Associating pedestrian crashes with demographic and socioeconomic factors

Deo Chimba[a,*], Abram Musinguzi[b], Emmanuel Kidando[c]

[a] Department of Civil Engineering, Tennessee State University, 3500 John A Merritt Blvd, Nashville, TN, 37209, United States
[b] Department of Civil Engineering, Tennessee State University, United States
[c] Florida State University, United States

ARTICLE INFO

ABSTRACT

In the last decade, the concept of walkable neighborhoods has emerged as a topic of great interest. However, it is still unclear about the influence of socioeconomic and demographic factors on pedestrian crashes. This study proposed a methodology for pedestrian crash analysis that combines Geographic Information System (GIS) methods and statistical analysis to study the influence of socioeconomic and demographic factors on the occurrence of pedestrian crashes. The analysis was based on statewide crash data collected in Tennessee from 2008 to 2012. First, GIS kernel density technique was proposed to identify high concentration of pedestrian crash clusters and results were presented using cases studies of Davidson and Hamilton counties. GIS analysis identified pedestrian crash clusters among block groups with a high population who walk to work and block groups with a high number of housing units with no vehicles. A negative binomial model was applied using a statewide data to test the statistical significance of explanatory variables. As expected, model results indicated that population density, population from 15 to 64 years of age, high population of neighborhoods commuting to work by walking (without adequate facilities supporting pedestrians such as sidewalks and crosswalks) and high population of neighborhoods of housing units with no vehicles significantly increase the number of pedestrian crashes. However, blocks whose streets have adequate presence of median, shoulders, and sidewalks had negative coefficients hence their presence tends to decrease pedestrian crashes. Furthermore population commuting to work by private cars and high median household income significantly reduces pedestrian crash frequency. The findings from Kernel density and statistical modeling are relatively identical in the sense that all found household vehicle availability to be a factor in influencing frequency of pedestrian crashes. The findings of this study can assist in implementation of proactive pedestrian safety strategies.

1. Background

In the United States, an increase in walking and bicycle use in the past decade has been well documented (Federal Highway Administration, 2010). As one would expect, this increase in usage has led to comparable rise in the number of pedestrian crashes. In 2012, there were 4743 pedestrian fatalities, accounting for 14 percent of all traffic fatalities, while an estimated 76,000 pedestrians were injured in traffic crashes in the United States (NHTSA, 2014). In other words, on average, a pedestrian was killed every 2 h and a pedestrian was injured every 7 min in traffic crashes. Therefore, the objective of this paper was to use Geographical Information Systems (GIS) and statistical modeling to evaluate influence of demographic and socioeconomic characteristics on pedestrian crashes using five years crash data (2008–2012) in Tennessee. Understanding the influence of demographic and socioeconomic characteristics on pedestrian safety may support safety implementation of proactive strategies to reduce pedestrian crashes.

Previous research applied statistical models and examined socioeconomic and demographic effects on pedestrian safety. For instance, pedestrian crash rates in the city of San Francisco were related to population density, age composition of the local population, unemployment, gender and education (LaScala et al., 2000). A study conducted in Kwun Tong District of Hong Kong found land mix and socio-economic deprivation index were more associated with the occurrence of serious and slight injuries (Yao and Loo, 2012). A method of frequency item sets was used to study characteristics of "black" zones and found that collisions with pedestrian involving young road users inside the built-up area is a typical accident pattern that frequently occurs inside a "black" zone (Geurts et al., 2005). Although most statistical models have been applied successfully to analyze pedestrian crashes, the questions of where are most of the crashes occurring and why, still remain demanding and this study proposes the application of GIS to answer these questions. Using GIS to geocode crash locations and plot them on dot maps is the most common first step in crash analysis

(Andaluz et al., 1997). GIS turns statistical data, such as crashes, and geographic data, such as roads and crash locations, into meaningful information for spatial analysis and mapping (FHWA, 2000). For instance, (Mitra, 2009) proposed that if GIS crash mapping is overlaid with other layers, it is possible to associate high-crash locations with spatial factors. (Delmelle and Thill, 2008) analyzed bicycle and pedestrian crash patterns and found that adult bicycle crashes formed a linear spatial pattern along major arterials and around business centers while the under 16 year old class of bicyclist showed smaller clusters or spots concentrated inside residential neighborhoods on local streets. Crash clusters suggest common causes and spatial dependence of traffic crashes. The detection of the spatial clusters can lead to the identification, and hence alleviation of the common cause to achieve improved road safety (Steenberghen et al., 2010).

Most GIS tools have great potential to locate crashes on a digital map. However, 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 would need to combine both GIS and statistical methods (Levine, 2006). This technique has been applied in previous research to analyze traffic crashes such as; (Anderson, 2009; Levine and Nitz, 1995; Steenberghen et al., 2004; Chimba et al., 2014). Using GIS, (Chimba et al., 2014) 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. Colorado Department of Transportation applied spatial statistics to develop safety performance for intersections (CDOT, 2009).

Several spatial statistics are used to analyze clusters of point features such as vehicle pedestrian crashes. Some of them such as Moran's I, Getis's G-statistic and Ripley's K-function are global measures that are used to identify whether the patterns of values across the study area tend to be clustered, random or dispersed and do not identify clusters on a map (Allen, 2010). However, kernel method or the local spatial autocorrelation methods can identify location of clusters and are commonly used techniques for determining hazardous locations in traffic crash analysis (Steenberghen et al., 2004; Erdogan et al., 2008). Therefore, for practical applications like high traffic crash location detection, local spatial statistics are better suited in that they can identify and examine where unusual clusters of events occur (Xie and Yan, 2013).

In this study, the key concept is that the frequent presence of pedestrian crash clusters arises from shared common causes. Kernel density is used due to its capability to spread the 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 (Anderson, 2009). 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 location. Kernel density estimation solves this problem by representing crashes per unit area using a density map feature in GIS (Pulugurtha et al., 2007). 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 (Bulajića et al., 2014).

Literature shows 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 (Pulugurtha et al., 2007; Loo et al., 2011). Furthermore, when calculating the density, only the points that fall within search neighborhood are considered. Hence, the search radius produces significant effects to the results that is included in 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, (Anderson, 2009) considered a search radius, which is twice the size of the grid cell. Using kernel density estimation and a search radius of 100m, (Bíl et al., 2013) identified hazardous road locations of traffic accidents. Blazquez and Celis (2013) identified critical areas with high child pedestrian crash risk using kernel density estimation and found no statistical significance of child pedestrian crashes with respect to gender, weekday, and month of the year. (Loo et al., 2011) 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 (100 m and 250m) preserves local features and is therefore more appropriate for identifying crash clusters at precise locations.

This paper applied KDE to create crash density maps in GIS and identified pedestrian high crash concentrations using 2008–2012 crash data from Tennessee. The analysis used a bandwidth of 200 m (based on a grid cell of 100 m) and considered more than 4000 blocks.

2. Data and methodology

2.1. Data

This study used crash data obtained from Tennessee Roadway Information Management System (TRIMS), a database managed by Tennessee Department of Transportation (TDOT). The final analysis sample comprised of 4816 pedestrian crashes that occurred between 2008 and 2012. This dataset contained 330 fatal crashes, 930 incapacitating injury crashes, 3275 non-incapacitating injury crashes and 281 property damage only crashes. The summary of crash statistics is presented in Table 1. The crash data was integrated with census block group data containing socioeconomic and demographic data by conducting spatial analysis in GIS to generate crash frequency per census block group. Block groups are geographical units created by US Census Bureau as clusters of census blocks and generally comprise between 600 and 3000 people. As expected, block group crash frequency was varying across the state with some census block groups having more crashes than others do. The distribution of pedestrian crashes across the census blocks indicated a high zero crash count as shown in Fig. 1. The data appear to be well approximated by a Poisson or negative binomial distribution.

Although the study was conducted throughout the entire state of Tennessee, two counties, Davidson and Hamilton, were considered as case studies to demonstrate crash patterns and for comparison purposes. Davidson County, in which the capital Nashville is located, covers 502 sqmi with a population of 668,347. Nashville city has a number of diverse physical environmental characteristics, ranging from inner-city neighborhoods with high density and well connected street networks to suburban neighborhoods with low density and low street connectivity. Hamilton on other hand covers an area of 576 sqmi and is the fourth most populous county in Tennessee with a population of 336,463 as of the 2010 census and its main city is Chattanooga. Both

Table 1
Crash Summary Statistics.

Crash Type	Minimum	Maximum	Sum	Mean	Std. Deviation
All crashes	0	58	4816	1.17	2.724
Fatal crashes	0	5	330	0.08	0.311
Incapacitating Injury crashes	0	8	930	0.23	0.604
Non-Incapacitating Injury crashes	0	51	3275	0.79	2.133
Property Damage crashes	0	4	281	0.07	0.304

Crashes per Census Block, 2008–2012.