**TRCLC 14-07**
**June 30, 2016**

# Development of Decision Support Tools to Assess Pedestrian and Bicycle Safety: Focus on Population, Demographic and Socio-economic Spectra

## FINAL REPORT

**Deo Chimba, PhD, PE., PTOE**
**Associate Professor**
**Civil Engineering Department**
**Tennessee State University**
**3500 John A. Merritt Blvd**
**Torrence Hall Bldg, Room 108B**
**Nashville, TN 37209**
**Phone: 615-963-5430**
**Fax: 615-963-5902**

**Abram Musinguzi**
**Graduate Research Assistant**
**Civil Engineering Department**
**Tennessee State University**

Transportation Research Center
*for Livable Communities*

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

**Technical Report**
**Documentation Page**

<table>
<tr><td>1. Report No.<br>TRCLC 14-07</td><td>2. Government Accession No.<br>N/A</td><td colspan="2">3. Recipient's Catalog No.<br>N/A</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle<br>Development of Decision Support Tools to Assess Pedestrian and Bicycle Safety: Focus on Population, Demographic and Socio-economic Spectra</td><td colspan="2">5. Report Date<br>June 30, 2016</td></tr>
<tr><td colspan="2">6. Performing Organization Code<br>N/A</td></tr>
<tr><td colspan="2">7. Author(s)<br>Deo Chimba, PhD, PE., PTOE<br>Abram Musinguzi</td><td colspan="2">8. Performing Org. Report No.<br>N/A</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address<br>Tennessee State University<br>3500 John A Merritt Blvd<br>Nashville, TN 37209</td><td colspan="2">10. Work Unit No. (TRAIS)<br>N/A</td></tr>
<tr><td colspan="2">11. Contract No.<br>TRCLC 14-07</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address<br>Transportation Research Center for Livable Communities (TRCLC)<br>1903 W. Michigan Ave., Kalamazoo, MI 49008-5316</td><td colspan="2">13. Type of Report & Period Covered<br>Final Report<br>7/1/2014 - 6/30/2016</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code<br>N/A</td></tr>
<tr><td colspan="4">15. Supplementary Notes</td></tr>
<tr><td colspan="4">16. Abstract<br>Despite the increase of these non-motorized trips, bicyclists and pedestrians remain vulnerable road users that are often over represented in traffic crashes. While the currently used methods that identify hazardous locations serve their purpose well, majority represent a reactive approach that seeks improvement after crashes happen. This research addressed these issues and proposed decision support tools to aid the implementation of bicycle and pedestrian safety strategies. This work developed an access based tool to predict the expected number of crashes at different neighborhood levels. This tool combines the traditional methods such as those provided in the Highway Safety manual to predict the expected number of bicycle and pedestrian crashes. First, a cluster analysis technique is proposed and developed a Geographic Information Systems (GIS) technique to facilitate the identification of high crash locations. Safety Performance Functions (SPFs) are developed in form of mathematical equations to relate the number of crashes to area socioeconomic and demographic characteristics. An integrated system consisting of access database and safety performance functions, and whose interface is designed to automatically compute the number of crashes given the input values is developed. Basing on crash value, the tool can be adopted as a framework to guide the appropriate allocation of safety improvement resources.</td></tr>
<tr><td colspan="2">17. Key Words<br>Key words up to five</td><td colspan="2">18. Distribution Statement<br>No restrictions.</td></tr>
<tr><td>19. Security Classification - report<br>Unclassified</td><td>20. Security Classification - page<br>Unclassified</td><td>21. No. of Pages</td><td>22. Price<br>N/A</td></tr>
</table>

## Disclaimer



## Acknowledgments

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

# Table of Contents

## List of Figures

## LIST OF TABLES



## ABBREVIATIONS

| | |
|---|---|
| TDOT | **T**ennessee **D**epartment of **T**ransportation |
| TRIMS | **T**ennessee **R**oadway **I**nformation **M**anagement **S**ystem |
| GIS | **G**eographical **I**nformation **S**ystem |
| SPF | **S**afety **P**erformance **F**unction |
| TIGER | **T**opologically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing |
| ACS | **A**merican **C**ommunity **S**urvey |
| BN | **B**ayesian **N**etwork |
| ANN | **A**rtificial **N**eural **N**etwork |

# CHAPTER 1: INTRODUCTION

## 1.1 Background

Tennessee is experiencing an exponential population growth. For instance, between 2000 and 2025 the population of the state of Tennessee is expected to grow by 32.9 percent. Yet, statistics indicate that walking and bicycle use has increased in the past decade in the United States [1]. Many communities are promoting walking and bicycling as alternative modes of transportation, while many states are supporting the use of these two non-motorized modes. The reasoning behind this new support for bicycling and walking is because of numerous benefits associated with these modes. For example, regular physical activity results in far reaching health benefits such as reduced risk of coronary heart disease, stroke, diabetes, and other chronic diseases. On contrary, neighborhoods with low physical activity such as areas with inaccessible or nonexistent sidewalks and bicycle or walking paths contribute to sedentary habits. These habits lead to poor health outcomes such as obesity, cardiovascular disease, diabetes, and some types of cancer [2, 3]. Moreover, walking is the most natural form of transportation in that almost everyone walks at some point during the day and it requires no equipment [4]. As one would expect, this increase in usage has led to analogous rise in the number of pedestrian and bicycle crashes. In 2012, there were 4,743 pedestrian fatalities while an estimated 76,000 pedestrians were injured in traffic crashes. Pedestrian deaths accounted for 14 percent of all traffic fatalities and made up 3 percent of all the people injured in traffic crashes. At the same time, 726 bicyclists were killed and an additional 49,000 were injured in motor vehicle traffic crashes. Bicyclist deaths accounted for 2 percent of all traffic fatalities and made up 2 percent of the people injured in traffic crashes during the year [5]. Despite the potential for reversing them with education, engineering and enforcement solutions, pedestrian and bicycle crashes remain a critical issue in the United States and other parts of the world. While the contributing factors to these crashes vary broadly, literature implicates socioeconomic and demographic as important factors contributing to these crashes.

The topic that this research seeks to put forward in discussions around pedestrian and bicycle safety more particularly with respect to socioeconomic and demographic factors. This concern stems from observations of pedestrian and bicycle safety falling on neighborhoods inhabited by groups more likely to be vulnerable such as low-income populations and racial minorities. These sociodemographic groups are also overrepresented among bicycle and pedestrian crash injuries and fatalities. Therefore, national , state and local agencies seeking to promote walking and bicycling should then look for solutions to make neighborhoods livable, where transportation, housing and commercial development investments can be coordinated so that people have access to adequate, affordable and environmentally sustainable travel options. This research seeks to answer the question of pedestrian and bicycle safety distribution varies among different socioeconomic and demographic groups with some groups more vulnerable to crashes than others. This works seeks to raise a similar question of what criteria can help decision making and planning in identifying sociodemographic groups likely to high pedestrian and bicycle crashes.

The above argument thus leads to the question of whether neighborhoods more vulnerable to high pedestrian and bicycle crashes receive their “fair share” of safety enforcement resources compared to other areas. For example, do roadways that serve as geographical boundaries of low-income census block groups and census tracts, communities with high poverty levels and racial minorities have a higher or lower share of pedestrian and bicycle safety projects enforcement resources than their counterparts? This research, thus seeks to address this concern

by developing a framework that would assist in allocation of safety enforcement resources. This is particularly important as many states and agencies are faced with challenge of limited funding resources and therefore are looking for ways of making effective use of available resources.
In summary, there is a dire need for research into the geographical distribution of bicycle and pedestrian safety, in particular, the importance that certain sociodemographic groups are often overrepresented in crashes. This research addresses these issues by exploring which locations and neighborhoods present disproportionate vulnerability for pedestrians and bicyclists. Potential associations between sociodemographic factors and pedestrian/bicycle related motor vehicle crashes are examined. Finally, a criterion is developed to identify high crash location that would assist in allocation of safety enforcement resources as well as developing effective safety countermeasures to alleviate bicycle and pedestrian crashes.

## 1.2 Problem statement

As the population continues to grow in Tennessee State, so does the number of cyclists, pedestrians and vehicles on Tennessee roadways. Moreover, with continuous fluctuation of fuel prices and economic uncertainties, many communities are seeing large increases in bicycling and walking as alternative modes of transportation. Although bicycling and walking are increasingly becoming popular, the users of these non-motorized modes are vulnerable and often exposed to severe injury traffic crashes. Tennessee has experienced an average of over 1000 pedestrian crashes and over 400 bicycle crashes per year over the last decade of which nearly 7 % of these crashes resulted in human fatalities. Despite the State's efforts to enforce countermeasures against crashes involving the users of these two modes, there is still a critical pedestrian and bicyclist safety issue on many roadways in Tennessee with some counties being ranked worse than others. In particular, Shelby, Davidson, Knox, Hamilton and Montgomery remain the worst performing with the majority of these crashes happening in these counties. The occurrences of these crashes result in significant health and human life consequences ranging from disability to fatality and such crashes impose significant monetary impacts not only to the state but also national economy. Most of these severe crashes are avoidable; as such it is very important to address bicycle and pedestrian safety as part of improving community livability. One way of addressing safety issue is to develop a framework for identifying possible high crash locations for bicycle and pedestrian related crashes. Therefore, this research aimed at developing a framework to identify bicycle and pedestrian high crash locations for safety improvement prioritization focusing on population, demographic and socioeconomic spectra with the state of Tennessee as a case study. This research comprised of in-depth analysis using existing data, conducts GIS cluster analysis and statistical modeling to examine and identify bicycle and pedestrian high crash locations. The study developed safety performance functions to identify magnitude and characteristics of variables associated with pedestrian and bicycle safety and hazardous locations. From the developed safety performance functions, this research developed criteria for identifying pedestrian and bicycle high crash locations and framework to prioritize allocation of safety improvement resources.

## 1.3 Research questions

This research was based on a stand point that planning and safety strategies of pedestrian and bicycle should aim at attaining livability objectives, whereby one of their goals should be to reduce sociodemographic disparities in terms of pedestrian and bicycle safety. This required an understanding of the relationship between the decision-making process that comprises the act of planning for pedestrian and bicycle safety improvements, the sociodemographic status of neighborhoods in Tennessee, and the outcomes in terms of the geographic distribution of

pedestrian and bicycle safety. And thus promoting bicycling and walkable neighborhoods must be achieved under the premise of providing safe environments for people to walk and that safety programs should give rise to policies that not only result in fewer but also less severe crashes.
To be more specific, this research labored to address the key question whether there is a geographical difference in the distribution of pedestrian and bicycle safety. Traffic safety is understood to prevail among certain sociodemographic groups while others are highly vulnerable to traffic crashes. The study used Tennessee as a case study and answered this question broken down into in three specific questions:

1. First, are there spatial variations in pedestrian and bicycle crashes with respect to socioeconomic and demographic characteristics? Spatial clustering was examined at the level of census block group and county.
2. Second, how do such socioeconomic and demographic factors associate with variations in the geographic distribution of pedestrian and bicycle crashes? The main characteristics examined at the level of census block group and county in this research are: median household income, percentage of population walking and bicycling to work, percentage of white and non-white populations, population density, and percentage of households with no access to a vehicle.
3. Third, what framework can decision makers use as a planning tool to assess bicycle and pedestrian safety? What criterion can be adopted by a decision maker in office or on site to identify high crash location that would assist in allocation of safety enforcement resources as well as developing effective safety countermeasures to alleviate crashes?

## 1.4 Objectives

The main objective of this research project was to develop "decision support tools to assess pedestrian and bicycle safety" in Tennessee. The tool will help in the development of pedestrian and bicycle safety programs that could be adopted assist not only Tennessee agencies but also nationally in better understanding of the causes of crashes and identifying appropriate operating strategies to enhance of pedestrian and bicycle safety.

1. To conduct cluster analysis in GIS to verify any spatial clustering and identify high crash locations within the spectra of socioeconomic and demographics.
2. To develop Safety peformance functions (SPFs) to examine relationships between bicycle/pedestrian crashes and associated factors.
3. To develop criteria for high crash location identification and a framework to prioritize funding of bicycle and pedestrian safety improvemnets.

## 1.5 Research Organization

This research is organized into seven chapters. The first chapter defines the research problem and explicitly states the study questions. In addition, it also states the general and specific objectives of this study. The second chapter is a literature review of previous efforts on bicycle and pedestrian high crash locations and explores both GIS and statistical methodologies. The third chapter describes data sources and the data collection process. In addition, it conducts descriptive statistics to analyze trends in the data. The fourth chapter describes cluster analysis procedure to verify any spatial clustering and identify high crash locations within the spectra of socioeconomic and demographics. The fifth chapter elaborates the development of safety performance functions to examine relationships between bicycle/pedestrian crashes and associated factors. The sixth chapter describes a criteria for high crash location identification and a framework to prioritize safety improvemnet resources. In the seventh chapter, which is also the final chapter, provides conclusions and based on the findings and discussions and gives recommendations for safety planning and future research.

# CHAPTER 2: LITERATURE REVIREW

## 2.1 Overview

The literature review summarizes methodologies and findings from previous and ongoing research related to pedestrian and bicycle safety analysis. Although several analytical tools are available to analyze crash data, literature review presented in this research focuses on the use of a Geographical Information System (GIS) methodology to study the spatial patterns of pedestrian and bicyclist crashes in order to identify high pedestrian crash zones. In addition, this review summarizes previous research on statistical modeling of pedestrian and bicycle crashes in order to understand the impact and significance of factors that influence the crash frequency and degree of injury severity.

## 2.2 GIS Application in Pedestrian and Bicyclist Safety Analysis

Recent safety research has widely applied GIS techniques to identify high crash zones. These techniques have been applied because they turn statistical data, such as traffic crashes and geographic data, such as roads and crash locations, into meaningful information for spatial analysis and mapping [6]**.** The identification of traffic crash hot spots provides insights into the casual factors and is an essential step for appropriate allocation of safety improvements resources [7]. Moreover, identification of high crash zones enhances better understanding of spatial patterns and clusters in crash data and this enhances the development of effective safety improvement strategies. For instance, corridors with a high incidence of truck crashes were identified using GIS [8]. This study further suggested that GIS techniques such as hotspot, cluster, and corridor analysis have great potential to improve crash location evaluation. Literature presents a number of GIS tools available to analyze pedestrian and bicyclist crashes. All these tools seek to answer the questions of “why and where most of the crashes occur?” GIS programs provide valuable tools to answer these questions [9]. However, using GIS requires the analyst to first geocode and represent crash locations on digital maps. Mapping crashes helps to identify spatial patterns, which gives an advance to exploring the causes of patterns observed as well as knowing where there is need to take action. Crashes are not just spatially random events since some areas experience a higher number, while others experience fewer or even none. There is an overall pattern to crashes which makes their locations more than just an irregular spatial distribution [10] and hence, there could be geographical factors responsible for these patterns and can be well studied using GIS. Generally, identifying spatial patterns in GIS can be categorized into two groups depending on their output. The first category consists of global measures such as Ripley’s K-function, Getis’s G-statistic and Moran’s I. These global measures are used to test whether a given point distribution differs from a random distribution. They only examine if there is a general tendency of features such as crashes to cluster but do not reveal the location of clusters within the distribution [11]. The second category consists of local measures such as kernel density and the local-autocorrelation methods. These identify exact position of a cluster within a section or a network. The methods from second category are more efficient as they are concerned with spatial dependencies on a localized scale [11]. To develop effective strategies, researchers in spatial modeling must not only account for dependence structure and spatial heteroskedasticity, but also assess the effects of spatial scale [12]. Therefore, kernel method or the local spatial autocorrelation methods are commonly used techniques for determining hazardous locations in traffic crash analysis [13, 14].

#### 2.2.1 Kernel Density technique

Kernel density is a GIS spatial statistic tool that can be used to calculate a magnitude per unit area from point features such as crash data. Pedestrian crash density calculations are performed using the simple method or kernel density Estimation (KDE) method [15]. The simple method divides the entire study area to predetermined number of cells and draws a circular neighborhood around each cell to calculate the individual cell density values, which is the ratio of number of features that fall within the search area to the size of the area. The kernel method divides the entire study area into predetermined number of cells but instead of considering a circular neighborhood around each cell as in simple method, the kernel method draws a circular neighborhood around each feature point (in this case a crash).

Kernel density calculates a magnitude per unit area from point or line features using a kernel function to fit a smoothly tapered surface to each point or line. Kernel density estimation involves placing a symmetrical surface over each point and then evaluating the distance from the point to a reference location based on a mathematical function and then summing the value for all the surfaces for that reference location as shown in Figure 1 This procedure is repeated for successive points and allows placing a kernel over each crash observation and by summing these individual kernels gives the density estimate for the distribution of crash points [7]. The kernel function can be expressed as;

$$\boldsymbol{f(x, y) = \frac{1}{n\tau^2} \sum_{i=1} k\left(\frac{d_i}{\tau}\right)} \tag{1}$$

Where;

*f (x, y):* density estimate at the location (x, y),

*n* : number of observations,

$\boldsymbol{\tau}$: Search radius or bandwidth,

*k:* kernel function, and

$d_i$: the distance between the location (x, y) and the location of the $i^{th}$ observation of features that fall within the search area.

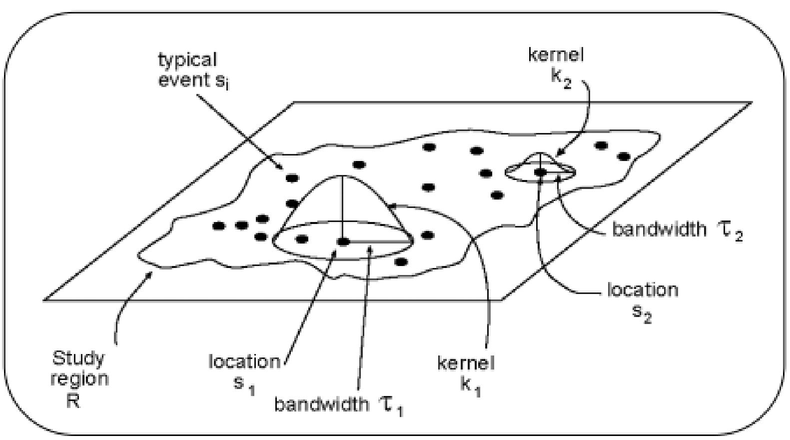


**Figure 1: Principle of kernel density**

(Source: Erdogan, et al. 2008)

Kernel density helps in determining the spread of risk of a crash. The spread of risk can be defined as the area around a defined cluster in which there is an increased likelihood for a crash to occur based on spatial dependency [7]. Furthermore, kernel density overcomes limitations of simple crash mapping using a dot map. The commonly used dot maps to represent crashes as one dot no matter how many occurred on a given location. Kernel density estimation solves this problem by representing crashes per unit area using a density map feature in GIS [16]. The result

of KDE analysis is a map with the intensity of pedestrian crashes represented on continuous surfaces, with darker shades representing locations characterized by the highest crash density and lighter shades representing locations with a lower crash density [17].
Literature implicates that the accuracy of kernel density depends on appropriate selection of search radius or bandwidth and the cell size. The search radius affects the resulting density map such that larger values of the search radius parameter produce a smoother, more generalized density raster while smaller values produce a raster that shows more detail [16, 18]. In addition, when calculating the density, only the points that fall within search neighborhood are considered. Hence, the search radius produces significant effects to the results of the density map. The choice of bandwidth and grid cell size is somewhat subjective and thus depends on the judgment of the analyst.

Different researchers adopted different criteria to select the appropriate search radius, for instance, [7] considered a search radius, which is two times the size of the grid cell. Using kernel density estimation and a search radius of 100m, [11] identified hazardous road locations of traffic accidents. [19] identified critical areas with high child pedestrian crash risk using kernel density estimation and found no statistical significance between child pedestrian crashes with respect to gender, weekday, and month of the year. [18] explored the effect of search radius when using kernel density and observed that bandwidth exerts great impacts on the network density pattern and proposed that narrow bandwidth (100m and 250m) preserves local features and is therefore more appropriate for identifying crash clusters at precise locations.
Although most GIS tools have great potential to locate crashes on a digital map, they have been criticized for not having statistical methods apart from means and standard deviations of variables. Therefore, to develop tools that are more robust, some researchers proposed combining GIS and statistical methods [20]. This technique has been applied in previous research to analyze traffic crashes such as; [7, 10, 13, 20]. For instance using GIS, [21] located crash clusters on roadway networks using geospatial tools and applied statistical methods to model the relationships of contributing factors. The results of this study revealed that, pedestrian and bicycle crash frequencies were correlated with percentage distribution of population by race, age groups, and mean household income, percentage in the state of Tennessee. Similarly, Colorado Department of Transportation applied spatial statistics to develop safety performance for intersections [22].

### 2.3 Statistical Modeling Pedestrian and Bicyclist Safety Analysis

There is a vast body of safety literature examining the factors affecting pedestrian and bicycle crash occurrence and their severity levels with motorized vehicles using statistical techniques. The results of these studies indicate that the probable causes of pedestrian and bicycle crashes vary widely. Regression analysis has been widely used to determine significant factors that influence crash occurrence. The most commonly used regression models in crash analysis are the logistic regression models [23, 24, 25, 21, 26, 19].

#### 2.3.1 Regression models

A mixed generalized ordered response logit (MGORL) model was applied to analyze the 2004 General Estimates System (GES) database and focused their analysis on crashes that involved pedestrians or bicyclists [27]. Their results found that elderly population, higher speed limits and darker times of the day lead to higher injury severity, while crashes that occur at signalized intersections are less severe than those that occur elsewhere. A combination of regression analysis and ordinary least squares (OLS) were applied to examine geographical relationships between environmental and demographic characteristics of the City and County of San Francisco [26]. This study observed that pedestrian injury rates were related to traffic flow, population density, age composition of the local population, unemployment, gender and education. A study conducted in in Kwun Tong District of Hong Kong found that land mix and socio-economic deprivation index were more associated with the occurrence of serious and slight injuries [28]. Furthermore, using a method of frequency item sets to study characteristics of "black" zones [29] found that a collision with pedestrian involving young road users inside the built-up area is a typical accident pattern that frequently occurs inside a "black" zone. Literature implicates that most safety programs have aimed at reducing the frequency of pedestrian–vehicle collisions, but few have focused specifically on reducing the risk of severe injury or death [30]. Studies indicate that the speed of cars, right turns on red at intersections, and geometric characteristics and are major risk factors for pedestrian injuries, particularly among young children and older adults [31, 32]. Despite these known hazards and the potential for reversing them with engineering solutions, pedestrian injuries remain a critical issue in the United States and other parts of the world. For instance, in 2012 there were 4,743 pedestrian fatalities, accounting for 14 percent of all traffic fatalities, and an estimated 76,000 injured in traffic crashes in the United States [5]. In 2006, nationwide U.S. pedestrian fatalities constituted 11 percent of total crash fatalities and pedestrian crash fatalities were decreasing but at a slow rate [33].

Reducing rates of pedestrian and bicyclist injuries seems timely, particularly in the as many communities and urbanizing areas are seeing large increases in walking and bicycling as an alternative mode of transportation. To deepen our understanding on injury severity analysis, we explored the vast literature available to uncover findings in existing studies to serve as evidence on the importance of previous studies and their significance in modeling traffic pedestrian and bicycle injury types and crash severity with a view to develop and adopt designs that can alleviate pedestrian injury and injury severity. Road environmental characteristics were positively associated with pedestrian fatalities at unsignalized zebra crosswalks in Poland [32]. This study found that roads with no street lighting, divided road, two-way roads, non-built-up area, and posted speed limit increase the probability of pedestrian death. The effect of road environmental characteristics on pedestrian severity was further explored in traffic safety facts 2012 data. These statistics indicate that in 2012, 73 percent of pedestrian fatalities occurred in an urban setting, 20 percent of pedestrian fatalities occurred at intersections while 89 percent of pedestrian fatalities occurred during clear/cloudy weather conditions [5]. Injuries in pedestrian crashes were greater in the areas with higher population density, average daily traffic, and number of cross-streets per kilometer roadway in the City of San Francisco, California [26] observed that injuries.

Studies have indicated that younger and older pedestrians are over represented in collisions ending in fatality [34, 30, 31, 32]. Another study observed that the number of fatal pedestrian crashes reduced when the speed of limit was reduced [35]. Similarly higher the posted speed limit was associated with higher probability of a pedestrian fatality [30]. The characteristics of the area or the neighborhood such a census block have been found associated with injury severity. [30] found 25% of fatal pedestrian crashes occurred in rural neighborhoods. A study showed that the characteristics of the local environment have a powerful influence on pedestrian casualties in England [36]. This study found that incidence of pedestrian causalities and injuries is higher in residential than economic zones. Additionally, the study found a quadratic relationship between urban density and pedestrian casualties with incidents diminishing for the most extremely dense wards. [37] examined the impacts of environmental attributes associated with pedestrian vehicular crashes near public schools and found that the presence of a driveway or turning bay on the school entrance decreases both crash occurrence and injury severity. This study further found that the presence of recreational facilities on the school site was positively associated with crash occurrence and injury severity of crashes.

### 2.3.2 Bayesian network techniques

Bayesian network are defined as a directed acyclic graph model annotated with probability that can express a joint probability distribution of a large set of variables [38]. These graphical structures are used to represent knowledge about an uncertain domain such effects of socio-demographic variables on pedestrian or bicycle crashes. Bayesian networks (BNs), also known as belief networks belong to the family of probabilistic graphical models. In particular, variables are represented as nodes of a graph and the interactions (direct dependences) as directed links (arcs) between the nodes as shown in Figure 2. Any pair of unconnected nodes of such a graph indicates (conditional) independence between the variables represented by these nodes under particular circumstances that can easily be read from the graph. Each node contains the states of the random variable and it represents a conditional probability table. The conditional probability table of a node contains the probabilities of the node being in a specific state, given the states of its parents. BNs combine principles from graph theory, probability theory, computer science, and statistics generate conditional dependencies the graph.


(a) (b)

**Figure 2: Example of BN; Divergent (a) and Convergent (b) diagram**

In Bayesian statistics, probability is not defined as the frequency of the occurrence of an event, but as the plausibility that a statement is true, given the information. This allows one to assign probabilities to propositions instead of only to random variables. The model is probabilistic which makes it possible to include factors that influence the frequency of events, but do not determine their occurrence. This is of great advantage when analyzing crash data to learn the causal relationships and hence can be used to predict the consequences of intervention.
Bayesian modeling is an approach to learning functions of form $f : X \rightarrow Y$ or P($Y$|$X$). Whereby $Y$ is a discrete valued random variable and $X$ is any vector containing discrete or continous variables. The model is used to classify groups of value in a data, typically a categorical data such as crash been severe or non-severe or occurrence of a crash in a particular block group or not. In general, Bayes rule can be expressed by the following equation

$$Posterior = \frac{likelihood * prior}{marginal\ likehood} \qquad (2)$$

$$p(y_i = 1 | \mathrm{x}_j) = \frac{p(y_i = 1) p(x_j | y_i = 1)}{p(y_i = 1) p(x_j | y_i = 1) + p(y_i = 0) p(x_j | y_i = 0)} \quad (3)$$

Whereby $p(y_i = 1 | \mathrm{x}_j)$ is the Conditional Probability of an event $y_i$ occurring, given the occurrence of an event $\mathrm{x}_j$.

Bayesian inference is an emerging approach in traffic safety investigation [39, 40, 38, 41]. [42] for instance, proposed a hierarchical Bayesian approach to evaluate effectiveness of safety seat belts in preventing both serious injury and reducing the likelihood of fatality. This study identified that, restrained occupants are less likely to suffer fatal or serious injuries compared to the unrestrained occupants. [43], proposed Bayesian modeling approach to evaluate cyclist injury occurrence and bicycle activity at signalized intersections and found that more cyclists at an intersection result into more cyclist injuries but lower injury rates due to the non-linear association between bicycle volume and injury occurrence. Bayesian inference offers an advantage of prior information about independent variables, which can be included in the inference procedures. Researchers in traffic safety suggest that such prior information can be used to improve the process of developing safety performance functions. For instance, [44] adopted hierarchical Bayesian and evaluated effects of incorporating informative priors in developing safety performance functions. [45], explored the application of Bayesian methods to derive an estimate for the expected safety of the treatment site. Related studies have suggested using posterior prediction capability of Bayesian approaches to estimate crash reduction factors. [46], for instance applied full Bayes approach to analyze the effect of road safety countermeasure. Bayesian methods can be extended into a “Bayesian network” which is a directed acyclic graph model annotated with probability that can express a joint probability distribution of a large set of variables. This offers a number of advantages as elaborated by [47] that, when the graphical model is used in conjunction with statistical approaches for data analysis it yield numerous advantages; First, that “BN can be used to learn causal relationships, and hence can be used to gain understanding about the problem domain and to predict the consequences of intervention”. Second, “because the model has both a causal and probabilistic semantics, it is an ideal representation for combining prior knowledge and data”.

Therefore, researchers in traffic safety have picked great interest in utilizing capabilities of BNs to discover the underlying patterns of crash incident data, to investigate the relationships between contributing variables and to make predictions using these relationships. [39], highlighted that the application of Bayesian Networks in road safety performance prediction is gaining acceptance because they model uncertainties involved in the factors that can lead to road crashes. By applying the knowledge of Bayesian network, [48] identified the top three crash contributors on Nigerian roads as road condition, driving under influence and reckless driving.

### 2.3.3 Application of Artificial Intelligence in safety analysis

Recently artificial intelligent techniques, such as neural network, fuzzy logic and genetic algorithms, have gained popularity in diagnostic task to correctly interpret accident data [49, 50, 51, 52]. However, each of artificial intelligence techniques has its own advantages and limitations. Artificial Neural Networks (ANN) are biologically inspired and highly sophisticated analytical techniques capable of modelling non-linear functions [53]. They have been widely applied as modelling approach because of their ability to learn nature of the data.

An ANN learns from examples through the process of updating its architecture and connection weights to perform the required task [50]. [54] for example utilized artificial neural network to analyze freeway accident frequencies. The study compared the prediction performance between the negative binomial regression model and ANN model and found ANN is a consistent alternative method for analyzing freeway accident frequency. [51], applied a series of ANNs to model the potentially non-linear relationships between the injury severity levels and crash-related factors.

Although ANN modeling has many advantages in comparison to analytical and statistical techniques, its major limitation is that ANN is a “black box” in a sense that it does not give the effect of each independent variable to the response variable [55]. Fuzzy logic modelling takes care of this limitation by representing data using rules rather than by specific equations. These rules have the form IF-THEN, based on the intuitive knowledge of experts and operators in the field. For instance, [56] developed two fuzzy logic models for predicting the risk of accidents that occurred on wet pavements using Mamdani and Sugeno inference methods respectively. The results of this study indicate that fuzzy logic model shows superiority over the probabilistic model and the nonlinear regression model. [57] developed a fuzzy logic prediction model for urban traffic accident with traffic and road conditions. In this study, Fuzzy logic proved a viable model by showing a good relationship between observed numbers and predicted numbers. However, fuzzy logic systems have also been criticized because of their difficulty to construct a complete fuzzy rule set for fuzzy logic and is time consuming. Another limitation is that, contrary to ANN, which learns the data by utilizing the already available output, fuzzy logic does not and this sometimes results in a big discrepancy between the observed and the predicted values. To overcome these limitations of the two techniques described above is to use adaptive neuro-fuzzy inference system (ANFIS) which combines the benefits of the two machine learning techniques (Fuzzy Logic and Neural Network) into a single technique [58]. Consequently, this combination compensates the limitations of the other and fully makes use of the excellent characteristics of neural network and fuzzy inference system. This method has been widely applied in traffic safety analysis [59, 60], health [61], power systems [62] and many other areas. However, literature offered barely any extensive application of ANFIS in pedestrian crash prediction.

#### 2.3.4 Conclusions from literature review

Despite the significance of previous studies, some research gaps still exist. In most emerging literature, pedestrian and bicycle injury occurrence and sociodemographics have not been extensively studied most especially at a smaller analysis unit. Pedestrian safety studies at analysis unit of a census block group have been rarely conducted in North America. Block groups are geographical units created by US Census Bureau as clusters of census blocks and generally comprise between 600 and 3,000 people. Local indicators of spatial association can be studied better at this analysis unit. While a few studies have been carried out in the United States, these have mainly focused on pedestrian injuries at county, city, or census tract level and did not focus on block group as the unit of study for example [21].

Although commendable efforts have been dedicated to alleviate these crashes through policy and funding, it is still unclear to what extent different communities will take advantage of this funding opportunity, or how they will modify their budgets to make better use of already available funding sources. A challenge still remains on criteria for allocate resources to implement pedestrian and bicycle safety strategies [63]. This research builds on this literature to extensively study the impacts of socioeconomic and demographic factors on pedestrian safety with a goal of developing a decision support tool for the implementation of pedestrian and bicycle safety strategies.

# CHAPTER 3: STUDY DATA

## 3.1 Overview

To achieve the objectives of this research, efforts to collect appropriate data are herein summarised in this chapter. The data collected is categorised in three types:

- Crash data
- Socioeconomic data
- Demographic data

The research team employed ArcGIS 10.1 and compiled all data in the GIS database. The GIS database enabled the research team to manipulate the data in forms that allowed appropriate analysis. The analysis was performed at three geographic levels including; census block group level, county level and State level. Therefore data collected had to be processed at these three levels.

## 3.2 Crash Data Collection

The crash data was obtained from Tennessee Roadway Information Management System (TRIMS) database maintained by TDOT. The data from the TRIMS database contained some micro-level information about crashes, such as beginning log mile, case number, person type, injury type, county, route, location, type of crash, year of crash, time of crash, total killed, total incapacitating injuries, manner of first collision, total injured, first harmful event, light conditions, weather conditions, relation to first junction, relation to first roadway, urban or rural and hit and run. The crash data was extracted from TRIMS database using search queries as illustrated in Figure 3. A non-motorized crash in this study is defined as any crash that involves a pedestrian or bicycle. This study initially collected all the 5845 pedestrian and 2,185 bicycle crashes that occurred in a five-year period from 2008 to 2012 in the entire State.

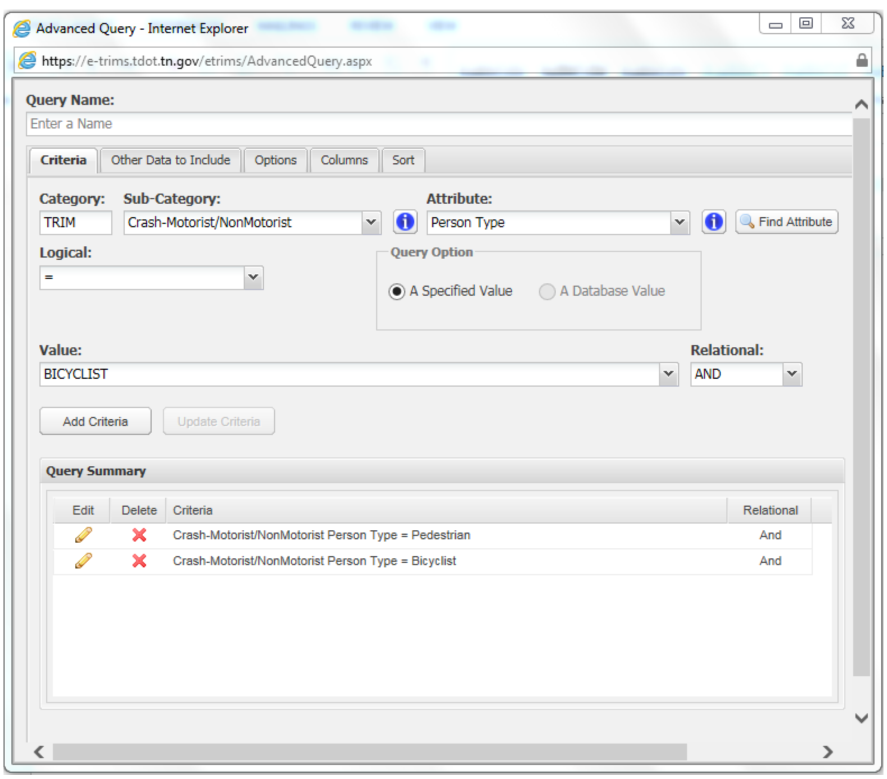


**Figure 3: Typical E-TRIMS Search Query**

### 3.3 Mapping of crash data

Geocoding is a very important process in crash analysis as it enables one to map locations from crash data that is readily available. Using GIS to geocode crash locations and plot the locations is the most common first step [64]. Geocoding is the process of automatically creating map features based on address, or similar information exploring the capabilities afforded by GIS software. Crashes can be geocoded using one of the three reference systems, street name or reference street name, milepost and address. The street name or reference street name and address are most commonly used in urban areas. For TDOT; the location code indicates county, route number, and log mile. The log mile makes it possible for TDOT to estimate the position of the crash along the street segment. Since crash data were readily available from TRIMS website in a shape file format, the tasks demanded by this research project were to verify geocoding of crashes. Out of the initially downloaded 5,845 pedestrian crashes, 4816 (approximately 82 %/) pedestrian crashes were accurately with geocoded. On other hand, out of 2,185 downloaded bicycle crashes, 1,808 (approximately 83%) were accurately Geocoded. Only Geocoded crashes were therefore maintained for subsequent analysis. Each crash is represented on the digital map by a symbol such as a dot such as maps are shown in Figures 4 and 5. Mapping pedestrian and bicycle crashes indicates that pedestrian and bicycle crash occurrence was varying across the State with some counties having more crashes than others. The counties with more crashes are; Shelby, Davidson, Knox, Hamilton, Montgomery and Sullivan.

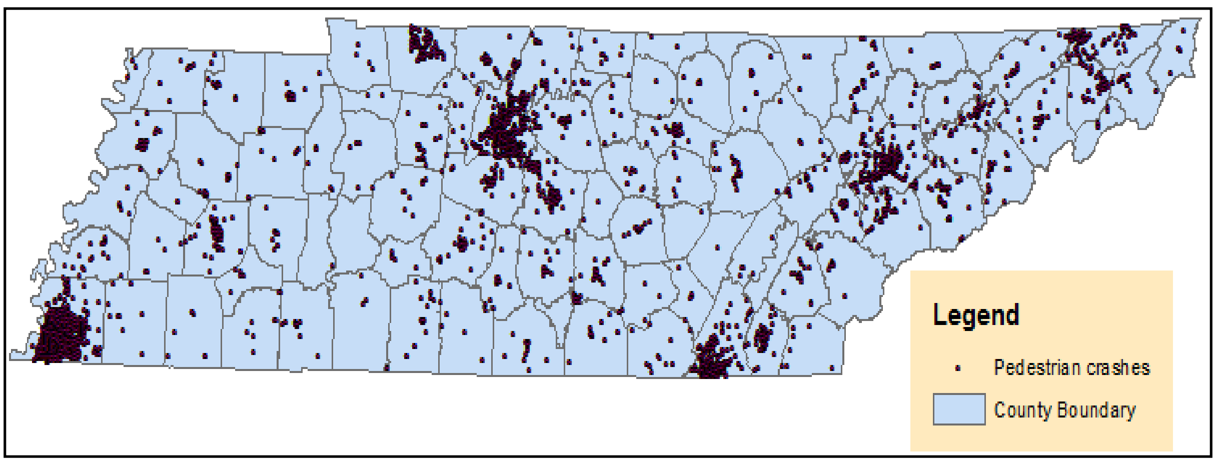


**Figure 4: Pedestrian crash distribution**

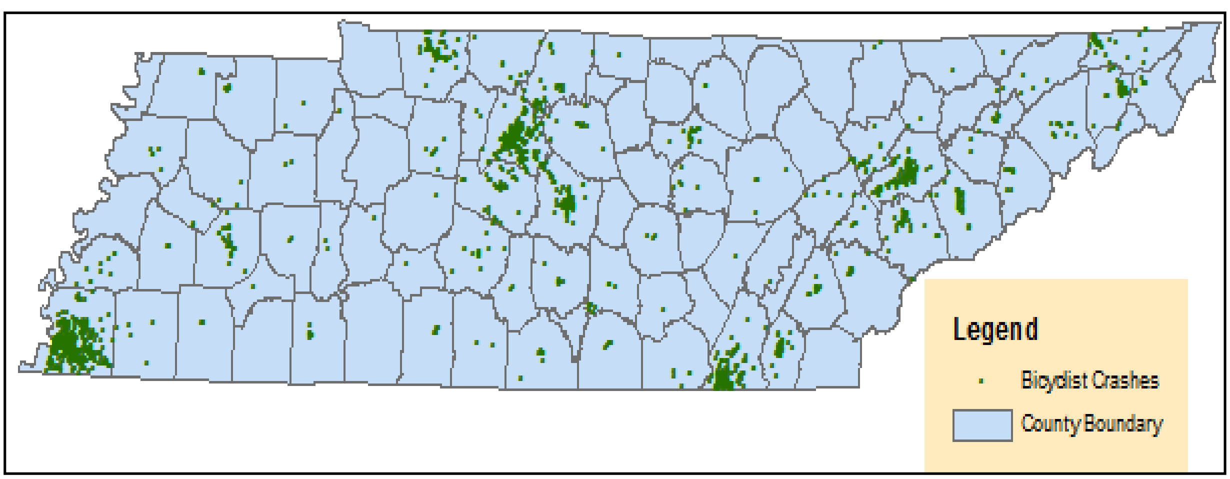


**Figure 5: Bicycle crash distribution**

### 3.4 Descriptive statistics

#### 3.4.1 Observed annual crash frequency

During the analysis period (2008 – 2012), there were 5845 pedestrian crashes and 2185 bicycle crashes in Tennessee. Figure 6 depicts pedestrian and bicycle crashes that occurred during a five year period. As shown in the figure, the trends in crash frequency between pedestrian and bicycle crashes were quite different in two aspects. First, the pedestrian crash frequency was considerably higher than that of bicycle crashes over the study period. Second, pedestrian crashes experienced a gradual increase from 2008 to 2012, while bicycle crashes fluctuate annually to conform to a natural variability in crash frequency, also called regression to the mean as indicated in the literature review.

**Table 1: Non-Motorized crashes in Tennessee from 2008 to 2012**

| Year of Crash | Pedestrian | Bicyclist |
|---|---|---|
| 2008 | 1091 | 450 |
| 2009 | 1101 | 405 |
| 2010 | 1185 | 385 |
| 2011 | 1241 | 487 |
| 2012 | 1227 | 458 |
| Grand Total | 5845 | 2185 |

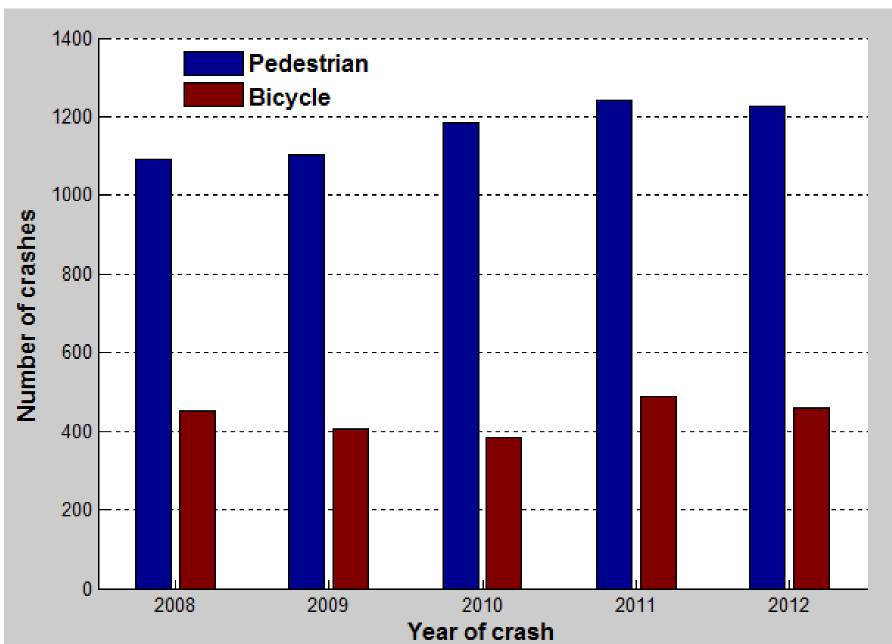


**Figure 6: Non-Motorized crashes in Tennessee from 2008 to 2012**

#### 3.4.2 County Wide Analysis

As illustrated in Figure 7, the number of pedestrian and bicycle crashes are different among counties. The highest number of crashes was observed in Shelby, Davidson, Knox, Hamilton, Montgomery and Rutherford, while overall; Shelby County experienced the highest number of pedestrian and bicycle crashes during a five year period. The non-motorized crash frequency pattern is the same for the counties with the highest number of crashes. It appears that crash frequency is higher for counties with major cities in Tennessee. This could reflect the high pedestrian and bicycle exposures in these cities.

**Table 2: Non-Motorized crashes by County**

| County | Pedestrian | Bicyclist |
|---|---|---|
| Shelby | 1809 | 517 |
| Davidson | 1189 | 381 |
| Knox | 404 | 196 |
| Hamilton | 389 | 186 |
| Montgomery | 158 | 74 |
| Rutherford | 135 | 153 |
| Others | 1761 | 678 |
| Grand Total | 5845 | 2185 |

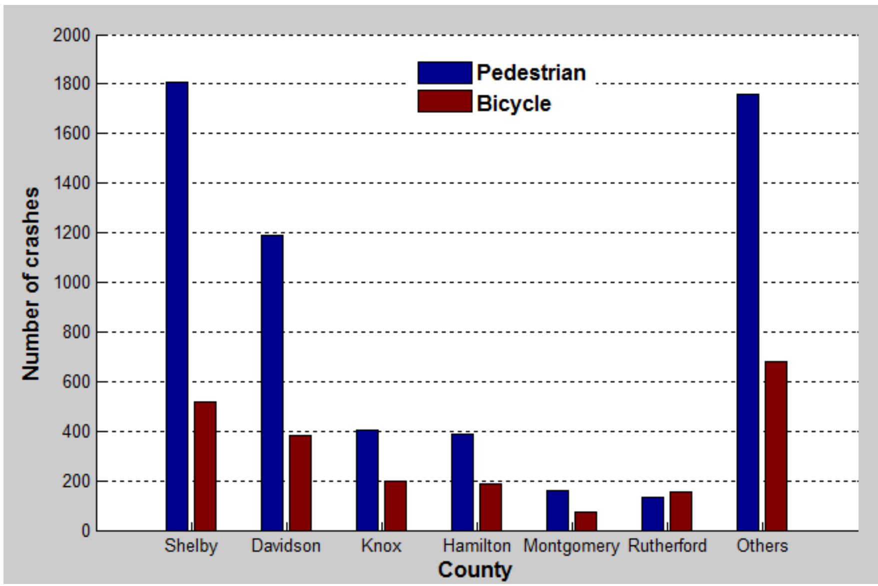


**Figure 7: Non-Motorized crashes by County**

### 3.4.3 Crash frequency by severity

Out of 5,845 pedestrian crashes, 389 were fatal, 1109 incapacitating, 4051 non-incapacitating, 118 Property damage (Over) and 178 were property damage (under). TDOT defines a Property damage (Over) crash as the one whose worth is $ 400 and above and property damage (under) as the one whose worth is below $ 400. Out of the 2185 total crashes, 33 were fatal, 279 incapacitating, 1603 non-incapacitating, 115 Property damage (Over) and 155 were property damage (under). As illustrated in Table 3 and Figure 8, the majority of non-motorized crashes are non-incapacitating crashes, representing approximately about 69% and 73% of pedestrian and bicyclist crashes respectively.

**Table 3: Non-Motorized crashes by County**

| Type of Crash | Pedestrian | Bicyclist |
|---|---|---|
| Fatal | 389 | 33 |
| Incapacitating Injury | 1109 | 279 |
| Non- Incapacitating Injury | 4051 | 1603 |
| Prop Damage (over) | 118 | 115 |
| Prop Damage (under) | 178 | 155 |
| Grand Total | 5845 | 2185 |

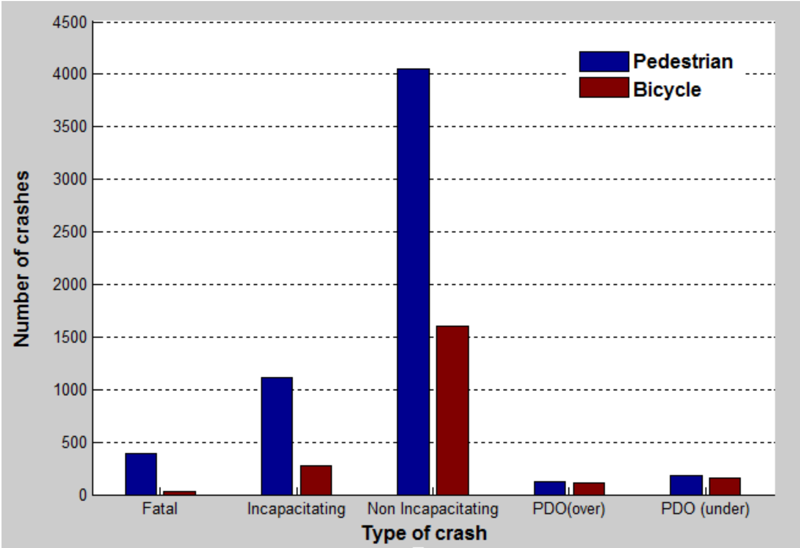


**Figure 8: Non-Motorized crashes by Type of crash**

### 3.4.4 Weekly crash frequency

As shown in Figure 9, the weekly crashes increase from Monday through Friday, reaching a peak on Friday, and drop on Saturday and Sunday.

**Table 4: Non-Motorized crashes by day of the week**

| Day of Week | Pedestrian | Bicyclist |
|---|---|---|
| Sunday | 565 | 223 |
| Monday | 866 | 324 |
| Tuesday | 840 | 319 |
| Wednesday | 888 | 316 |
| Thursday | 885 | 349 |
| Friday | 1004 | 372 |
| Saturday | 797 | 282 |
| Grand Total | 5845 | 2185 |

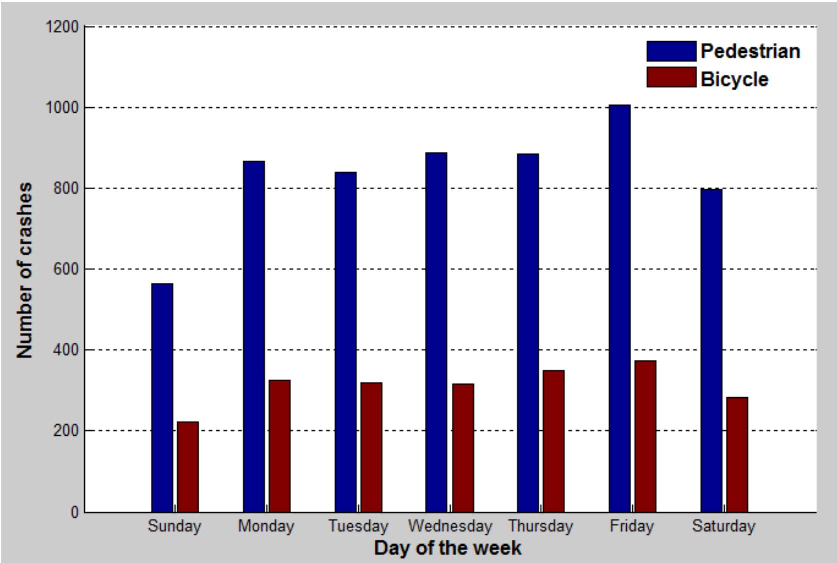


**Figure 9: Non-Motorized crashes by day of the week**

### 3.4.5 Crash frequency by road location

Table 5 and Figure 10 indicate more than half of the crashes occurred at intersections while relatively larger portion occurred along the roadway. Majority of crashes occurred at intersection and along the roadway. It can be seen that 52.32% of the pedestrian crashes occurred at intersections while 346.84% of them occurred along the roadway. Likewise 66.91 % of the bicycle crashes occurred at intersections as compared to 32.59 % that occurred along the roadways. Therefore, it is observed that a large proportion of bicycle crashes occur at intersection compared to pedestrian crashes.

**Table 5: Non-Motorized crashes by Road Location**

| Location | Pedestrian | Bicyclist |
|---|---|---|
| Along Roadway | 46.84 | 32.59 |
| At an Intersection | 52.32 | 66.91 |
| Others | 0.84 | 0.50 |
| Grand Total | 100.00 | 100.00 |

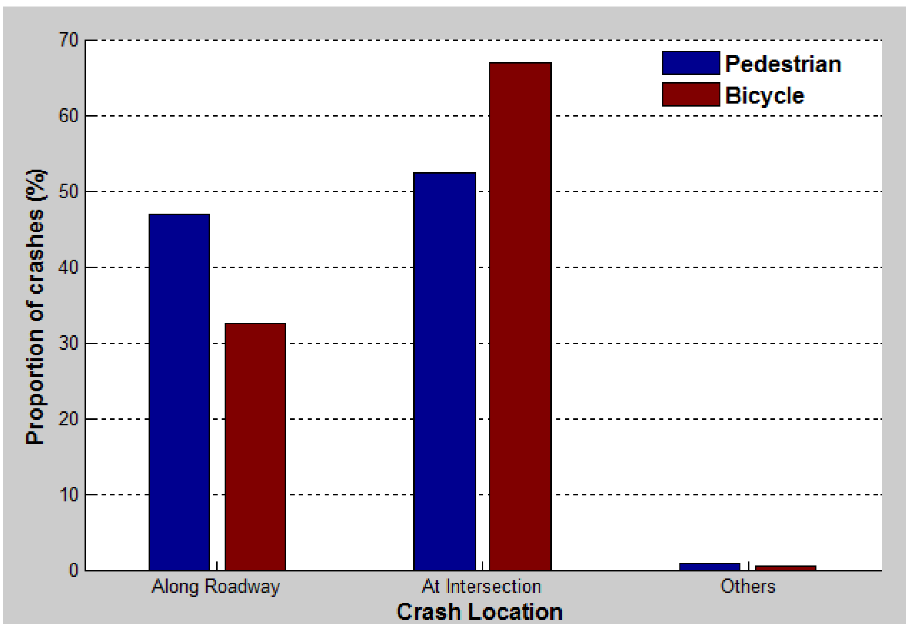


**Figure 10: Non-Motorized crashes by Road Location**

### 3.4.6 Crash frequency by weather condition

Experience has indicated that adverse weather conditions pose road safety risks. As illustrated in Table 6 and Figure 11, it can be seen that most of crashes occurred in clear weather conditions. About 84.77% pedestrian crashes and 90.51% bicycle crashes occurred under clear weather conditions.

**Table 6: Non-Motorized crashes by weather condition**

| Weather Condition | Pedestrian | Bicyclist |
|---|---|---|
| Clear | 84.77 | 90.51 |
| Cloudy | 3.92 | 3.43 |
| Fog | 0.36 | 0.14 |
| Rain | 10.13 | 5.69 |
| Snow | 0.82 | 0.23 |
| Grand Total | 100 | 100.00 |

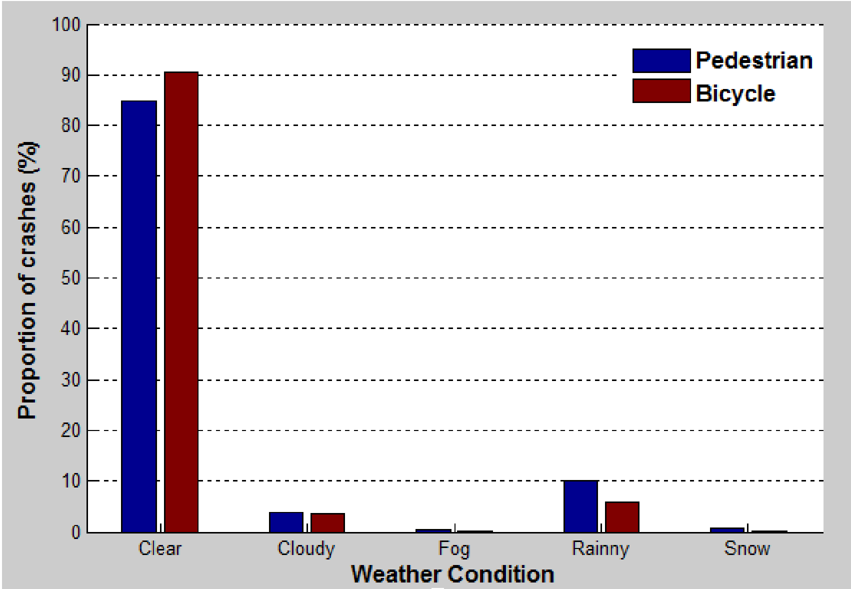

**Figure 11: Non-Motorized crashes by weather condition**

### 3.4.7 Crash frequency by light condition

Light condition has been cited to influence crashes. Table 7 indicates that majority of non-motorized crashes occurred during day light. About 55.32% pedestrian crashes and 75.76% bicycle crashes occurred during day light conditions.

**Table 7: Non-Motorized crashes by light condition**

| Light Conditions | Pedestrian | Bicyclist |
|---|---|---|
| Dark-Lighted | 28.04 | 15.80 |
| Dark-Not Lighted | 13.02 | 4.49 |
| Dark-Unknown Lighting | 0.31 | 0.19 |
| Dawn | 1.21 | 0.93 |
| Daylight | 55.32 | 75.76 |
| Dusk | 2.08 | 2.83 |
| Grand Total | 100.00 | 100.00 |

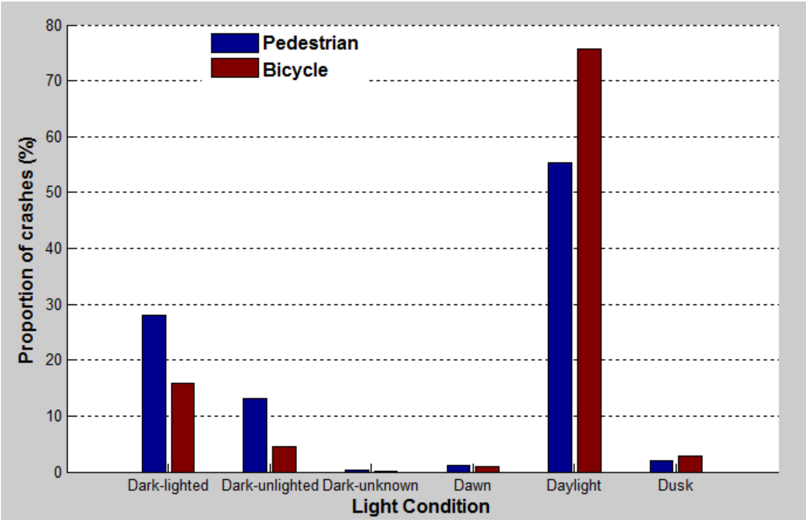

**Figure 12: Non-Motorized crashes by light condition**

### 3.4.8 Crash rate by Census block group

With the help of ArcGIS 10.1, this research integrated crash data and census data to generate crash frequencies per block group. This Figures 13 and Figures 14 indicate a five year period non-motorized crash frequency per census block group. It can be seen that the trend appears to be the same for both pedestrian and bicycle crashes. Using a specific case of Shelby County, it is quite clear that same block blocks are predominantly high crash areas for both pedestrian and bicyclists.

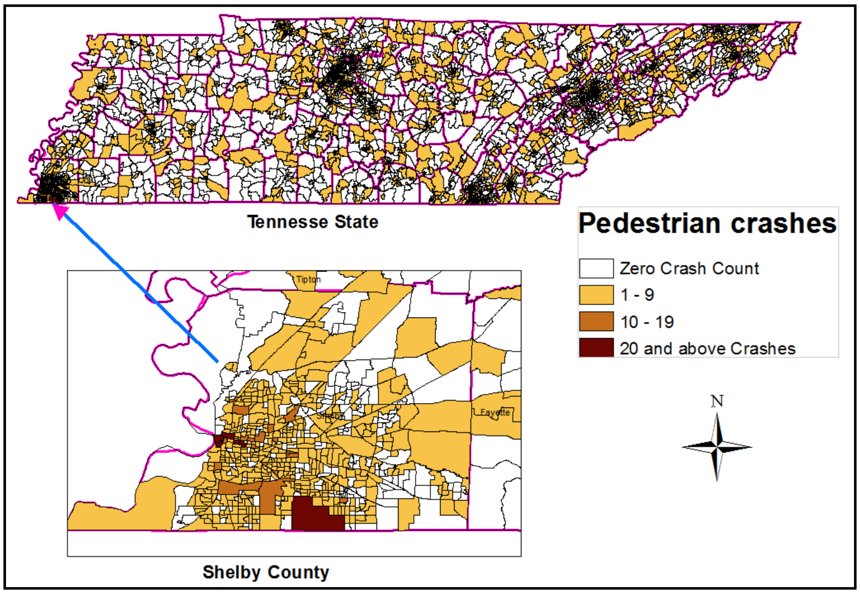


**Figure 13: Pedestrian crash rate by census block group**

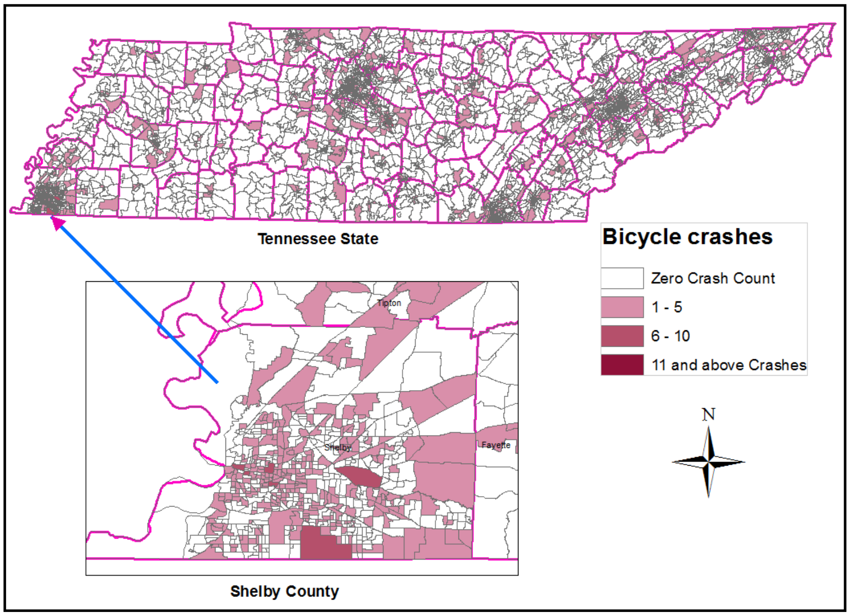


**Figure 14: Pedestrian crash rate by census block group**

Figure 15 provides a probability density distribution, which illustrate the percentage of census block groups that experienced specific crash frequencies during the analysis period. The data appear to be well approximated by a Poisson or negative binomial distribution, with roughly 54 percent and 75 percent of the census block groups experiencing zero pedestrian and bicycle crashes respectively during the analysis period.

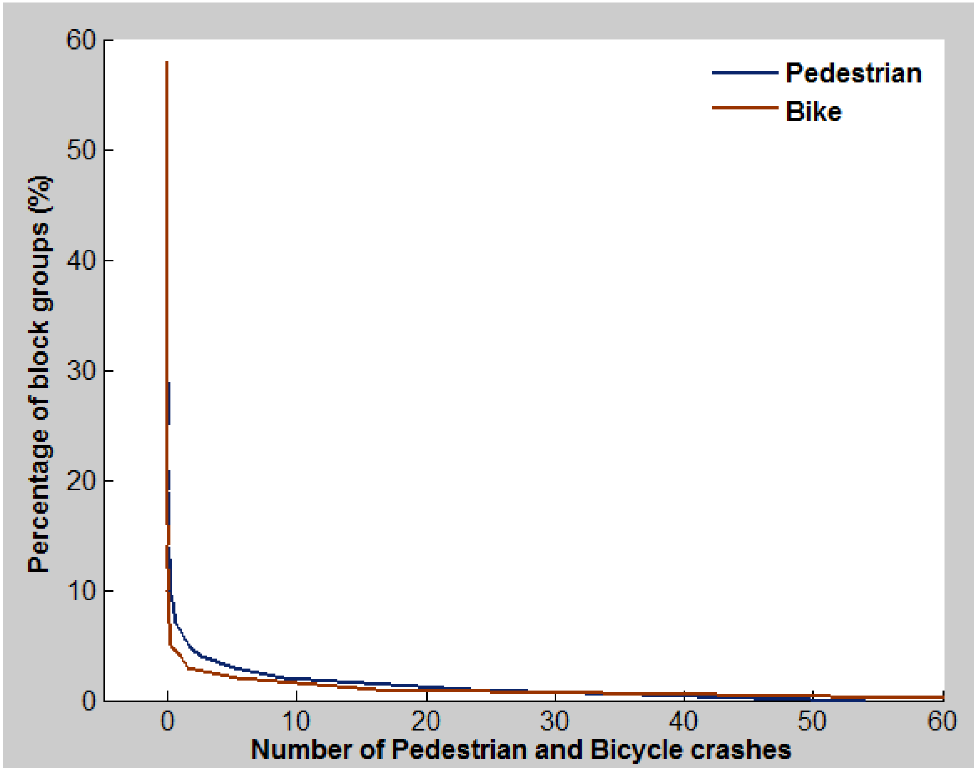


**Figure 15: Probability distributions of Non-Motorized by Census Block group**

### 3.5 Socioeconomic and demographic data

Studies indicate that socioeconomic and demographic data have been found to be contributing factors to non- motorized crashes. American Community Survey 2006-2010 five year estimates at census block group level were obtained from the Census Bureau website. All the data were available in GIS database. This dataset contain 4,125 census block groups in Tennessee and the list of data available at the census block level included the following;

- Area
- Population count
- Total population by age distribution
- Median Age by gender
- Total population by race
- Mode of transport to work
- Travel time to work
- Educational attainment
- Households below poverty level
- Households at or above poverty level
- Median household income in the past 12 months
- Housing unit car ownership

To allow for further analysis, more data processing was performed by creating other relevant variables such as population density. The overall state wide population age distribution indicates that the majority of population is aged between 35 and 65 years as shown in Figure 16. Statistics of indicate that the majority is white population who comprise of about 75.19% followed by African American comprising about 16.32%. Statewide statistics of transport mode to work indicate that 96.38% of the people in the workforce commute to work by private mode or driving their own cars. The proportion of workers who walk to work is

0.69%, while those who commute to work by bicycling is 0.06%. Although walking and bicycling have been largely promoted and encouraged, the rates are still lower compared to motorized traffic. Figure 20 shows household car ownership statistics which indicate that 61.52 % of housing units own two or more vehicles while 6.17% of housing units do not own any vehicles. A reasonable proportion of households live at or below poverty level. For example Figure 21 data indicates that 15.85% of households are at or below poverty level. However, 84.15% of households in Tennessee live above poverty level. The effect of these variables on bicycle and pedestrian safety was investigated at a block group and county level in subsequent chapters.

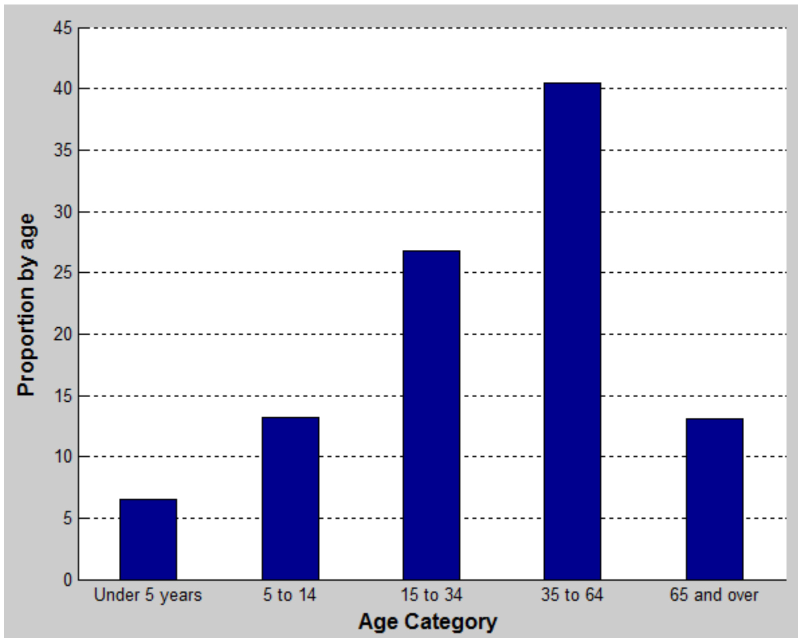


**Figure 16: Proportion of Population by Age distribution**

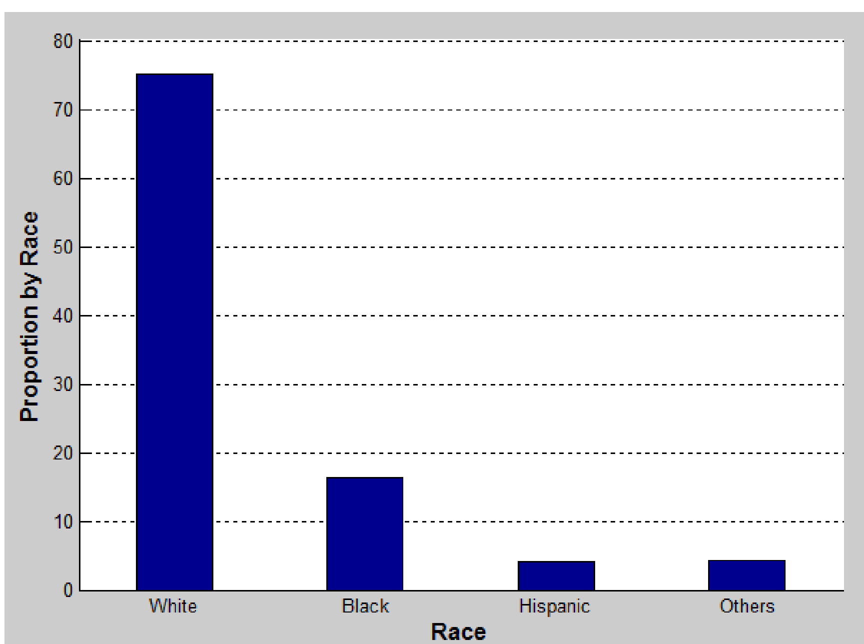


**Figure 17: Proportion of Population by Race**

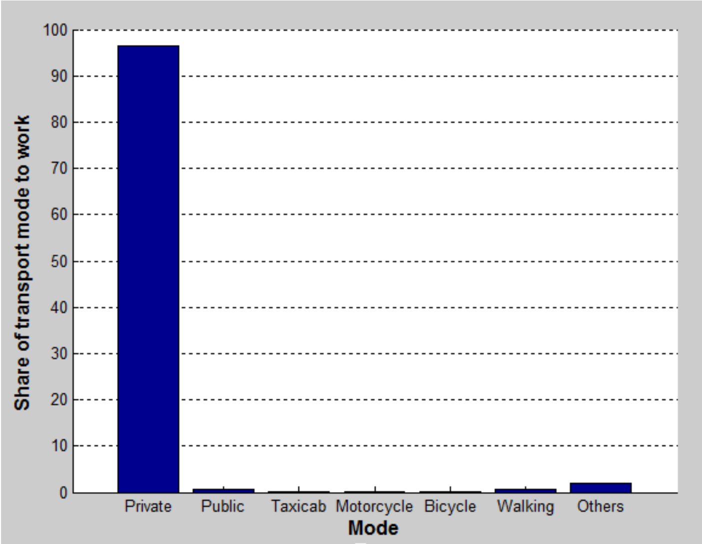


**Figure 18: Proportion of Population in the workforce transport mode to work**

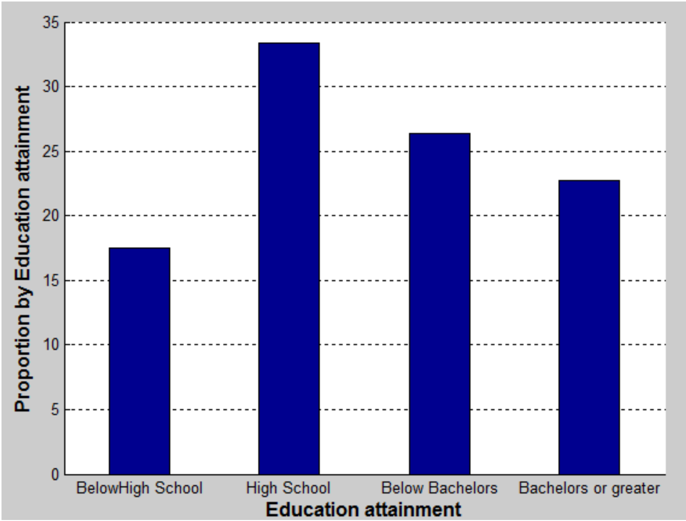


**Figure 19: Proportion of Population by Education Attainment**

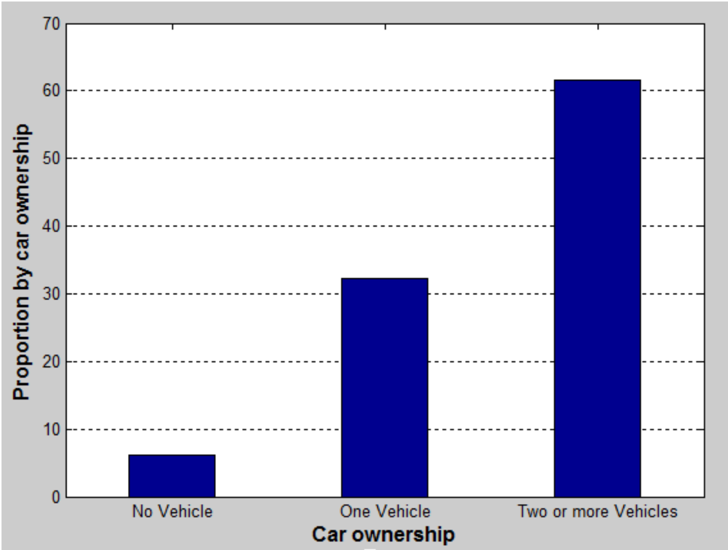


**Figure 20: Proportion of housing units by Car ownership**

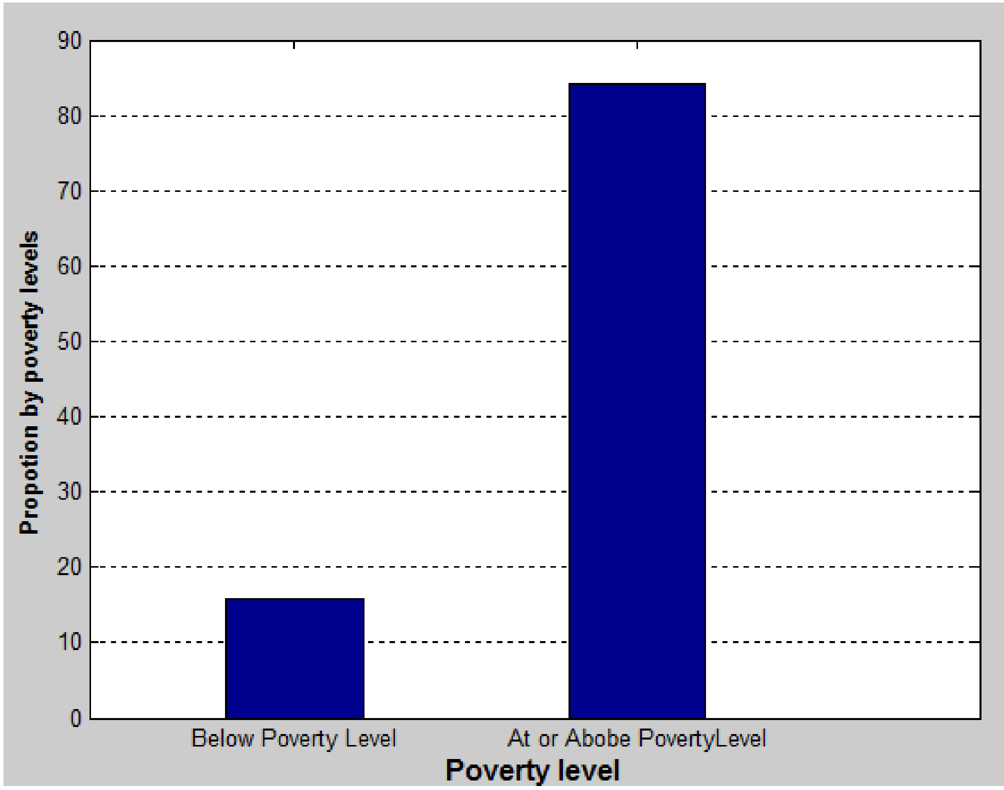

**Figure 21: Proportion of Household by Poverty Level**

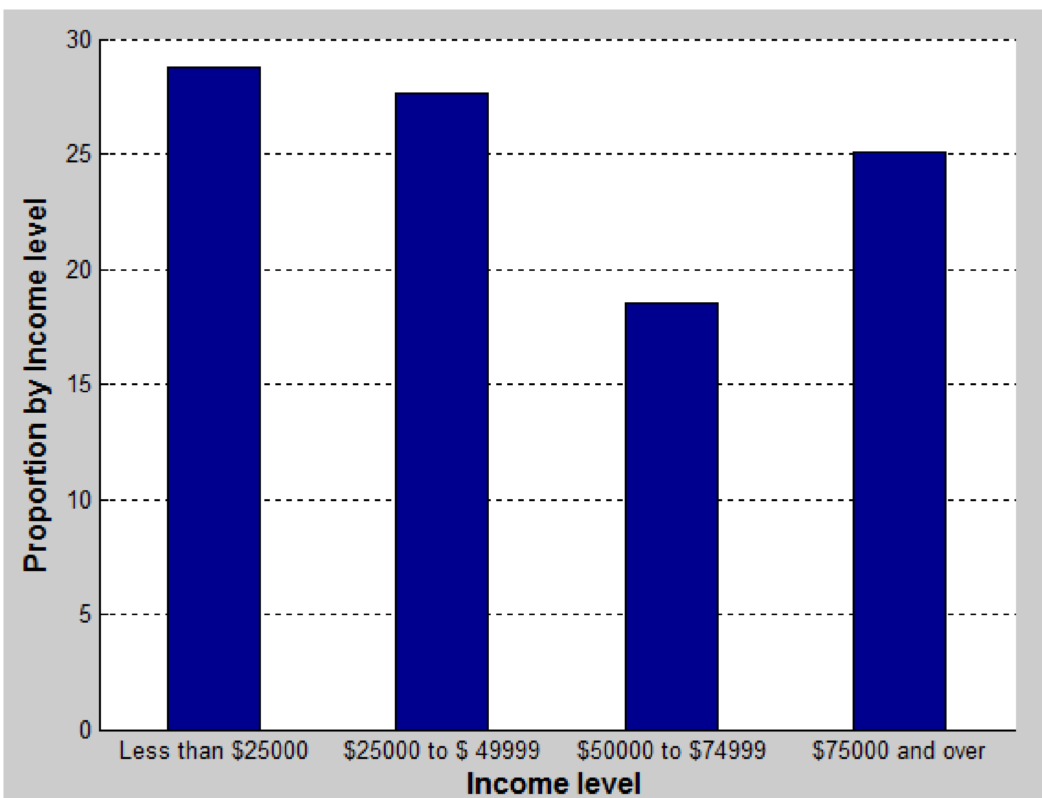

**Figure 22: Proportion of Households by Income level**

# CHAPTER 4: CLUSTER ANALYSIS

## 3.1 Introduction

The objective of this cluster analysis was to identify locations that experience a significantly higher proportion of pedestrian/bicycle crashes and that the occurrence of these crashes was not a random or chance event. Consequently the attributes (crash, demographic and socio-economic attributes) associated with such high crash clusters were extracted for further analysis. For pedestrian and bicycle crashes that are influenced by geographical factors, it is important to analyze the spatial dependences of crash data spread in space.

## 4.2 Procedure of Anselin Local Moran's I

Anselin local Moran's I tool identifies areas of clustering by location as well as by values of similar magnitude. This tool was used to show the block group significant clustering of crashes across the state.

The local moran's I statistic of spatial association is given by;

$$I_i = \frac{x_i - \overline{X}}{S_i^{\,2}} \sum_{j=1}^{n} w_{i,j}\left(x_j - \overline{X}\right) \qquad (4)$$

Where $x_i$ is a crash frequency for a block group i, $\overline{X}$ is the mean of the crashes, $w_{i,j}$ is the spatial weight between block groups i and j, and:

$$S_i^{\,2} = \sum_{j=1}^{n} \frac{\left(x_j - \overline{X}\right)^2}{n-1} - \overline{X}^2 \qquad (5)$$

With *n* equating to the total number of block groups

The Z-score for the statistics are computed as:

$$Z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \qquad (6)$$

Where:

$$E[I_i] = -\frac{\sum_{j=1}^{n} w_{i,j}}{n-1} \qquad (7)$$

$$V[I_i] = E\left[I_i^{\,2}\right] - E[I_i]^2 \qquad (8)$$

A positive value for ***I*** indicates that a feature has neighboring features with similarly high or low attributes values; this feature is part of a cluster. A negative value for **I** indicates that a feature has neighboring features with dissimilar values; this feature is an outlier. The local Moran's I index (I) is a relative measure and can only be interpreted within the context of its computed z-score or p-value. Therefore in either instance, the p-value for the feature must be small enough for the cluster or outlier to be considered statistically significant. A block group with a high positive z-score greater than +1.96 (Z> +1.96) represents areas that are part of statistically significant clusters of similar high or low pedestrian or bicycle fatalities and injuries at 95 percent confidence level. Such a block group is assigned a cluster type called "HH" for a statistically significant (0.05 level) cluster of high pedestrian/bicycle fatalities and injuries and "LL" for a statistically significant (0.05 level) cluster of low values. A block group with a low negative z-score smaller than -1.96 (Z< -1.96) represents areas that are part of statistically significant outlier of dissimilar pedestrian or bicycle fatalities and injuries at 95 percent confidence level. Such a block group is assigned an outlier type called "HL" if the block group has high pedestrian or bicycle fatalities and injuries, and is surrounded by block groups with low pedestrian fatalities and injuries; or it is assigned an outlier type called "LH" if the block group has a low value and is surrounded by block groups with high values. Areas where there was no significant clustering of values do not get a cluster type and therefore do not appear on the map

### 4.3 Results

Cluster analysis results are discussed in chapter 4.2.3. Census block groups where total crashes have *Z*-scores of ≥1.96 are statistically significant crash clusters at α = 0.05 level or 95% confidence interval. The results of Local Moran's I analysis identified 444 Significant block group clusters of high or low pedestrian values. These clusters were observed in Six counties of the state which were; Shelby, Davidson, Knox, Hamilton, Sevier, and Montgomery. While 546 Significant clusters of high and low bicycle crashes were identified. These clusters were observed in ten counties across the state which were; Shelby, Davidson, Williamson Rutherford, Knox, Hamilton, Marion Sevier, Montgomery and Blount. The detection of these clusters occurred at a maximum clustering distance which was determined using multi-distance spatial clustering (Ripley's K function) tool in GIS. Ripley's K function measures the distance between features to determine clustering by generating a hypothetical random distribution using the same number of features and the same area. The difference between the observed index value and the index value generated by the hypothetical random data indicates the degree of clustering. Figure 23 and Figure 24; indicate most significant clustering occurred at 2.3 miles for pedestrian clusters and 2.5 miles for bicycle crashes.

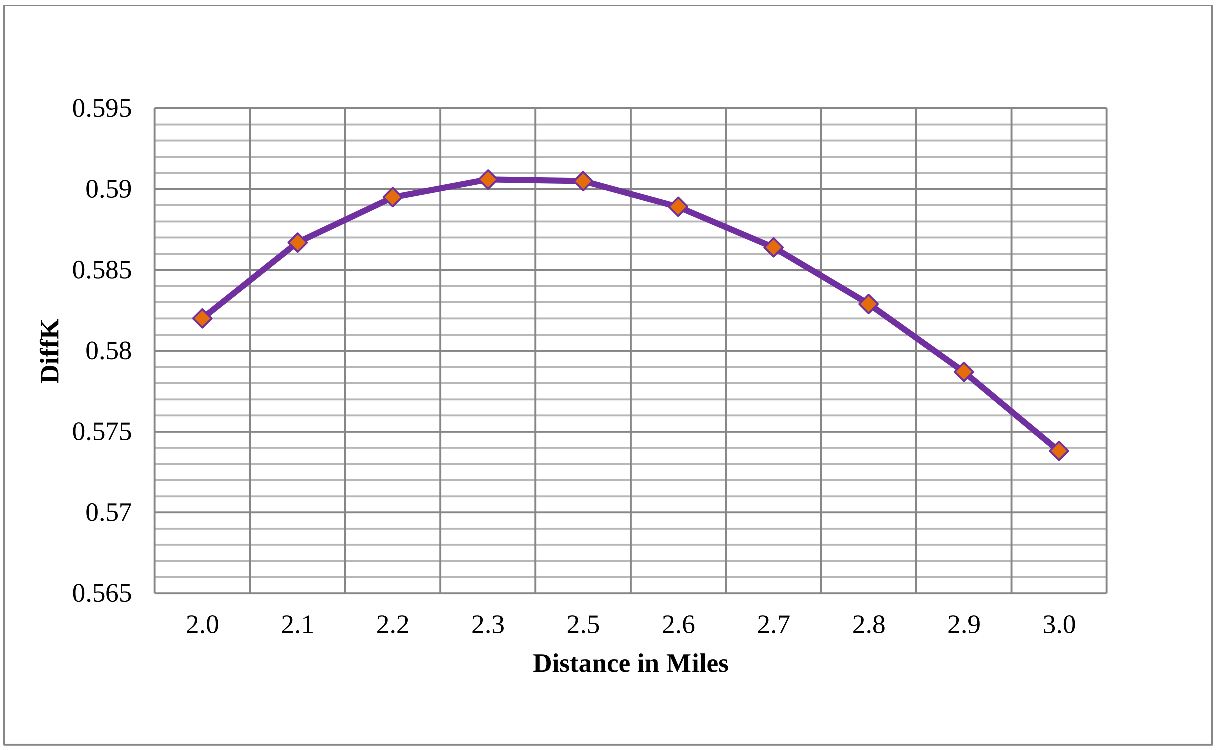


**Figure 23: Clustering of Pedestrian crashes at different distances**


**Figure 24: Clustering of Bicycle crashes at different distances**

Generally, clusters were observed in the main cities of the counties across the state. For instance Memphis, Nashville, Knoxville, Chattanooga and Clarksville. These areas are high-density zones with mixed commercial, residential and administrative activities and hence often involve a lot pedestrian and bicycle exposures.

Block groups belonging to significant clusters of pedestrian and bicycle crashes are often adjacent to each other. This tendency to cluster may be an indicator of local dependencies of crashes and also reflect that these areas are most likely associated with some inherent risks, which needs to be investigated.

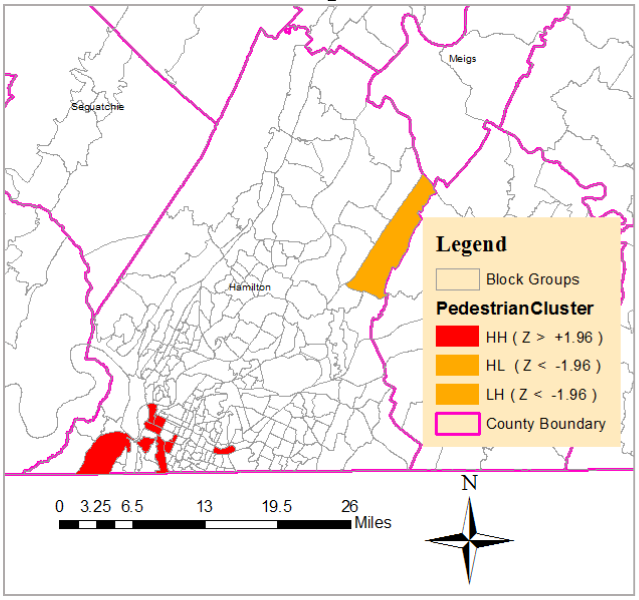


**Figure 25: Pedestrian crash clusters in Hamilton County**

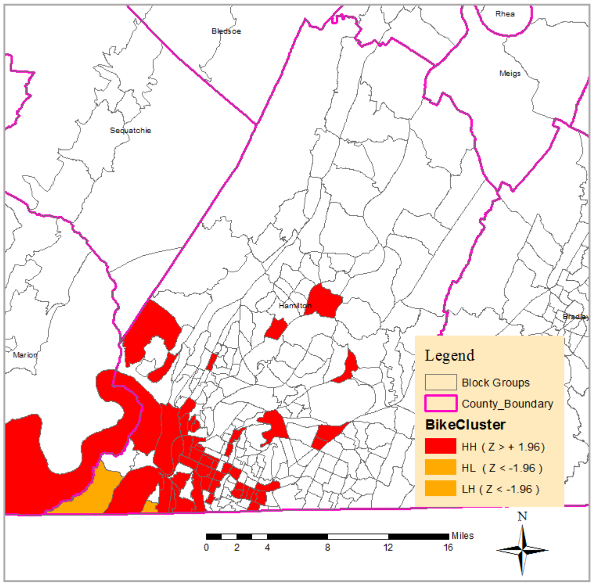


**Figure 26: Bicycle crash clusters in Hamilton County**

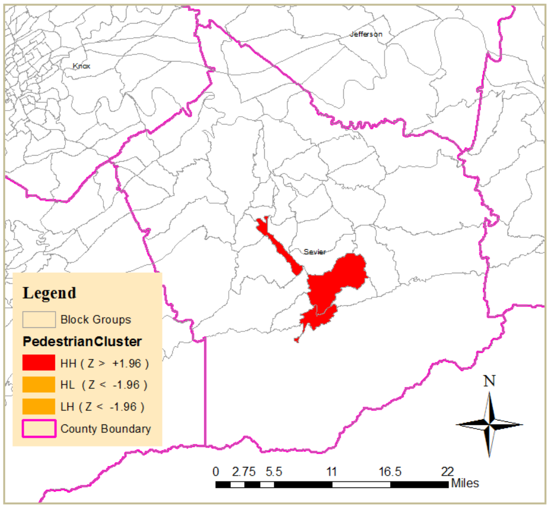


**Figure 27: Pedestrian crash clusters in Sevier County**

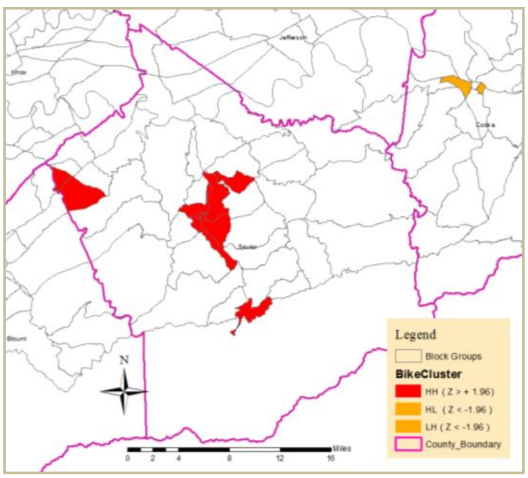


**Figure 28: Bicycle crash clusters in Sevier County**

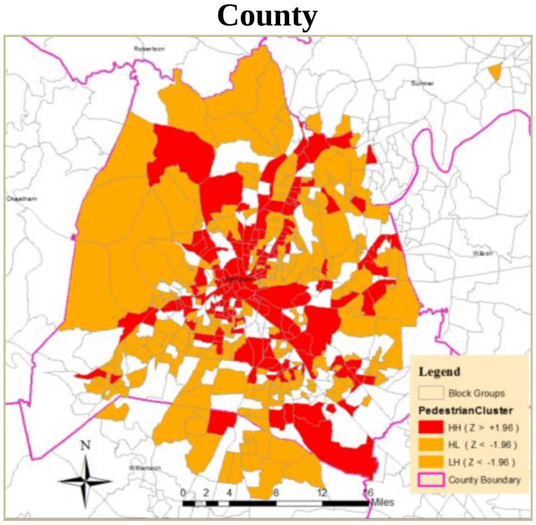


**Figure 29: Pedestrian clusters in Davidson County**

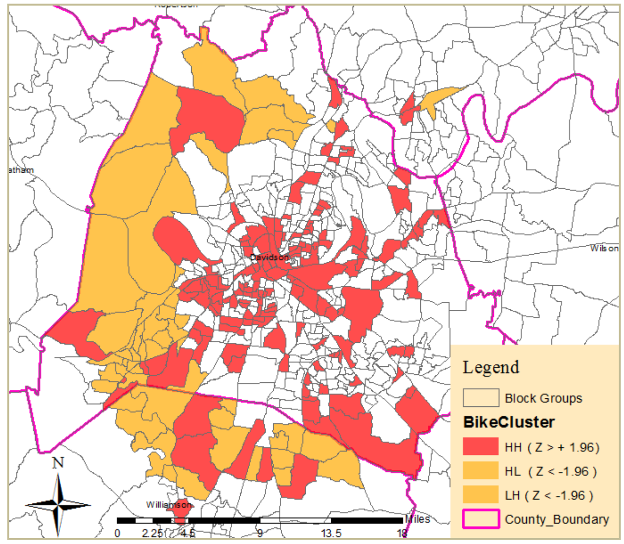


**Figure 30: Bicycle clusters in Davidson County**

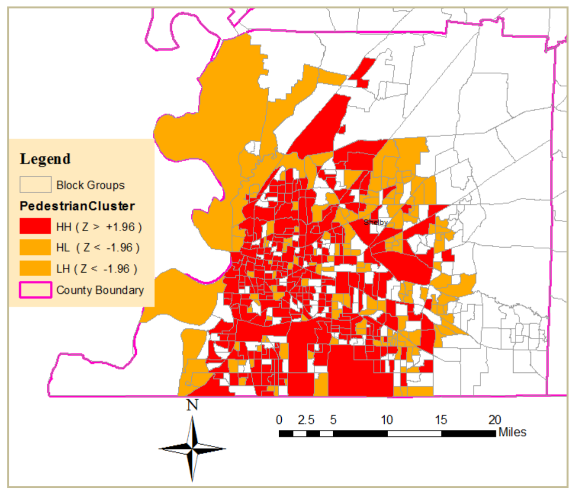


**Figure 31: Pedestrian crash clusters in Shelby County**

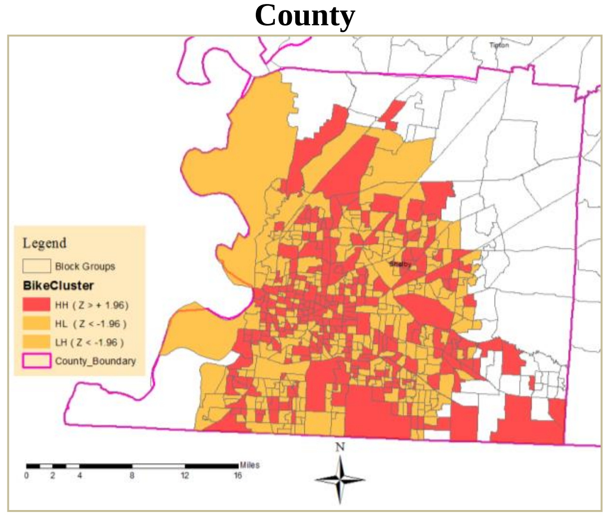


**Figure 32: Bicycle crash clusters in Shelby County**

## 4.4 Identifying crash associated factors

First the location of high and low census blocks was identified to investigate their built environment characteristics. The built environment around the crash location influences the frequency of crashes. For example, Figure 33 indicates mapping of the four high crash and four low crash census block groups in Davidson County. The result indicates that high crash areas are inner-city neighborhoods with high density and well-connected street networks, while low crash areas are the suburban neighborhoods with low density and low street connectivity. Through spatial analysis, our study identified that both the high crash census block groups and the low crash census block groups formed a cluster. This might be an indicator of local spatial autocorrelation stated by [12]. For instance, it is possible that there exist spatial variations in pedestrian crashes with respect to socioeconomic and demographic characteristics at high and low crash locations on the map.

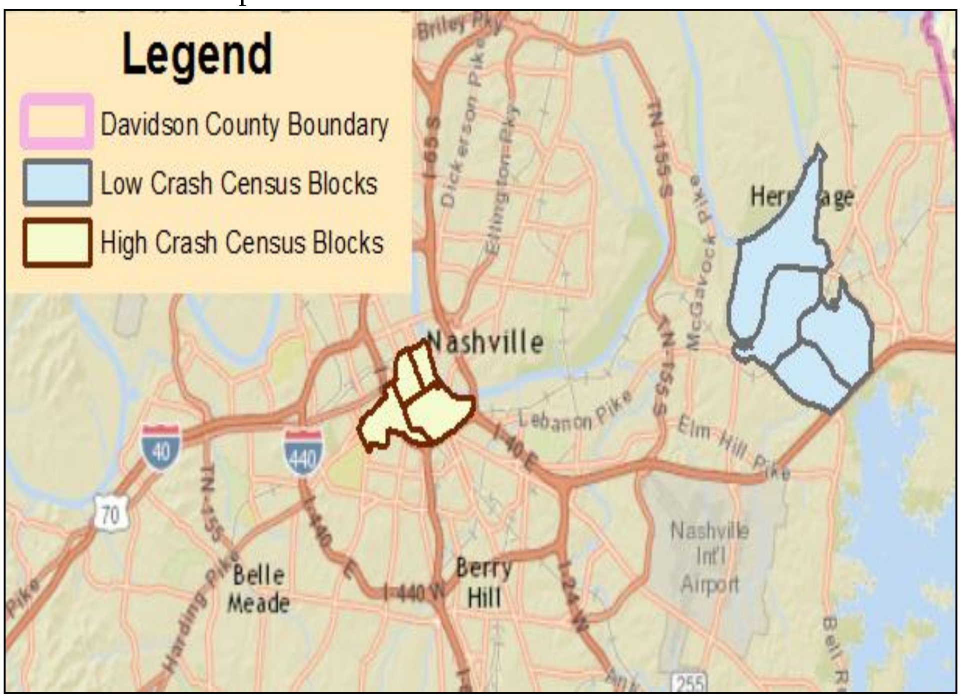


**Figure 33: Location of low crash and high crash census block groups in Davidson**

In order to understand the effects of the built environment, this work considered road environmental characteristics, individual crash characteristics and neighborhood characteristics around the crash location. To demonstrate this, we applied spatial analysis in GIS and probed a certain buffer distance around the crash within a census block group. Typically, Figure 34 illustrates the four high crash census blocks. It is clear that, these block groups have a dense street network with numerous intersections where majority of pedestrian crashes tend to occur. Increasing the buffer distance, results in high number of intersection and hence more crashes. Therefore, neighborhoods with high intersection density may associate strongly with high pedestrian injuries. The average sociodemographic values in the surrounding of a crash were calculated, the results show increasing values as the area spread wider over the crash location. The insight for this is that, increase or decrease in such demographics may have an effect on pedestrian safety. On other hand, Figure 35 illustrates the four low crash census blocks, it is clear that, these block groups have a low street connectivity with fewer intersections. The results from mapping of crashes also indicated very low crashes occurred in these block groups. The average sociodemographic values in the surrounding of a crash, the results show increasing values as the area spread wider over the crash location. However, these characteristic vary widely from those of the four high crash census blocks.

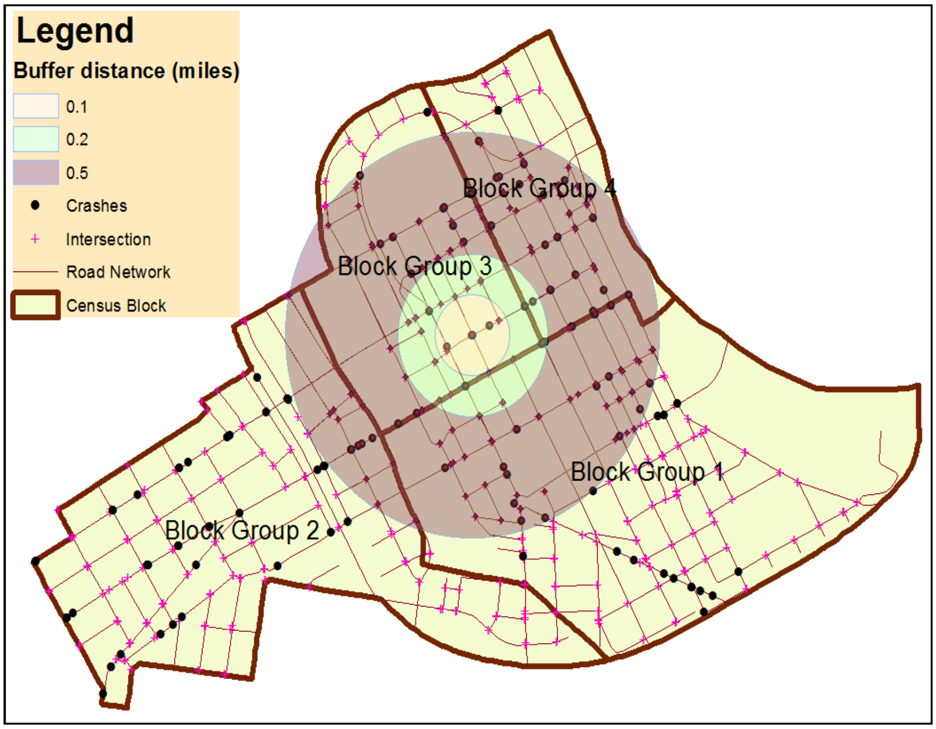


**Figure 34: Effects of built environment on four high crash census blocks in Davidson**

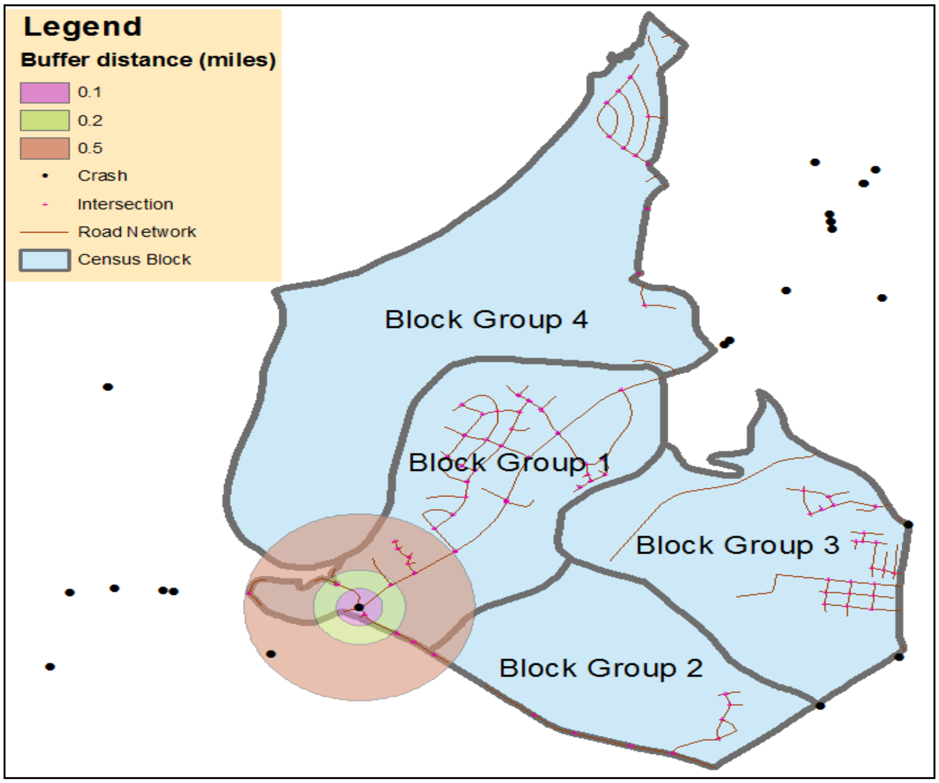


**Figure 35: Effects of built environment on four high low crash census blocks in Davidson**

Kernel density was used to identify high concentrations of pedestrian crashes in Tennessee that occurred in a period of five years. This study used a bandwidth of 200m and the grid cell size of 100m to identify high crash density areas. Census dataset was overlaid onto high crash density areas to investigate the influence of socioeconomic and demographic factors on pedestrian safety. Such techniques have been proposed in previous studies such as [65]. Results for kernel density analysis are presented in Figures 36-43. Although the study analyzed statewide crashes, only results for Davidson and Hamilton counties are presented as cases studies for simplicity and comparison purposes. Results presented in Fig. 36 and 37 indicate a pronounced high crash concentration among block groups with a high number of housing units with no vehicles, and absence of crash clusters in block groups with high housing units with 2 or more vehicles areas in Davidson County. This means that vehicle availability plays an important role in the mode choice where people without vehicles may choose to walk or use other modes other than driving, which makes them vulnerable to vehicle-pedestrian crashes. This pattern of association was consistent in both Davidson and Hamilton (Figures 38 and 39). These results may further help to understand the safety risks among income groups because vehicle availability depends on income. For instance, low-income populations travel less frequently, have the lowest income groups and are much less likely to own an automobile [66].

This study also investigated the influence of transport mode to work on pedestrian crash occurrence. Results of kernel density analysis shown in Figures 40 and 41 indicated high-density clusters occur among populations who commute to work by walking, while block groups that predominately inhabited by populations that commute to work by private cars were weakly associated with high crash concentrations in Davidson County. Results for Hamilton County showed similar patterns in Figures 42 and 43. These results are important in providing improved information about high crash risk locations.

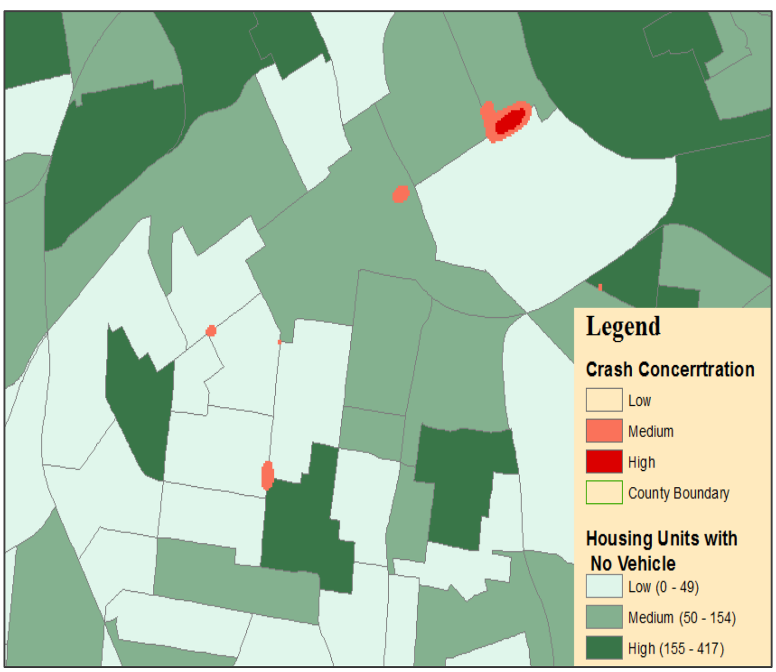


**Figure 36: Housing units with no vehicle in Davidson County**

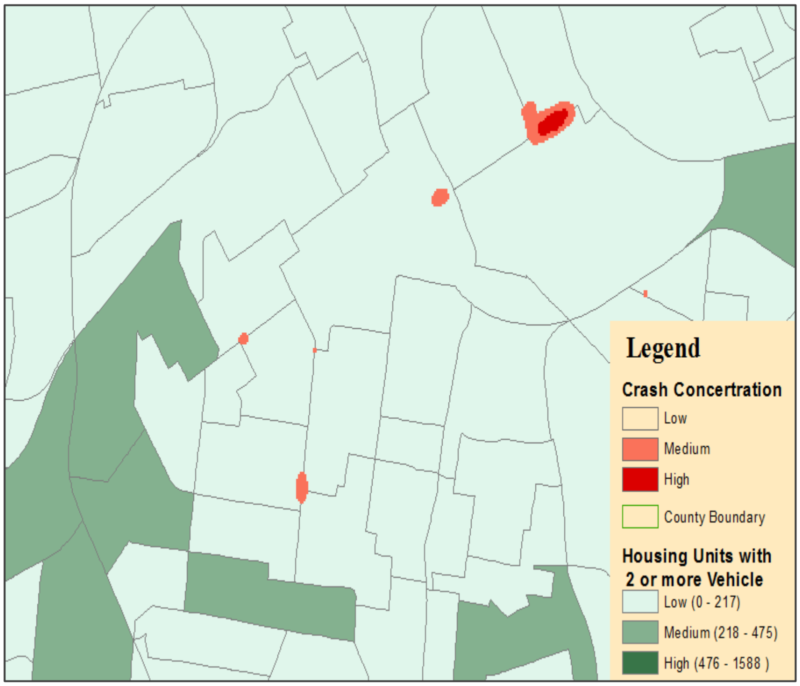

Figure 37: Housing units with 2 or more vehicles in Davidson County

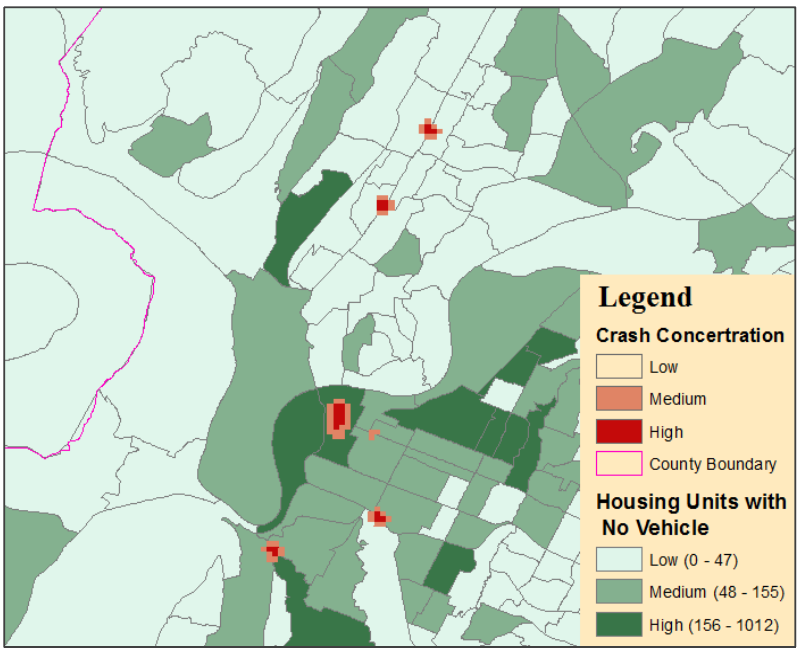

Figure 38: Housing units with no vehicle in Hamilton County

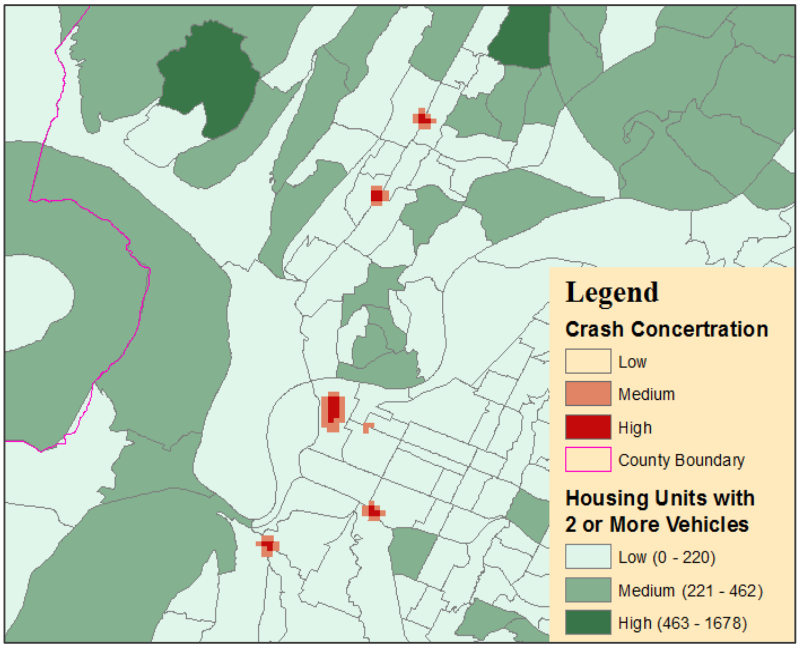


**Figure 39: Housing units with 2 or more vehicles in Hamilton County**

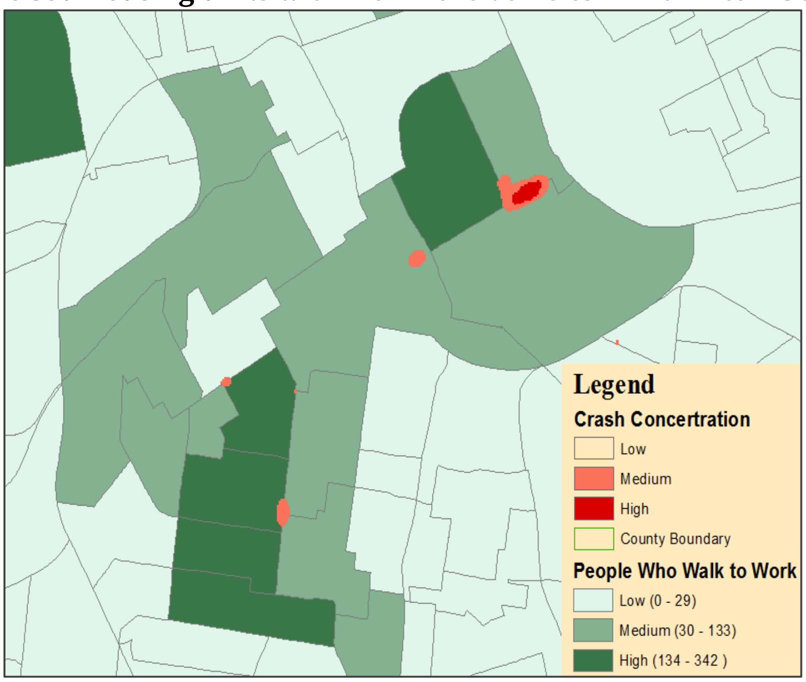


**Figure 40: population commuting to work by walking in Davidson County**

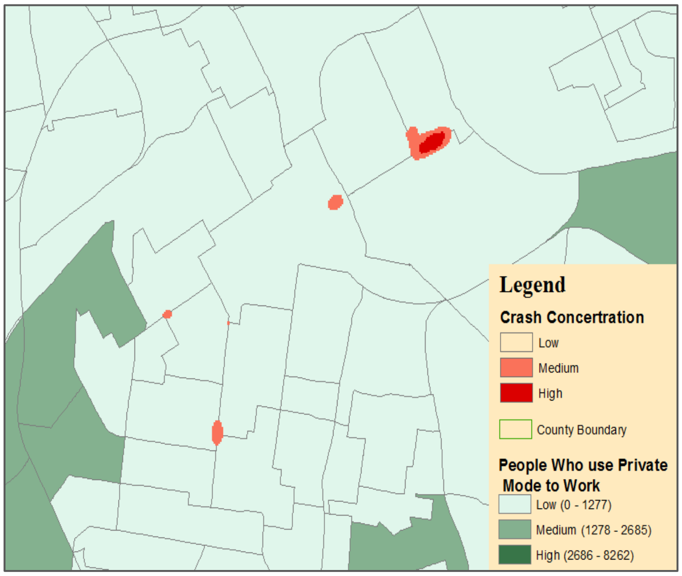


**Figure 41: Commuting to work by private cars in Davidson County**

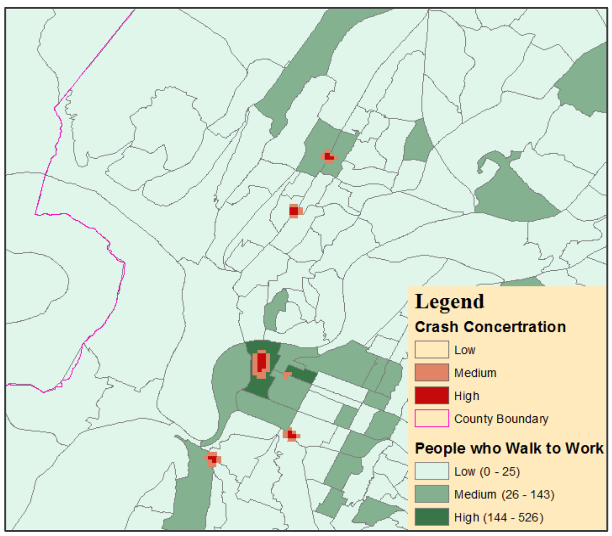


**Figure 42: population commuting to work by walking in Hamilton County**

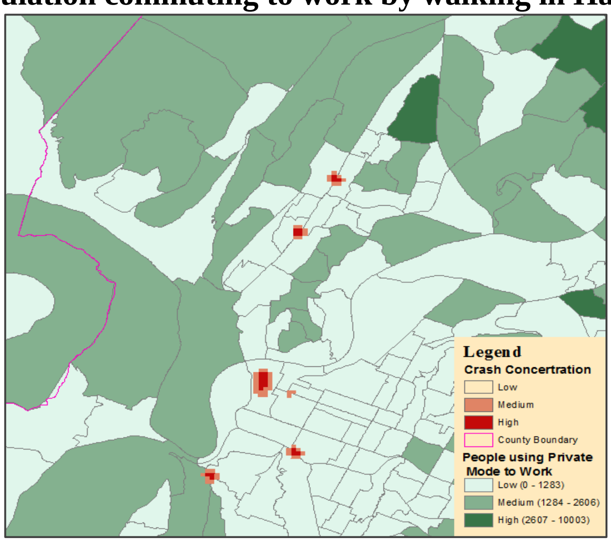


**Figure 43: Population commuting to work by private mode in Hamilton County**

# CHAPTER 5: DEVELOPMENT OF SAFETY PERFORMANCE FUNCTIONS

## 5.1 Introduction

SPFs are crash prediction models inform of mathematical equations that relate the number of crashes of different types to site characteristics [67]. Therefore, SPFs are also known as safety crash prediction models and so he term is used interchangeably with safety performance functions in this study. The dependent variable of such equations is the number of crashes of a specific type. The primary purpose of this SPF from Safety Analyst is to assist an agency in their network screening process, i.e., to identify sites that may benefit from a safety treatment.

Highway safety manual indicates two types of SPFs that can possibly be developed (Level I and level II SPF's) depending on the variables used [68, 67]. Level I SPFs determine crash frequencies based only on traffic volumes (AADT) and segment length while Level II SPFs uses as many variables as possible. Safety Analyst software that has been developed by FHWA can only develop Level I SPFs [68]. Level II SPFs includes several variables other than just traffic volume like weather conditions, roadway geometry, traffic data and human factors. This report is based on Level II SPFs which are the focus of this report are used whenever a detail of factors influencing crash occurrence on a given roadway segment needs be determined. FHWA published Safety Performance Function Development Guide [67] a report, which provides guidance in the process of development of jurisdiction-specific SPFs. The report analyzes different methodologies that have been used in development of crash prediction models and shows what approach can be used. This research incorporates the guidance from the report together with other sources from literature.

## 5.2 Count models

Crashes are examples of "count data" and are properly modeled using a specific family of statistical models called count data models. The most popular count data models for rare events are Poisson and negative binomial regression models.

### 5.2.1 Poisson regression model

Consider $y_i$ number of crashes occurring in a certain period at a site *i*. In a Poisson regression model, the probability of having $y_i$ crashes in that period is given by:

$$\boldsymbol{P(y_i) = \frac{exp(-\lambda_i) * \lambda_i^{y_i}}{y_i!}} \qquad (9)$$

Where:

$\lambda_i$ is the Poisson parameter for site *i*, which is equal to site *i*'s expected number of crashes at a period, $E\ (y_i)$.

In Poisson regression models, the intent is to express the expected number of crashes as $f(X_i)$ function of site characteristics. In other words,$\lambda_i = f(\beta X_i)$, where f is a function, $X_i$ is a vector of explanatory variables, and $\beta$ is a vector of estimable parameters (coefficients of $X_i$) [67] . The relationship between the expected number of crashes$\lambda_i$ and explanatory variables can be expressed in a log linear form as;

$$\lambda_i = \exp(\beta X_i) \text{ or } \ln(\lambda_i) = (\beta X_i) \qquad (10)$$

Expressing their relatioship in a linear ensures that the Poisson parameter is always positive and yields a linear combination of the predictor variables on the right-hand side by taking the log on both sides of the equation. This is often prefered in modeling count count data such as crashes. This type of model form belongs to a category of models called generalized linear models (GLM) in which the regression coefficients and their standard error are estimated by maximizing the likelihood or log likelihood of the parameters for the observed data. This procedure is called estimating by the maximum likelihood method.

### 5.2.2 Negative binomial model and Overdispersion

The Poisson models assume that the mean and variance are equal, that is, *E* ($y_i$) = *VAR* ($y_i$). However, this is contrary to the most likely situation often associated with crash data in which the variance is larger than the mean. This phenomenon is called overdispersion [69], i.e., *VAR* ($y_i$) > *E* ($y_i$). Negative binomial regression model relaxes the Poisson model assumption of the mean being equal to the variance and is used in to account for overdispersion in modeling crash counts [70]. The probability function of Negative binomial regression model can be written as follows:

$$P(y_i) = \frac{\tau(y_i+\alpha^{-1})}{\tau(\alpha^{-1})\tau(y_i+1)}\left[\frac{1}{1+\alpha\mu}\right]^{1/\alpha}\left[\frac{\alpha\mu}{1+\alpha\mu}\right]^{y_i} \quad (11)$$

Where;

$\mu = E(y_i) = \exp(X_i\beta)$: the mean,

$X_i$ : the value of independent variable

$\beta$ : the coefficient of independent variable

$\alpha$: the overdispersion factor.

The calculation of overdispersion parameter ($\alpha$) uses maximum likelihood approach [22]. This method estimates the most likely value of the dispersion parameter by calculating the log-likelihood for a range of possible values of $\alpha$, and selecting the value of $\alpha$ with the largest log-likelihood. This process is often tedious, but with development of new statistical software packages such Stata which can estimate this parameter eliminating necessity of intensive calculations. Two other terms exist in modeling count data; underdispersion which occurs when the variance is less than mean [71] and equidispersion where the variance and mean are the same. However, these cases rarely occur in crash data [69].

## 5.3 Correlation Between variables

The use of correlated variables can result in errors during determination of the coefficients. Therefore, care must be taken not to use correlated variables regardless of being significant. One approach of avoiding this is to perform the correlation test between variables and eliminate one of the two variables that are highly correlated [72]. For example, variable housing units with two or more vehicles was significantly correlated with household median income and was therefore removed.

## 5.4 Adopted Model form

In this study for handling the over-dispersion of exposures' data a Negative Binomial Regression Model was used for pedestrian and bicycle crash frequency in order to develop SPF from significant variables. The model adopted considers population as exposure variable. This implies that, no crash is expected for a block group or county without population. The functional form of the prediction model is as shown in equation 12.

$$\mu_i = (P)_i * \exp(\alpha_1 X_1 + \alpha_2 X_2 + \cdots \ldots . + \alpha_n X_n) \quad (12)$$

Where;

| | | |
|---|---|---|
| $\mu i$ | = | 5 year predicted crash frequency |
| $\alpha_1, \alpha_2 \ldots \alpha_n$ | = | regression parameters |
| $P$ | = | Population (exposure variable) |
| $X_1, X_2, \ldots X_n$ | = | Explanatory variables |

The function above can be re-arranged to be in exponential form and hence obtaining a utility function shown in equation 13

$$\mu_i = \exp(lnP + \alpha_1 X_1 + \alpha_2 X_2 + \cdots \ldots . + \alpha_n X_n) \quad (13)$$

| | | |
|---|---|---|
| $\mu i$ | = | predicted crash frequency for 5 years |
| $\alpha_1, \alpha_2 \ldots \alpha_n$ | = | regression parameters |
| $X_1, X_2, \ldots X_n$ | = | Explanatory variables |
| $P$ | = | Population (exposure variable) |

## 5.5 Discussion of Results

STATA software package was used to model crash data using Negative binomial model. A Z-test was used at a 95% confidence level of significance for the independent variables. Socio-economic and demographic variables at a census block and county levels were analyzed to investigate their effects on non-motorized crash frequency. This data is summarized in the appendix A and B. The population is entered as an exposure variable with a constant coefficient of 1 (one) in exponential equation 12. The overdispersion parameter is modeled as a constant value and is provided as a beta value. Crash prediction began with determining the relationship between variables relevant to non-motorized crash frequency and consequently led to development of safety performance functions (crash prediction models).

### 5.5.1 Pedestrian crash model at census block group

Among the many variables processed for the non-motorized crash analysis, the variables shown in Table 8 were found to be significant in terms of statistical measures for the pedestrian crash frequency model at census block group level.

**Table 8: Summary statistics of variables in pedestrian crash model at census block group level**

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Population density (1000 per sq. Mile) | 1.62 | 2.53 | 0.00 | 89.44 |
| Population below 15 years of age (%) | 19.02 | 7.76 | 0.00 | 59.33 |
| Population from 15 to 64 years of age (%) | 66.98 | 8.36 | 11.80 | 100.00 |
| Population commuting to work by private cars (%) | 95.84 | 5.81 | 0.00 | 100.00 |
| Population commuting to work by walking (%) | 0.83 | 2.89 | 0.00 | 100.00 |
| Median household income ("000" $) | 45.42 | 24.35 | 0.00 | 247.36 |
| Housing units with no vehicles (%) | 6.94 | 9.47 | 0.00 | 83.97 |

When developing a crash prediction model, a number of factors were investigated to determine their impact and significance on pedestrian crashes. Table 9 below shows the relationship between a number of these factors and the number of observed pedestrian crashes. The factors of “population density”, “population from 15 to 64 years of age,” “population commuting to work by walking,” and “housing units with no vehicles” show a positive association with the number of pedestrian crashes. However, “population below 15 years of age”, “population commuting to work by private cars,” and “median household income” displays a negative association with the number of pedestrian crashes. The effect of variables on number of pedestrian crashes is indicated by the coefficient in the model results where the magnitude of coefficient shows its impact and the sign of coefficient indicates its significance. For example a variable with a large coefficient in magnitude indicates that the number of pedestrian crashes is more sensitive to that factor than others. However, variables with positive coefficients indicate that they increase the number of pedestrian crashes while variables with negative coefficients indicate that they decrease the number of pedestrian crashes. Among each of these factors, the number of pedestrian crashes appears to be most sensitive to the population density and least sensitive to population below 15 years of age. Population commuting to work by walking and number of housing units with no vehicles appear to increase the number of pedestrian crashes by the same magnitude. The Z-statistic or p-value indicates the level of significance. For example variables with Z-statistic of 1.96 and over, means that such variables have significant relationships pedestrian crash frequency at 95% confidence interval.

**Table 9: Pedestrian crash model at census block group level**

| **All Crashes** | **Coefficient** | **Z** | **p-value** |
|---|---|---|---|
| Population density (1000 per sq. mile) | 0.1169 | 7.77 | 0.000 |
| Population below 15 years of age (%) | -0.0083 | -2.08 | 0.037 |
| Population from 15 to 64 years of age (%) | 0.0143 | 3.76 | 0.000 |
| Population commuting to work by private cars (%) | -0.0379 | -7.12 | 0.000 |
| Population commuting to work by walking (%) | 0.0294 | 2.34 | 0.019 |
| Median household income ("000" $) | -0.0101 | -7.34 | 0.000 |
| Housing units with no vehicles (%) | 0.0303 | 8.86 | 0.000 |
| Constant | -4.4193 | -7.14 | 0.000 |
| Population | Exposure | | |
| alpha | 1.586 | | |

Coefficients of significant variables can be used to formulate crash prediction models that can be used to predict the total number of pedestrian crashes within a census block group based on the factors affecting pedestrian crashes. The pedestrian crash prediction model developed takes the following form;

$$\mu=\exp[\ln(P)+0.1169A-0.0083B+0.0143C-0.0379D+0.0294E-0.0101F+0.0303G-4.4193]$$

Where;

$\mu$: Number of pedestrian crashes
P: Population of a block group
A: Population density (1000 per sq. mile)
B: Population below 15 years of age (%)
C: Population from 15 to 64 years of age (%)
D: Population commuting to work by private cars (%)
E: Population commuting to work by walking (%)
F: Median household income ("000" $)
G: Housing units with no vehicles (%)

### 5.5.2 Bicycle crash model at census block group

Similar to the pedestrian crash prediction model, a number of factors were investigated with regard to significant factors causing bicycle crashes. The variables summarized in Table 10 were found to be significant for the bicycle crash frequency model.

**Table 10: Summary of variables used in bicycle crash model at census block group level**

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Population density (1000 per sq. mile) | 1.62 | 2.53 | 0.00 | 89.44 |
| Population below 15 years of age (%) | 19.02 | 7.76 | 0.00 | 59.33 |
| Population from 15 to 64 years of age (%) | 66.98 | 8.36 | 11.80 | 100.00 |
| Population commuting to work by private cars (%) | 95.84 | 5.81 | 0.00 | 100.00 |
| Population commuting to work by bicycling (%) | 0.08 | 0.51 | 0.00 | 10.94 |
| Average household income ("000" $) | 45.42 | 24.35 | 0.00 | 247.36 |
| Housing units with no vehicles (%) | 36.81 | 50.62 | 0.00 | 1012.00 |

Table 11 below shows the relationship between a number of these factors and the number of observed pedestrian crashes. The factors of "population density", "population from 15 to 64 years of age," "population commuting to work by bicycling," and "housing units with no vehicles" show a positive association with the number of pedestrian crashes. However, "population below 15 years of age", "population commuting to work by private cars," and "median household income" indicate a negative association with the number of pedestrian crashes. Unlike pedestrian crashes, the number of bicycle crashes appears to be most sensitive to population commuting to work by bicycling and least sensitive to number of housing units with no vehicles.

**Table 11: Bicycle crash model at census block group level**

| **All Crashes** | **Coefficient** | **Z** | **p-value** |
|---|---|---|---|
| Population density (1000 per sq. mile) | 0.1264 | 6.29 | 0.000 |
| Population below 15 years of age (%) | -0.0238 | -4.29 | 0.000 |
| Population from 15 to 64 years of age (%) | 0.0074 | 1.48 | 0.139 |
| Population commuting to work by private cars (%) | -0.0456 | -6.86 | 0.000 |
| Population commuting to work by bicycling (%) | 0.1993 | 2.82 | 0.005 |
| Average household income ("000" $) | -0.0048 | -2.93 | 0.003 |
| Housing units with no vehicles (%) | 0.0028 | 3.54 | 0.000 |
| Constant | -3.9830 | -4.88 | 0.000 |
| Population | Exposure | | |
| alpha | 2.329 | | |

The bicycle crash prediction model developed can be used to predict the total number of bicycle crashes (all crashes) within a census block group and it takes the following form;

$\mu$=Exp (lnP + 0.1264A-0.0238B+0.0074C-0.0456D+0.1993E-0.0048F+0.0028G-3.9830)

Where;

$\mu$: Number of bicycle crashes
P: Population of a block group
A: Population density (1000 per sq. mile)
B: Population below 15 years of age (%)
C: Population from 15 to 64 years of age (%)
D: Population commuting to work by private cars (%)
E: Population commuting to work by bicycling (%)
F: Median household income ("000" $)
G: Housing units with no vehicles (%)

### 5.5.3 Pedestrian crash model at County level

A pedestrian crash prediction models were developed at county and a number of factors were investigated with regard to impact and significance of factors causing crashes. The variables summarized in Table 12 were found to be significant for the pedestrian crash frequency model.

**Table 12: Pedestrian crash model at county level**

| Variable | Coefficient | z | P>z |
|---|---|---|---|
| Population below 15 years of age (%) | -0.0281 | -0.91 | 0.362 |
| Population from 15 to 64 years of age (%) | 0.0231 | 0.91 | 0.364 |
| Population of White (%) | -0.0461 | -2.08 | 0.038 |
| Population of African American (%) | -0.0368 | -1.6 | 0.109 |
| Population of Hispanic (%) | 0.0546 | 1.64 | 0.101 |
| Population commuting to work by private cars (%) | -0.0705 | -1.13 | 0.257 |
| Population commuting to work by walking (%) | -0.2909 | -1.64 | 0.102 |
| Median household income ("000" $) | -0.0025 | -1.91 | 0.056 |
| Housing units with no vehicles (%) | 0.0848 | 2.37 | 0.018 |
| Constant | 1.9170 | 0.3 | 0.768 |
| Population | Exposure | | |
| Number of observations | 95 | | |
| alpha | 0.11 | | |

The pedestrian crash prediction model developed can be used to prediction the total number of pedestrian crashes of a county and it takes the following form;

$\mu = Exp\ [\ln(P) - 0.028A + 0.023B - 0.046C - 0.037D + 0.055E - 0.071F - 0.291G - 0.003H + 0.085I + 1.917]$

Where;

$\mu$: Number of pedestrian crashes

P: Population of a County

A: Population below 15 years of age (%)

B: Population from 15 to 64 years of age (%)

C: Population of White (%)

D: Population of African American (%)

E: Population of Hispanic (%)

F: Population commuting to work by private cars (%)

G: Population commuting to work by walking (%)

H: Median household income ("000" $)

I: Housing units with no vehicles (%)

### 5.5.4 Bicycle crash model at County level

Similarly, a bicycle crash prediction models were developed at county and a number of factors were investigated with regard to impact and significance of factors causing crashes. The variables summarized in Table 13 were found to be significant for the bicycle crash frequency model.

**Table 13: Bicycle crash model at county level**

| Variable | Coefficient | z | P>z |
|---|---|---|---|
| Population below 15 years of age (%) | -0.102 | -1.86 | 0.063 |
| Population from 15 to 64 years of age (%) | 0.104 | 2.33 | 0.02 |
| Population of White (%) | -0.063 | -1.58 | 0.113 |
| Population of African American (%) | -0.044 | -1.08 | 0.278 |
| Population of Hispanic (%) | 0.095 | 1.59 | 0.113 |
| Population commuting to work by private cars (%) | 0.049 | 0.5 | 0.62 |
| Population commuting to work by Bicycling (%) | 0.241 | 0.18 | 0.854 |
| Median household income ("000" $) | 0.0003 | 0.12 | 0.903 |
| Housing units with no vehicles (%) | -0.064 | -0.99 | 0.322 |
| Constant | -12.792 | -1.2 | 0.23 |
| Population | Exposure | | |
| Number of observations | 95 | | |
| alpha | 0.29 | | |

The Bicycle crash prediction model developed can be used to prediction the total number of bicycle crashes of a county and it takes the following form;

$$\mu = Exp\ [\ln(P) - 0.102A + 0.104B - 0.063C - 0.044D + 0.095E + 0.049F + 0.241G + 0.0003H - 0.064I - 12.792]$$

Where;
$\mu$: Number of Bicycle crashes
P: Population of a County
A: Population below 15 years of age (%)
B: Population from 15 to 64 years of age (%)
C: Population of White (%)
D: Population of African American (%)
E: Population of Hispanic (%)
F: Population commuting to work by private cars (%)
G: Population commuting to work by Bicycling (%)
H: Median household income ("000" $)
I: Housing units with no vehicles (%)

# CHAPTER 6: DEVELOPING CRITERIA FOR HIGH CRASH IDENTIFICATION

## 6.1 Overview

Crash prediction models are very important tools in predicting the number of crashes. By crash models we mean equations where the expected number of injury accidents on a road section or at a junction is expressed as a function of the environmental characteristics, traffic and geometric characteristics on that section or junction. For non-negative, count data such as crashes poison and negative binomial models would be the natural way of modeling the data [73]. Negative binomial model was used to developed safety performance functions (SPFs) described in chapter four. This research developed an access decision support tool in conjunction with crash prediction models to simply the prediction using models that are often complicated and time consuming. The tool offers the following advantages.

i. It is built in form of a database: With huge amounts of data now available, local and national agencies are now building their database.
ii. User friendly: Unlike crash prediction models that are expressed in form of complicated equations and time consuming, this tool simplifies this work by developing a platform that could be used by any risk assessor either in the office or on site.
iii. It helps users to gain more insight into the relationships between crashes and sociodemographic factors by varying the values of contributing factors.

## 6.2 Data preparation

Two data sets of 5,845 pedestrian crashes and 2185 bicycle crashes were obtained from Tennessee Roadway Information Management System (TRIMS) which is a database maintained by TDOT. Crash data was imported into both geodatabase with ArcGIS and Microsoft excel. The 2006-2010 Census Block group TIGER shape file of America Community Survey (ACS) was imported into both geodatabase with ArcGIS and Microsoft excel. This data set consisted of 4125 block groups aggregated in Tennessee's 95 counties. This dataset provided all the socioeconomic and demographic information. The original crash data was checked to remove any duplicates. After removing any duplicates, geocoding of crash data was validated. Our final data sample consisted 4,816 and 1,808 pedestrian and bicycle crashes respectively. Crash data was then integrated with socioeconomic and demographic data using the 2006-2010 Census Block group TIGER shape file of America Community Survey as base layer. The integrated data set was imported to Microsoft excel for further processing. The clean files were imported to Microsoft access database.

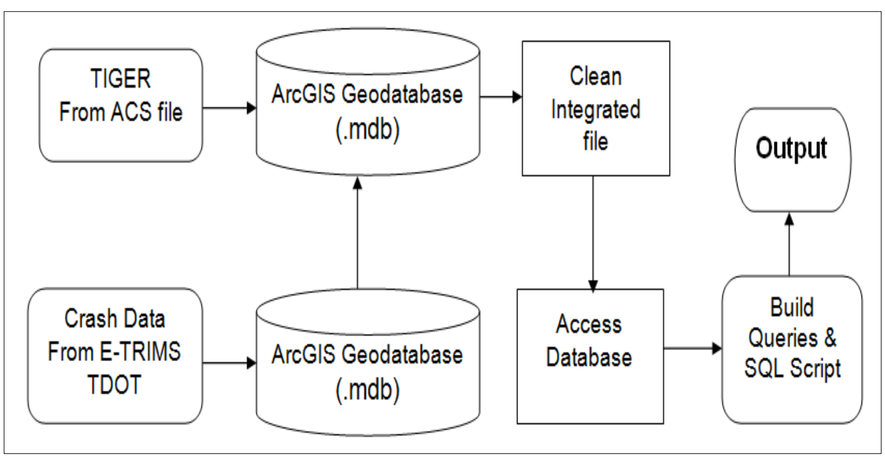


**Figure 44: Crash prediction tool development flowchart**

### 6.3 Developing the tool

We start by determining the prediction accuracy of the Safety performance functions (SPFs) developed. [74] Suggested that a common and simple approach to evaluate models is to regress predicted vs. observed values (or vice versa) and compare slope and intercept parameters against the 1:1 line. The analysis of the coefficient of determination ($R^2$), the slope and the intercept of the line fitted to the data provides elements for judging and building confidence on model performance. The $R^2$ shows the proportion of total variance explained by the regression model (and also how much of the linear variation in the observed values is explained by the variation in the predicted values) [75]. The $R^2$ can be interpreted by graphically plotting the observed values against model predictions. If predicted values are similar to the observed values, the points would be roughly a straight line with a slope 1.0 and the scatter points about the regression line would be small. Figure 45 shows a correction between actual and observed values. The analysis of the coefficient of determination ($R^2$) for the county crash prediction model. Indicates $R^2$ value of 0.9628 which is excellent prediction of the data by the model. The second part of this process was to incorporate the SPF in Microsoft access. The MS access database comprised of three main items; the table, queries and forms. The table saves raw data. Queries are used to linked different table properties using structural query language (SQL) commands while forms display results in form of reports. The process of crash prediction tool development is show in a flowchart in Figure 44.

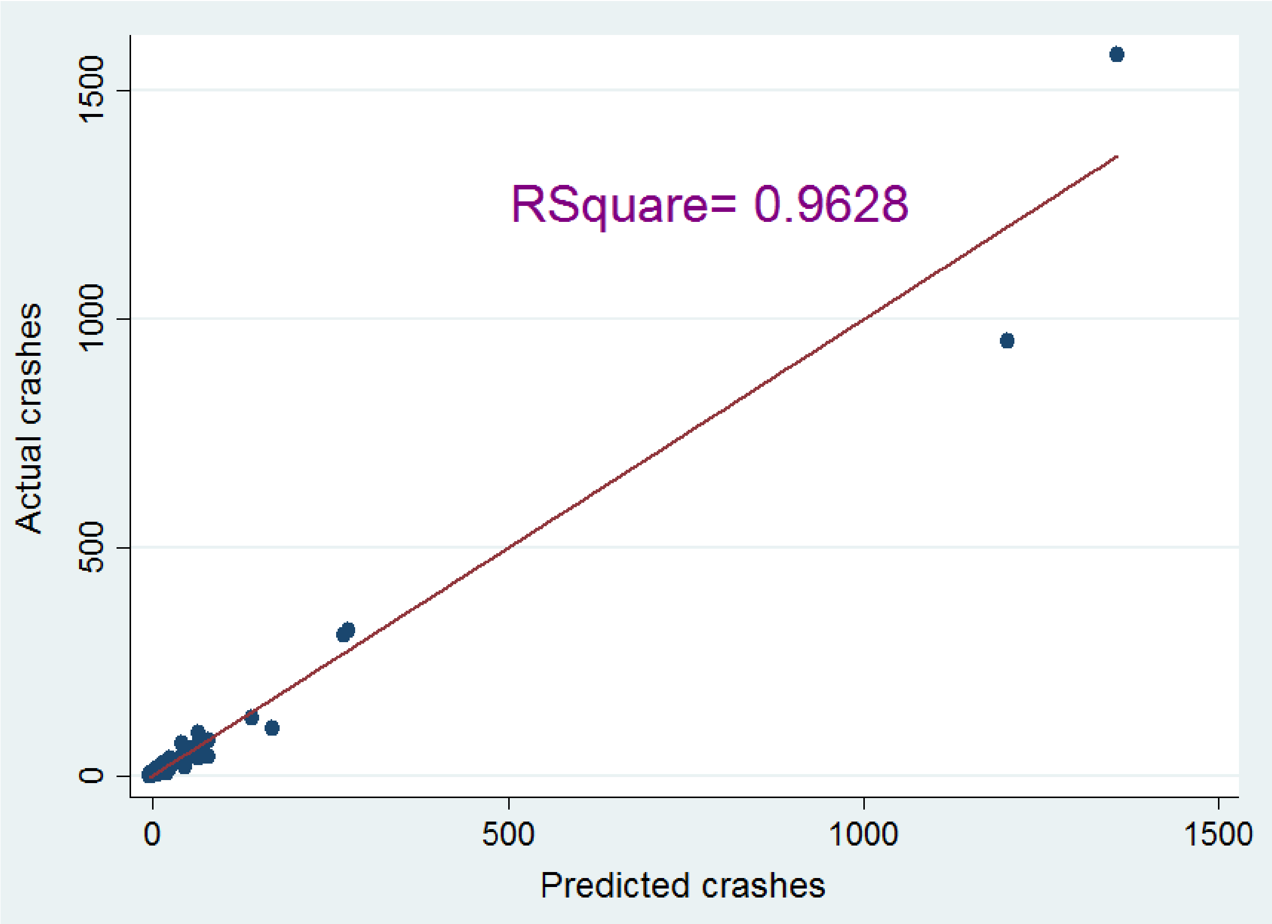


**Figure 45: Actual vs. Predicted pedestrian crashes**

### 6.4 The decision support system and a planning scenario

Consider a scenario in which a state planner in office or on site is required to allocate safety enforcement resources per county. It is assumed that the planner has all the funding and resources required in order to improve pedestrian and or bicycle safety, and the task now is to develop a criterion of allocating these safety enforcement resources. From the available statistics and it is established that the probability of crash occurrence is strongly associated with the mean values of certain socioeconomic and demographic variables within a county. The state planner is to decide how counties will take advantage of this funding to make better use of the already available resources?

We develop an access decision support system shown in Figure 46 whose interface is designed to automatically compute the number of crashes for a given county given the input values. To make the tool more flexible and user friendly, we develop another interface shown in Figure 47 that allows varying the input variables. Changing the values of variables in the decision support system results in a crash value. The resulting crash value is an indicator of the safety risk and can guide the State planner's appropriate allocation of safety improvement resources.

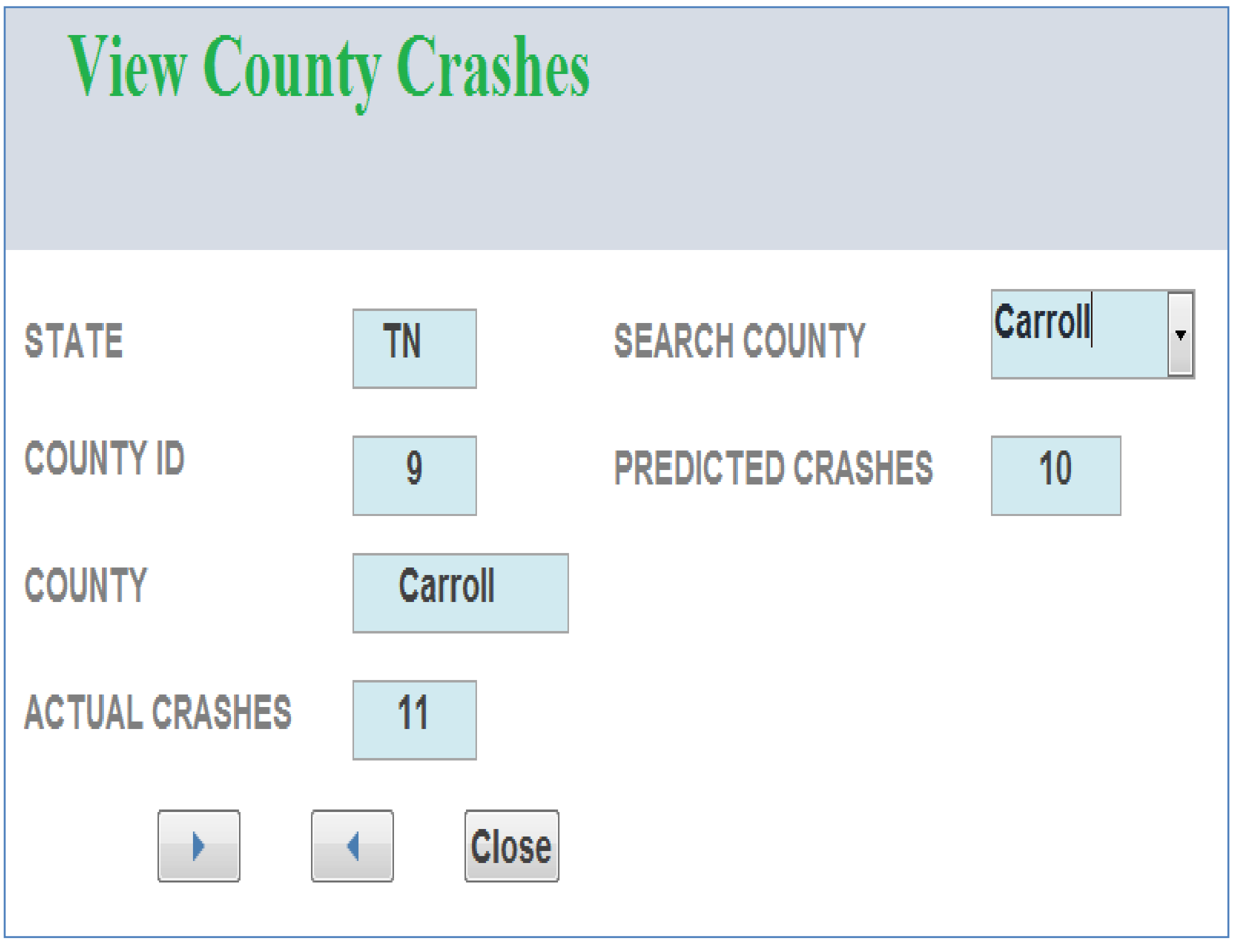


**Figure 46: Interface of Decision Support system**

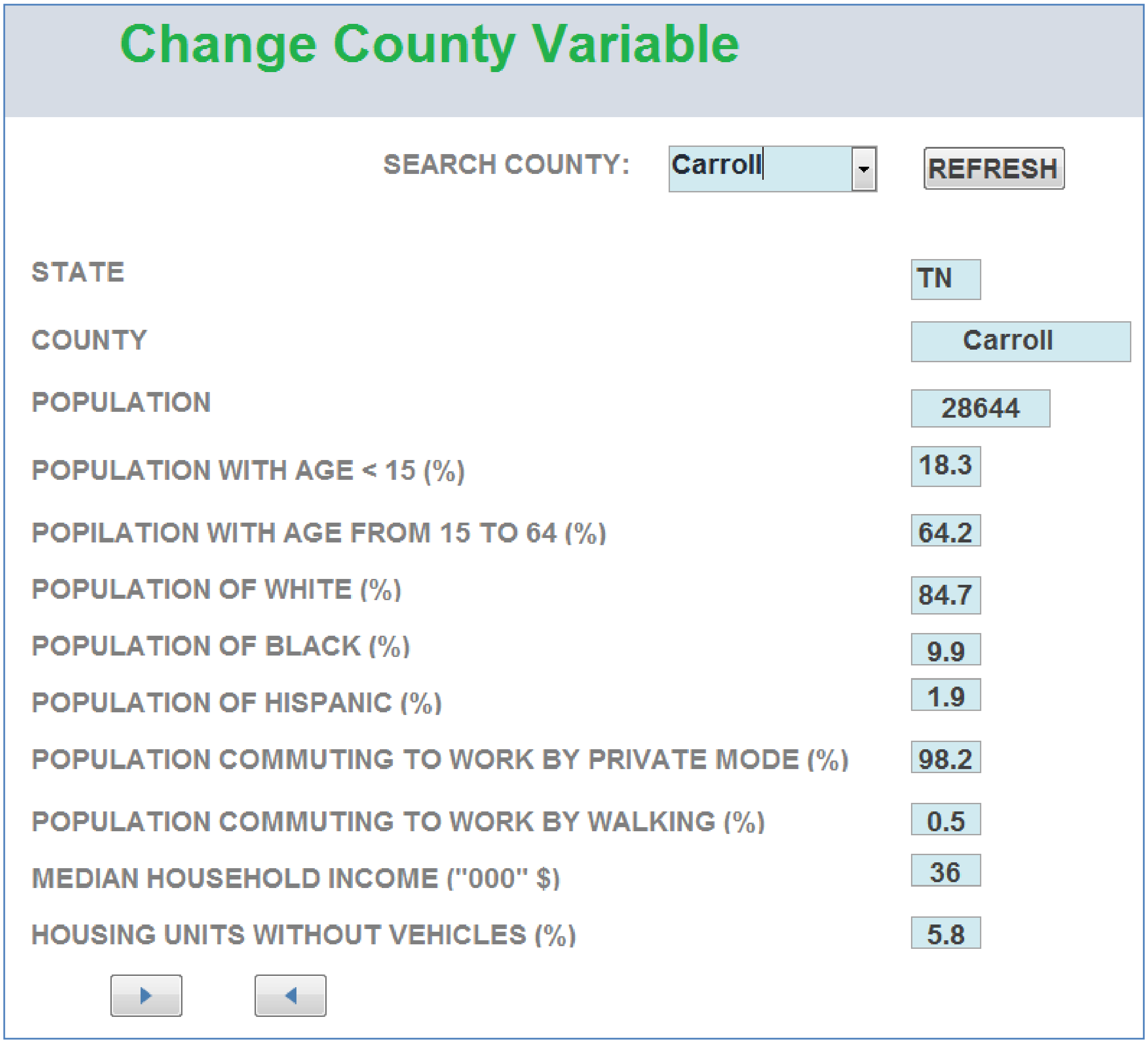


**Figure 47: Variables associated with crash occurrence**

# CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

## 7.1 Conclusions

The goal of this research was to develop decision support tools for the implementation of bicycle and pedestrian safety strategies. This tool combines the traditional methods such as those provided in the Highway Safety manual to predict the expected number of bicycle and pedestrian crashes. The work is based on a five year (2008-2012) statewide crash data obtained from Tennessee Roadway Information Management System (TRIMS) database maintained by TDOT. Socioeconomic and demographic data was obtained from Census Bureau website. Data from American Community Survey (ACS) 2006-2010 five year estimates at census block group level was used in the study. Three tools are proposed and developed by this research. First, a cluster analysis technique is proposed and developed a Geographic Information Systems (GIS) technique to facilitate the identification of crash clusters. A GIS Anselin local Moran's I tool was proposed to identify areas of clustering by location as well as by values of similar magnitude. This tool was used to identify significant clusters at a level of a block group and ultimately gave us improved information about high risk locations. Safety Performance Functions (SPFs) were then developed inform of mathematical equations to relate the number of crashes to area socioeconomic and demographic characteristics. The SPFs were developed using Negative Binomial Functions proposed in Highway capacity manual. The SPF gave improved information about associated factors. An integrated system consisting of access database and safety performance functions, and whose interface is designed to automatically compute the number of crashes given the input values is developed. Basing on crash value, the tool can be adopted as a framework to guide the appropriate allocation of safety improvement resources. Recommendations

## 7.2 Recommendations

Although this research successfully developed decision support tools to predict the number of crashes, future search need to develop similar tools that would take into consideration of crash severity. This would give rise to decisions that do not only reduce the number of crashes but also the degree of severity. Secondly, this research assessed pedestrian and bicycle safety with respect to socioeconomic and demographic factors. In order to develop a more comprehensive tools, future research should include other variables. For example, geometric elements and environmental characteristics of roadways that make geographical boundaries of block groups that are significant clusters of pedestrian and bicycle crashes should be investigated. Moreover, the study was based on population living in census block as exposure; however this may not represent the actual pedestrian activity. Similar future studies should use pedestrian volumes to achieve improved predictions. Much more work is needed to develop more robust and user friendly software. Future work can incorporate the current tool to a GIS in order to visualize the predicted number of crashes on a map.

## REFERENCES


[1] Federal Highway Administration, "The National Bicycling and Walking Study:15–Year Status Report," U.S Department of Transportation, Washington, DC, 2010.

[2] M. Hamer and Y. Chida, "Walking and primary prevention: a meta-analysis of prospective cohort studies," *Br J Sports Med* , vol. 42, p. 238–243, 2008.

[3] M. H. Murphy, A. M. Nevill, E. M. Murtagh and R. L. Holder, "The effect of walking on fitness, fatness and resting blood pressure: A meta-analysis of randomised, controlled trials," *Preventive Medicine* , vol. 44, p. 377 – 385, 2007.

[4] D. A. Sleet, K. Pollack and F. Rivara, "It wouldn't hurt to walk: promoting pedestrian injury research," *Injury Prevention,* vol. 16, pp. 211-212, 2010.

[5] NHTSA, "Traffic Safety Facts 2012 Data," National Center for Statistics and Analysis, Washington, DC, 2014.

[6] FHWA, "GIS Tools for Improving Pedestrian & Bicycle Safety TechBrief, FHWA RD-00-153, Federal Highway Administration (FHWA)," U.S. Department of Transportation (U.S. DOT), 2000.

[7] T. K. Anderson, "Kernel density estimation and K-means clustering to profile road accident hotspots," *Accident Analysis and Prevention ,* vol. 41, p. 359–364, 2009.

[8] D. L. Harkey, "Development of a GIS-Based Crash Referencing and Analysis System," in *Enhancing Transportation Safety in the 21st Century ITE International Conference*, Washington, 1999.

[9] A. Mitchell, The ESRI Guide to GIS Analysis, Volume 1: Geographic Patterns and Relationships, 1st ed., Redlands,CA: Esri Press, 1999.

[10] N. Levine, K. E. Kim and L. H. Nitz, "Spatial Analysis of Honolulu Motor Vehicle Crashes: I. Spatial Patterns," *Accident Analysis and Prevention,* vol. 27, no. 5, pp. 663-674, 1995.

[11] M. Bíl, R. Andrásik and Z. Janoska, "Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation," *Accident Analysis and Prevention,* vol. 55, p. 265–273, 2013.

[12] A. Getis and J. K. Ord, "The Analysis of Spatial Association by Use of Distance Statistics," *Geographical Analysis,* vol. 24, no. 3, p. 189–206, 1992.

[13] D. T. T. I. ,. F. B. Steenberghena T, "Intra-urban location and clustering of road accidents:a Belgian example," *International Journal of Geographical Information Science,* vol. 18, no. 2, p. 169–181, 2004.

[14] S. Erdogan, I. Yilmaz, T. Baybura and M. Gullu, "Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar," *Accident Analysis and Prevention* , vol. 40 , p. 174–181, 2008.

[15] ESRI, Arc/Info Documentation—Version 8.2, 2nd ed., Redlands, CA: Environmental Scientific and Research Institute, Inc, 2002b.

[16] S. S. Pulugurtha, V. K. Krishnakumar and S. S. Nambisan, "New methods to identify and rank high pedestriancrash zones: An illustration," *Accident Analysis and Prevention* , vol. 39 , p. 800–811, 2007.

[17] A. Bulajića, D. Jovanovićb, B. Matovićc and S. Bačkalićd, "Identification of High-Density Locations with Homogeneous attributes of Pedestrian crashes in the Urban area of Novi Sad," in *Road Accidents Prevention* , Borsko Jezero, 2014.

[18] B. P. Loo, S. Yao and J. Wu, "Spatial Point Analysis of Road Crashes in Shanghai A GIS-based Network Kernel Density Method," in *Geoinformatics, 2011 19th International Conference* , Shanghai, 2011.

[19] C. A. Blazquez and M. S. Celis, "A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile," *Accident Analysis and Prevention,* vol. 50, p. 304–311, 2013.

[20] N. Levine, "Crime Mapping and the Crimestat Program," *Geographical Analysis,* vol. 38, p. 41–56, 2006.

[21] D. Chimba, D. Emaasit, C. R. Cherry and Z. Pannell, "Patterning Demographic and Socioeconomic Characteristics Affecting Pedestrian and Bicycle Crash Frequency," in *Transportation Research Board*, Washington, 2014.

[22] CDOT, "Safety Performance Functions for intersections. Report No. CDOT-2009-10," Colorado Department of Transportation, Denver, CO, 2009.

[23] D. Kravetz and R. B. Noland, "A Spatial Analysis of Income Disparities in Pedestrian Safety in Northern New Jersey: Is there an Environmental Justice Issue?," in *Transportation Research Board*, Washington,DC, 2012.

[24] R. J. Schneider, R. M. Ryznar and A. J. Khattak, "An accident waiting to happen: a spatial approach to proactive pedestrian planning," *Accident Analysis and Prevention,* vol. 36, p. 193–211, 2004.

[25] E. Hauer, "Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation," *Accident Analysis & Prevention,* vol. 33, no. 6, p. 799–808, 2001.

[26] E. A. LaScala, D. Gerber and P. J. Gruenewald, "Demographic and environmental correlates of pedestrian injury collisions: A spatial analysis," *Accident Analysis and Prevention,* vol. 32, p. 651–658, 2000.

[27] C. R. B. D. A. H. Naveen Eluru, "A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes," *Accident Analysis and Prevention,* vol. 40, no. 3, p. 1033–1054, 2008.

[28] S. Yao and B. P. Y. Loo, "Identification of Hazardous Road Locations for Pedestrians," *Procedia Engineering,* vol. 45, p. 815 – 823, 2012.

[29] K. Geurts, I. Thomas and G. Wets, "Understanding spatial concentrations of road accidents using frequent item sets," *Accident Analysis and Prevention* , vol. 37, p. 787–799, 2005.

[30] B. J. Campbell, C. V. Zegeer, H. H. Huang and M. J. Cynecki, "A Review of Pedestrian Safety Research in the United States and Abroad," Federal Highway Administration, Washington,DC, 2004.

[31] H. Fontaine and Y. Gourlet, "Fatal Pedestrian accidents in France: A typological analysis," *Accident Analysis and Prevention,* vol. 29, no. 3 , pp. 303-312, 1997.

[32] P. Olszewski, P. Szagała, M. Wolanski´ and A. Zielinska´, "Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in Poland," *Accident Analysis and Prevention* , vol. 84, p. 83–91, 2015.

[33] National Center for Statistics and Analysis, "National Pedestrian Crash Report," U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC, 2008.

[34] M. Abdel-Aty and J. Keller, "Exploring the overall and specific crash severity levels at signalized intersections," *Accident Analysis and Prevention* , vol. 37, p. 417–425, 2005.

[35] R. W. G. Anderson, A. J. McLean, M. J. B. Farmer, B. H. Lee and C. G. Brooks, "Vehicle

Travel Speeds and the incidence of fatal pedestrian crashes," *Accident Analysis and Prevention* , vol. 29, no. 5, p. 667–674, 1997.

[36] D. J. Graham and S. Glaister, "Spatial Variation in Road Pedestrian Casualties:The Role of Urban Scale, Density and Land-use Mix," *Urban Studies,* vol. 40, no. 8, p. 1591–1607, 2003.

[37] K. J. Clifton and K. Kreamer-Fults, "An examination of the environmental attributes associated with pedestrian–vehicular crashes near public schools," *Accident Analysis and Prevention,* vol. 39, p. 708–715, 2007.

[38] J. Sun and J. Sun, "A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data," *Transportation Research Part C* , vol. 54 , p. 176–186, 2015.

[39] A. Gregoriades, "Towards a user-centred road safety management method based on road traffic simulation," in *Simulation Conference*, Washington, DC, 2007.

[40] R. Yu and M. Abdel-Aty, "Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data," *Accident Analysis and Prevention,* vol. 62, p. 161–167, 2014.

[41] C. Chen, G. Zhang, R. Tarefder, J. Ma, H. Wei and H. Guan, "A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes," *Accident Analysis and Prevention* , vol. 80 , p. 76–88, 2015.

[42] I. M. Abdalla, "Effectiveness of safety belts and Hierarchical Bayesian analysis of their relative use," *Safety Science* , vol. 43, p. 91 –103, 2005.

[43] J. Strauss, L. F. Miranda-Moreno and P. Morency, "Cyclist activity and injury risk analysis at signalized intersections:A Bayesian modelling approach," *Accident Analysis and Prevention,* vol. 59 , p. 9–17, 2013.

[44] R. Yu and M. Abdel-Aty, "Investigating different approaches to develop informative priors in hierarchical Bayesian safety performance functions," *Accident Analysis and Prevention,* vol. 56, p. 51–58, 2013.

[45] B. Lan, B. Persaud, C. Lyon and R. Bhim, "Validation of a Full Bayes methodology for observational before–after road safety studies and application to evaluation of rural signal conversions," *Accident Analysis and Prevention* , vol. 41, p. 574–580, 2009.

[46] O. Yanmaz-Tuzel and K. Ozbay, "A comparative Full Bayesian before-and-after analysis and application to urban road safety countermeasures in New Jersey," *Accident Analysis and Prevention* , vol. 42 , p. 2099–2107, 2010.

[47] D. Heckerman, "A Tutorial on Learning With Bayesian Networks," Microsoft Corporation, Redmond,WA, 1995.

[48] A. C. Mbakwe, A. A. Saka, K. Choi and Y.-J. Lee, "Modeling Highway Traffic Safety in Nigeria Using Delphi Technique and Bayesian Network," in *Transportation Research Board*, Washington,DC, 2014.

[49] L. Chong, M. M. Abbas, A. M. Flintsch and B. Higgs, "A rule-based neural network approach to model driver naturalistic behavior in traffic," *Transportation Research Part C 3,* vol. 2, p. 207–223, 2013.

[50] L. Deka and M. Quddus, "Network-level accident-mapping: Distance based pattern matching using artificial neural network," *Accident Analysis and Prevention,* vol. 65, p. 105–113, 2014.

[51] D. Delen, R. Sharda and M. Bessonov, "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks," *Accident Analysis and*

*Prevention* , vol. 38, p. 434–444, 2006.

[52] G. QiangZen and G. HelaiHuang, "A stable and optimized neural network model for crash injury severity prediction," *Accident Analysis and Prevention,* vol. 73, p. 351 –358, 2014.

[53] A. P. Engelbrecht, Computational Intelligence: An introduction;second edition, London: John Wiley & Sons Ltd, 2007.

[54] L.-Y. Chang, "Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network," *Safety science,* vol. 43, no. 8, p. 541–557, 2005.

[55] K. Kumara, M. Paridab and V. K. Katiyarc, "Short term traffic flow prediction for a non urban highway using Artificial Neural Network," *Procedia - Social and Behavioral Sciences,* vol. 104, p. 755 – 764, 2013.

[56] X. Jie, K. B. T and E.-G. Moustafa, "Prediction of Risk of Wet-Pavement Accidents: Fuzzy Logic Model," *Journal of the Transportation Research Board,* vol. 1717, pp. 28-36, 2000.

[57] M. Xiang-hai, Z. Lai and Q. Guan-ming, "Traffic accidents prediction and prominent influencing factors analysis based on fuzzy logic," *Journal of Transportation Systems and Information Technology,* vol. 9, no. 2, pp. 87-92, 2009.

[58] J.-S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy InferenceSystem," *IEEE Trans Syst Man Cybern Syst Hum,* vol. 23, no. 3, pp. 665-684, 1993.

[59] M. H. Hosseinlou and M. S. Moshtaghin, "Predicting and identifying traffic hot spots applying neuro-fuzzy systems in intercity roads," *International Journal of Enviornmental Science and Technology,* vol. 6, no. 2, pp. 309-314, 2009.

[60] M. Effati, M. A. Rajabi, F. Samadzadegan and S. Shabani, "A geospatial neuro-fuzzy approach for identification of hazardous zones in regional transportation corridors," *International Journal of Civil Engineering,* vol. 12, no. 3, pp. 289-303, 2014.

[61] A. T. Agrawal and P. S. Ashtankar, "Adaptive Neuro-Fuzzy Inference System for Health Monitoring at Home," *International Journal of Advanced Science and Technology,* vol. 55, pp. 89-100, 2013.

[62] D. R and P. S, "ANFIS based Neuro-Fuzzy Controller in LFC of Wind-Micro Hydro-Diesel Hybrid Power System," *International Journal of Computer Applications,* vol. 42, no. 6, 2012.

[63] Charles Zegeer; Dan Nabors; Dan Gelinne; Nancy Lefler; Max Bushell, "Pedestrian Safety Strategic Plan: Recommendations for Research and Product Development," Federal Highway Administration, Washington D.C, 2010.

[64] D. Andaluz, T. Robers and S. Siddall, "GIS Adds A New Dimension to Crash Analysis," *Journal of the Urban and Regional Information Systems Association*, vol. 9, no. 1, pp. 56-59, 1997.

[65] S. Mitra, "Spatial Autocorrelation and Bayesian Spatial Statistical Method for Analyzing Intersections Prone to Injury Crashes.Transportation Research Record," in *Journal of the Transportation Research Board 2136* , Washington, DC, 2009.

[66] J. Pucher and J. L. Renne, "Socioeconomics of Urban Travel: Evidence from the 2001 NHTS," Eno Transportation Foundation, Inc, Washington, DC, 2006.

[67] R. Srinivasan and K. Bauer, "Safety Performance Function Development Guide: Developing Jurisdiction-Specific SPFs," Publication FHWA-SA-14-005. FHWA, US Department of Transportation, Washington DC, 2013.

[68] R. A. Tegge, J.-H. Jo and Y. Ouyang, "Development and Application of Safety Performance

Functions for Illinois," Illinois Department of Transportation, Urbana, IL, 2010.

[69] D. Chimba, D. Emaasit, S. Allen, B. Hurst and M. Nelson, "Safety Effectiveness Evaluation of the Cable Rail Systems in Tennessee," in *Transportation Research Board 92nd Annual Meeting ( 13-0528)*, Washington, DC, 2013.

[70] S. R. Geedipally, D. Lord and B.-J. Park, "Analyzing Different Parameterizations of the Varying Dispersion Parameter as a Function of Segment Length," *Transportation Research Record*, vol. 2, no. 2013, pp. 108-118, 2009.

[71] D. Lord, S. R. Geedipally and S. D. Guikema, "Extension of the Application of Conway-Maxwell-Poisson Models: Analyzing Traffic Crash Data Exhibiting Underdispersion," *Risk Analysis*, vol. 30, no. 8, pp. 1268-1276, 2010.

[72] A. Kusumawati, D. P. Upahita and S. Hendarto, "Development of Accident Prediction Model for Rural Toll Road Sections," in *Proceedings of the Eastern Asia Society for Transportation Studies, Vol.9*, Bangdun, 2013.

[73] C. Zhang, X. Yan, L. Ma and M. An, "Crash Prediction and Risk Evaluation Based on Traffic Analysis Zones," *Mathematical Problems in Engineering*, vol. 2014 , p. 9, 2014.

[74] G. Pineiro, S. Perelman, J. P. Guerschman and J. M. Paruelo, "How to evaluate models: Observed vs. predicted or predicted vs. observed?," *Ecological Modelling*, vol. 216, pp. 316-322, 2008.

[75] E. P. Smith and K. A. Rose, "Model goodness-of-fit analysis using regression and related techniques," *EcologicalModelling*, vol. 77, pp. 49-64, 1995.

# APPENDICES

## Appendix A: Pedestrian crash modeling data

| ID | STATE | COUNTY | AREA | Population | Density | Age<15 | Age_15to64 | Age>=65 | Pop_White | Pop_Black | Pop_Hispanic | Race_others | Mode_Private | Mode_Walking | Mode_Others | Income | No_vehicle | Crashes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | TN | Anderson | 337.162 | 74257 | 0.220 | 18.16 | 64.96 | 16.88 | 89.76 | 3.98 | 2.20 | 4.06 | 96.98 | 0.77 | 2.25 | 45.35 | 6.94 | 35 |
| 2 | TN | Bedford | 473.635 | 44172 | 0.093 | 22.45 | 65.09 | 12.45 | 77.60 | 7.69 | 10.60 | 4.11 | 97.30 | 0.22 | 2.48 | 39.94 | 5.45 | 18 |
| 3 | TN | Benton | 394.142 | 16456 | 0.042 | 16.37 | 64.24 | 19.39 | 93.68 | 2.27 | 1.53 | 2.52 | 96.36 | 0.45 | 3.19 | 35.52 | 7.52 | 6 |
| 4 | TN | Bledsoe | 406.425 | 12946 | 0.032 | 16.58 | 68.67 | 14.75 | 92.11 | 3.51 | 1.99 | 2.39 | 97.27 | 0.56 | 2.17 | 30.77 | 5.42 | 1 |
| 5 | TN | Blount | 558.706 | 121140 | 0.217 | 18.48 | 66.24 | 15.27 | 91.19 | 2.84 | 2.52 | 3.45 | 97.72 | 0.34 | 1.94 | 46.38 | 4.24 | 39 |
| 6 | TN | Bradley | 328.762 | 97192 | 0.296 | 19.45 | 66.89 | 13.66 | 88.44 | 4.24 | 4.25 | 3.07 | 96.60 | 0.60 | 2.80 | 39.80 | 4.74 | 60 |
| 7 | TN | Campbell | 480.191 | 40623 | 0.085 | 18.22 | 65.11 | 16.67 | 96.19 | 0.38 | 1.10 | 2.33 | 97.68 | 0.36 | 1.96 | 29.16 | 8.63 | 29 |
| 8 | TN | Cannon | 265.635 | 13631 | 0.051 | 18.60 | 65.90 | 15.49 | 93.62 | 0.75 | 0.80 | 4.83 | 96.05 | 0.40 | 3.54 | 36.71 | 4.98 | 2 |
| 9 | TN | Carroll | 599.252 | 28644 | 0.048 | 18.26 | 64.24 | 17.50 | 84.71 | 9.93 | 1.88 | 3.47 | 98.18 | 0.47 | 1.35 | 35.90 | 5.78 | 11 |
| 10 | TN | Carter | 341.203 | 57710 | 0.169 | 16.69 | 66.70 | 16.62 | 94.53 | 1.20 | 1.57 | 2.70 | 95.96 | 0.74 | 3.31 | 32.01 | 6.08 | 16 |
| 11 | TN | Cheatham | 302.437 | 38809 | 0.128 | 20.76 | 68.73 | 10.51 | 93.76 | 1.53 | 2.29 | 2.42 | 96.06 | 0.29 | 3.65 | 54.89 | 2.56 | 9 |
| 12 | TN | Chester | 285.736 | 16793 | 0.059 | 19.60 | 66.05 | 14.35 | 85.47 | 8.94 | 1.73 | 3.85 | 96.08 | 1.33 | 2.59 | 39.64 | 5.11 | 5 |
| 13 | TN | Claiborne | 434.580 | 31901 | 0.073 | 17.21 | 67.54 | 15.24 | 95.85 | 0.85 | 0.90 | 2.40 | 97.39 | 0.37 | 2.24 | 31.14 | 6.84 | 8 |
| 14 | TN | Clay | 236.536 | 7888 | 0.033 | 14.55 | 63.70 | 21.74 | 97.50 | 1.77 | 0.56 | 0.16 | 95.29 | 1.34 | 3.37 | 33.77 | 4.28 | 2 |
| 15 | TN | Cocke | 434.565 | 35473 | 0.082 | 17.92 | 66.51 | 15.57 | 93.02 | 1.77 | 1.73 | 3.48 | 98.19 | 0.79 | 1.02 | 29.23 | 6.80 | 21 |
| 16 | TN | Coffee | 428.957 | 52344 | 0.122 | 20.51 | 64.21 | 15.28 | 88.70 | 2.61 | 3.61 | 5.09 | 97.12 | 0.62 | 2.26 | 39.59 | 5.92 | 38 |
| 17 | TN | Crockett | 265.535 | 14524 | 0.055 | 20.17 | 63.63 | 16.20 | 77.00 | 14.18 | 7.92 | 0.90 | 97.31 | 0.79 | 1.90 | 36.83 | 4.09 | 2 |
| 18 | TN | Cumberland | 681.025 | 54977 | 0.081 | 16.01 | 59.22 | 24.77 | 95.36 | 0.37 | 2.16 | 2.12 | 96.86 | 0.42 | 2.72 | 36.10 | 4.36 | 24 |
| 19 | TN | Davidson | 504.033 | 612884 | 1.216 | 18.54 | 70.99 | 10.46 | 57.74 | 27.25 | 8.83 | 6.18 | 94.29 | 0.90 | 4.80 | 49.83 | 7.54 | 951 |
| 20 | TN | Decatur | 333.845 | 11716 | 0.035 | 17.67 | 62.62 | 19.71 | 91.92 | 2.33 | 2.36 | 3.40 | 98.27 | 0.28 | 1.45 | 29.41 | 7.92 | 1 |
| 21 | TN | DeKalb | 304.347 | 18569 | 0.061 | 19.53 | 65.35 | 15.12 | 90.48 | 1.57 | 5.70 | 2.25 | 96.94 | 1.27 | 1.79 | 34.24 | 6.24 | 7 |
| 22 | TN | Dickson | 489.896 | 48712 | 0.099 | 20.65 | 66.85 | 12.50 | 89.11 | 4.00 | 2.85 | 4.04 | 97.72 | 0.40 | 1.88 | 45.40 | 5.69 | 24 |
| 23 | TN | Dyer | 512.327 | 38174 | 0.075 | 20.48 | 65.63 | 13.89 | 80.70 | 14.33 | 2.47 | 2.50 | 98.03 | 0.33 | 1.64 | 37.47 | 8.87 | 27 |
| 24 | TN | Fayette | 704.786 | 37458 | 0.053 | 20.08 | 65.86 | 14.06 | 67.57 | 28.37 | 2.22 | 1.84 | 96.93 | 0.82 | 2.25 | 54.52 | 5.09 | 9 |
| 25 | TN | Fentress | 498.612 | 17777 | 0.036 | 19.06 | 64.81 | 16.13 | 97.00 | 0.06 | 0.92 | 2.03 | 95.57 | 0.44 | 3.99 | 28.96 | 4.83 | 7 |
| 26 | TN | Franklin | 554.542 | 41054 | 0.074 | 18.44 | 64.81 | 16.75 | 87.07 | 3.15 | 2.42 | 7.36 | 96.47 | 1.01 | 2.52 | 42.60 | 4.70 | 12 |
| 27 | TN | Gibson | 602.742 | 49015 | 0.081 | 20.50 | 62.97 | 16.53 | 76.97 | 18.97 | 2.00 | 2.06 | 97.66 | 0.72 | 1.62 | 36.47 | 7.91 | 15 |
| 28 | TN | Giles | 610.927 | 29558 | 0.048 | 18.07 | 65.55 | 16.38 | 84.73 | 10.68 | 1.64 | 2.94 | 97.99 | 0.43 | 1.58 | 38.05 | 5.51 | 5 |
| 29 | TN | Grainger | 280.600 | 22419 | 0.080 | 18.07 | 66.78 | 15.15 | 95.44 | 0.30 | 2.04 | 2.22 | 95.71 | 0.85 | 3.44 | 30.01 | 6.19 | 6 |
| 30 | TN | Greene | 622.165 | 68172 | 0.110 | 17.36 | 65.92 | 16.72 | 93.51 | 2.02 | 2.35 | 2.12 | 97.31 | 0.51 | 2.18 | 37.69 | 4.31 | 25 |
| 31 | TN | Grundy | 360.534 | 13910 | 0.039 | 19.35 | 63.95 | 16.71 | 76.40 | 0.47 | 0.80 | 22.33 | 94.21 | 0.95 | 4.85 | 27.24 | 5.52 | 5 |

| ID | STATE | COUNTY | AREA | Population | Density | Age<15 | Age_15to64 | Age>=65 | Pop_White | Pop_Black | Pop_Hispanic | Race_others | Mode_Private | Mode_Walking | Mode_Others | Income | No_vehicle | Crashes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 33 | TN | Hamilton | 542.431 | 328960 | 0.606 | 17.97 | 67.61 | 14.43 | 71.67 | 19.91 | 3.90 | 4.53 | 96.23 | 0.96 | 2.81 | 49.33 | 7.79 | 316 |
| 34 | TN | Hancock | 222.340 | 6782 | 0.031 | 17.40 | 66.06 | 16.54 | 97.11 | 0.20 | 0.29 | 2.40 | 97.64 | 1.68 | 0.69 | 26.27 | 10.80 | 0 |
| 35 | TN | Hardeman | 667.768 | 27655 | 0.041 | 17.49 | 68.96 | 13.55 | 55.41 | 40.59 | 1.34 | 2.66 | 96.28 | 1.06 | 2.67 | 31.28 | 10.32 | 6 |
| 36 | TN | Hardin | 577.318 | 25995 | 0.045 | 17.87 | 64.34 | 17.80 | 91.21 | 2.98 | 1.76 | 4.04 | 97.98 | 0.37 | 1.66 | 31.83 | 5.26 | 16 |
| 37 | TN | Hawkins | 486.975 | 56562 | 0.116 | 18.58 | 65.53 | 15.89 | 95.43 | 1.16 | 1.22 | 2.20 | 98.39 | 0.31 | 1.30 | 35.71 | 5.41 | 16 |
| 38 | TN | Haywood | 533.112 | 19010 | 0.036 | 21.61 | 64.84 | 13.54 | 44.58 | 50.21 | 3.30 | 1.91 | 98.39 | 0.20 | 1.41 | 476.82 | 11.73 | 5 |
| 39 | TN | Henderson | 520.073 | 27518 | 0.053 | 20.12 | 65.43 | 14.45 | 87.08 | 7.74 | 1.62 | 3.56 | 97.03 | 0.15 | 2.83 | 38.47 | 4.29 | 10 |
| 40 | TN | Henry | 562.096 | 32042 | 0.057 | 18.43 | 62.60 | 18.97 | 86.90 | 8.00 | 1.64 | 3.45 | 97.32 | 0.48 | 2.20 | 36.65 | 7.00 | 15 |
| 41 | TN | Hickman | 612.499 | 24506 | 0.040 | 18.79 | 67.97 | 13.24 | 90.81 | 4.41 | 1.70 | 3.09 | 97.21 | 0.36 | 2.43 | 41.41 | 5.22 | 10 |
| 42 | TN | Houston | 200.286 | 8286 | 0.041 | 19.78 | 62.89 | 17.33 | 92.56 | 1.57 | 1.66 | 4.20 | 97.46 | 0.67 | 1.87 | 34.97 | 8.11 | 2 |
| 43 | TN | Humphreys | 530.980 | 18416 | 0.035 | 19.06 | 64.31 | 16.63 | 93.96 | 2.62 | 1.27 | 2.15 | 98.65 | 0.19 | 1.16 | 41.07 | 4.78 | 5 |
| 44 | TN | Jackson | 308.320 | 11491 | 0.037 | 16.86 | 66.38 | 16.76 | 95.42 | 0.30 | 0.94 | 3.34 | 97.26 | 0.52 | 2.22 | 31.98 | 6.09 | 3 |
| 45 | TN | Jefferson | 274.078 | 50600 | 0.185 | 18.46 | 66.24 | 15.30 | 92.30 | 2.29 | 2.90 | 2.50 | 97.02 | 0.99 | 1.99 | 39.47 | 4.51 | 11 |
| 46 | TN | Johnson | 298.475 | 18190 | 0.061 | 14.80 | 67.81 | 17.39 | 94.44 | 2.09 | 1.32 | 2.15 | 97.50 | 0.70 | 1.79 | 30.21 | 4.65 | 6 |
| 47 | TN | Knox | 508.215 | 423748 | 0.834 | 18.25 | 68.91 | 12.84 | 83.43 | 8.82 | 3.03 | 4.72 | 95.99 | 0.81 | 3.19 | 50.69 | 5.71 | 309 |
| 48 | TN | Lake | 165.784 | 7827 | 0.047 | 13.67 | 72.97 | 13.36 | 67.67 | 27.63 | 1.29 | 3.41 | 98.84 | 0.13 | 1.02 | 24.35 | 10.26 | 3 |
| 49 | TN | Lauderdale | 471.992 | 27745 | 0.059 | 20.53 | 67.32 | 12.15 | 60.77 | 34.35 | 1.75 | 3.12 | 98.35 | 0.18 | 1.46 | 33.56 | 7.88 | 5 |
| 50 | TN | Lawrence | 617.128 | 41319 | 0.067 | 21.00 | 63.43 | 15.57 | 94.34 | 1.64 | 1.66 | 2.36 | 97.66 | 0.33 | 2.02 | 34.98 | 6.51 | 21 |
| 51 | TN | Lewis | 282.089 | 12003 | 0.043 | 21.12 | 63.75 | 15.13 | 94.40 | 1.16 | 2.56 | 1.88 | 97.20 | 0.72 | 2.08 | 36.29 | 5.85 | 5 |
| 52 | TN | Lincoln | 570.338 | 32885 | 0.058 | 19.39 | 64.03 | 16.59 | 86.28 | 6.02 | 2.43 | 5.27 | 97.64 | 0.43 | 1.93 | 41.41 | 5.20 | 14 |
| 53 | TN | Loudon | 229.216 | 47102 | 0.205 | 16.85 | 62.82 | 20.33 | 89.72 | 0.74 | 6.11 | 3.43 | 97.50 | 0.65 | 1.86 | 49.65 | 4.25 | 16 |
| 54 | TN | McMinn | 430.125 | 52075 | 0.121 | 18.97 | 65.00 | 16.03 | 89.32 | 3.89 | 2.73 | 4.05 | 96.60 | 1.18 | 2.22 | 36.67 | 5.56 | 17 |
| 55 | TN | McNairy | 562.860 | 25760 | 0.046 | 19.11 | 64.07 | 16.82 | 89.74 | 5.77 | 1.40 | 3.09 | 97.47 | 0.54 | 1.99 | 34.68 | 6.28 | 7 |
| 56 | TN | Macon | 307.144 | 21934 | 0.071 | 21.22 | 65.45 | 13.33 | 93.99 | 0.51 | 3.47 | 2.03 | 96.96 | 0.66 | 2.39 | 33.03 | 5.50 | 10 |
| 57 | TN | Madison | 557.117 | 97378 | 0.175 | 20.16 | 67.19 | 12.65 | 58.34 | 35.18 | 3.17 | 3.31 | 96.88 | 0.64 | 2.48 | 39.35 | 7.69 | 77 |
| 58 | TN | Marion | 498.160 | 28123 | 0.056 | 18.10 | 66.79 | 15.12 | 90.50 | 1.65 | 1.20 | 6.66 | 97.35 | 0.52 | 2.13 | 38.97 | 5.43 | 11 |
| 59 | TN | Marshall | 375.460 | 29902 | 0.080 | 20.65 | 66.70 | 12.65 | 85.76 | 6.79 | 4.35 | 3.10 | 97.23 | 0.46 | 2.32 | 41.06 | 4.02 | 13 |
| 60 | TN | Maury | 613.138 | 79029 | 0.129 | 20.41 | 67.02 | 12.57 | 79.22 | 12.75 | 4.61 | 3.42 | 97.60 | 0.42 | 1.98 | 44.05 | 4.48 | 40 |
| 61 | TN | Meigs | 195.122 | 11581 | 0.059 | 18.41 | 66.80 | 14.79 | 95.26 | 1.07 | 1.25 | 2.42 | 95.23 | 2.53 | 2.25 | 36.43 | 6.34 | 0 |
| 62 | TN | Monroe | 635.565 | 44015 | 0.069 | 19.24 | 65.34 | 15.42 | 91.17 | 1.95 | 3.13 | 3.75 | 97.67 | 0.86 | 1.47 | 35.96 | 4.58 | 12 |
| 63 | TN | Montgomery | 539.177 | 163603 | 0.303 | 23.65 | 68.33 | 8.02 | 65.97 | 18.37 | 7.29 | 8.36 | 97.00 | 0.81 | 2.18 | 46.58 | 3.95 | 127 |
| 64 | TN | Moore | 129.223 | 6266 | 0.048 | 18.72 | 64.27 | 17.01 | 91.84 | 1.33 | 0.51 | 6.32 | 97.89 | 0.30 | 1.81 | 43.80 | 3.71 | 0 |

| ID | STATE | COUNTY | AREA | Population | Density | Age<15 | Age_15to64 | Age>=65 | Pop_White | Pop_Black | Pop_Hispanic | Race_others | Mode_Private | Mode_Walking | Mode_Others | Income | No_vehicle | Crashes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 65 | TN | Morgan | 522.180 | 21664 | 0.041 | 18.04 | 68.52 | 13.44 | 92.97 | 2.62 | 0.79 | 3.63 | 96.57 | 1.04 | 2.39 | 36.66 | 5.27 | 5 |
| 66 | TN | Obion | 544.728 | 31905 | 0.059 | 19.08 | 64.69 | 16.23 | 84.46 | 10.17 | 2.99 | 2.38 | 97.84 | 0.50 | 1.66 | 40.23 | 7.92 | 9 |
| 67 | TN | Overton | 433.483 | 21777 | 0.050 | 19.26 | 64.38 | 16.35 | 95.68 | 0.56 | 0.18 | 3.58 | 97.54 | 0.17 | 2.29 | 36.36 | 4.45 | 6 |
| 68 | TN | Perry | 414.731 | 7778 | 0.019 | 18.37 | 64.25 | 17.38 | 94.68 | 2.43 | 0.91 | 1.97 | 96.55 | 0.72 | 2.73 | 31.36 | 4.31 | 0 |
| 69 | TN | Pickett | 162.979 | 5072 | 0.031 | 19.42 | 62.95 | 17.63 | 93.51 | 0.00 | 2.51 | 3.98 | 97.63 | 1.30 | 1.07 | 30.23 | 5.11 | 0 |
| 70 | TN | Polk | 434.676 | 16690 | 0.038 | 18.27 | 65.43 | 16.29 | 95.56 | 0.09 | 1.12 | 3.22 | 96.19 | 0.70 | 3.11 | 34.23 | 5.83 | 7 |
| 71 | TN | Putnam | 401.103 | 70570 | 0.176 | 17.90 | 67.98 | 14.12 | 89.42 | 1.94 | 5.02 | 3.63 | 97.41 | 0.72 | 1.87 | 35.73 | 4.30 | 32 |
| 72 | TN | Rhea | 315.377 | 31215 | 0.099 | 19.45 | 65.42 | 15.13 | 91.55 | 2.09 | 3.31 | 3.05 | 96.70 | 0.76 | 2.54 | 34.93 | 5.23 | 5 |
| 73 | TN | Roane | 360.708 | 54156 | 0.150 | 17.33 | 64.87 | 17.79 | 92.81 | 2.42 | 1.37 | 3.40 | 97.30 | 0.47 | 2.22 | 43.16 | 4.79 | 18 |
| 74 | TN | Robertson | 476.287 | 64347 | 0.135 | 21.80 | 66.66 | 11.53 | 84.17 | 7.28 | 5.53 | 3.02 | 96.69 | 0.27 | 3.03 | 51.98 | 5.53 | 36 |
| 75 | TN | Rutherford | 619.364 | 250517 | 0.404 | 22.06 | 69.96 | 7.98 | 75.77 | 12.00 | 6.08 | 6.14 | 97.19 | 0.61 | 2.20 | 55.21 | 3.20 | 103 |
| 76 | TN | Scott | 532.297 | 22171 | 0.042 | 20.97 | 65.96 | 13.06 | 97.67 | 0.00 | 0.93 | 1.40 | 98.04 | 0.22 | 1.74 | 29.52 | 8.49 | 3 |
| 77 | TN | Sequatchie | 265.858 | 13814 | 0.052 | 20.12 | 65.09 | 14.79 | 84.46 | 0.00 | 2.48 | 13.06 | 96.93 | 1.11 | 1.97 | 35.28 | 4.70 | 6 |
| 78 | TN | Sevier | 592.500 | 87507 | 0.148 | 18.14 | 66.98 | 14.88 | 91.41 | 0.74 | 4.42 | 3.43 | 96.72 | 0.80 | 2.47 | 40.37 | 4.32 | 71 |
| 79 | TN | Shelby | 763.174 | 922696 | 1.209 | 22.16 | 67.79 | 10.04 | 39.24 | 50.66 | 5.04 | 5.06 | 95.60 | 0.83 | 3.57 | 47.40 | 9.53 | 1578 |
| 80 | TN | Smith | 314.289 | 19035 | 0.061 | 20.06 | 66.56 | 13.38 | 92.73 | 1.85 | 1.92 | 3.49 | 96.62 | 1.02 | 2.37 | 42.01 | 4.75 | 8 |
| 81 | TN | Stew art | 459.330 | 13133 | 0.029 | 19.33 | 64.61 | 16.07 | 93.22 | 1.57 | 1.72 | 3.49 | 96.39 | 0.43 | 3.19 | 31.38 | 7.28 | 4 |
| 82 | TN | Sullivan | 413.363 | 155815 | 0.377 | 17.03 | 65.10 | 17.88 | 93.71 | 1.87 | 1.38 | 3.04 | 97.71 | 0.55 | 1.74 | 40.29 | 5.91 | 94 |
| 83 | TN | Sumner | 529.449 | 155592 | 0.294 | 21.02 | 66.85 | 12.13 | 86.73 | 6.51 | 3.56 | 3.20 | 96.44 | 0.38 | 3.18 | 56.54 | 3.50 | 55 |
| 84 | TN | Tipton | 458.366 | 59689 | 0.130 | 21.99 | 67.29 | 10.72 | 75.94 | 18.65 | 2.02 | 3.39 | 97.89 | 0.50 | 1.61 | 46.05 | 3.47 | 18 |
| 85 | TN | Trousdale | 114.193 | 7751 | 0.068 | 19.60 | 67.17 | 13.24 | 85.09 | 7.07 | 2.30 | 5.54 | 96.64 | 0.74 | 2.62 | 45.12 | 4.26 | 5 |
| 86 | TN | Unicoi | 186.165 | 18257 | 0.098 | 17.02 | 63.99 | 18.99 | 94.91 | 0.78 | 3.46 | 0.85 | 95.24 | 0.70 | 4.06 | 34.35 | 6.30 | 5 |
| 87 | TN | Union | 223.549 | 19088 | 0.085 | 19.56 | 67.21 | 13.23 | 96.90 | 0.00 | 1.03 | 2.06 | 97.27 | 0.46 | 2.27 | 31.01 | 4.59 | 5 |
| 88 | TN | Van Buren | 273.415 | 5503 | 0.020 | 17.35 | 66.67 | 15.97 | 93.05 | 1.76 | 0.62 | 4.57 | 96.63 | 0.82 | 2.55 | 27.96 | 3.34 | 1 |
| 89 | TN | Warren | 432.680 | 39539 | 0.091 | 20.35 | 64.95 | 14.70 | 86.67 | 2.84 | 7.69 | 2.81 | 96.91 | 0.72 | 2.37 | 34.87 | 4.48 | 14 |
| 90 | TN | Washington | 326.465 | 119768 | 0.367 | 17.07 | 68.17 | 14.76 | 89.47 | 3.83 | 2.68 | 4.02 | 96.88 | 0.63 | 2.48 | 40.96 | 5.64 | 70 |
| 91 | TN | Wayne | 734.100 | 17016 | 0.023 | 15.58 | 69.26 | 15.16 | 90.31 | 5.73 | 1.42 | 2.54 | 97.90 | 0.72 | 1.38 | 34.56 | 5.89 | 3 |
| 92 | TN | Weak ley | 580.364 | 34557 | 0.060 | 16.99 | 68.07 | 14.94 | 86.86 | 6.98 | 1.90 | 4.26 | 96.31 | 1.42 | 2.27 | 34.73 | 5.38 | 13 |
| 93 | TN | White | 376.673 | 25460 | 0.068 | 18.57 | 64.62 | 16.81 | 93.90 | 1.53 | 1.57 | 3.00 | 96.59 | 0.36 | 3.05 | 34.05 | 6.72 | 5 |
| 94 | TN | Williamson | 582.599 | 174260 | 0.299 | 24.15 | 66.68 | 9.16 | 86.11 | 4.20 | 4.15 | 5.53 | 94.82 | 0.32 | 4.85 | 92.00 | 2.10 | 41 |
| 95 | TN | Wilson | 570.826 | 109563 | 0.192 | 21.03 | 67.45 | 11.52 | 87.05 | 6.43 | 2.94 | 3.57 | 96.42 | 0.33 | 3.25 | 63.41 | 3.51 | 32 |

## Appendix B: Bicycle crash modeling data

| COUNTY ID | STATE | COUNTY | AREA | Population | Density | Age<15 | Age_15to64 | Age>=65 | Pop_White | Pop_Black | Pop_Hispanic | Race_others | Mode_Private | Mode_Bicycling | Mode_Others | Income | No_vehicle | Crashes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | TN | Anderson | 337.162 | 74257 | 0.220 | 18.16 | 64.96 | 16.88 | 89.76 | 3.98 | 2.20 | 4.06 | 96.98 | 0.00 | 2.25 | 45.35 | 6.94 | 18 |
| 2 | TN | Bedford | 473.635 | 44172 | 0.093 | 22.45 | 65.09 | 12.45 | 77.60 | 7.69 | 10.60 | 4.11 | 97.30 | 0.00 | 2.48 | 39.94 | 5.45 | 7 |
| 3 | TN | Benton | 394.142 | 16456 | 0.042 | 16.37 | 64.24 | 19.39 | 93.68 | 2.27 | 1.53 | 2.52 | 96.36 | 0.00 | 3.19 | 35.52 | 7.52 | 1 |
| 4 | TN | Bledsoe | 406.425 | 12946 | 0.032 | 16.58 | 68.67 | 14.75 | 92.11 | 3.51 | 1.99 | 2.39 | 97.27 | 0.00 | 2.17 | 30.77 | 5.42 | 1 |
| 5 | TN | Blount | 558.706 | 121140 | 0.217 | 18.48 | 66.24 | 15.27 | 91.19 | 2.84 | 2.52 | 3.45 | 97.72 | 0.05 | 1.94 | 46.38 | 4.24 | 30 |
| 6 | TN | Bradley | 328.762 | 97192 | 0.296 | 19.45 | 66.89 | 13.66 | 88.44 | 4.24 | 4.25 | 3.07 | 96.60 | 0.06 | 2.80 | 39.80 | 4.74 | 32 |
| 7 | TN | Campbell | 480.191 | 40623 | 0.085 | 18.22 | 65.11 | 16.67 | 96.19 | 0.38 | 1.10 | 2.33 | 97.68 | 0.06 | 1.96 | 29.16 | 8.63 | 2 |
| 8 | TN | Cannon | 265.635 | 13631 | 0.051 | 18.60 | 65.90 | 15.49 | 93.62 | 0.75 | 0.80 | 4.83 | 96.05 | 0.00 | 3.54 | 36.71 | 4.98 | 0 |
| 9 | TN | Carroll | 599.252 | 28644 | 0.048 | 18.26 | 64.24 | 17.50 | 84.71 | 9.93 | 1.88 | 3.47 | 98.18 | 0.02 | 1.35 | 35.90 | 5.78 | 2 |
| 10 | TN | Carter | 341.203 | 57710 | 0.169 | 16.69 | 66.70 | 16.62 | 94.53 | 1.20 | 1.57 | 2.70 | 95.96 | 0.02 | 3.31 | 32.01 | 6.08 | 8 |
| 11 | TN | Cheatham | 302.437 | 38809 | 0.128 | 20.76 | 68.73 | 10.51 | 93.76 | 1.53 | 2.29 | 2.42 | 96.06 | 0.02 | 3.65 | 54.89 | 2.56 | 2 |
| 12 | TN | Chester | 285.736 | 16793 | 0.059 | 19.60 | 66.05 | 14.35 | 85.47 | 8.94 | 1.73 | 3.85 | 96.08 | 0.21 | 2.59 | 39.64 | 5.11 | 1 |
| 13 | TN | Claiborne | 434.580 | 31901 | 0.073 | 17.21 | 67.54 | 15.24 | 95.85 | 0.85 | 0.90 | 2.40 | 97.39 | 0.00 | 2.24 | 31.14 | 6.84 | 1 |
| 14 | TN | Clay | 236.536 | 7888 | 0.033 | 14.55 | 63.70 | 21.74 | 97.50 | 1.77 | 0.56 | 0.16 | 95.29 | 0.00 | 3.37 | 33.77 | 4.28 | 0 |
| 15 | TN | Cocke | 434.565 | 35473 | 0.082 | 17.92 | 66.51 | 15.57 | 93.02 | 1.77 | 1.73 | 3.48 | 98.19 | 0.00 | 1.02 | 29.23 | 6.80 | 9 |
| 16 | TN | Coffee | 428.957 | 52344 | 0.122 | 20.51 | 64.21 | 15.28 | 88.70 | 2.61 | 3.61 | 5.09 | 97.12 | 0.09 | 2.26 | 39.59 | 5.92 | 9 |
| 17 | TN | Crockett | 265.535 | 14524 | 0.055 | 20.17 | 63.63 | 16.20 | 77.00 | 14.18 | 7.92 | 0.90 | 97.31 | 0.12 | 1.90 | 36.83 | 4.09 | 3 |
| 18 | TN | Cumberland | 681.025 | 54977 | 0.081 | 16.01 | 59.22 | 24.77 | 95.36 | 0.37 | 2.16 | 2.12 | 96.86 | 0.08 | 2.72 | 36.10 | 4.36 | 4 |
| 19 | TN | Davidson | 504.033 | 612884 | 1.216 | 18.54 | 70.99 | 10.46 | 57.74 | 27.25 | 8.83 | 6.18 | 94.29 | 0.14 | 4.80 | 49.83 | 7.54 | 304 |
| 20 | TN | Decatur | 333.845 | 11716 | 0.035 | 17.67 | 62.62 | 19.71 | 91.92 | 2.33 | 2.36 | 3.40 | 98.27 | 0.00 | 1.45 | 29.41 | 7.92 | 2 |
| 21 | TN | DeKalb | 304.347 | 18569 | 0.061 | 19.53 | 65.35 | 15.12 | 90.48 | 1.57 | 5.70 | 2.25 | 96.94 | 0.00 | 1.79 | 34.24 | 6.24 | 0 |
| 22 | TN | Dickson | 489.896 | 48712 | 0.099 | 20.65 | 66.85 | 12.50 | 89.11 | 4.00 | 2.85 | 4.04 | 97.72 | 0.00 | 1.88 | 45.40 | 5.69 | 7 |
| 23 | TN | Dyer | 512.327 | 38174 | 0.075 | 20.48 | 65.63 | 13.89 | 80.70 | 14.33 | 2.47 | 2.50 | 98.03 | 0.00 | 1.64 | 37.47 | 8.87 | 5 |
| 24 | TN | Fayette | 704.786 | 37458 | 0.053 | 20.08 | 65.86 | 14.06 | 67.57 | 28.37 | 2.22 | 1.84 | 96.93 | 0.01 | 2.25 | 54.52 | 5.09 | 5 |
| 25 | TN | Fentress | 498.612 | 17777 | 0.036 | 19.06 | 64.81 | 16.13 | 97.00 | 0.06 | 0.92 | 2.03 | 95.57 | 0.00 | 3.99 | 28.96 | 4.83 | 0 |
| 26 | TN | Franklin | 554.542 | 41054 | 0.074 | 18.44 | 64.81 | 16.75 | 87.07 | 3.15 | 2.42 | 7.36 | 96.47 | 0.42 | 2.52 | 42.60 | 4.70 | 5 |
| 27 | TN | Gibson | 602.742 | 49015 | 0.081 | 20.50 | 62.97 | 16.53 | 76.97 | 18.97 | 2.00 | 2.06 | 97.66 | 0.00 | 1.62 | 36.47 | 7.91 | 3 |
| 28 | TN | Giles | 610.927 | 29558 | 0.048 | 18.07 | 65.55 | 16.38 | 84.73 | 10.68 | 1.64 | 2.94 | 97.99 | 0.00 | 1.58 | 38.05 | 5.51 | 2 |
| 29 | TN | Grainger | 280.600 | 22419 | 0.080 | 18.07 | 66.78 | 15.15 | 95.44 | 0.30 | 2.04 | 2.22 | 95.71 | 0.00 | 3.44 | 30.01 | 6.19 | 1 |
| 30 | TN | Greene | 622.165 | 68172 | 0.110 | 17.36 | 65.92 | 16.72 | 93.51 | 2.02 | 2.35 | 2.12 | 97.31 | 0.02 | 2.18 | 37.69 | 4.31 | 10 |
| 31 | TN | Grundy | 360.534 | 13910 | 0.039 | 19.35 | 63.95 | 16.71 | 76.40 | 0.47 | 0.80 | 22.33 | 94.21 | 0.00 | 4.85 | 27.24 | 5.52 | 1 |
| 32 | TN | Hamblen | 161.179 | 61857 | 0.384 | 19.57 | 65.10 | 15.32 | 82.96 | 3.79 | 9.75 | 3.50 | 97.15 | 0.06 | 2.57 | 40.94 | 5.84 | 8 |

| COUNTY ID | STATE | COUNTY | AREA | Population | Density | Age<15 | Age_15to64 | Age>=65 | Pop_White | Pop_Black | Pop_Hispanic | Race_others | Mode_Private | Mode_Bicycling | Mode_Others | Income | No_vehicle | Crashes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 33 | TN | Hamilton | 542.431 | 328960 | 0.606 | 17.97 | 67.61 | 14.43 | 71.67 | 19.91 | 3.90 | 4.53 | 96.23 | 0.11 | 2.81 | 49.33 | 7.79 | 156 |
| 34 | TN | Hancock | 222.340 | 6782 | 0.031 | 17.40 | 66.06 | 16.54 | 97.11 | 0.20 | 0.29 | 2.40 | 97.64 | 0.00 | 0.69 | 26.27 | 10.80 | 1 |
| 35 | TN | Hardeman | 667.768 | 27655 | 0.041 | 17.49 | 68.96 | 13.55 | 55.41 | 40.59 | 1.34 | 2.66 | 96.28 | 0.03 | 2.67 | 31.28 | 10.32 | 2 |
| 36 | TN | Hardin | 577.318 | 25995 | 0.045 | 17.87 | 64.34 | 17.80 | 91.21 | 2.98 | 1.76 | 4.04 | 97.98 | 0.00 | 1.66 | 31.83 | 5.26 | 7 |
| 37 | TN | Hawkins | 486.975 | 56562 | 0.116 | 18.58 | 65.53 | 15.89 | 95.43 | 1.16 | 1.22 | 2.20 | 98.39 | 0.04 | 1.30 | 35.71 | 5.41 | 5 |
| 38 | TN | Haywood | 533.112 | 19010 | 0.036 | 21.61 | 64.84 | 13.54 | 44.58 | 50.21 | 3.30 | 1.91 | 98.39 | 0.00 | 1.41 | 476.82 | 11.73 | 2 |
| 39 | TN | Henderson | 520.073 | 27518 | 0.053 | 20.12 | 65.43 | 14.45 | 87.08 | 7.74 | 1.62 | 3.56 | 97.03 | 0.00 | 2.83 | 38.47 | 4.29 | 2 |
| 40 | TN | Henry | 562.096 | 32042 | 0.057 | 18.43 | 62.60 | 18.97 | 86.90 | 8.00 | 1.64 | 3.45 | 97.32 | 0.04 | 2.20 | 36.65 | 7.00 | 2 |
| 41 | TN | Hickman | 612.499 | 24506 | 0.040 | 18.79 | 67.97 | 13.24 | 90.81 | 4.41 | 1.70 | 3.09 | 97.21 | 0.00 | 2.43 | 41.41 | 5.22 | 1 |
| 42 | TN | Houston | 200.286 | 8286 | 0.041 | 19.78 | 62.89 | 17.33 | 92.56 | 1.57 | 1.66 | 4.20 | 97.46 | 0.00 | 1.87 | 34.97 | 8.11 | 0 |
| 43 | TN | Humphreys | 530.980 | 18416 | 0.035 | 19.06 | 64.31 | 16.63 | 93.96 | 2.62 | 1.27 | 2.15 | 98.65 | 0.00 | 1.16 | 41.07 | 4.78 | 0 |
| 44 | TN | Jackson | 308.320 | 11491 | 0.037 | 16.86 | 66.38 | 16.76 | 95.42 | 0.30 | 0.94 | 3.34 | 97.26 | 0.00 | 2.22 | 31.98 | 6.09 | 0 |
| 45 | TN | Jefferson | 274.078 | 50600 | 0.185 | 18.46 | 66.24 | 15.30 | 92.30 | 2.29 | 2.90 | 2.50 | 97.02 | 0.00 | 1.99 | 39.47 | 4.51 | 2 |
| 46 | TN | Johnson | 298.475 | 18190 | 0.061 | 14.80 | 67.81 | 17.39 | 94.44 | 2.09 | 1.32 | 2.15 | 97.50 | 0.15 | 1.79 | 30.21 | 4.65 | 0 |
| 47 | TN | Knox | 508.215 | 423748 | 0.834 | 18.25 | 68.91 | 12.84 | 83.43 | 8.82 | 3.03 | 4.72 | 95.99 | 0.12 | 3.19 | 50.69 | 5.71 | 145 |
| 48 | TN | Lake | 165.784 | 7827 | 0.047 | 13.67 | 72.97 | 13.36 | 67.67 | 27.63 | 1.29 | 3.41 | 98.84 | 0.00 | 1.02 | 24.35 | 10.26 | 0 |
| 49 | TN | Lauderdale | 471.992 | 27745 | 0.059 | 20.53 | 67.32 | 12.15 | 60.77 | 34.35 | 1.75 | 3.12 | 98.35 | 0.00 | 1.46 | 33.56 | 7.88 | 0 |
| 50 | TN | Lawrence | 617.128 | 41319 | 0.067 | 21.00 | 63.43 | 15.57 | 94.34 | 1.64 | 1.66 | 2.36 | 97.66 | 0.00 | 2.02 | 34.98 | 6.51 | 4 |
| 51 | TN | Lewis | 282.089 | 12003 | 0.043 | 21.12 | 63.75 | 15.13 | 94.40 | 1.16 | 2.56 | 1.88 | 97.20 | 0.00 | 2.08 | 36.29 | 5.85 | 1 |
| 52 | TN | Lincoln | 570.338 | 32885 | 0.058 | 19.39 | 64.03 | 16.59 | 86.28 | 6.02 | 2.43 | 5.27 | 97.64 | 0.00 | 1.93 | 41.41 | 5.20 | 7 |
| 53 | TN | Loudon | 229.216 | 47102 | 0.205 | 16.85 | 62.82 | 20.33 | 89.72 | 0.74 | 6.11 | 3.43 | 97.50 | 0.02 | 1.86 | 49.65 | 4.25 | 6 |
| 54 | TN | McMinn | 430.125 | 52075 | 0.121 | 18.97 | 65.00 | 16.03 | 89.32 | 3.89 | 2.73 | 4.05 | 96.60 | 0.08 | 2.22 | 36.67 | 5.56 | 13 |
| 55 | TN | McNairy | 562.860 | 25760 | 0.046 | 19.11 | 64.07 | 16.82 | 89.74 | 5.77 | 1.40 | 3.09 | 97.47 | 0.00 | 1.99 | 34.68 | 6.28 | 0 |
| 56 | TN | Macon | 307.144 | 21934 | 0.071 | 21.22 | 65.45 | 13.33 | 93.99 | 0.51 | 3.47 | 2.03 | 96.96 | 0.00 | 2.39 | 33.03 | 5.50 | 2 |
| 57 | TN | Madison | 557.117 | 97378 | 0.175 | 20.16 | 67.19 | 12.65 | 58.34 | 35.18 | 3.17 | 3.31 | 96.88 | 0.03 | 2.48 | 39.35 | 7.69 | 30 |
| 58 | TN | Marion | 498.160 | 28123 | 0.056 | 18.10 | 66.79 | 15.12 | 90.50 | 1.65 | 1.20 | 6.66 | 97.35 | 0.00 | 2.13 | 38.97 | 5.43 | 4 |
| 59 | TN | Marshall | 375.460 | 29902 | 0.080 | 20.65 | 66.70 | 12.65 | 85.76 | 6.79 | 4.35 | 3.10 | 97.23 | 0.00 | 2.32 | 41.06 | 4.02 | 2 |
| 60 | TN | Maury | 613.138 | 79029 | 0.129 | 20.41 | 67.02 | 12.57 | 79.22 | 12.75 | 4.61 | 3.42 | 97.60 | 0.03 | 1.98 | 44.05 | 4.48 | 14 |
| 61 | TN | Meigs | 195.122 | 11581 | 0.059 | 18.41 | 66.80 | 14.79 | 95.26 | 1.07 | 1.25 | 2.42 | 95.23 | 0.00 | 2.25 | 36.43 | 6.34 | 0 |
| 62 | TN | Monroe | 635.565 | 44015 | 0.069 | 19.24 | 65.34 | 15.42 | 91.17 | 1.95 | 3.13 | 3.75 | 97.67 | 0.00 | 1.47 | 35.96 | 4.58 | 7 |
| 63 | TN | Montgomery | 539.177 | 163603 | 0.303 | 23.65 | 68.33 | 8.02 | 65.97 | 18.37 | 7.29 | 8.36 | 97.00 | 0.04 | 2.18 | 46.58 | 3.95 | 65 |
| 64 | TN | Moore | 129.223 | 6266 | 0.048 | 18.72 | 64.27 | 17.01 | 91.84 | 1.33 | 0.51 | 6.32 | 97.89 | 0.37 | 1.81 | 43.80 | 3.71 | 0 |

| COUNTY ID | STATE | COUNTY | AREA | Population | Density | Age<15 | Age_15to64 | Age>=65 | Pop_White | Pop_Black | Pop_Hispanic | Race_others | Mode_Private | Mode_Bicycling | Mode_Others | Income | No_vehicle | Crashes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 65 | TN | Morgan | 522.180 | 21664 | 0.041 | 18.04 | 68.52 | 13.44 | 92.97 | 2.62 | 0.79 | 3.63 | 96.57 | 0.07 | 2.39 | 36.66 | 5.27 | 0 |
| 66 | TN | Obion | 544.728 | 31905 | 0.059 | 19.08 | 64.69 | 16.23 | 84.46 | 10.17 | 2.99 | 2.38 | 97.84 | 0.00 | 1.66 | 40.23 | 7.92 | 1 |
| 67 | TN | Overton | 433.483 | 21777 | 0.050 | 19.26 | 64.38 | 16.35 | 95.68 | 0.56 | 0.18 | 3.58 | 97.54 | 0.00 | 2.29 | 36.36 | 4.45 | 1 |
| 68 | TN | Perry | 414.731 | 7778 | 0.019 | 18.37 | 64.25 | 17.38 | 94.68 | 2.43 | 0.91 | 1.97 | 96.55 | 0.00 | 2.73 | 31.36 | 4.31 | 0 |
| 69 | TN | Pickett | 162.979 | 5072 | 0.031 | 19.42 | 62.95 | 17.63 | 93.51 | 0.00 | 2.51 | 3.98 | 97.63 | 0.00 | 1.07 | 30.23 | 5.11 | 0 |
| 70 | TN | Polk | 434.676 | 16690 | 0.038 | 18.27 | 65.43 | 16.29 | 95.56 | 0.09 | 1.12 | 3.22 | 96.19 | 0.07 | 3.11 | 34.23 | 5.83 | 0 |
| 71 | TN | Putnam | 401.103 | 70570 | 0.176 | 17.90 | 67.98 | 14.12 | 89.42 | 1.94 | 5.02 | 3.63 | 97.41 | 0.00 | 1.87 | 35.73 | 4.30 | 22 |
| 72 | TN | Rhea | 315.377 | 31215 | 0.099 | 19.45 | 65.42 | 15.13 | 91.55 | 2.09 | 3.31 | 3.05 | 96.70 | 0.09 | 2.54 | 34.93 | 5.23 | 1 |
| 73 | TN | Roane | 360.708 | 54156 | 0.150 | 17.33 | 64.87 | 17.79 | 92.81 | 2.42 | 1.37 | 3.40 | 97.30 | 0.00 | 2.22 | 43.16 | 4.79 | 6 |
| 74 | TN | Robertson | 476.287 | 64347 | 0.135 | 21.80 | 66.66 | 11.53 | 84.17 | 7.28 | 5.53 | 3.02 | 96.69 | 0.10 | 3.03 | 51.98 | 5.53 | 9 |
| 75 | TN | Rutherford | 619.364 | 250517 | 0.404 | 22.06 | 69.96 | 7.98 | 75.77 | 12.00 | 6.08 | 6.14 | 97.19 | 0.00 | 2.20 | 55.21 | 3.20 | 123 |
| 76 | TN | Scott | 532.297 | 22171 | 0.042 | 20.97 | 65.96 | 13.06 | 97.67 | 0.00 | 0.93 | 1.40 | 98.04 | 0.00 | 1.74 | 29.52 | 8.49 | 0 |
| 77 | TN | Sequatchie | 265.858 | 13814 | 0.052 | 20.12 | 65.09 | 14.79 | 84.46 | 0.00 | 2.48 | 13.06 | 96.93 | 0.09 | 1.97 | 35.28 | 4.70 | 0 |
| 78 | TN | Sevier | 592.500 | 87507 | 0.148 | 18.14 | 66.98 | 14.88 | 91.41 | 0.74 | 4.42 | 3.43 | 96.72 | 0.07 | 2.47 | 40.37 | 4.32 | 39 |
| 79 | TN | Shelby | 763.174 | 922696 | 1.209 | 22.16 | 67.79 | 10.04 | 39.24 | 50.66 | 5.04 | 5.06 | 95.60 | 0.00 | 3.57 | 47.40 | 9.53 | 453 |
| 80 | TN | Smith | 314.289 | 19035 | 0.061 | 20.06 | 66.56 | 13.38 | 92.73 | 1.85 | 1.92 | 3.49 | 96.62 | 0.00 | 2.37 | 42.01 | 4.75 | 0 |
| 81 | TN | Stewart | 459.330 | 13133 | 0.029 | 19.33 | 64.61 | 16.07 | 93.22 | 1.57 | 1.72 | 3.49 | 96.39 | 0.01 | 3.19 | 31.38 | 7.28 | 0 |
| 82 | TN | Sullivan | 413.363 | 155815 | 0.377 | 17.03 | 65.10 | 17.88 | 93.71 | 1.87 | 1.38 | 3.04 | 97.71 | 0.05 | 1.74 | 40.29 | 5.91 | 34 |
| 83 | TN | Sumner | 529.449 | 155592 | 0.294 | 21.02 | 66.85 | 12.13 | 86.73 | 6.51 | 3.56 | 3.20 | 96.44 | 0.00 | 3.18 | 56.54 | 3.50 | 35 |
| 84 | TN | Tipton | 458.366 | 59689 | 0.130 | 21.99 | 67.29 | 10.72 | 75.94 | 18.65 | 2.02 | 3.39 | 97.89 | 0.00 | 1.61 | 46.05 | 3.47 | 10 |
| 85 | TN | Trousdale | 114.193 | 7751 | 0.068 | 19.60 | 67.17 | 13.24 | 85.09 | 7.07 | 2.30 | 5.54 | 96.64 | 0.00 | 2.62 | 45.12 | 4.26 | 0 |
| 86 | TN | Unicoi | 186.165 | 18257 | 0.098 | 17.02 | 63.99 | 18.99 | 94.91 | 0.78 | 3.46 | 0.85 | 95.24 | 0.00 | 4.06 | 34.35 | 6.30 | 2 |
| 87 | TN | Union | 223.549 | 19088 | 0.085 | 19.56 | 67.21 | 13.23 | 96.90 | 0.00 | 1.03 | 2.06 | 97.27 | 0.00 | 2.27 | 31.01 | 4.59 | 2 |
| 88 | TN | Van Buren | 273.415 | 5503 | 0.020 | 17.35 | 66.67 | 15.97 | 93.05 | 1.76 | 0.62 | 4.57 | 96.63 | 0.00 | 2.55 | 27.96 | 3.34 | 0 |
| 89 | TN | Warren | 432.680 | 39539 | 0.091 | 20.35 | 64.95 | 14.70 | 86.67 | 2.84 | 7.69 | 2.81 | 96.91 | 0.03 | 2.37 | 34.87 | 4.48 | 3 |
| 90 | TN | Washington | 326.465 | 119768 | 0.367 | 17.07 | 68.17 | 14.76 | 89.47 | 3.83 | 2.68 | 4.02 | 96.88 | 0.00 | 2.48 | 40.96 | 5.64 | 43 |
| 91 | TN | Wayne | 734.100 | 17016 | 0.023 | 15.58 | 69.26 | 15.16 | 90.31 | 5.73 | 1.42 | 2.54 | 97.90 | 0.05 | 1.38 | 34.56 | 5.89 | 6 |
| 92 | TN | Weak ley | 580.364 | 34557 | 0.060 | 16.99 | 68.07 | 14.94 | 86.86 | 6.98 | 1.90 | 4.26 | 96.31 | 0.16 | 2.27 | 34.73 | 5.38 | 8 |
| 93 | TN | White | 376.673 | 25460 | 0.068 | 18.57 | 64.62 | 16.81 | 93.90 | 1.53 | 1.57 | 3.00 | 96.59 | 0.00 | 3.05 | 34.05 | 6.72 | 6 |
| 94 | TN | Williamson | 582.599 | 174260 | 0.299 | 24.15 | 66.68 | 9.16 | 86.11 | 4.20 | 4.15 | 5.53 | 94.82 | 0.03 | 4.85 | 92.00 | 2.10 | 30 |
| 95 | TN | Wilson | 570.826 | 109563 | 0.192 | 21.03 | 67.45 | 11.52 | 87.05 | 6.43 | 2.94 | 3.57 | 96.42 | 0.06 | 3.25 | 63.41 | 3.51 | 13 |