, Gordon R, and Tarek Sayed. 2006. "Macro-level collision prediction models for evaluating neighbourhood traffic safety." *Canadian Journal of Civil Engineering* 33 (5): 609-621.

Males, Mike A. 2009. "Poverty as a determinant of young drivers' fatal crash risks." *Journal of safety research* 40 (6): 443-448.

Mannering, Fred L, Venky Shankar, and Chandra R Bhat. 2016. "Unobserved heterogeneity and the statistical analysis of highway accident data." *Analytic methods in accident research* 11: 1-16.

Marsden, Peter V, and Noah E Friedkin. 1993. "Network studies of social influence." *Sociological Methods & Research* 22 (1): 127-151.

Marshall, Wesley E, and Nicholas N Ferenchak. 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.

Massie, Dawn L, Kenneth L Campbell, and Allan F Williams. 1995. "Traffic accident involvement rates by driver age and gender." *Accident Analysis & Prevention* 27 (1): 73-87.

Mayrose, James, Dietrich Jehle, Marc Hayes, Dylan Tinnesz, Gina Piazza, and Gregory E Wilding. 2005. "Influence of the unbelted rear-seat passenger on driver mortality:"the backseat bullet"." *Academic emergency medicine* 12 (2): 130-134.

Mayrose, James, and Dietrich VK Jehle. 2002. "An analysis of race and demographic factors among motor vehicle fatalities." *Journal of Trauma and Acute Care Surgery* 52 (4): 752-755.

McAndrews, Carolyn, Kirsten Beyer, Clare E Guse, and Peter Layde. 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.

Merlin, Louis A., Christopher R. Cherry, A.M. Hezaveh, and Eric Dumbaugh. 2019. "Residential Accessibility's Relationships with Crash Rates per Capita." Transportation Research Board 98th Annual Meeting.

Miaou, Shaw-Pin, Joon Jin Song, and Bani K Mallick. 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, Mehdi, Gholamali Shafabakhsh, and Ali Naderan. 2018. "Effects of modal shares on crash frequencies at aggregate level." *Accident Analysis & Prevention* 120: 295-303.
Mohan, Dinesh. 2002. "Social cost of road traffic crashes in India." Proceedings First Safe Community Conference on Cost of Injury. Viborg, Denmark.
Moran, Patrick AP. 1950. "Notes on continuous stochastic phenomena." *Biometrika* 37 (1/2): 17-23.
Naderan, Ali, and Jalil Shahi. 2010. "Aggregate crash prediction models: Introducing crash generation concept." *Accident Analysis & Prevention* 42 (1): 339-346.
Nakaya, Tomoki. 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, Tomoki, M Charlton, P Lewis, S Fortheringham, and Chris Brunsdon. 2012. "Windows Application for Geographically Weighted Regression Modeling." *Ritsumeikan University, Kyoto, Japan.*
Nakaya, Tomoki, Alexander S Fotheringham, Chris Brunsdon, and Martin Charlton. 2005. "Geographically weighted Poisson regression for disease association mapping." *Statistics in medicine* 24 (17): 2695-2717.
NHTSA. 2013. "Traffic safety facts 2013." *Washington, DC: Department of Transportation, National Highway Traffic Safety Administration*.
---. 2014. "The economic and societal impact of motor vehicle crashes, 2010." *Report DOT HS* 812: 013.
Nordfjærn, Trond, Amin Mohamadi Hezaveh, and Amir Reza Mamdoohi. 2015. "An analysis of reported driver behaviour in samples of domestic and expatriate Iranians." *Journal of Risk Research* 18 (5): 566-580.
O'brien, Robert M. 2007. "A caution regarding rules of thumb for variance inflation factors." *Quality & quantity* 41 (5): 673-690.
Ord, Keith. 1975. "Estimation methods for models of spatial interaction." *Journal of the American Statistical Association* 70 (349): 120-126.
Özkan, Türker, Timo Lajunen, Joannes El Chliaoutakis, Dianne Parker, and Heikki Summala. 2006. "Cross-cultural differences in driving skills: A comparison of six countries." *Accident Analysis & Prevention* 38 (5): 1011-1018. https://doi.org/http://dx.doi.org/10.1016/j.aap.2006.04.006. http://www.sciencedirect.com/science/article/pii/S000145750600056X.
Parker, Jennifer D, Nataliya Kravets, Keeve Nachman, and Amir Sapkota. 2012. *Linkage of the 1999-2008 National Health and Nutrition Examination Surveys to traffic indicators from the National Highway Planning Network*. Vol. 45: Citeseer.
Pérez-Marín, Ana M, and Montserrat Guillen. 2019. "Semi-autonomous vehicles: Usage-based data evidences of what could be expected from eliminating speed limit violations." *Accident Analysis & Prevention* 123: 99-106.
Pirdavani, Ali, Tom Bellemans, Tom Brijs, Bruno Kochan, and Geert Wets. 2014. "Assessing the road safety impacts of a teleworking policy by means of geographically weighted regression method." *Journal of transport geography* 39: 96-110.
Pirdavani, Ali, Tom Brijs, Tom Bellemans, Bruno Kochan, and Geert Wets. 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.
---. 2012b. "Developing zonal crash prediction models with a focus on application of different exposure measures."
---. 2013. "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, Ali, Tom Brijs, Tom Bellemans, and Geert Wets. 2013. "Spatial analysis of fatal and injury crashes in Flanders, Belgium: application of geographically weighted regression technique."
Pollution, Health Effects Institute. Panel on the Health Effects of Traffic-Related Air. 2010. *Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects*. Vol. 17: Health Effects Institute.
Pucher, John, and John L Renne. 2005. "Rural mobility and mode choice: Evidence from the 2001 National Household Travel Survey." *Transportation* 32 (2): 165-186.
Pulugurtha, Srinivas S, Venkata Ramana Duddu, and Yashaswi Kotagiri. 2013. "Traffic analysis zone level crash estimation models based on land use characteristics." *Accident Analysis & Prevention* 50: 678-687.
Quddus, Mohammed 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, Kyoung-Ah, Joon-Ki Kim, Young-Ihn Lee, and Gudmundur F Ulfarsson. 2016. "Spatial regression analysis of traffic crashes in Seoul." *Accident Analysis & Prevention* 91: 190-199.

Romano, Eduardo O, A Scott Tippetts, and Robert B Voas. 2006. "Language, income, education, and alcohol-related fatal motor vehicle crashes." *Journal of ethnicity in substance abuse* 5 (2): 119-137.

Saha, Dibakar, Priyanka Alluri, Albert Gan, and Wanyang Wu. 2018. "Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models." *Accident Analysis & Prevention*.

Schepers, JP, and Ethan Heinen. 2013. "How does a modal shift from short car trips to cycling affect road safety?" *Accident Analysis & Prevention* 50: 1118-1127.

Schiff, Melissa, and Thomas Becker. 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, Chowdhury, Mohamed Abdel-Aty, and Keechoo Choi. 2012. "Macroscopic spatial analysis of pedestrian and bicycle crashes." *Accident Analysis & Prevention* 45: 382-391.

Stamatiadis, Nikiforos, and Giovanni Puccini. 2000. "Socioeconomic descriptors of fatal crash rates in the Southeast USA." *Injury control and safety promotion* 7 (3): 165-173.

Tiwari, Geetam, Deepty Jain, and Kalaga Ramachandra Rao. 2016. "Impact of public transport and non-motorized transport infrastructure on travel mode shares, energy, emissions and safety: Case of Indian cities." *Transportation research part D: transport and environment* 44: 277-291.

Ukkusuri, Satish, Samiul Hasan, and H Aziz. 2011. "Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency." *Transportation Research Record: Journal of the Transportation Research Board* (2237): 98-106.

Wang, Yiyi, and Kara M Kockelman. 2013. "A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods." *Accident Analysis & Prevention* 60: 71-84.

Washington, Simon, Jeff Metarko, Ignatius Fomunung, Rob Ross, Frank Julian, and Erick Moran. 1999. "An inter-regional comparison: fatal crashes in the southeastern and non-southeastern United States: preliminary findings." *Accident Analysis & Prevention* 31 (1): 135-146.

Washington, Simon P, Matthew G Karlaftis, and Fred Mannering. 2010. *Statistical and econometric methods for transportation data analysis*. CRC press.

Wegman, Fred, and Siem Oppe. 2010. "Benchmarking road safety performances of countries." *Safety science* 48 (9): 1203-1211.

WHO. 2015. *Global status report on road safety 2015*. Edited by World Health Organization. World Health Organization.

Wier, Megan, June Weintraub, Elizabeth H Humphreys, Edmund Seto, and Rajiv Bhatia. 2009. "An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning." *Accident Analysis & Prevention* 41 (1): 137-145.

Wijnen, Wim, and Henk Stipdonk. 2016. "Social costs of road crashes: An international analysis." *Accident Analysis & Prevention* 94: 97-106.

Williams, Allan F, and Oliver Carsten. 1989. "Driver age and crash involvement." *American Journal of Public Health* 79 (3): 326-327.

Woghiren-Akinnifesi, Efomo L. 2013. "Residential proximity to major highways—United States, 2010." *CDC Health Disparities and Inequalities Report—United States, 2013* 62 (3): 46.

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

Xia, Jianhong Cecilia, Joshua Nesbitt, Rebekah Daley, Arfanara Najnin, Todd Litman, and Surya Prasad Tiwari. 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, Pengpeng, and Helai Huang. 2015. "Modeling crash spatial heterogeneity: random parameter versus geographically weighting." *Accident Analysis & Prevention* 75: 16-25.

Xu, Pengpeng, Helai Huang, Ni Dong, and SC Wong. 2017. "Revisiting crash spatial heterogeneity: a Bayesian spatially varying coefficients approach." *Accident Analysis & Prevention* 98: 330-337.

Yee, WY, Peter A Cameron, and MJ Bailey. 2006. "Road traffic injuries in the elderly." *Emergency Medicine Journal* 23 (1): 42-46.

Zofío, Jose L, Ana M Condeço-Melhorado, Andrés Maroto-Sánchez, and Javier Gutiérrez. 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, C Peek-Asa, PS Whitten, Sung-Won Choi, NL Sprince, and Michael P Jones. 2005. "Fatal motor vehicle crashes in rural and urban areas: decomposing rates into contributing factors." *Injury Prevention* 11 (1): 24-28.

Collaborative Sciences Center for
ROAD SAFETY

730 Martin Luther King Jr. Blvd.
Suite 300
Chapel Hill, NC 27599-3430
info@roadsafety.unc.edu

www.roadsafety.unc.edu