Journal of Transport Geography 24 (2012) 206–214

Contents lists available at SciVerse ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Identifying clusters and risk factors of injuries in pedestrian–vehicle crashes in a GIS environment

Dajun Dai *

Department of Geosciences, Georgia State University, Atlanta, GA 30302, USA
Partnership for Urban Health Research, Institute of Public Health, Georgia State University, Atlanta, GA 30302, USA

ARTICLE INFO

Keywords:
Pedestrian injury
Geographic Information Systems (GISs)
Spatiotemporal clustering
SatScan
Atlanta

ABSTRACT

There is a growing concern about pedestrian injuries on road. Using pedestrian crash records (2000–2007) from the Georgia Department of Transportation as a case study, this paper applied a spatiotemporal clustering technique to identify clusters of injured pedestrians and then investigated the influence of personal and environmental factors on pedestrian injuries. The Bernoulli model in SatScan was used to detect the roadway segments, where pedestrian injuries were significantly clustered. Descriptive statistics and temporal (yearly, monthly, day of week, and hourly) trends of the injuries were explored, respectively. The logistic regression model was used to assess the injury risk associated with pedestrian factors (gender, age, intoxication, and maneuvers), driver's factors (gender, age, and intoxication), and environmental factors (light conditions, surface conditions, and weather conditions). The results showed that suburban high-activity corridors, where state highways intersect local streets, significantly elevated injury risks in crashes compared to other areas. The percentage of injuries was pronounced in summers, on the weekends, and from evenings to early mornings. Age, pedestrian maneuvers, and inadequate lighting were significant risk factors for pedestrian injuries. Walking/driving under influence and male pedestrians/motorists showed an increased risk of injuries when only the main effect of the factors was considered. The high-risk roadway segment and the risk factors highlight the need for heightened investigation and education in the specific areas and populations.

1. Introduction

For motor vehicle collisions involving pedestrians, crash severity is critical in affecting the health outcome, treatment complications, and long-term health and wellbeing of pedestrians. Severe injuries increase the likelihood of lifelong disability and fatality (Lefler and Gabler, 2004; Yoganandan et al., 2010). Significantly severe injuries are pronounced in areas on high-speed roadways, without sidewalks, or with poor lighting (Clifton et al., 2009; Coate and Markowitz, 2004) and among children or senior pedestrians (Gorrie et al., 2008). Using spatial clustering techniques and statistical methods, this study aims to detect roadway segments presenting an elevated injury risk given a crash and the personal and environmental factors influencing the injury severity.

Pedestrian injuries and fatalities have been significant health and financial burdens in the United States. Pedestrian–vehicle collisions take more than 4000 lives each year in the US (National Highway Traffic Safety Administration, NHTSA, 2009). The estimated annual economic impact is nearly $29 billion (National Safety Council; NSC, 2009). A question arises as to, where these severe crashes occur and their contributing factors. One might argue that the goal of pedestrian safety is to reduce pedestrian crashes, not to reduce the collision severity given a crash. These severe pedestrian injures, however, are more likely to involve emergency medical services, hospitalizations, long-term health care, and life-long disabilities and fatalities (Bell and Schuurman, 2010; Zajac and Ivan, 2003). Understanding these severe injuries therefore deserves special attention because of the significant costs to the victims and the society (Chang and Wang, 2006; Xie et al., 2009).

A variety of methods have been proposed to define clusters of injured pedestrians, yet spatiotemporal clustering techniques on crashes are still needed. One approach defines clusters as sites with highest number of injured pedestrians or highest density using kernel estimation (Anderson, 2009; Pulugurtha et al., 2007; Warden et al., 2011). Yet these methods may be affected by the size effect and provide no insight about the underlying factors affecting the distribution (Yamada and Thill, 2010). Another approach, to consider the size effect, uses the injury rate, such as the number of injuries per vehicular volume, per pedestrian volume, or per residing population in a buffer of a crash site (Pulugurtha et al., 2007). The issue is that the population exposed to the injury risk

* Address: Department of Geosciences, 33 Gilmer Street, SE, Georgia State University, Atlanta, GA 30303, USA. Tel.: +1 404 413 5797; fax: +1 404 413 5768.
E-mail address: ddai@gsu.edu

0966-6923/$ - see front matter
doi:10.1016/j.jtrangeo.2012.02.005

is related to both pedestrian volume and vehicle volume, yet pedestrian counts are not readily available. In this research, it is assumed that all pedestrians who experience crashes are at risk of injury, which is in line with previous studies (Marshall et al., 2000; Warden, 2008). In addition, it is necessary to evaluate the statistical significance of the injury risk in a cluster. The statistically significant clusters suggest that the injuries are less likely to occur "by chance" and factors other than sizes may play a role in elevating the injury risk. Once identified, these factors may be addressed through intervention programs.

Given the unique feature of network-constrained pedestrian crashes, a variety of network-based clustering techniques have been proposed but challenges remain. The network-based local indicators (Yamada and Thill, 2010) depend on the pre-specified spatial weight matrix. The network clustering method (Steenberghen et al., 2010) needs to identify an influence range of a focus location *a priori*. Besides, they are limited in accounting for space–time interactions which is important in traffic accidents (Brugge et al., 2002). So a method is desirable if it can account for the spatially variable populations experiencing crashes and space–time interactions, as well as to be less sensitive to the network-constrained feature of pedestrian crashes.

Besides the geographic factors of pedestrian injuries, the non-geographic factors play a critical role on the injury risk given a crash. A number of personal factors have been linked to the elevated risk of injury for pedestrians, such as age, gender, substance use (Elliott et al., 2009; Gorrie et al., 2008). Pedestrian maneuver is also a concern. Severer injuries in crashes are associated with playing on streets and crossing against traffic lights or without traffic control devices (Christie et al., 2007; King et al., 2009). For environmental conditions, studies (Clifton et al., 2009; Coate and Markowitz, 2004) showed darkness may increase the likelihood of pedestrian injuries, but weather (e.g., rain or snow) was not a significant risk factor in a study on Baltimore, Maryland (Clifton et al., 2009). From the perspective of city planning and injury prevention, it is vital to learn, where pedestrians are injured—the so called "black spot" and what factors contribute to the injuries, thus providing critical information for intervention and prevention (Warden, 2008).

Using pedestrian crash records in metropolitan Atlanta, this research has two objectives: (1) to identify clusters of injured pedestrians using Geographic Information Systems (GISs) and spatiotemporal clustering techniques and (2) to investigate the effect of personal and environmental factors on pedestrian injury risk using a logistic regression model. The contributions are twofold.

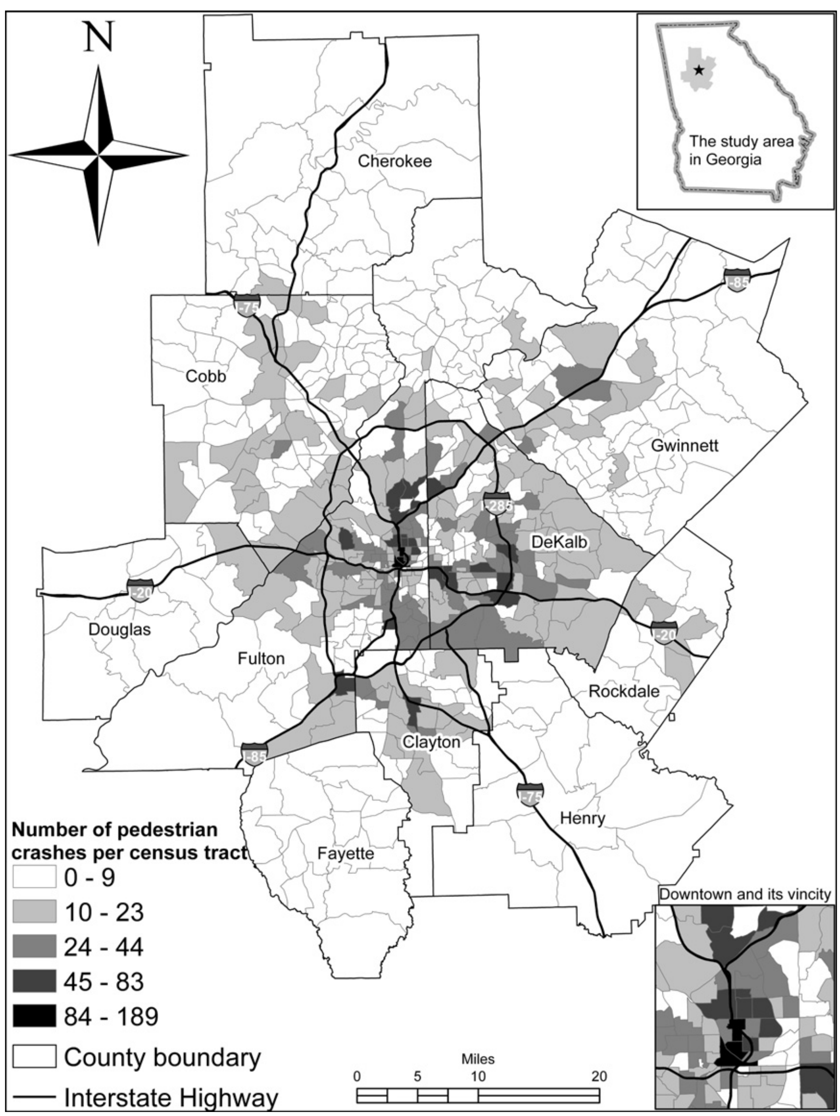

Fig. 1. Number of pedestrian crashes in the study area at census tract level.

First, it detects the black spots, where a significantly elevated injury risk given a crash exists which can be targeted for intervention with a higher priority. Besides, these sites provide targets for understanding factors in the neighborhoods in order to improve pedestrian safety. Second, it identifies factors related to pedestrian injuries from personal and environmental aspects. Findings allow injury prevention professionals to target the specific populations and environmental factors. Although this research focuses on pedestrian-vehicle crashes in Atlanta, the methodologies can be applied to other regions and other types of events (e.g., automobile crashes) for traffic safety.

2. Data description

The study area centers on the ten-county urban region of metropolitan Atlanta in Georgia (Fig. 1). As a fast growing city experiencing urban sprawl, Atlanta grew 15% between 2000 and 2005 (Ewing et al., 2003). The selection of this study area is appropriate as some studies (Ewing et al., 2003; Paulozzi, 2006) suggest that the injury risk for pedestrians is greater in rapidly growing cities characterized by urban sprawl. The appropriateness of this study area is also supported by a previous study (Beck et al., 2007) reporting an increasing fatality trend from 2000 to 2004 in Atlanta given a declined rate on average in the US.

This study obtained 8-year (January 1, 2000–December 31, 2007) vehicle–pedestrian collision records ($n = 8403$) in the study area from the Georgia Department of Transportation (GDOT). The GDOT compiles the traffic accident records regularly based on the police reports which provide detailed facts used for legal purposes and for identifying traffic safety hazards (GDOT, 2010a). To make accurate and consistent reports, the law enforcement agencies follow the rigorous instruction from a uniform vehicle accident report (GDOT, 2010a), a primary source of the official Georgia Accident Reporting System (www.dot.ga.gov/statistics/CrashData). For each crash record, locational information is represented by a route ID and the distance from the accident scene to the nearest intersection or to the nearest hundredth mile log, which allows the GDOT to identify the accident's location (latitude and longitude) using the dynamic segmentation technique in GIS. Of the 8403 crashes, 92.4% ($n = 7763$) were provided with geographic coordinates. Each record describes a pedestrian with respect to the age, gender, substance use, the maneuver, and the injury type. It also includes the involved drivers (age, gender, and intoxication) and the environment in terms of weather, road surface, and light conditions. Following the uniform vehicle accident report instruction, police officers classified the 7763 crashes into fatalities ($n = 390$), serious injuries ($n = 1042$), visible injuries ($n = 2692$), complaints ($n = 2682$), and no injuries ($n = 957$). Serious injuries indicate pedestrians cannot walk or normally continue their activities. Visible injuries indicate wounds or bleeding which are visible to people other than the victims. Complaints suggest that victims complain being hurt without any wounds visible to others. Due to privacy protection, detailed severity scores and injured body parts evaluated by on-scene health care providers were unavailable. One concern is that some crashes might not be reported to the police. This study assumes that this under-reporting issue is minor given that all motorists operating government-owned and commercial vehicles (e.g., delivery services) are required to report any accidents regardless of crash severity; besides, many private vehicle drivers (or witnesses) will notify the police of any crashes. Unfortunately the reporting rate for private vehicles has not been documented. Of the 7763 crashes with coordinates, there were 568 (7.3%) crashes missing either pedestrian or motorist information which therefore were removed from the subsequent analyses.

To reduce the bias in the spatial and statistical analysis due to the relatively small numbers in some categories, this study classified all crashes into two broad groups—injuries (fatalities, serious injuries, and visible injuries) and no injuries (complaints and crashes without injuries), which is in line with previous studies (Clifton et al., 2009; Kim et al., 2008). Fatalities are grouped into the injury category as they include those being killed on the scene and those who died of severe injuries on the way to or at hospitals. Complaints are classified into the non-injury group because they claimed possible injuries which were invisible to others. It is assumed that complaints were affected less severely than pedestrians with visible injuries, which is supported by the fact that nearly 50% of the complaints did not take any medical treatment. It is possible that some complaints developed injuries after leaving the scene. Yet medical reports are unavailable to validate the complaints due to the privacy protection.

3. Detecting clusters of pedestrian injuries

This section explores spatial and spatiotemporal clustering of pedestrian injuries using SatScan (developed jointly by Kulldorff M., Boston, Massachusetts and Information Management Services, Inc, Silver Spring, Maryland). Locating clusters of pedestrian injuries given a crash allows one to target specific areas for intervention. It provides areas of focus to examine the factors that could contribute to the clusters other than what the traffic volumes of pedestrians and vehicles could explain. Although crashes in the non-injury group shall be reduced as well, segments where the injury risk is elevated given a pedestrian crash shall be in a higher priority for intervention, such as placing warning signs or evaluate the built environment. As a widely accepted software, SatScan has reasonable sensitivity and specificity compared to other cluster detection methods (Song and Kulldorff, 2003) and has been used in a variety of studies. Yet few studies have applied it to detect clusters of pedestrian injuries.

This study employed the Bernoulli model in SatScan for cluster detection. Under the null hypothesis, the Bernoulli model assumes that cases (i.e., crashes in the injury group) have the same geographical distribution as the controls (i.e., crashes in the non-injury groups), which is consistent with previous studies (Marshall et al., 2000; Warden, 2008). The alternative hypothesis is that there are some geographical areas that have a higher proportion of pedestrian injuries given a crash. To test the null hypothesis, SatScan imposes a search window on the study area, which is elliptic in pure spatial clustering and is conic in spatiotemporal clustering, where its height corresponds to a time interval (Kulldorff, 2009). The window varies continuously in size and location. SatScan then evaluates whether the distribution of cases is significantly different from controls by using a likelihood ratio test and the Monte Carlo simulation.

The clusters of injuries were searched in both space–time and space. For space–time clustering, only space–year interaction was considered because very few sites experienced two or more crashes in the same month or on the same day. Space–time clustering using temporal scales shorter than a year has few cases and controls in each search window, which is less statistically reliable given the large variance resulted from the "small population" (Mu and Wang, 2008). This research conducted 999 Monte Carlo replications and clusters with their p-values less than 0.05 were significant.

This study used small search windows in order to reduce the impact of the difference between Euclidean distance and network-distance. For all models, the maximum minor axis of the elliptic search windows was set as one tenth of a mile, which is consistent with a previous study on crash analysis (Yamada and

Thill, 2010). The ratio of major to minor axis ranged from one to five (default setting in SatScan). Limiting the window sizes allows the search to focus on each short roadway segment and its close proximity (e.g., vehicles driving off or onto a road), thus identifying specific streets (or intersections) as clusters of injuries. Large search windows may return super clusters mixing many streets, where heterogeneous volumes of pedestrians and vehicles are often present. Besides, these super clusters cannot clearly tell the high-risk roadway segments. It is admitted that small search windows may return clusters covering multiple streets too. Such clusters require cautions for interpretation.

Mapping the number of pedestrian crashes in each census tract (Fig. 1) shows that more crashes occurred within the I-285 perimeter compared to counties further away. The spatiotemporal clustering detected one roadway segment (Inset 1 in Fig. 2) on the Buford Highway from Cliff Valley Way to the east of Briarwood Road, where injured pedestrians (2001–2004) were significantly clustered ($p = 0.01$). As shown in Table 1, there were 19 pedestrians hit by vehicles in the 5 years (2001–2004) on this segment, all of whom were injured ranging from fatalities to visible wounds. Crashes recurred on some spots, such as the two intersections on this segment. A field trip reveals that a large shopping plaza and a variety of ethnic restaurants, convenient stores, gas stations, and other businesses are present, thus formulating a high-activity corridor along the Highway. Such a corridor in suburbs presents a typical setting of high-risk zones for pedestrian injuries, where pedestrians have to negotiate with heavy fast-speed traffic on wide streets. As an arterial road, the Buford Highway consists of seven lanes with estimated daily traffic count up to 31,400 vehicles (GDOT, 2010b) and with a speed limit of 45 miles per hour (mph) in this zone. The Highway bisects the neighborhoods, where there are many apartment complexes, but only on the south side exist discontinuous sidewalks.

Another four separated clusters of pedestrian injuries were detected but none of them was statistically significant ($\alpha = 0.05$). State Highway 19, an arterial road connecting north suburbs and the downtown, formulated two clusters (Inset 2 in Fig. 2), where it intersects local streets. The fourth cluster (Inset 3 in Fig. 2) revealed a dangerous zone, where more than half of the cases resulted in fatalities or serious injuries in an approximate 0.4-mile roadway segment from 2004 to 2007. The area includes seven intersections between the arterial road (Norcross Tucker Road with a minimum speed limit of 45 mph) and local streets, thus increasing the travel complexity for both pedestrians and drivers and the

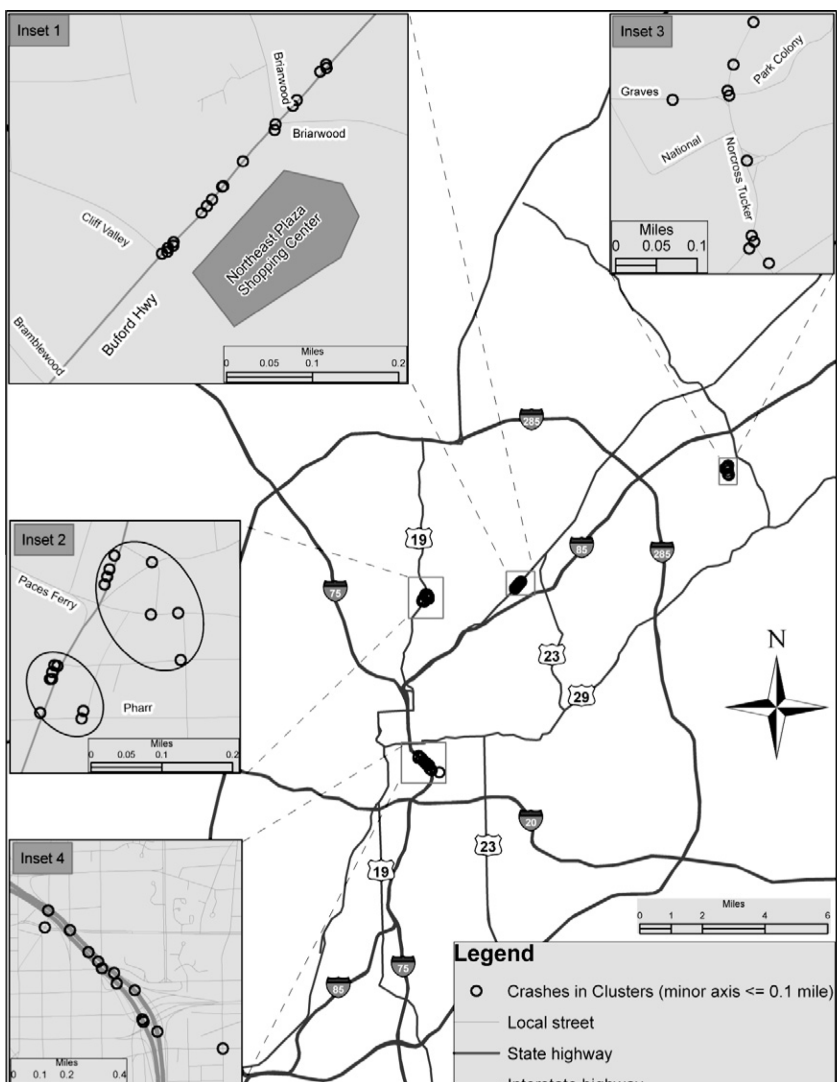

Fig. 2. Clusters of injured pedestrians. Inset 1 shows the first most likely clusters ($p < 0.05$). Clusters in the other three insets have p-values larger than 0.05. *Note*: pedestrian crashes outside of the clusters were masked out for simplicity.

Table 1
Crash description in the clusters.

	Cluster 1*	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Year of the clusters	01–04	01–04	00–03	04–07	01–04
Crashes in the clusters	19	12	9	11	13
Fatality	2	1	2	2	5
Serious injury	5	5	2	4	1
Visible injury	12	6	5	5	7
Complaint	0	0	0	0	0
No injury	0	0	0	0	0
Pedestrian characteristics					
Male	19	10	3	10	6
Female	0	2	6	1	7
Crossing not at Crosswalk	12	7	3	7	6
Driver's characteristics					
Male	17	9	5	8	8
Female	2	3	4	3	5
Environmental conditions					
Rainy/Snowy/Icy	5	2	1	0	2
Darkness	14	12	7	6	11

Note: Clusters 2 (upper right) and 3 (lower left) are in Inset 2 of Fig. 2.
* Significant ($\alpha = 0.05$).

Table 2
Descriptive statistics of the pedestrian crashes ($n = 7195$) in Atlanta, Georgia (2000–2007).

Variables	# Pedestrian crash	% Pedestrian crash	Variables	# Pedestrian crash	% Pedestrian crash
Gender (Pedestrian)			*Intoxication (Driver)*		
Female	2654 (2423)	36.9 (36.1)	Not	5599 (5256)	77.8 (78.4)
Male	4541 (4282)	63.1 (63.9)	Not known	1402 (1294)	19.5 (19.3)
			Yes	186 (147)	2.6 (2.2)
Gender (Driver)			Others[a]	8 (8)	0.1 (0.1)
Female	2541 (2365)	35.3 (35.3)			
Male	4654 (4340)	64.7 (64.7)	*Pedestrian maneuver*		
			Crossing at crosswalk	1593 (1491)	22.1 (22.2)
Age (in years; Pedestrian)			Crossing, not at crosswalk	2676 (2549)	37.2 (38.0)
14 or less	1239 (1125)	17.2 (16.8)	Walking with traffic	420 (394)	5.8 (5.9)
15–34	2878 (2648)	40.0 (39.5)	Walking against traffic	193 (185)	2.7 (2.8)
35–54	2279 (2174)	31.7 (32.4)	Working on road/vehicle	273 (252)	3.8 (3.8)
55 or older	799 (758)	11.1 (11.3)	Playing/standing on road	560 (506)	7.8 (7.5)
			Darting into traffic	1260 (1115)	3.1 (3.2)
Age (in years; Driver)			Others[a]	220 (213)	17.5 (16.6)
25 or less	1302 (1207)	18.1 (18)			
25–50	3127 (2886)	43.5 (43)	*Surface conditions*		
51 or older	2766 (2612)	38.4 (39)	Dry	6184 (5775)	85.9 (86.1)
			Wet/snowy/icy	999 (918)	13.9 (13.7)
Intoxication (Pedestrian)			Other[a]	12 (12)	0.2 (0.2)
Not	5035 (4684)	70.0 (69.9)			
Not known	1857 (1723)	25.8 (25.7)	*Weather conditions*		
Yes	284 (279)	3.9 (4.2)	Clear	5072 (4720)	70.5 (70.4)
Others[a]	19 (19)	0.3 (0.3)	Cloudy	1436 (1355)	20.0 (20.2)
			Rain/snow/sleet/fog	674 (617)	9.4 (9.2)
Light conditions			Other[a]	13 (13)	0.2 (0.2)
Daylight	4345 (4072)	60.4 (60.7)			
Darkness	2850 (2633)	39.6 (39.3)			

[a] Rare events that do not clearly fall into the rest of the categories with no further explanation. In bracket are the values when only single-pedestrian crashes were considered ($n = 6705$).

likelihood of pedestrian injuries. The space–time clustering revealed a cluster (Inset 4) around the intersections of the local streets with ramps of interstate highways (I-75 and I-85) in downtown Atlanta. Nearly half of the cases resulted in fatalities or serious injuries (Table 3) from 2001 to 2004. Interpreting the clusters 2–5, however, requires caution because of their statistical insignificance and inclusion of multiple heterogeneous streets.

To evaluate how sensitive the clustering results are, this study alternated the search window threshold. The model consistently detected the same roadway segment on the Buford Highway as the significant cluster of pedestrian injuries when smaller settings (the maximum minor axis is 0.05 and 0.025 mile, respectively) were used. When larger settings (the maximum minor axis is 0.2 and 0.25 mile, respectively) were employed, the roadway segment from east of Briarwood Road to Bramblewood Drive (Inset 1 in Fig. 2) on the Buford Highway was identified, where pedestrian injuries (2002–2004) were significantly clustered compared to non-injury crashes nearby. At last, when only single-pedestrian crashes—one pedestrian per crash—were considered ($n = 6705$), the clusters remained consistent compared to the findings above. This research also explored the spatial clustering (without taking into account temporal variations) using the aforementioned settings, but did not find any significant clusters. This suggests apparent yearly variations in pedestrian injuries in the study area. The fatality risk given a crash was explored using the Bernoulli model (i.e., cases were pedestrian crashes with fatalities and controls were pedestrian crashes without fatalities) in the above settings. No fatalities were significantly clustered. The clusters of pedestrian injuries identified in Fig. 2 help investigate the road features around these black spots. The characteristics of the associated pedestrians and motorists as well as the environmental conditions (see Table 1), however, indicate that the risk of injury may vary among these factors. One question arises as to how personal and environmental factors affect the risk of injury.

4. Exploring personal and environmental factors elevating the injury risk among pedestrian crashes

This study explored the personal and environmental factors contributing to the risk of injury among pedestrians who experience crashes based on the aforementioned injury classification. The personal factors included age, sex, and intoxication of pedestrians and drivers as well as pedestrian maneuvers. The environmental factors included weather, light, and surface conditions when crashes occurred. The selection of these factors was based on studies showing their importance in affecting the injury risk (Coate and Markowitz, 2004; Gorrie et al., 2008). The analysis did not include road conditions and vehicle types. Most of the crashes (over 98%) occurred on roadways without any defect. Information on vehicle types is not available in the publicly available database; therefore interpreting the results shall take cautions considering the impact of vehicles.

The initial analysis was descriptive statistics of the pedestrian crashes (see Table 2) when all crashes and only single-pedestrian crashes were considered respectively. Of these crashes, more males were involved in pedestrian crashes than females for both pedestrians and motorists. The distribution of the crashes among different age groups reflects the high incidence of crashes involving child pedestrians after adults. Intoxication affected more pedestrians than motorists. Nearly 60% of all crashes occurred when pedestrians crossed the streets and more than half of them were not at crosswalks. Over 60% of the crashes occurred during daylight hours.

The second analysis was to analyze the temporal variation in the percentage of injuries among crashes at four dimensions (yearly, monthly, day of a week, and hourly) based on all crashes (see Fig. 3). The single-pedestrian crashes exhibited the similar trends and thus were not shown. The yearly variation (Fig. 3a) shows a generally declining trend in the 8 years regarding the percentage of injuries but 2007 was still high. The monthly trend (Fig. 3b) reveals that the percentage of injuries was higher in summers between May and July than the other seasons and was highest in June. Weekends (Fig. 3c) encountered a higher percentage of injuries compared to weekdays. Fridays had a higher percentage of injuries compared to the other weekdays. The hourly trend high-

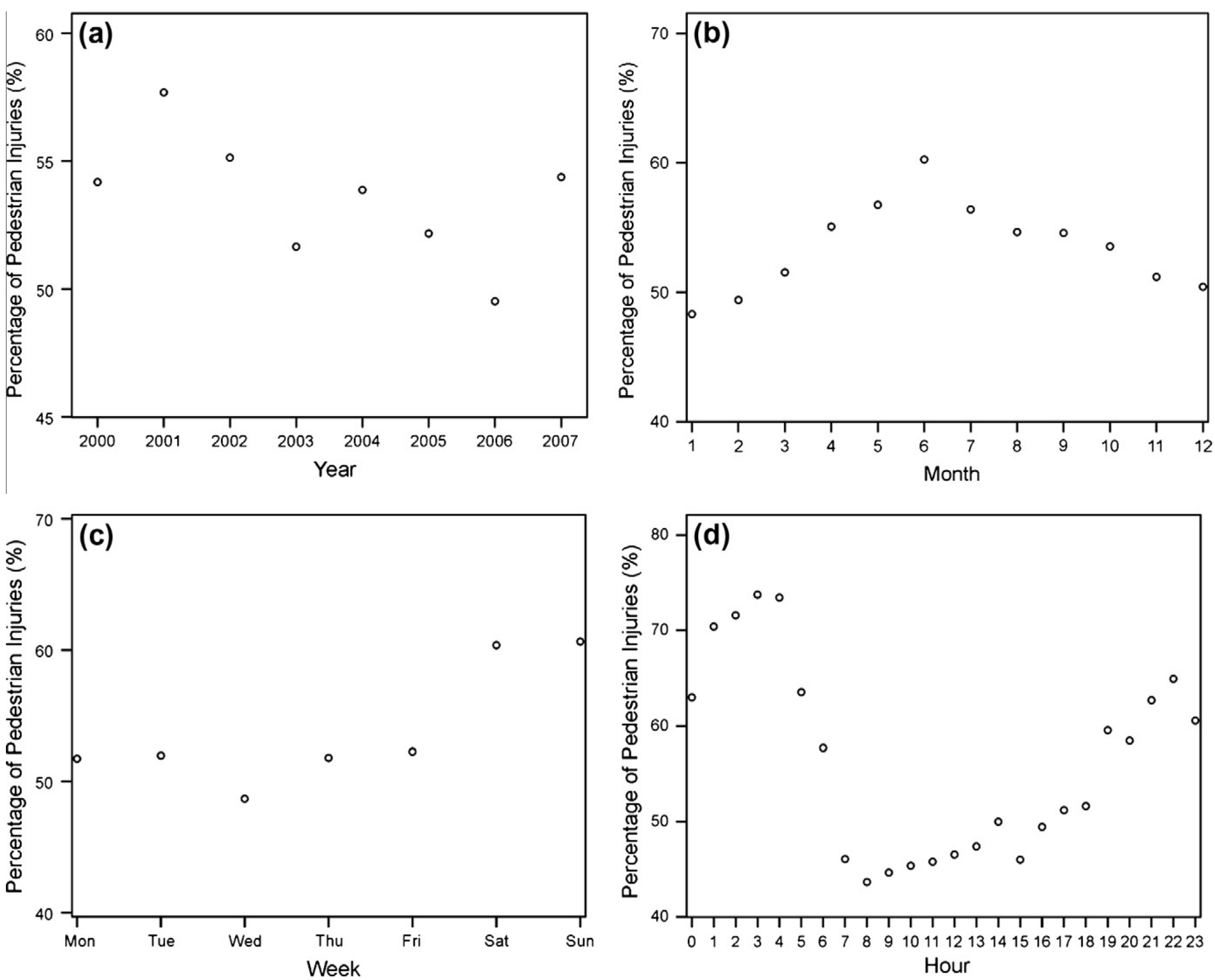

Fig. 3. Pedestrian injury rate by year (a), by month (b), by day of a week (c), and by time of a day (d).

Table 3
Definitions of independent variables.

	Variable	Definition
Personal factor	Gender[a]	0 = female and 1 = male
	Pedestrian age (in years)	0 = 15–34; 1 = 14 or less; 2 = 35–54; and 3 = 55 or older
	Driver's age (in years)	0 = 25 or less; 1 = 26–50; 2 = 51 or older
	Intoxication[a]	0 = no intoxication; 1 = not known; 2 = intoxicated by alcohol and/or drugs; and 3 = others[b]
	Pedestrian maneuver	0 = crossing, not at cross walk; 1 = crossing at crosswalk; 2 = walking with traffic; 3 = walking against traffic; 4 = working on road/vehicle; 5 = playing/standing on road; 6 = darting into traffic; and 7 = others[b]
Environmental factor	Light conditions	0 = daylight; and 1 = dusk, dawn, or darkness
	Surface conditions	0 = dry; 1 = wet/snowy/icy; and 2 = others[b]
	Weather conditions	0 = clear; 1 = cloudy; 2 = rain/snow/sleet/fog; and 3 = others[b]

[a] Both pedestrians and drivers fall into the same classification.
[b] Events that do not clearly fall into the rest of the categories.

Table 4
Influence of personal and environmental factors on the risk of injury.[a]

	Injury rate[b]	Model 1 odds ratio (95% CI[c])	Injury rate[b]	Model 2 odds ratio (95% CI[c])	Model 3 odds ratio (95% CI[c])
Gender (Pedestrian)					
Female[d]	491	1	489	1	1
Male	560	1.182* (1.067, 1.308)	564	1.211** (1.089, 1.346)	1.271 (0.986, 1.639)
Age (in years; Pedestrian)					
15–34[d]	514	1	517	1	1
14 or less	600	1.430** (1.238, 1.652)	613	1.461** (1.256, 1.700)	1.946** (1.246, 3.041)
35–54	523	1.040 (0.927, 1.167)	521	1.010 (0.896, 1.138)	1.113 (0.844, 1.468)
55 or older	544	1.298** (1.100, 1.532)	540	1.258* (1.061, 1.492)	1.402 (0.980, 2.005)
Intoxication (Pedestrian)					
Not[d]	493	1	497	1	1
Not known	630	1.558** (1.389, 1.748)	627	1.525** (1.354, 1.718)	1.376 (0.991, 1.912)
Yes	669	1.524* (1.169, 1.986)	670	1.521** (1.163, 1.989)	1.959 (0.714, 5.379)
Others	368	0.664 (0.254, 1.741)	368	0.657 (0.250, 1.723)	1.711 (0.071, 41.053)
Pedestrian Maneuver					
Crossing at crosswalk[d]	387	1	386	1	1
Crossing, not at crosswalk	625	2.225** (1.946, 2.543)	628	2.232** (1.943, 2.563)	1.966** (1.508, 2.563)
Walking with traffic	500	1.391* (1.112, 1.742)	500	1.373** (1.089, 1.733)	1.225 (0.770, 1.950)
Walking against traffic	446	1.108 (0.812, 1.510)	454	1.153 (0.840, 1.582)	1.137 (0.527, 2.457)
Working on road/vehicle	495	1.441* (1.104, 1.880)	480	1.358* (1.030, 1.791)	1.394 (0.704, 2.764)
Playing/standing on road	589	1.868** (1.521, 2.293)	587	1.820** (1.467, 2.257)	2.476** (1.577, 3.887)
Darting into traffic	705	3.333** (2.433, 4.566)	718	3.502** (2.531, 4.845)	1.609** (1.168, 2.216)
Others	512	1.482** (1.269, 1.731)	514	1.495** (1.271, 1.760)	4.010** (2.022, 7.952)
Gender (Driver)					
Female[d]	516	1	524	1	1
Male	545	1.133* (1.021, 1.258)	544	1.106 (0.993, 1.232)	1.191 (0.938, 1.514)
Age (in years; Driver)					
25 or less[d]	616	1	623	1	1
25–50	545	0.744** (0.649, 0.854)	546	0.716** (0.621, 0.826)	0.793* (0.639, 0.984)
51 or older	486	0.601** (0.519, 0.696)	487	0.581** (0.499, 0.677)	0.728* (0.568, 0.932)
Intoxication (Driver)					
Not[d]	530	1	534	1	1
Not known	528	1.085 (0.945, 1.246)	526	1.088 (0.942, 1.257)	1.380 (0.828, 2.299)
Yes	763	2.446** (1.718, 3.482)	769	2.382** (1.598, 3.551)	3.325 (0.980, 11.283)
Others	375	0.632 (0.142, 2.825)	375	0.647 (0.145, 2.881)	0.520 (0.052, 5.164)
Light conditions					
Daylight[d]	484	1	489	1	1
Darkness	612	1.504** (1.354, 1.671)	612	1.486** (1.332, 1.658)	1.439** (1.268, 1.632)
Surface conditions					
Dry[d]	536	1	539	1	1
Wet/snowy/icy	529	0.955 (0.759, 1.201)	527	0.989 (0.778, 1.257)	0.526 (0.274, 1.008)
Other	417	0.285 (0.063, 1.281)	417	0.279 (0.062, 1.257)	0.294 (0.029, 2.963)
Weather conditions					
Clear[d]	534	1	537	1	1
Cloudy	542	1.070 (0.938, 1.220)	545	1.069 (0.933, 1.224)	1.069 (0.904, 1.265)
Rain/snow/sleet/fog	522	0.946 (0.719, 1.245)	514	0.879 (0.660, 1.171)	0.650 (0.284, 1.487)
Other	769	4.923* (1.008, 24.041)	769	4.952* (1.013, 24.211)	2.584 (0.219, 30.493)

Note: Model 1 was based on all crashes ($n = 7195$). Models 2 and 3 were based on only single-pedestrian crashes ($n = 6705$).
[a] The Hosmer and Lemeshow test in the three models returned p-values of 0.792, 0.058 and 0.590 respectively. The corresponding Chi-square values were 4.670, 15.076, and 6.513, indicating the number of injuries are not significantly different from those predicted by the models and that the overall fit of the three models are good.
[b] Number of pedestrian injuries per 1000 crashes in a category when all crashes and single-pedestrian crashes were considered respectively.
[c] CI, confidence interval.
[d] Reference category in the model serves as the basis for comparison.
* $\alpha = 0.05$.
** $\alpha = 0.01$.

lights the danger of the time from 7 pm to 6 am. The percentage of injuries was the highest from 1 am to 4 am and was the lowest from 7 am to 10 am.

To examine the influence of the factors from the three main categories (pedestrian, driver, and environment), this research used the logistic regression model in keeping with previous studies on pedestrian crashes (Sze and Wong, 2007; Tay et al., 2008). The dependent variable used binary coding to represent being injured (code 1, injured) for a pedestrian, including fatalities, serious injuries, and visible injuries. This study, therefore, models the risk of injury for a pedestrian who experienced a crash. To account for the impact of multiple pedestrians per crash, models 1 and 2 considered main effects of the factors based on all and single-pedestrian crashes respectively. Model 3 considered both main effects and the interaction between every two factors within each category based on single-pedestrian crashes.

As shown in Table 3, independent variables included four pedestrian factors (age, gender, intoxication, and maneuver), three driver's factors (age, gender, and intoxication), and three environmental factors (weather, surface, and darkness). Consistent with the previous study (Beck et al., 2007) on pedestrian crashes in Atlanta, this study classified ages (in years) of pedestrians into four groups: younger than or equal to 14, 15–34, 35–54, and older than or equal to 55, reflecting children, younger adults, adults, and senior pedestrians, respectively. The ages of the motorists involving pedestrian crashes were grouped into three categories: 25 or less, 26–50, and 51 or older in line with a prior study (Abdel-Aty et al., 2007) and the GDOT report (GDOT, 2008), reflecting young drivers, adult drivers, and old drivers, respectively. The coefficients for these variables describe the difference in injury risk in pedestrian crashes in a particular category relative to the risk in the reference category defined in Table 4.

The regression result (Table 4) sheds light on the influence of personal and environmental factors on injury risk. Compared to younger adults, the injury risk significantly increased ($p < 0.05$) in children (by at least 43%) and senior pedestrians (by at least 25.8%) when only main effects were considered. The relationships remained consistent when adult pedestrians were alternatively used as the reference. Compared to female pedestrians, the odds of injuries (versus no injuries) increased by a factor of at least 1.182 in males in models 1 and 2. Walking under influence by alcohol and/or drugs significantly increased the injury risk by 52% in models 1 and 2. Regarding pedestrian maneuvers, compared to crossing at a crosswalk, crossing not at a crosswalk, playing or standing on road, and darting into traffic were consistently the significant risk factors in three models. In addition, walking with traffic and working on road/vehicle were significant risk factors in the first two models. For the motorists, younger drivers less than 25 years old significantly increased the injury risk compared to older drivers in three models. The impact of driver's age remained when it was reclassified into four groups (25 or less, 26–34, 35–64, and 65 or older). The risk for injuries given a crash was significantly increased when drivers were male or intoxicated (see models 1 and 2). Among the three environmental factors, only the light condition was significantly associated with the injury risk. Severe surface (wet, snowy, and icy) and weather (rain, snow, sleet, and fog) conditions were in fact associated with a slightly decreased injury risk given a crash, but none of the associations were significant. Model 3 shows three interactions presenting significantly lower injury risks given a pedestrian crash. Their odds ratios are 0.535 for male pedestrians playing/standing on road (versus females crossing not at crosswalk), 0.690 for children aged 14 or less with unknown intoxication (versus pedestrians aged 15–34 without intoxication), and 0.673 for pedestrians aged 55 or older with unknown intoxication (versus pedestrians aged 15–34 with unknown intoxication).

5. Discussions and conclusions

In order to implement intervention and prevention programs effectively, pedestrian safety professions need to make decisions on which roadway segments to target and which population groups to focus. This study makes contributions in this regard by employing a spatial clustering technique to detect roadway segments with significantly elevated injury risks given a crash. It also evaluates the factors significantly contributing to the injury risk from personal and environmental aspects.

Using the spatial epidemiological model, this study provides insights into the clustering of pedestrian injuries in the suburban high-activity corridors. The clusters consistently reported in different settings reveal the extreme danger of these zones. The geographic characteristics of these high-risk corridors in general include the mixture of state highways carrying heavy fast-speed traffic and local roads in suburbs. Along state highways and their intersections with local streets are commonly businesses (e.g., restaurants or gas stations) mixed with residential properties, where pedestrians are likely to be present and negotiate with traffic. These suburban corridors commonly have huge blocks and multiple lanes with limited availability of sidewalks, making them difficult for walking and crossing (Ewing et al., 2003). Given the importance of built environment (e.g., crosswalks and sidewalks) and the land-use pattern for pedestrian safety (Dai et al., 2010; Graham and Glaister, 2003), it is necessary to examine the built environment and the land use within these clusters. On the contrary, no pedestrian injuries were significantly clustered in downtown Atlanta despite the high crash incidence, which is consistent with a previous study (Zajac and Ivan, 2003). The unique urban settings, such as slow traffic given the high pedestrian volume and various traffic control devices (e.g., low speed limit and dense stop lights), may reduce the severity of collision (Graham and Glaister, 2003).

The apparent temporal variations in the percentage of pedestrian are consistent with findings from previous studies (Brugge et al., 2002; Coate and Markowitz, 2004), which suggests more prevention programs shall be conducted on weekends, Fridays, and evenings (Clifton et al., 2009; Wanvik, 2009). The effect of pedestrian factors including age, walking under influence, crossing a street not at a crosswalk, and darting into traffic is in line with previous findings (Clifton et al., 2009; Zajac and Ivan, 2003). Children, seniors, and intoxicated pedestrians are at a higher risk given their limited cognitive ability to respond effectively to traffic (Gorrie et al., 2008; Hasselberg and Laflamme, 2004), which calls on attention for education efforts aiming to these special populations. Darting into traffic and crossing not at crosswalks highlight the importance of pedestrian compliance to traffic rules and the availability of crosswalk signs to improve pedestrian safety (Clifton et al., 2009; Gorrie et al., 2008). In line with previous studies (Helai et al., 2008; Zajac and Ivan, 2003), younger, male, or intoxicated drivers pose a significant high risk for injuries to pedestrian in a crash, which requires particular attention to these populations. This study reveals the significant influence of lighting, not surface or weather conditions, to the injury risk for pedestrians in a collision, which is consistent with findings from previous studies (Clifton et al., 2009; Coate and Markowitz, 2004). Improving light conditions may be necessary to enhance pedestrian safety in poorly lit areas. The findings of surface and weather conditions might reflect the context of Atlanta which has short snowy and icy seasons and therefore warrant further investigation in other regions.

Given that most pedestrian–vehicle crashes occurred on road, network distance is better than Euclidean distance to describe the spatial separations between events. This study restricted the search window sizes to account for the difference. Network-based clustering methods, such as SANET (Okabe et al., 2006), are still in early stages (Warden, 2008) for statistically evaluating injury risks while considering the at-risk background population in both space and time. Network-based clustering would be logically the next step to extend this study.

The results of this study are subject to several limitations. First, it is unlikely that all crashes were reported to the police and thus reporting bias may influence the research findings. A pedestrian hit by a car may leave the scene without calling 911 in spite of bruises and scrapes (Brugge et al., 2002). Further investigation is necessary to study the influence of the reporting bias. Second, errors may arise from the original crash reports. A police officer may not be able to make accurate records of all relevant information because of causalities, coordinating firemen and ambulance men, and guiding traffic (Loo, 2006). This study was unable to validate the original reports given their restricted access. Third, the cases and controls in the Bernoulli model may differ in some contexts, such as age, gender, or light conditions, which may influence the results. This may be mitigated by covariate adjustment of these factors using the crashes in the same category. But the small number of crashes in each category may compromise the power of the statistical analysis. Fourth, traffic volumes for vehicles and pedestrians shall be taken into account. It is necessary in future research to investigate how to survey the vast and dynamic pedestrian traffic in the vicinity of vehicles in an urban setting. The cases and controls in this study, which were drawn from the same database that recorded crashes in the same study period, were able to reflect the pedestrians exposed to injury risk. Finally, this study did not control the effect of contact points of collisions on the injury risk. This requires access to information on pedestrians and vehicles which were not available in the publicly accessible database. Accessing such restricted data is warranted in future research in order to better understand the high-risk segments and factors.

In summary, this research provides insights into pedestrian injuries using spatial and statistical techniques, which has important implications to public health professions to improve traffic safety. The spatial technique enables us to identify high risky road segments which highlight the need for intervention and prevention. Statistical analysis then reveals the injury risk associated with time, pedestrians, motorists, and the environment. The integration of the two allows us to fully understand pedestrian crashes and develop more focused programs to improve pedestrian safety.

Acknowledgement

Financial support from Georgia State University Injury and Violence Seed Program is gratefully acknowledged. Suggestions from the editors and anonymous reviewers are greatly appreciated.

References

Abdel-Aty, M., Chundi, S.S., Lee, C., 2007. Geo-spatial and log-linear analysis of pedestrian and bicyclist crashes involving school-aged children. Journal of Safety Research 38, 571–579.
Anderson, T.K., 2009. Kernel density estimation and K-means clustering to profile road accident hotspots. Accident Analysis and Prevention 41, 359–364.
Beck, L.F., Paulozzi, L.J., Davidson, S.C., 2007. Pedestrian fatalities, Atlanta metropolitan statistical area and United States, 2000–2004. Journal of Safety Research 38, 613–616.
Bell, N., Schuurman, N., 2010. GIS and injury prevention and control: history, challenges, and opportunities. International Journal of Environmental Research and Public Health 7, 1002–1010.
Brugge, D., Lai, Z., Hill, C., Rand, W., 2002. Traffic injury data, policy, and public health: lessons from Boston Chinatown. Journal of Urban Health-Bulletin of the New York Academy of Medicine 79, 87–103.
Chang, L.Y., Wang, H.W., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accident Analysis and Prevention 38, 1019–1027.
Christie, N., Ward, H., Kimberlee, R., Towner, E., Sleney, J., 2007. Understanding high traffic injury risks for children in low socioeconomic areas: a qualitative study of parents' views. Injury Prevention 13, 394–397.
Clifton, K.J., Burnier, C.V., Akar, G., 2009. Severity of injury resulting from pedestrian–vehicle crashes: what can we learn from examining the built environment. Transportation Research Part D 14, 425–436.
Coate, D., Markowitz, S., 2004. The effects of daylight and daylight saving time on US pedestrian fatalities and motor vehicle occupant fatalities. Accident Analysis and Prevention 36, 351–357.
Dai, D., Taquechel, E., Steward, J., Strasser, S., 2010. The impact of built environment on pedestrian crashes and the identification of crash clusters on an urban university campus. Western Journal of Emergency Medicine 11, 294–301.
Elliott, S., Woolacott, H., Braithwaite, R., 2009. The prevalence of drugs and alcohol found in road traffic fatalities: a comparative study of victims. Science & Justice 49, 19–23.
Ewing, R., Schieber, R.A., Zegeer, C.V., 2003. Urban sprawl as a risk factor in motor vehicle occupant and pedestrian fatalities. American Journal of Public Health 93, 1541–1545.
Georgia Department of Transportation, 2008. Crash Analysis, Statistics & Information. <www.dot.state.ga.us/statistics/crashdata/pages/casi.aspx> (accessed 28.01.12).
Georgia Department of Transportation, 2010a. Georgia Uniform Vehicle Accident Report Instruction Guide. <www.dot.state.ga.us/statistics/crashdata/documents/georgiauniformvehicleaccidentreport.pdf> (accessed 7.08.10).
Georgia Department of Transportation, 2010b. Traffic Count. <www.dot.state.ga.us/statistics/trafficdata/pages/default.aspx> (accessed 7.08.10).
Gorrie, C.A., Brown, J., Waite, P.M.E., 2008. Crash characteristics of older pedestrian fatalities: dementia pathology may be related to 'at risk' traffic situations. Accident Analysis and Prevention 40, 912–919.
Graham, D.J., Glaister, S., 2003. Spatial variation in road pedestrian casualties: the role of urban scale, density and land-use mix. Urban Studies 40, 1591–1607.
Hasselberg, M., Laflamme, L., 2004. Children at risk in traffic: improvement potentials in the Swedish context. Acta Paediatrica 93, 113–119.
Helai, H., Chor, C.H., Haque, M.M., 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis. Accident Analysis and Prevention 40, 45–54.
Kim, J.K., Ulfarsson, G.F., Shankar, V.N., Kim, S., 2008. Age and pedestrian injury severity in motor-vehicle crashes: a heteroskedastic logit analysis. Accident Analysis and Prevention 40, 1695–1702.
King, M.J., Soole, D., Ghafourian, A., 2009. Illegal pedestrian crossing at signalised intersections: incidence and relative risk. Accident Analysis and Prevention 41, 485–490.
Kulldorff, M., 2009. SatScan TM User Guide for Version 8.0. <www.satscan.org> (accessed 7.08.10).
Lefler, D.E., Gabler, H.C., 2004. The fatality and injury risk of light truck impacts with pedestrians in the United States. Accident Analysis and Prevention 36, 295–304.
Loo, B.P.Y., 2006. Validating crash locations for quantitative spatial analysis: a GIS-based approach. Accident Analysis and Prevention 38, 879–886.
Marshall, S.W., Waller, A.E., Loomis, D.P., Langlois, J.A., 2000. Selection of controls in injury case-control studies. Annals of Epidemiology 10, 455.
Mu, L., Wang, F., 2008. A scale-space clustering method: mitigating the effect of scale in the analysis of zone-based data. Annals of the Association of American Geographer 98, 85–101.
National Highway Traffic Safety Administration, 2009. Traffic Safety Facts. <www-nrd.nhtsa.dot.gov/Pubs/811402.pdf> (accessed 15.01.10).
National Safety Council, 2009. Estimating the Costs of Unintentional Injuries. <www.nsc.org> (accessed 30.01.10).
Okabe, A., Okunuki, K., Shiode, S., 2006. SANET: a toolbox for spatial analysis on a network. Geographic Analysis 38, 57–66.
Paulozzi, L.J., 2006. Is it safe to walk in the Sunbelt? Geographic variation among pedestrian fatalities in the United States, 1999–2003. Journal of Safety Research 37, 453–459.
Pulugurtha, S.S., Krishnakumar, V.K., Nambisan, S.S., 2007. New methods to identify and rank high pedestrian crash zones: an illustration. Accident Analysis and Prevention 39, 800–811.
Song, C., Kulldorff, M., 2003. Power evaluation of disease clustering tests. International Journal of Health Geographics 2, 9.
Steenberghen, T., Aerts, K., Thomas, I., 2010. Spatial clustering of events on a network. Journal of Transport Geography 18, 411–418.
Sze, N.N., Wong, S.C., 2007. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. Accident Analysis and Prevention 39, 1267–1278.
Tay, R., Rifaat, S.M., Chin, H.C., 2008. A logistic model of the effects of roadway, environmental, vehicle, crash and driver characteristics on hit-and-run crashes. Accident Analysis and Prevention 40, 1330–1336.
Wanvik, P.O., 2009. Effects of road lighting: an analysis based on Dutch accident statistics 1987–2006. Accident Analysis and Prevention 41, 123–128.
Warden, C.R., 2008. Comparison of Poisson and Bernoulli spatial cluster analyses of pediatric injuries in a fire district. International Journal of Health Geographics 7, 51.
Warden, C.R., Duh, J.D., Lafrenz, M., Chang, H., Monsere, C., 2011. Geographical analysis of commercial motor vehicle hazardous materials crashes on the Oregon state highway system. Environmental Hazards 10, 171–184.
Xie, Y.C., Zhang, Y.L., Liang, F.M., 2009. Crash injury severity analysis using Bayesian Ordered Probit models. Journal of Transportation Engineering 135, 18–25.
Yamada, I., Thill, J.-C., 2010. Local indicators of network-constrained clusters in spatial patterns represented by a link attribute. Annals of the Association of American Geographer 100, 269–285.
Yoganandan, N., Baisden, J.L., Maiman, D.J., Gennarelli, T.A., Guan, Y.B., Pintar, F.A., Laud, P., Ridella, S.A., 2010. Severe-to-fatal head injuries in motor vehicle impacts. Accident Analysis and Prevention 42, 1370–1378.
Zajac, S.S., Ivan, J.N., 2003. Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut. Accident Analysis and Prevention 35, 369–379.