Journal of Transport Geography 58 (2017) 71–91

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Spatial investigation of aging-involved crashes: A GIS-based case study in Northwest Florida

Mehmet Baran Ulak [a,*], Eren Erman Ozguven [a], Lisa Spainhour [a], Omer Arda Vanli [b]

[a] Department of Civil and Environmental Engineering, FAMU–FSU College of Engineering, Tallahassee, FL 32310, United States
[b] Department of Industrial and Manufacturing Engineering, FAMU–FSU College of Engineering, Tallahassee, FL 32310, United States

ARTICLE INFO

Article history:
Received 29 December 2015
Received in revised form 24 May 2016
Accepted 22 November 2016
Available online 6 December 2016

Keywords:
Crash analysis
Aging populations
Geographic information systems
Spatial density analysis
Spatial dependence

ABSTRACT

This study attempts to understand the unique nature of crashes involving aging drivers, unlike many previous crash-focused traffic safety studies mostly focusing on the general population. The utmost importance is given to answering the following question: How do the crashes involving aging drivers vary compared to crashes involving other age groups? To achieve this objective, a three-step spatial analysis was conducted using geographic information systems (GIS) with a case study application on three urban counties in the Northwest Florida region, based on crash data obtained from the Florida Department of Transportation (FDOT). First, crash clusters were investigated using a kernel density estimation (KDE) approach. Second, a crash density ratio difference (DRD) measure was proposed for comparing maxima-normalized crash densities for two different age groups. Third, a population factor (PF) was developed in order to investigate effect of spatial dependency by incorporating the effect of both number and percent of 65 + populations in a region. This spatial analysis was followed by a logistic regression-based approach in order to identify the statistically significant factors that can help investigate the distinct patterns of crashes involving aging drivers. Results of this study indicate that crashes involving aging drivers differ from other age group crashes both spatially and temporally. Further, the DRD and PF factors are useful metrics to identify and investigate important regions of study. The GIS-based knowledge gained from this research can contribute to the development of more reliable aging-focused safety plans and models.

1. Introduction

Traffic crashes have vital, social and economic consequences for many developed and developing regions that heavily depend on a safe and reliable traffic network. According to the Federal Highway Administration of the U.S. Department of Transportation (NHTSA, 2014), in the U.S., roadway crashes are one of the leading cause of death and injury, and the total societal cost of roadway crashes is more than $230 billion annually. Therefore, studies on traffic crashes are of critical importance to reduce the drastic effects of those crashes on the society. Over the last 25 years, many researchers have recognized the necessity of delving into the nature of the traffic crashes. This necessity arises from the fact that developing methodologies to reduce crashes are vital to provide the public with safe and reliable transportation. From a transportation safety perspective, this problem becomes even more challenging and complex when older drivers are considered, since research has shown that they are more vulnerable to roadway crashes than other age groups (Abdel-Aty et al., 1998; Alam and Spainhour, 2014). Physical and cognitive limitations, slower reflexes, deteriorated visions, and other health conditions due to aging contribute to the escalation of this crash risk for aging drivers (Hellinga and Macgregor, 1999; Merat et al., 2005; Sandler et al., 2015). A thorough review of the literature shows that there is still a gap in terms of investigating the spatial properties of crashes involving aging drivers as well as investigating the significant factors that specifically affect the occurrence of these crashes. Note that by "crashes involving aging drivers," we mean the crashes involving at least one driver 65 years and older, whether the driver is at-fault or not.

Given the limitations of existing traffic crash studies focused on the aging drivers, this paper develops a GIS-based spatial analysis methodology with the following objectives: (a) to identify the clustering behavior of crashes on a given roadway network, and (b) to discover the geospatial differences between aging drivers and other age groups based on the comparison of high risk crash locations. For this purpose, we first proposed a "density ratio difference" (DRD) approach. Next, spatial dependence of the crashes involving aging drivers was investigated in the form of the effect of spatial distribution of aging populations on the crash occurrences. We investigated this spatial dependency by developing a novel metric called "population factor" (PF). To the authors' knowledge, these have not been done before in the traffic safety/engineering field, which represent the main contributions of this paper. This spatial analysis methodology is applied on three urban counties

* Corresponding author at: 2525 Pottsdamer Street, A-219, Tallahassee, FL 32310, United States.
E-mail address: mulak@fsu.edu (M.B. Ulak).

http://dx.doi.org/10.1016/j.jtrangeo.2016.11.011
0966-6923/

in the Northwest Florida region, namely Leon, Bay and Escambia counties. These counties are documented to have high crash rates with respect to their aging population by the Safe Mobility for Life Coalition in Florida (Sandler et al., 2015). The spatial analysis followed by a statistical analysis to identify the significant factors influencing the crashes involving aging drivers spatially and temporally using a logistic regression-based approach. Aging population living in the vicinity of high crash risk locations are also included as a factor in this analysis. The proposed methodology can enable transportation officials to analyze the significant factors that influence the crashes involving aging drivers, along with the prevalence of occurrence locations of these crashes on the roadway network. This can help decision makers understand the unique nature of crashes involving aging drivers, and differentiate the spatial properties and contributing factors from other age group-involved crashes.

2. Literature review

2.1. Effects of aging on crash involvement

Several researchers have investigated the effects of aging on crashes and traffic safety (Please see Bayam et al. (2005) for an exhaustive review of these studies). A study by Abdel-Aty et al. (1998) showed that driver age was significantly correlated with crash-related factors such as average daily traffic and roadway characteristics. In this study, relative crash frequencies were used instead of actual total crash numbers since the total number of crashes varied with respect to the number of drivers in each age group. The relative crash frequency measure performed better than the actual crash numbers while comparing crashes of different age groups.

Several studies verify that the driving behavior of aging drivers is substantially different than that of other age group drivers (Abdel-Aty et al., 1999a; Boyce and Geller, 2002; Krahe and Fenske, 2002). For example, reckless and aggressive driving was found to be uncommon in older drivers (Jang, 2006; Krahe and Fenske, 2002; Rong et al., 2011). This finding implies that aging drivers tend to avoid risky situations, and adopt a risk-averse driving behavior. Aging drivers were also found to avoid driving at night, during adverse weather, on slippery roadways, during peak hours, long distances, highly congested roadways, and roadways with high speed limits (Baker et al., 2003; Bayam et al., 2005; Collia et al., 2003). Furthermore, Mori and Mizohata (1995) showed that aging drivers drove slowly and preferred large gaps in the traffic. Due to these cautious and risk-averse driving behaviors, aging drivers are often regarded as safer drivers than younger drivers in terms of speeding, speed variation, focusing on roadway, signal usage, and keeping distance gaps (Boyce and Geller, 2002). Furthermore, healthy aging drivers are found to be better or at least at the same level with younger drivers in terms of driving skills (Carr et al., 2016). However, even though aging drivers are experienced, cautious, and risk-averse, challenges such as deterioration in health, reflexes, vision and cognitive skills impose a substantial crash risk on aging drivers.

Aging drivers were also found to have different crash involvement characteristics than drivers from other age groups. For example, aging drivers were found more prone to be involved in a crash while turning left or right, changing lanes, merging into traffic, and approaching intersections than drivers of other age groups (McGwin and Brown, 1999). Similarly, failing to yield the right of way, disregarding unseen objects, unintentionally neglecting stop signs and signals were influential on the crashes involving aging drivers (Abdel-Aty et al., 1999a; McGwin and Brown, 1999). However, some factors such as tiredness (fatigue), adverse weather conditions, speeding, road curvatures, and driving at night were identified as non-influential factors for crashes involving aging drivers in the literature (McGwin and Brown, 1999). Moreover, aging drivers were also less likely to be involved in crashes due to alcohol-drug abuse (Abdel-Aty and Abdelwahab, 2000) than other age groups. Note that these factors are not less influential because aging drivers are more resistant to fatigue, night and/or bad weather conditions, nor they are more agile than drivers from other age groups. The reason is that they tend to drive more attentively than other drivers in those conditions, or they simply avoid driving under these adverse conditions. In sum, unique characteristics of driving behavior and crashes involving aging drivers compel devising specific measures addressing their specific problems and ensuring their safety (Bédard et al., 2002).

2.2. Geospatial crash analysis

There are several studies that consider the use of geostatistical and geospatial methods in order to analyze spatial data such as roadway crashes, and to illustrate them visually on maps. Methods such as Ripley's K-function, Getis's G-statistics or Moran's-I are widely used to investigate whether crashes are randomly distributed or not (Steenberghen et al., 2010). Ripley's K-function can test whether spatially distributed points form statistically significant clusters or not whereas Local Indicators of Spatial Association (LISA) methods like Getis's G-statistics or Moran's-I can disclose specific locations of the statistically significant clusters of points (Getis and Ord, 1992). However, these methods either do not illustrate crash clusters (e.g., Ripley's K-function), or present only specific cluster locations in space (e.g., Getis's G-statistics and Moran's-I) rather than showing the clustering behavior on the roadway network. Kernel density estimation (KDE), on the other hand, is capable of highlighting the clustering behavior of the crashes visually on a roadway network, which is of interest in this study. This ability is important in order to investigate the density of spatial distribution of crashes instead of only identifying the cluster locations or determining the existence of statistically significant clusters. Moreover, a comparison of the local spatial autocorrelation with the kernel methods provided by Flahaut et al. (2003) show that these two methods produce similar results in terms of locations of crash clusters.

KDE has been a widely used method to create density surfaces using spatially distributed point data (Brunsdon, 1995). The KDE method has been used to identify 'hot spots' (i.e. locations with an unusually high occurrence of a particular phenomenon), across a range of disciplines. One such example is identifying crime-related clusters by geographical proximity or hot spots in boroughs or cities (Chainey et al., 2008). Similarly, GIS can be used to analyze roadway crashes in order to identify high risk locations, hotspots and/or clusters (NHTSA, 2015). Indeed, this approach has been implemented for the geo-spatial crash analysis successfully (Erdogan et al., 2008; Kilamanua et al., 2011; Mohaymany et al., 2013). Recently, researchers have also been focusing on the effects of spatial dependence and spatial heterogeneity on the crash occurrences since spatial factors are also very important determinants of crashes (Delmelle et al., 2011; Effati et al., 2015). For instance, Effati et al. (2015) showed that crash severity was highly influenced by the urban development and geographic elevation along a highway corridor.

Two types of KDE approaches can be applied to cluster analysis. Planar (Euclidean) KDE uses the straight-line, "as the crow flies," distance between two incidents to measure proximity. For instance, a planar KDE-based cluster analysis was developed by Sabel et al. (2005) for the spatial evaluation of roadway crashes using GIS. The authors combined the spatial data for average traffic flows with the crash locations, and identified significant crash clusters. Similarly, most of the existing studies use the planar distance-based KDE approach. Because vehicles must follow road networks, relying on planar KDE may cause the following problems: (a) Overestimation: Some roadways that do not possess high risk are shown to be risky, (b) Underestimation: Since multiple roadways are shown as critical locations rather than the actual roadways that have high crash risk, agencies may not show the extra attention needed for the actual high risk locations (Yamada and Thill, 2004). Therefore, a network distance-based KDE approach was developed especially for estimating spatial density of data points distributed along networks (Okabe and Sugihara, 2012; Okabe et al., 2009; Steenberghen et al., 2010; Xie and Yan, 2013). This approach solves the estimation-related problems using the actual roadway distances between the crashes.

A comparison between planar and network-based KDE models for pedestrian-involved crashes is presented by Dai et al. (2010). Results showed that network-based KDE performed better than the planar approach for data distributed over networks. Other studies showed similar results (Mohaymany et al., 2013); however, there have not been any studies in the literature that focused on the KDE analysis specifically for crashes involving aging drivers. An important issue related to the KDE analysis is the choice of kernel function and the smoothing parameter called bandwidth. Generally, in an urban network, bandwidths with a range of 50 m to 300 m are considered to be appropriate (Steenberghen et al., 2010; Xie and Yan, 2013). For rural areas, on the other hand, the selection of bandwidths around 1000 m was suggested as an appropriate choice (Blazquez and Celis, 2013). In addition, there are various kernel functions such as uniform, normal, biweight, and Epanechnikov that can be used while estimating the kernel densities. However, the choice of kernel function has little effect on the density estimation compared to the effect of bandwidth selection as shown by Xie and Yan (2008). Therefore, the shape of the resultant density function is essentially defined by the bandwidth rather than the kernel function.

2.3. Statistical analysis of crash influencing factors

Several researchers have also recognized the need to study and investigate significant factors associated with the roadway crashes. Please refer to Lord and Mannering (2010) and Mannering and Bhat (2014) for an extensive review of the statistical methods used for this purpose. To the authors' knowledge, most of the existing studies focused only on the effects of the aging driver behavior, characteristics and limitations on the crashes (Abdel-Aty and Abdelwahab, 2000; Abdel-Aty and Radwan, 2000; Abdel-Aty et al., 1999a; Alam and Spainhour, 2008; Eby and Molnar, 2009; Evans, 2000; Mayhew et al., 2006; McGwin and Brown, 1999; Ryan et al., 1998; Stamatiadis et al., 1991). Studies that specifically focus on the traffic, roadway and environment-related factors influencing the crashes involving aging drivers, however, are very limited. A pioneering age-focused study showed that aging roadway users were more vulnerable to crashes than other age groups (Abdel-Aty et al., 1998). Several researchers have statistically shown that various environmental, roadway and driver-related factors had different impacts on the crash involvement likelihood of different age groups (Abdel-Aty and Abdelwahab, 2000; Abdel-Aty et al., 1999a; Alam and Spainhour, 2008). In Ryan et al. (1998), the effects of age and gender in the crash likelihood were investigated, where the crash rate of 70+ drivers appeared to be one of the highest groups when crash rate per distance traveled was considered as the basis for the methodology. The effect of non-resident drivers on the crashes has also been investigated by Abdel-Aty et al. (1999b), where it was clearly shown that resident and non-resident drivers have similar crash involvement across different age groups. This is also a critical issue for states like Florida, which attracts substantial numbers of visitors (tourists) as well as nonpermanent aging residents (in-flow of retired adults from other states).

Among the regression models used in the literature, logistic models are one of the widely used binary choice models (Al-Ghamdi, 2002; Karacasu et al., 2013; Kononen et al., 2011; Ma et al., 2009). Moreover, machine learning techniques have also been popular recently (Effati et al., 2015). At-fault drivers' age was investigated as a contributing factor in fatal crashes by implementing a logit-based analysis by Alam and Spainhour (2014). This study revealed that the probability of a fatal crash is higher for younger and older drivers than other age groups. Passenger/vehicle and crash data were investigated to produce crash involvement rates per vehicle-mile of travel (Massie et al., 1995). They observed that the fatality rates were elevated for drivers aged over 75. Similarly, the effects of aging on driving were examined for the injuries in single-vehicle crashes by Islam and Mannering (2006), which shows that aging drivers had higher crash and fatality ratios with respect to other age groups. However, none of these studies specifically focused on crashes involving aging drivers, using GIS-based models and statistical methods.

3. Methodology

The main objective of this paper is to understand the unique nature of the crashes involving aging (65+) drivers, using the following approaches: (a) a geo-spatial analysis of crashes involving 65+ drivers in order to identify high risk locations, and (b) a statistical analysis to identify the significant factors that affect those crashes. A descriptive flowchart displaying the overall geo-spatial analysis methodology is provided in Fig. 1. This methodology was applied on three urban counties in Northwest Florida: Leon, Bay and Escambia counties. These counties are documented by the Safe Mobility for Life Coalition in Florida to have high number of crashes involving aging drivers compared to their populations (Sandler et al., 2015).

3.1. Demographics data

Aging demographics data was obtained from the U.S. Census Bureau (2015) databases. Currently, 65+ population constitutes 14% of total population in U.S. and 18% of total population of Florida (Administration on Aging, 2014; Bureau of Economic and Business Research, 2014). Moreover, the aging population growth is faster in Florida. Consequently, the number of aging road users and crashes involving aging drivers on Florida roadways will increase. This increase makes studies on crashes involving aging drivers even more critical. Table 1 and Fig. 2 show the population demographics of 65 years and older and 50–64 years old age groups in the Northwest Florida region. The studied counties are among the counties with highest amount of 65+ population. Persons 50–64 years old, also referred as Baby Boomers, are also of critical importance since the aging of the Baby Boom generation is expected to produce a 79% increase of the 65+ population in the next 20 years (Koffman et al., 2010).

3.2. Roadway crash data

Roadway network and crash data (2008–2012) were obtained from TIGER Geodatabase of U.S. Census Bureau (2010) and Florida Department of Transportation databases (FDOT, 2015), respectively. In spite of the availability of a substantial number of attributes within the crash database, not all of them were suitable to be candidate variables for the regression analysis. Therefore, the available attributes were investigated thoroughly to discover more critical ones for the crashes involving aging drivers. Potential explanatory variables were identified by examining crash metadata tables and conducting statistical investigations, supported by the literature review disclosing the most significant attributes used in the crash analysis. Many variables were condensed from multivalued to binary, which was found to improve the accuracy of the analysis. For instance, for the road surface condition variable, several similar and even overlapping effects, such as wet, slippery, and icy, were combined into one "negative" effect ("1" in the analysis), and labelled slippery. The selected binary continuous variables are presented in Table 2.

Note that every crash in the database is represented by a point that has these listed attributes. To make the proposed approach possible, crashes were grouped into three main categories based on age of drivers: (a) teen drivers (16–24), (b) aging drivers (65+) and (c) other drivers (25–64). This grouping made it possible to create density maps for each age group, and consequently to make a comparison possible between different age groups.

A data matrix is given in below Table 2 in order to show the number of crashes in each age group for studied counties. Note that every crash has been taken into account in the analysis including both single and

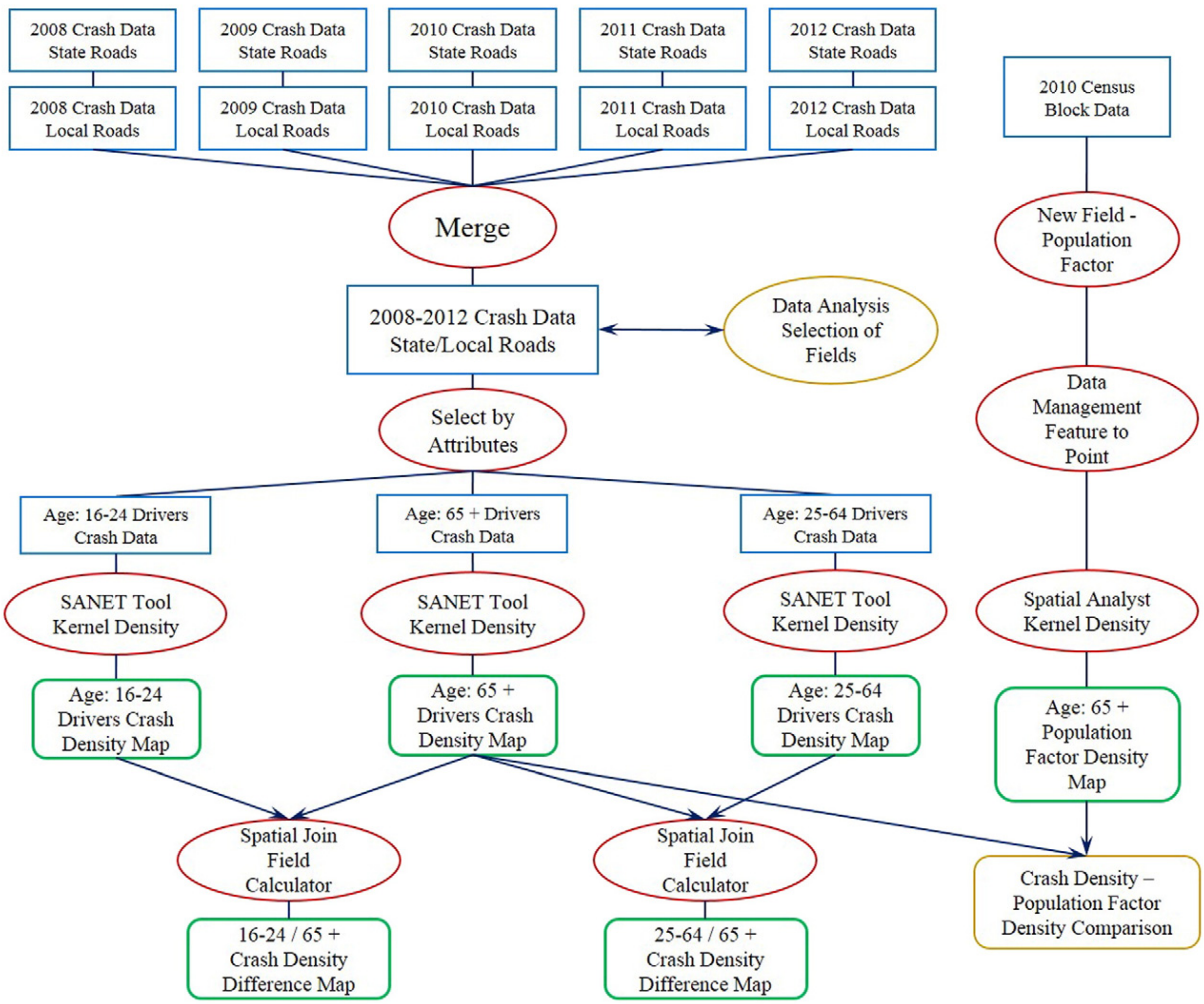

Fig. 1. Geo-spatial crash density analysis methodology.

multiple vehicle crashes as well as the aging driver-involved vehicle crashes with pedestrians and bicyclists. However, since the focus of this study is the aging (65+) drivers, we did not consider the crashes that only involve 65+ pedestrians and/or bicyclists where the driver of the vehicle belongs to another age group. As such, investigating those crashes involving only 65+ pedestrians and/or bicyclists is beyond the scope of this work. However, extending this research towards including both 65+ occupants in a vehicle as well as 65+ pedestrians and bicyclists is a very interesting future research direction. To avoid undercounting and miscalculations, the following approach was implemented: a) all crashes involving at least one aging driver were included in the 65+ crash density analysis, b) all crashes involving at least one driver aged 25–64 were included in the 25–64 crash density analysis, c) all crashes involving at least one driver aged 16–24 were included in the 16–24 crash density analysis. Note that if a crash involved both aging and teen drivers, we regarded this crash both as an aging crash and as a teen crash. Therefore, the same crash was included separately in both aging and teen crash density maps. This was necessary since excluding that crash from one of the maps would result in misleading density maps. Crashes with parked cars were also included; however, such vehicles are typically coded as driverless, and thus only the driver of the moving vehicle would be included in the data set.

Table 1
Population figures for 50–64 and 65+ age groups in Northwest Florida (U.S. Census Bureau, 2010)

County	65+ Population	65+ Percentage	50–64 Population	50–64 Percentage
Bay	24,559	14.5%	33,830	20.0%
Calhoun	2258	15.4%	2884	19.7%
Escambia	42,929	14.4%	58,661	19.7%
Franklin	2015	17.4%	2506	21.7%
Gadsden	6323	13.6%	9936	21.4%
Gulf	2587	16.3%	3408	21.5%
Holmes	3425	17.2%	4097	20.6%
Jackson	7801	15.7%	10,248	20.6%
Jefferson	2432	16.5%	3614	24.5%
Leon	25,980	9.4%	46,886	17.0%
Liberty	888	10.6%	1545	18.5%
Okaloosa	25,218	13.9%	35,370	19.6%
Santa Rosa	19,460	12.9%	30,391	20.1%
Wakulla	3339	10.8%	6336	20.6%
Walton	8943	16.2%	12,253	22.3%
Washington	3827	15.4%	5022	20.2%

Note: The data was retrieved from U.S. Census Bureau, 2015. 2010 US Census Blocks in Florida. Florida Geographic Data Library. (URL http://www.fgdl.org/metadataexplorer/explorer.jsp.) (accessed 15.1.15).

3.3. Geo-spatial crash density analysis

The most hazardous locations for each age group were detected by using age group-specific geo-spatial crash density maps. These maps were created by network distance-based KDE analysis, which was implemented via the SANET tool (Okabe and Sugihara, 2012) in ArcGIS. KDE is a non-parametric method to estimate probability density function of a random variable, which was crash points in the case of this

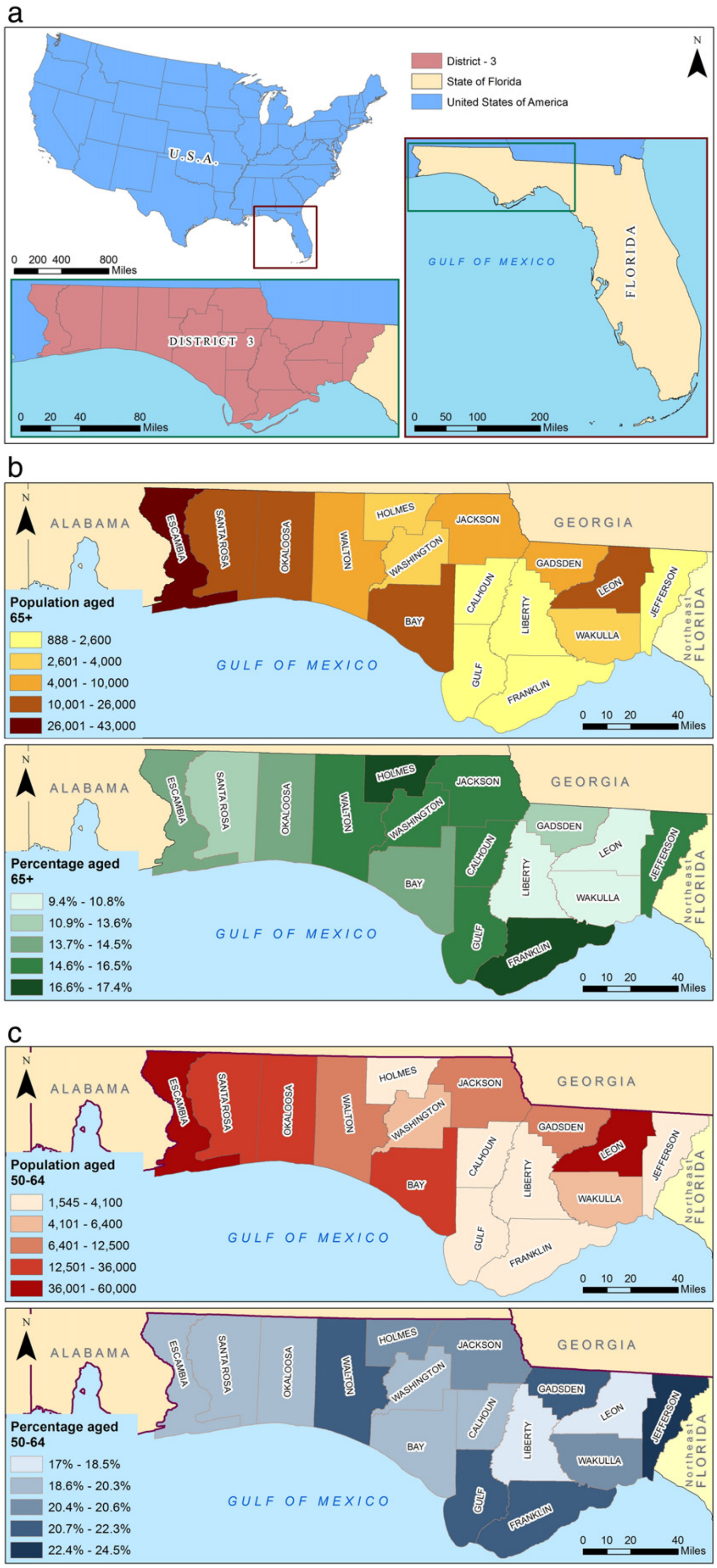

Fig. 2. Location of study area and comparative demographics of Northwest Florida: (a) Study area (b) Population counts and percentages for 50–64 (c) Population counts and percentages for 65+ age groups.

research. For this analysis, a 300 m bandwidth has been selected. Please remind that, generally, in an urban network, bandwidths with a range of 50 m to 300 m are considered to be appropriate (Steenberghen et al., 2010; Xie and Yan, 2013), whereas 1000 m was suggested for rural regions (Blazquez and Celis, 2013). Since the studied counties also have more lightly populated areas and less congested roadway networks than other big metropolitan areas such as Miami and Tampa, Florida, selection of 300 m was appropriate for these counties. Moreover, the SANET tool requires a selection of "cell width", which has been chosen to be 30 m based on the suggestion (one-tenth of the bandwidth) presented in the Okabe and Sugihara (2012).

Network distance-based type of KDE approach was chosen due the following reasons: (a) This is a very effective method to conduct density analysis for networks that possess spatially distributed data points along the roadways, and (b) the distance between individual crashes is calculated using the actual roadway network distance rather than the Euclidean distance (Dai et al., 2010; Okabe and Sugihara, 2012; Okabe et al., 2009; Steenberghen et al., 2010; Xie and Yan, 2013). This is particularly important because of the overestimation and underestimations problems that can be caused by the planar KDE approach.

3.4. GIS-based crash density ratio difference

A comparative crash density analysis, which was first developed by Ulak et al. (2015), was conducted to observe the aging-specific spatial patterns and to detect age group-specific hot spots. A comparison based strictly on the number or density of crashes would disguise age-group specific hot spots due to the uneven number of crashes between different age groups (Abdel-Aty et al., 1998). Therefore, the "density ratio difference" (DRD) parameter was developed. The DRD is the difference between the maxima-normalized crash densities for two different age groups. The formula of DRD is shown in Eq. (1).

$$DRD_{ij} = \frac{D_i}{max(D_i)} - \frac{D_j}{max(D_j)} \tag{1}$$

where DRD_{ij} is the "density ratio difference" between the compared maps i and j whereas D_i and D_j are the density values of the corresponding roadway sections, and $\max(D_i)$ and $\max(D_j)$ are the maximum crash density values of the compared maps, respectively.

Based on this approach, the aforementioned age groups' crash density maps were normalized with their own maximum density value. Dividing each crash density map by its own maximum density ensures that the maximum values would be equal to 1 on every map. In other words, the highest value of normalized density for each map (16–24, 25–64 and 65+ drivers) is equal to 1, whereas the lowest value of normalized density is equal to 0. The resultant normalized values of crash densities convey a clear illustration of the scale of the crash density at a location since these values reflect a percentage with respect to the maximum crash density. Moreover, obtaining density ratio values between 0 and 1 in all maps is also necessary to be able compare the crash density maps of two different age groups. Following the calculation of normalized crash densities, crash density maps of 16–24 and 25–64 age groups were subtracted from the crash density map of 65+ drivers individually. After normalization, if the crash density of a specific location in aging drivers' was equal to the crash density of that location in the map of the other age group, that location did not reflect an aging-specific spatial significance. On the contrary, that location would have common spatial characteristics in terms of crash frequencies for both age groups. Therefore, subtracting two values from each other revealed the relatively different locations on the network.

The GIS maps of the density ratio difference function are useful in graphically exploring the regions with a large difference in crash densities. As an objective test for the statistical significance of the difference, we conducted a non-parametric Kruskal-Wallis test to investigate whether two normalized crash densities were significantly different from each other. For this purpose, we randomly selected 100 points on the roadway network, and then compared normalized crash density of an age group at those points with the normalized crash density of other age group (i.e. 65+ drivers). The number 100 was selected based on the results of Moran's I spatial autocorrelation analysis to find the maximum sample size resulting in non-significant spatial correlation between the crash densities at roadway locations. A total of six comparison results were reported: 16–24 vs. 65+ and 25–64 vs. 65+ for Leon County, Bay County and Escambia County. The Kruskal-Wallis test provided information on the statistical significance of the difference between the normalized density values at the 100 locations based on the null hypothesis that two normalized density values are originated from the same distribution. A Monte Carlo study was also conducted to repeat the Kruskal-Wallis test 10,000 times for each age group comparison and the average p-value for the tests and the standard error of the p-value were reported.

3.5. Evaluation of the crash density and population relationship

Several studies show that crash occurrences and crash severity are affected by spatial factors such as urban development and geographic elevation (Delmelle et al., 2011; Effati et al., 2015). These factors can be evaluated in the context of spatial dependence and spatial heterogeneity. In this regard, we hypothesize that spatial distribution of aging population affects the likelihood of crashes involving aging drivers. This effect is due to the aging drivers' tendency to avoid traveling long distances (Collia et al., 2003), which implies that they might be involved in crashes more frequently close to their houses. This phenomenon would in turn increase the density and number of hotspots of crashes involving aging drivers in the proximity of areas with high number of 65+ populations. We developed a comparative evaluation approach, in order to visually and statistically explore if the crashes involving aging drivers and areas with high aging populations are correlated. For this purpose, a specific parameter, namely the "population factor" (PF) shown in Eq. (2), was developed. Population factor is basically a measure that incorporates the effect of the number of households with the 65+ population counts and percentages in each population block.

$$PF_{ij} = \frac{A_{ij}}{H_i} * \frac{A_{ij}}{\sum_i A_{ij}} * 10,000 \tag{2}$$

where PF_{ij} is the "population factor" of the age group j for the census block i, A_{ij} is the number of people that belong to jth age group in the ith census block, H_i is the number of households in the ith census block, and $\sum_i A_{ij}$ is the total population count that belongs to jth age group (a county for the case study presented). A multiplication factor of 10,000 is used to avoid substantially small PF values.

We needed this new factor because neither number of people nor percentage of these people in the block is adequate to properly reflect the possible effect of the population on the crashes. Evaluating only 65+ population count would disguise those sparsely populated blocks that have high percentage of aging residents. On the other hand, evaluating only the aging population percentages (65+/Total Population) would disguise those highly populated blocks that have relatively fewer aging residents.

We proposed that using the number of households is more appropriate than using the total population count for a traffic crash-focused analysis due to the following reasons: First, number of households is a better indicator of the total travel generated in a census block than the total population count since the latter also includes non-drivers and minors. Indeed, U.S. Department of Transportation and U.S. Census Bureau assess the travel behavior and commuting trends based on the households per block rather than the population in the block (FHA, 2011; McKenzie and Rapino, 2011).

Second, a 2012 report of U.S. Census Bureau (2012) about households and families in 2010 indicates that aging (65+) people live in

Table 2
Data matrix: (a) Number of crashes in each age group (FDOT, 2015), and (b) Independent variables used in the regression analysis.

Number of crashes in each group						
Crash data matrix	Leon County		Bay County		Escambia County	
Driver age	Total crashes: 24,020		Total crashes: 14,892		Total crashes: 24,527	
	Count	%	Count	%	Count	%
16–24	11,702	49%	5842	39%	9989	41%
25–64	18,586	77%	12,391	83%	20,477	83%
65+	2434	10%	2306	15%	4060	17%

Independent variables for the regression analysis

Binary variables: "Day of Week" (1:Weekend, 0:Weekday), "At Peak Hour" (1:Yes, 0:No), "Alcohol-Drug Abuse" (1:Yes, 0:No), "Intersection Presence" (1:Yes, 0:No), "Traffic Control Unit Presence" (1:Yes, 0:No), "Work Zone Presence" (1:Yes, 0:No), "Weather Condition" (1:Bad, 0:Clear), "Light Condition" (1:Night, 0:Else), "Road Condition" (1: Defected, 0:Good), "Road Surface Condition" (1:Slippery, 0:Dry), "Visibility" (1:Bad, 0:Clear) and "Lane Departure Action" (1:Yes, 0:No).
Continuous Variables: "AADT", "Speed Limit" and "Aging Population Factor Density".

Total crash number of 16–24, 25–64, and 65+ age groups indicates crashes involving at least one 16–24, at least one 25–64, and at least one 65 driver, respectively. Summation of crash numbers does not correspond to total crash number due to overlaps in between age groups.

households with 1.4 members on average, and people aged 64 and under live in households with 3.04 members on average, in the State of Florida. This data shows that it is more likely for aging (65+) residents to stay in small households comprising of 1 or 2 dwellers. On the other hand, it is more probable that teens, young adults, and middle aged residents live in larger households. Therefore, using the number of households in the block instead of the population per block prevents us from underestimating the effect of aging populations since the influence of their travels on crashes is not excessively reduced (which would have been the case if we had divided by the population count).

Third, there are also a group of aging residents who prefer to live in independent senior living communities where usually only 65+ residents stay. (Note that nursing homes and assisted living facilities were excluded in this study, because 65+ residents in these facilities usually do not drive due to their medical, physical and cognitive limitations). Due to the uneven number of people per household figure mentioned above (1.4 vs. 3.04), the population of these senior communities remain lower than the population of a regular community when both communities have the same number of households. Because of this difference in dwelling types, the household choice while calculating the population factor also prevents overestimating the effect of senior living communities since the influence of other regions are not unduly reduced.

It is critical to note that the population density maps developed herein are not akin to the common density maps generated via dividing the population figure by the area. The proposed PF methodology is unique in the sense that it can be used to estimate the possible effects of the population living in the vicinity on the roadway crashes through the density analysis.

Planar kernel density estimation (KDE) was adopted as the density analysis method to create the PF density map for aging residents. We aimed to achieve two goals by implementing a density analysis rather than using the areal census data directly. First, the spatial relationship between neighboring blocks was also taken into account by adopting the density analysis method. Without the density analysis, this spatial relationship would be disregarded, and the effect of census blocks in the vicinity of a crash might be found by averaging or summing their population values, which would not reflect a joint spatial effect. Second, the effect of the census block area was intrinsically taken into account by the density analysis in terms of the distance between the block centroids. A larger block will provide a larger distance between the centroids, and therefore a lower contribution to the density. Thus, rather than simply dividing the number of aging population to the census block area, the distance parameter was also taken into the account as a representative measure of the spatial relationship. Note that, different KDE parameters were used for the PF density analysis than those for crash density analysis. For this purpose, a sensitivity analysis was conducted based on the bandwidth values ranging from 500 m up to 5000 m. At the end, 2000 m was selected as the bandwidth for the PF density analysis, and 10 m for the cell width. However, please note that, there is no exact procedure available for determining the optimum bandwidth.

Consequently, each crash point was assigned a corresponding PF density value as a new attribute for statistical analysis purposes. Note that these values for 65+ were assigned to all crash points whether that point was related to a crash involving aging drivers or not, since the objective was to evaluate the spatial effect of aging population on the occurrence of the crashes through a regression analysis. As a result, all crash points were given a PF density value, which led to using PF values as one of the factors for the aging-focused regression analysis.

3.6. Regression analysis

In this section, aging-focused logistic regression analysis will be presented in order to comprehend the effects of different factors on crashes involving aging drivers, including the time-based factors such as the time of the day as well as the aging PF values that were assigned to the crash points. For this purpose, a regression analysis was proposed based on the selected attributes, which enabled us to interpret the significance of each factor on the likelihood of crashes involving aging drivers. There is a vast literature available on the generalized linear regression models (Maddala, 1986). Please refer to Lord and Mannering (2010) and Mannering and Bhat (2014) for a review of these models. Unlike classical regression analysis where response and predictor variables are continuous, binary response and predictor variables can be modeled by the binary choice models. Since the crash database is mostly composed of binary attributes, widely used logistic regression was implemented for the regression analysis. The log-likelihood function that was solved for estimating coefficients of the predictors of the logit model can be written as follows:

$$lnL(\boldsymbol{\delta}) = \sum_{i=1}^{n} \{ J_i * [\Psi(\boldsymbol{X}_i\boldsymbol{\delta})] + (1-J_i) * [1-\Psi(\boldsymbol{X}_i\boldsymbol{\delta})] \} \quad (3)$$

where J_i is the response variable that has a binary value (0 or 1), $\Psi(\boldsymbol{X}_i\boldsymbol{\delta})$ is the cumulative distribution function of logistic function, and $\boldsymbol{X}_i$ is the row vector of predictors for ith observation, and $\boldsymbol{\delta}$ is the vector of coefficients of the predictors. Note that, the response variable, $J_i$is equal to 1 if the crash involves an aging driver, or 0 otherwise.

4. Case study application results

In this section, a case study application is presented for Leon, Bay and Escambia counties, respectively. With this application, we aim to achieve the following: (a) to conduct a comparative crash density analysis, (b) to identify the spatial crash patterns for different age groups, (c) to present this knowledge in terms of visual illustrations, and (d) to conduct a logistic regression analysis in order to determine the significant factors that differentiate the crashes involving aging drivers from those involving other age groups.

4.1. GIS-based visual illustrations

4.1.1. Crash density maps

Resultant crash density maps in Fig. 3 illustrate the geo-spatial crash patterns of three different age groups, namely teenage drivers (16–24), adult drivers (25–64) and aging drivers (65+) for Leon, Bay and Escambia counties, respectively. These counties also include the three major cities in the region, namely Tallahassee, Panama City and Pensacola. Note that we conducted the analyses for the entire Escambia, Bay and Leon counties, although maps illustrate the densely populated regions only in order to show a better picture of the high risk locations. Fig. 3 shows that the crash hot spot locations are substantially different for each age group. This difference is not equally explicit

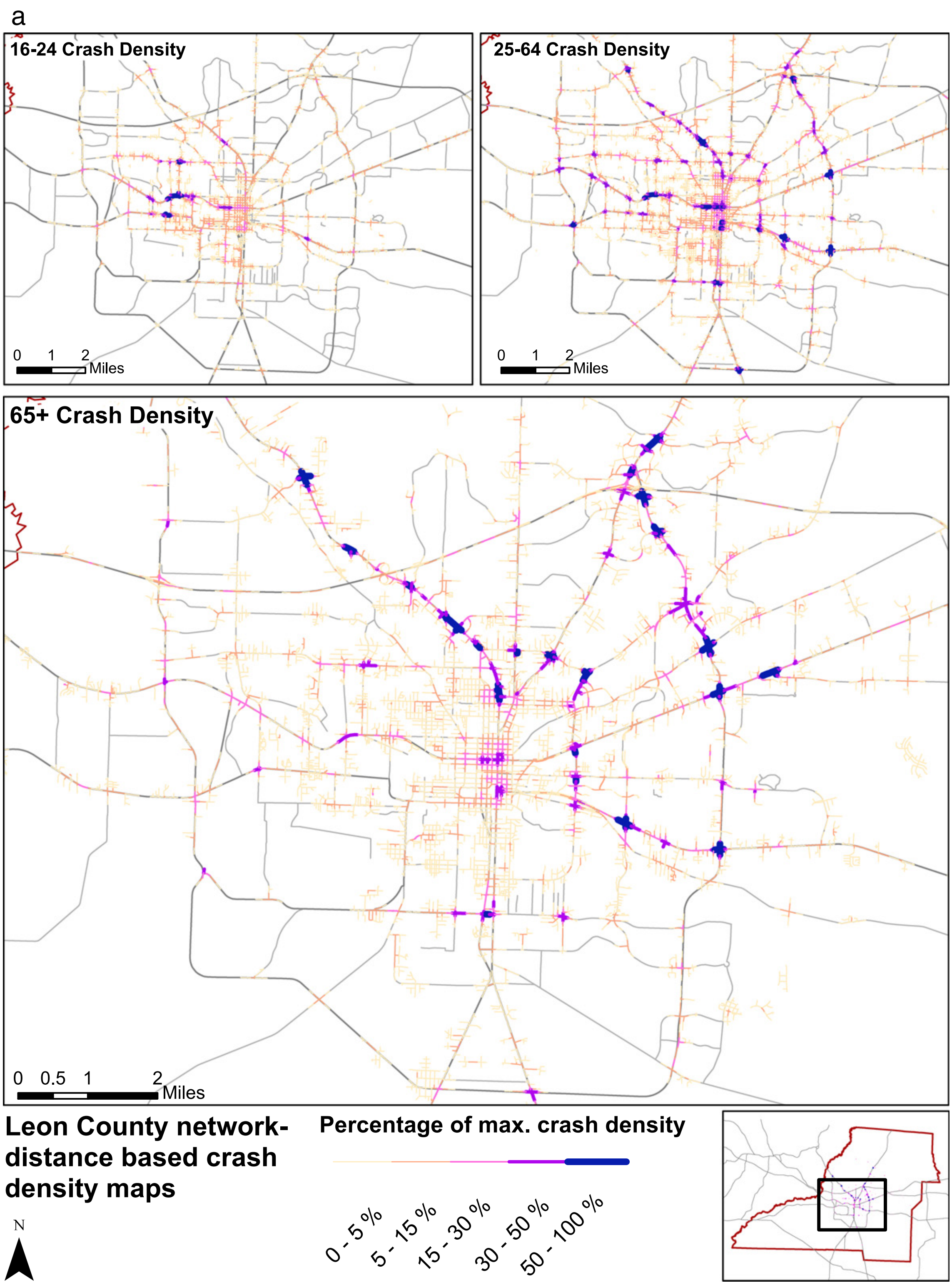

Fig. 3. Crash density maps (2008–2012). Left: 16–24 age, Middle: 25–64 age, and Right: 65+. a) Leon County, b) Bay County, c) Escambia County.

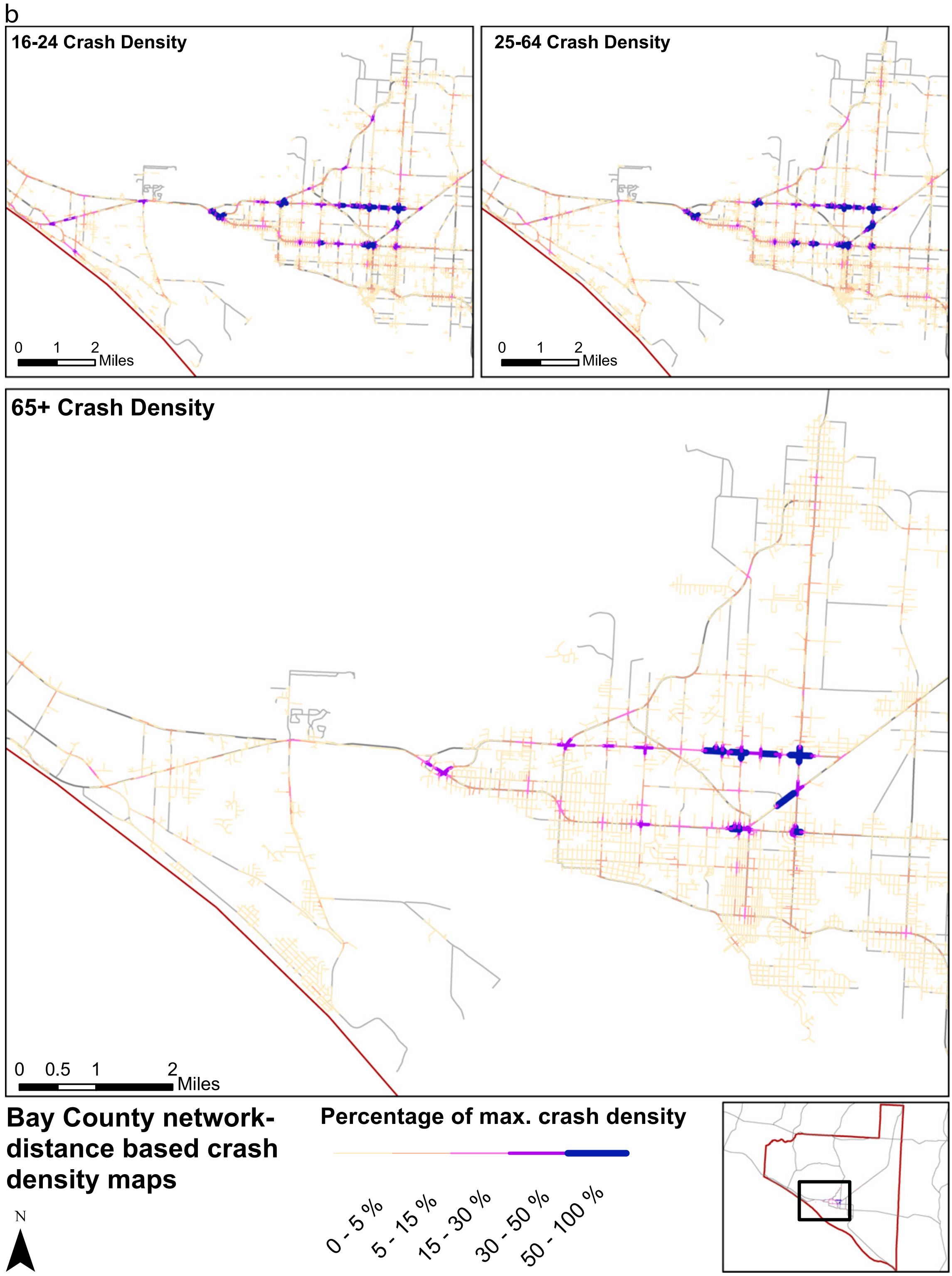

Fig. 3 (*continued*).

in all networks because of the unique population and roadway characteristics of these networks. However, age specific crash patterns are still observable on all three networks. For example, in Escambia County (Fig.3. c), crashes involving aging drivers are clustered more in the northwest of Pensacola than in other parts of the county. On the other hand, clusters of crashes involving other age groups, particularly those involving the driver group aged 25–64, are substantially closer to the Downtown Pensacola. Visual inspection of these crash density maps is critical in order to create a primary perception of age-specific crash hot spots, yet a robust comparison analysis is necessary to recognize the geo-spatial patterns of crash clusters, in which aging drivers were involved.

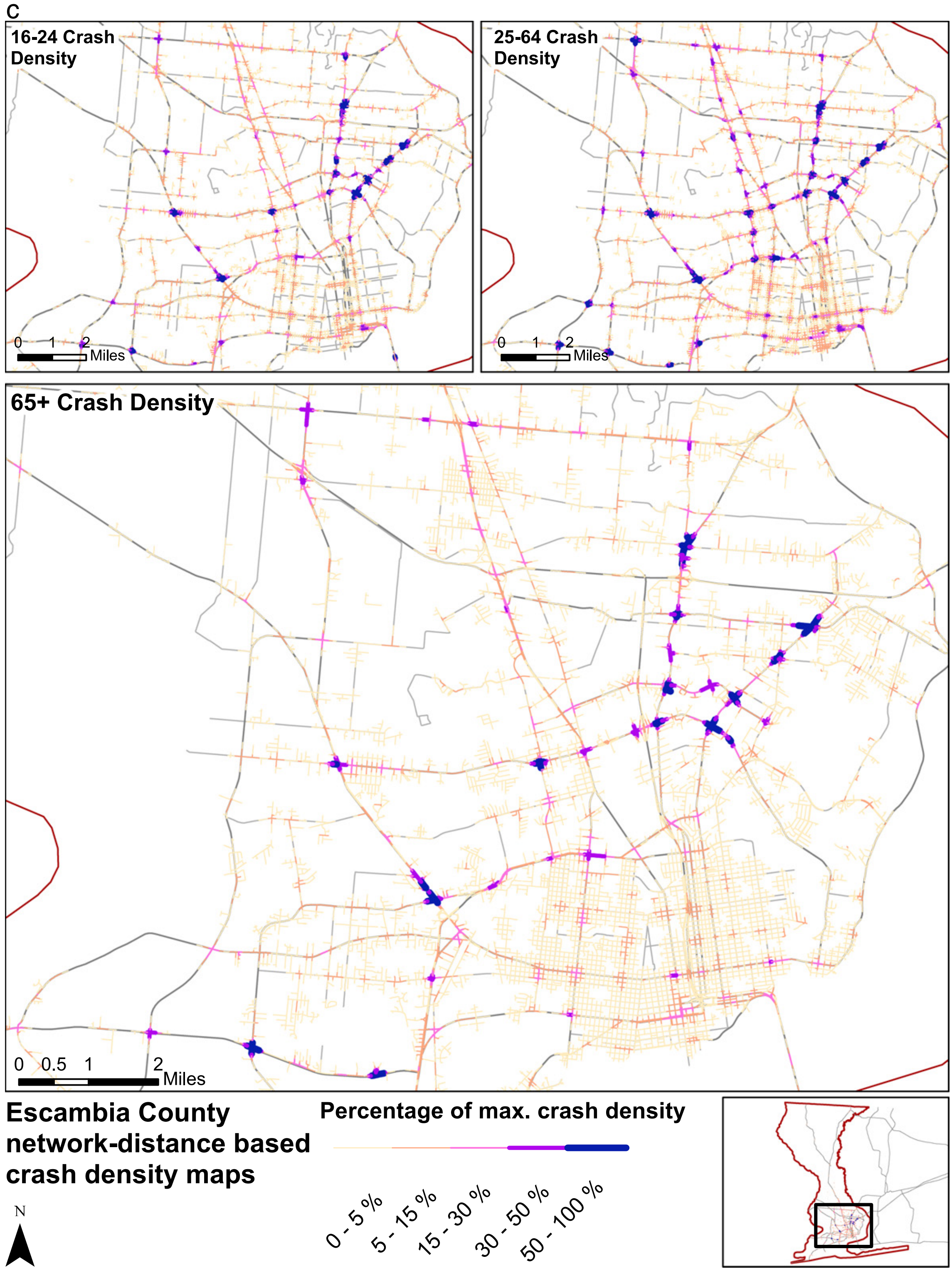

Fig. 3 (*continued*).

4.1.2. Crash density ratio difference (DRD) maps and statistical significance analysis

Figs. 5, 6 and 7 show density ratio difference maps created according to the approach explicitly described in the Methodology section. The resultant numerical value of this analysis is the difference of the normalized crash densities, which can also be expressed as a percentage difference. Note that the density ratio difference provides visually exploratory results rather than providing statistical significance. Therefore, we conducted the Kruskal-Wallis test to investigate whether two normalized crash density maps are significantly different (detailed explanation for this test is provided in the Methodology section). The Kruskal-Wallis test assumes that the observations are independently distributed. Since the normalized density values from the map were spatially autocorrelated, we could not use the entire data in performing the test. To eliminate the spatial autocorrelation, permutation resamples were taken without replacement from the density profiles with the sizes of 1000, 500, 250, 100, 50, and the Moran's I coefficient was calculated for each resample. Since small resampling sizes correspond to large pairwise spatial distances between the resampled points, they are expected to reduce spatial autocorrelation. Moran's I coefficient was used to test whether there was tendency towards clustering with positive values or dispersion with negative values (Getis and Ord, 1992). The *p*-value of the coefficient was used to determine whether we can reject the null hypothesis that states that feature values are randomly distributed across the study area. The Moran's I results from the resampling tests are shown in Table 3, which shows that a resampling size of about 100 is sufficient to overcome spatial autocorrelation. This indicates that the maximum sample size satisfying non-significant spatial correlation between crash density values is equal to 100.

The Kruskal-Wallis (KW) test was used to test the null hypothesis that the two crash density groups have identical distribution functions (Kvam and Vidakovic, 2007). We applied the KW test to the resampled normalized densities for Escambia, Leon and Bay counties. Fig. 4 shows the cumulative distribution functions of one of the samples (size of 100) for comparing age groups 16–24 and 65+ and comparing the 25–64 and 65+ age groups. The titles of the figures show the *p*-values of the KW test, which measures the separation between the functions. It can be seen that the difference between 25 and 64 and 65+ age groups is more significant (smaller *p*-value).

A Monte Carlo simulation was conducted to repeat the KW tests 10,000 times and the average and the standard error of the average *p*-values are reported in Table 4. As it can be seen, there is a significant difference between the normalized crash density values of the 16–24 age group and the 65+ age group and between the 25–64 age group and to 65+ age group at the three counties (at the 0.05 significance level). The confidence intervals obtained from the Monte Carlo simulation indicates that the *p*-values are statistically significant at the 95% confidence level, with the largest difference observed in Leon and Bay Counties, between the 25–64 and 65+ age groups.

As a result of DRD analysis, we obtained six maps for the three selected counties based on the following age groups: 16–24 versus 65+ (Figs. 5a, 6a, 7a) and 25–64 versus 65+ (Figs. 5b, 6b, 7b). On these maps, DRD values were classified according to their standard deviations that illustrate the variation of the values from the mean. Figs. 5, 6 and 7 indicate that every network has unique age-specific crash pattern, and each age group-involved crashes occurred more frequently at separate locations when compared to crashes of other groups. For instance, crashes involving aging drivers are more clustered at the northeastern part of Tallahassee, Leon County (Fig. 5), whereas crashes of other age groups are concentrated relatively in the central or slightly western part of the city. Note that crashes involving 16–24 old drivers are mostly in the vicinity of the Florida State University. Similarly, in Bay County (Fig. 6), we observe more teenage, young and middle adulthood crashes on the Panama City Beach connector roadways when compared to older adult crashes. The comparative analysis using the DRD method provides a much more explicit means of identifying geo-spatial differences in age-related crash patterns than the crash density approach alone.

4.1.3. Crash density and 65+ population factor (PF) density maps

The effect of 65+ population on geo-spatial characteristics of crashes involving aging drivers can be investigated via the comparative inspection of 3-D map that shows aging PF density and network-based density of crashes involving aging drivers together (Fig. 8). Since the focus is only on the 65+ populations in this section, density ratio difference maps are not used. Briefly, grayscale-colored base map represents the aging PF density where the darker areas indicate higher PF densities. On the other hand, the 3-D distribution represents the density of crashes involving aging drivers, where higher and redder peaks represent the higher crash densities.

This visual comparison provides a simple yet unique approach for evaluating the aging population effect on the crashes involving aging drivers. Fig. 8 shows that crashes involving aging drivers are more frequent in those highly 65+ populated areas than other locations such as the downtown. This pattern reveals an interesting aging-specific behavior since older drivers may prefer to visit those places such as grocery stores and pharmacies that are closer to their homes (Charness and Schaie, 2003; Staplin et al., 2012). The visual comparison of aging driver-involved crash densities and the spatial distribution of aging populations provides a strong indication of the spatial dependency effect. This effect was also observed through the regression analysis presented in the following section. However, note that some of these crashes involving aging drivers may not be merely due to those aging people living close by. Especially for the interstate highways such as I-10, crashes may involve aging nonresident travelers from other counties and states.

Table 3
Results of Moran's I spatial autocorrelation tests for testing clustering of density of aging crashes. Tests were conducted by different random samples collected from spatial data of Leon County crash density analysis.

Random sample size	1000	500	250	100	50
Moran's index	0.106	0.046	0.164	0.061	0.020
z Score	13.060	3.955	7.037	1.223	1.051
p value	0.000	0.000	0.000	0.221	0.293
Distribution ($\alpha = 0.05$)	Clustered	Clustered	Clustered	Random	Random

4.2. Regression analysis

In this section, we conduct a regression analysis to identify the significant factors affecting the crashes involving aging drivers. Note that only crash points with Annual Average Daily Traffic (AADT) information were used, which corresponds to 90% of the total data points approximately. Moreover, Santa Rosa County was added to the analysis for Escambia County in order to avoid misleading results that might arise due to ignoring the major roadways that connect these two counties. The models for Leon, Bay and Escambia counties were found to be significant based on the substantially small *p* values (Table 5). For this analysis, "*Aging Population Factor Density*" was also included in the analysis as a continuous variable. Information on the predictors is provided at the end of Tables 2 and 5. Recall that the dependent variable of the model is a binary type, which is either equal to 1 if the crash involves an aging driver, or 0 otherwise.

According to the results, AADT increases the probability of a crash to involve an aging driver(s); however, for Leon and Bay counties, we do not have any significant evidence showing that an increase in the AADT causes more crashes for older drivers. Similarly, results indicate that higher speed limits increase the possibility of a crash to involve an aging driver(s) (insignificant for Escambia and Santa Rosa counties). This finding contradicts results from some previous studies arguing that crashes involving aging drivers were overrepresented when speed limits are lower (Baker et al., 2003; Bayam et al., 2005; McGwin and Brown, 1999). Hourly average speeds can be used instead of the posted speed limits in order to obtain more accurate results.

The "Day of Week" variable shows that aging drivers are less prone to having crashes during weekends than other age groups. This can be due to the fact that aging populations may prefer to drive and/or complete their daily activities like shopping mostly on weekdays rather than weekends. Similarly, highly significant "At Peak Hour" factor indicates that aging drivers tend to involve in less crashes than other groups at AM (morning) and PM (evening) rush hours. This finding confirms and supports the results of previous studies (Collia et al., 2003). These temporal results separate crashes involving aging drivers from crashes involving other age groups.

Results show that the spatial allocations of aging populations have a significant effect on the likelihood of a crash to involve an aging driver(s). Since the 65+ Floridians are mostly comprised of retirees, they may prefer to complete their daily activities within a reasonable distance (Charness and Schaie, 2003; Collia et al., 2003). Transportation

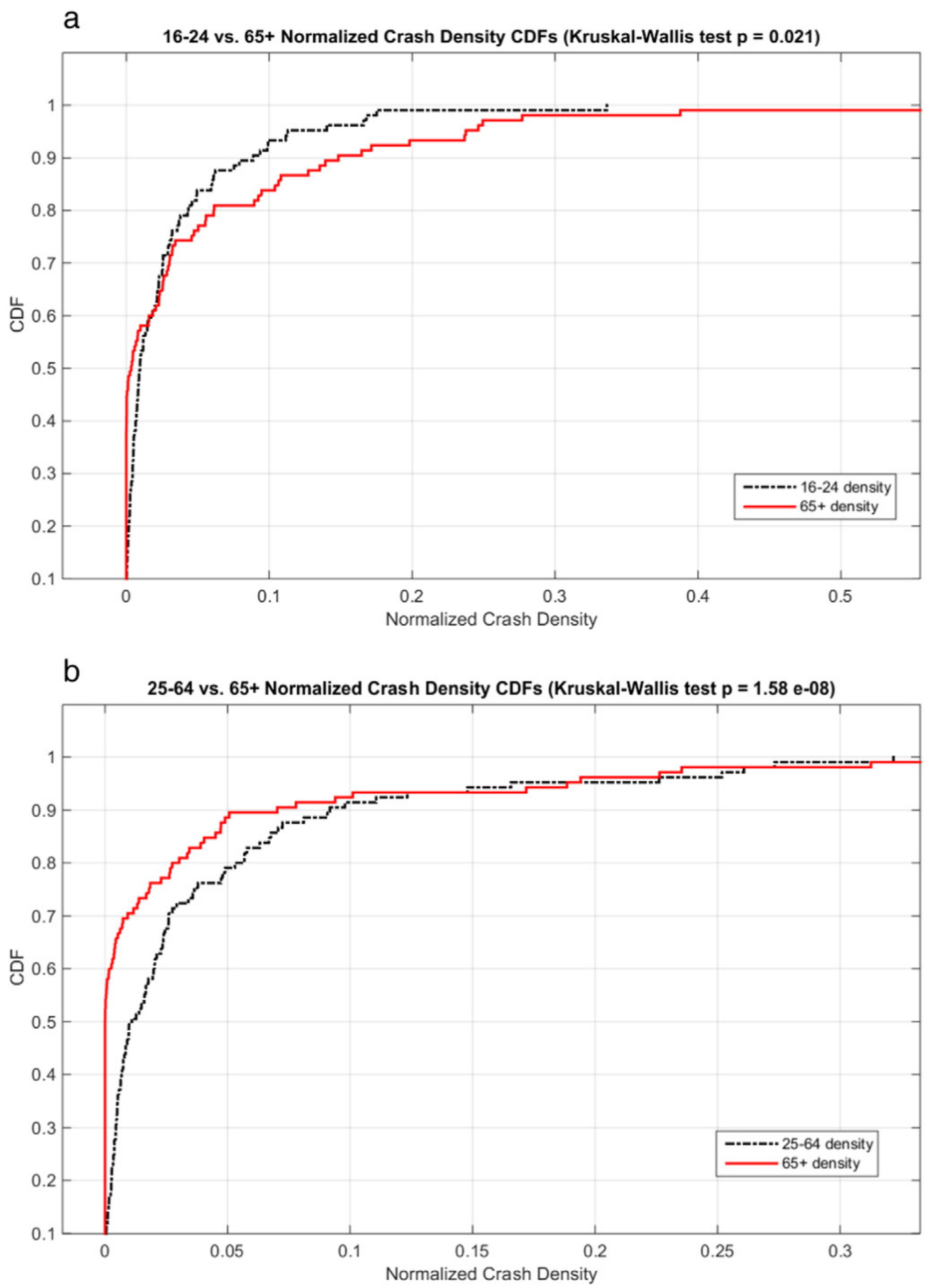

Fig. 4. Cumulative distribution function of the normalized crash density (for a sample size of 100) observed on the roadway networks in Leon County for age groups (a) 16–24 and 65+ and (b) 25–64 and 65+.

Table 4
Testing statistical significance of difference between two normalized crash densities: results of repeated Kruskal-Wallis tests from 10,000 Monte Carlo simulations. Null hypothesis is that two crash densities are from the same distribution.

Kruskal-Wallis test	Escambia County		Leon County		Bay County	
Age groups	25–64 vs. 65+	16–24 vs. 65+	25–64 vs. 65+	16–24 vs. 65+	25–64 vs. 65+	16–24 vs. 65+
$\overline{p}$	0.00042	0.01540	0.00003	0.04630	0.00023	0.00520
SEM	0.00002	0.00037	0.00000	0.00100	0.00002	0.00016
$\overline{p}+2\sigma/\sqrt{N}$	0.00038	0.01470	0.00002	0.04430	0.00020	0.00490
$\overline{p}+2\sigma/\sqrt{N}$	0.00047	0.01620	0.00004	0.04840	0.00027	0.00550

Abbreviations and formulas: $\overline{p}$: average p value ($\sum$ p/N), SEM: standard error of mean $\sigma/\sqrt{N}$, σ: standard deviation of p values, N: number of runs (10,000), $\overline{p} \pm 2\sigma/\sqrt{N}$: confidence intervals.

officials, especially those that operate the roadways close to senior communities, should be aware of the consequences of this high crash risk for aging populations on those roadways. One strategy that can help solving this problem is providing better information with regards to the design and operational characteristics of the critical and risky locations.

The strongly significant negative coefficient of "Alcohol-Drug Abuse" indicates that aging drivers are less prone to be involved in a crash while impaired by a substance (Abdel-Aty et al., 1999a; McGwin and Brown, 1999). "Intersection presence" and "Traffic Control Unit Presence" tend to increase the crash risk for aging drivers, implying the vulnerability of aging drivers to distractions and complex situations on the roadways. These results are also consistent with the findings of previous studies (Bayam et al., 2005; McGwin and Brown, 1999; Stamatiadis et al., 1991). The high crash risk at intersections is particularly important for older drivers because popular places like grocery stores and pharmacies are often located in the commercial areas that often contain

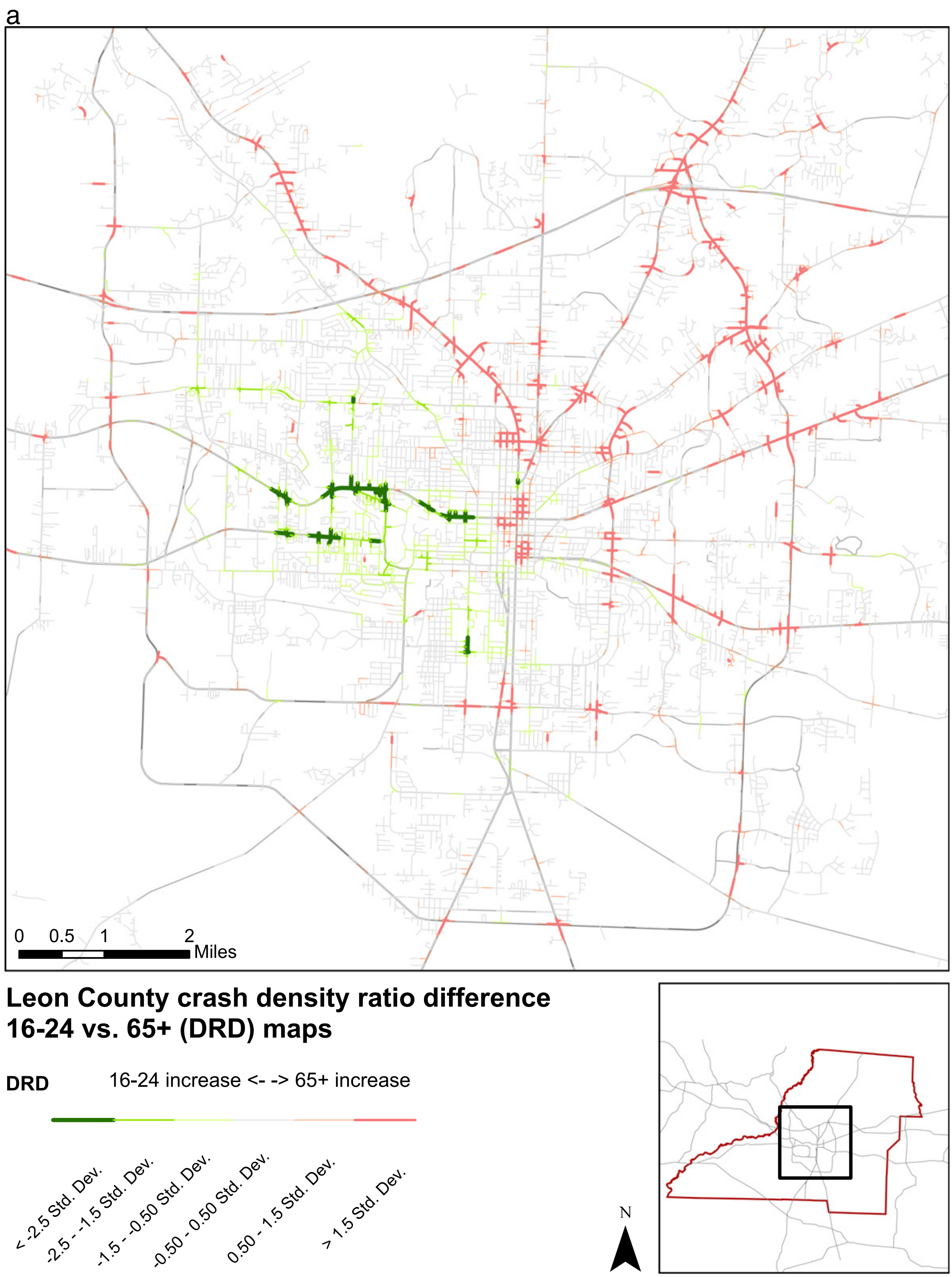

Fig. 5. Crash density ratio difference maps (2008–2012), Leon County. a) 16–24 versus 65+, and b) 25–64 versus 65+.

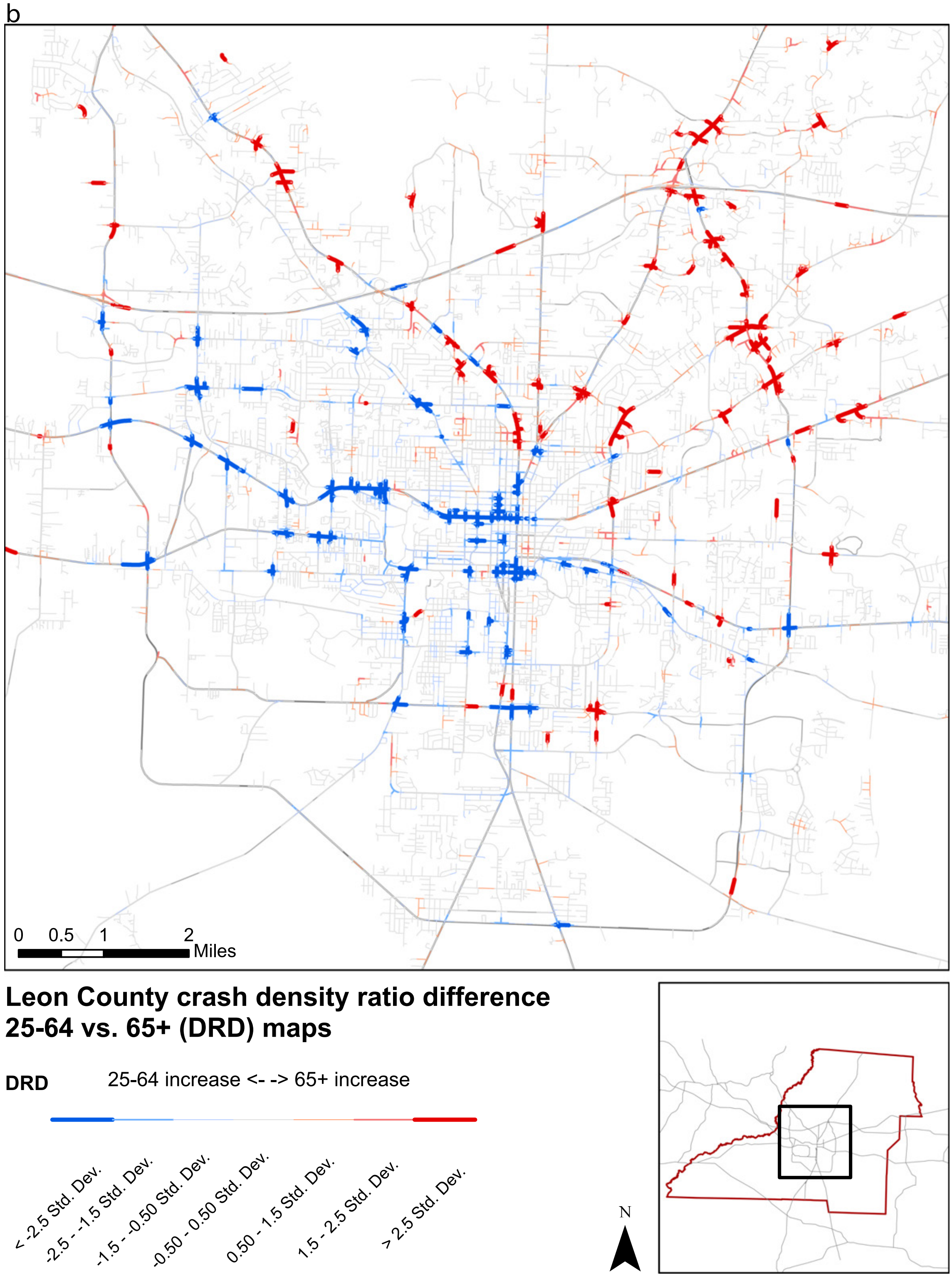

Fig. 5 (*continued*).

numerous intersections for access (Charness and Schaie, 2003; Staplin et al., 2012). Moreover, multiple intersections, complex signalizations and unexpected design features increase the complexity for aging drivers.

"Weather Condition" was not found to be significant, implying that adverse weather conditions affect both 65 + and 65- drivers similarly. "Light Condition," "Road Condition," "Road Surface Condition," "Visibility," and "Lane Departure Action" have a negative correlation with the crashes involving aging drivers. These results confirm and support findings of previous studies, which state that older drivers may prefer to avoid or do not prefer to drive at night, they may not necessarily drive on roads with defects or with slippery surfaces, they may prefer to drive with full visibility, and reckless driving is less common for aging drivers (Baker et al., 2003; Collia et al., 2003; Jang, 2006; Krahe and Fenske, 2002; McGwin and Brown, 1999; Rong et al., 2011).

5. Conclusions

In this study, we present a methodology to evaluate and analyze the roadway crashes involving aging drivers via simple yet novel techniques. These techniques can help officials determine the critical roadway segments and intersections for aging drivers, and understand the effect of spatial distribution of the aging population on the crashes involving aging drivers. Results indicate that the crashes involving aging drivers differ from other crashes both spatially and temporally. Density maps of crashes involving aging drivers have a unique geo-spatial pattern that differentiates them from density maps of the crashes involving other age groups. We observed that the "density ratio

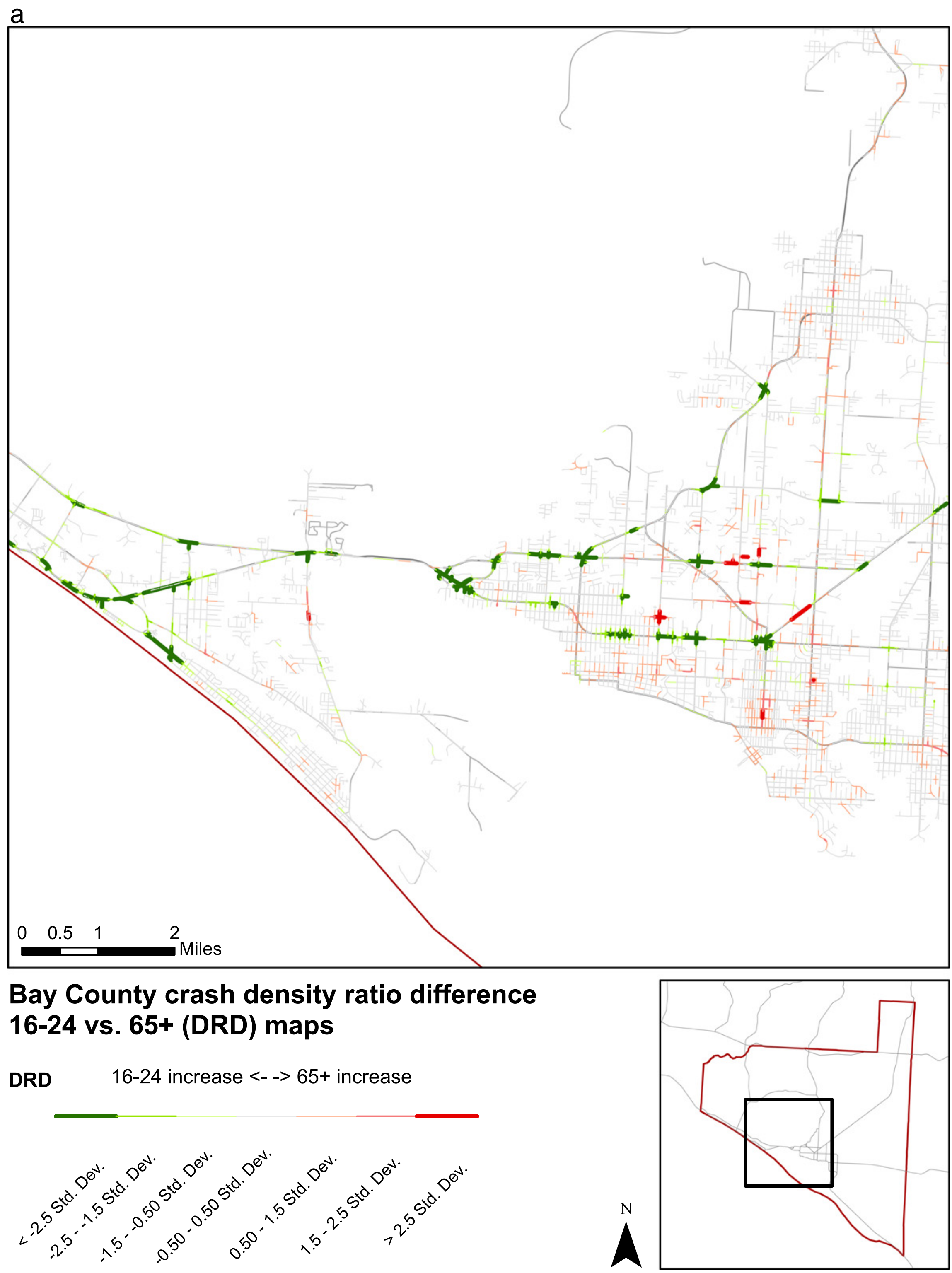

Fig. 6. Crash density ratio difference maps (2008–2012), Bay County. a) 16–24 versus 65+, and b) 25–64 versus 65+.

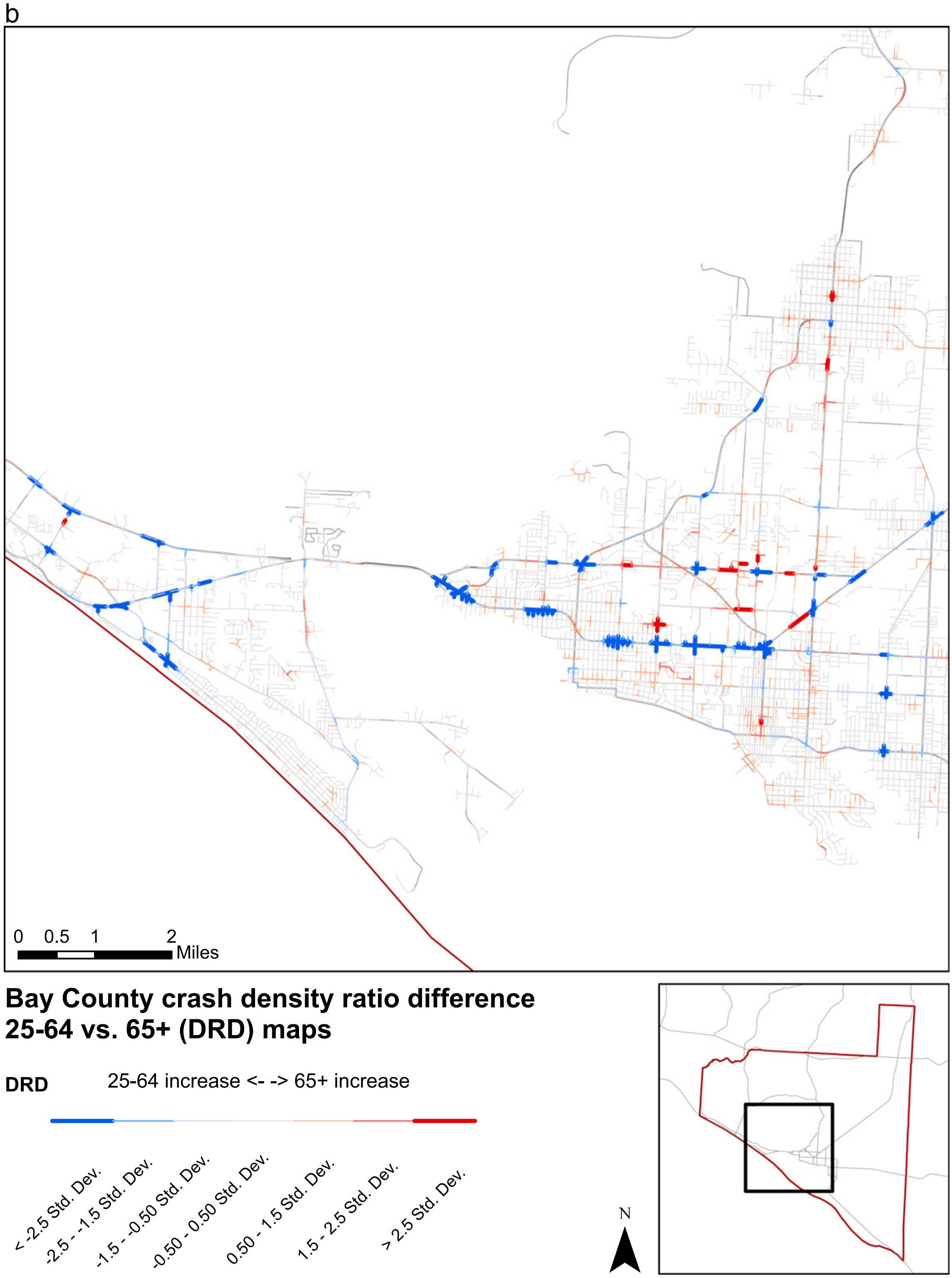

Fig. 6 (*continued*).

difference" method provides a more explicit way to visually identify the geo-spatial differences in age-related crash patterns than directly using the crash densities. We also developed the "population factor" parameter, which provides a more meaningful way to investigate spatial dependence, since neither number nor percentage of population is enough to entirely represent the actual effect of population on the crashes. Regression analysis results present a strong correlation between the spatial allocation of aging populations and crashes involving aging drivers. This correlation implies that the aging population living in the vicinity of a roadway can have a substantial effect on the likelihood of a 65 + driver crash on that roadway. This phenomenon can be related to the following aging population-specific behaviors consistent with the previous studies: a) aging people tend to avoid traveling long distances (Collia et al., 2003), and b) they may want to visit places like grocery stores and pharmacies with the most familiar routes and that are closer to their home (Charness and Schaie, 2003). The population factor approach can be very useful to identify the critical roadway sections that possess an elevated crash risk for aging drivers. Supporting the results of previous studies, key findings of the regression analysis reveal that the likelihood of crashes involving aging drivers is influenced by several factors such as demographic, social, spatial, temporal, behavioral, environmental, and traffic factors. However, we showed that while some factors are significant for every analyzed location, the significance of others depends on the specific characteristics of the location.

There are a number of limitations associated with this effort that suggest several areas of future work. First of all, some of the results of this research may be location specific and therefore may not be transferable to other locations, including those with a lower percentage of

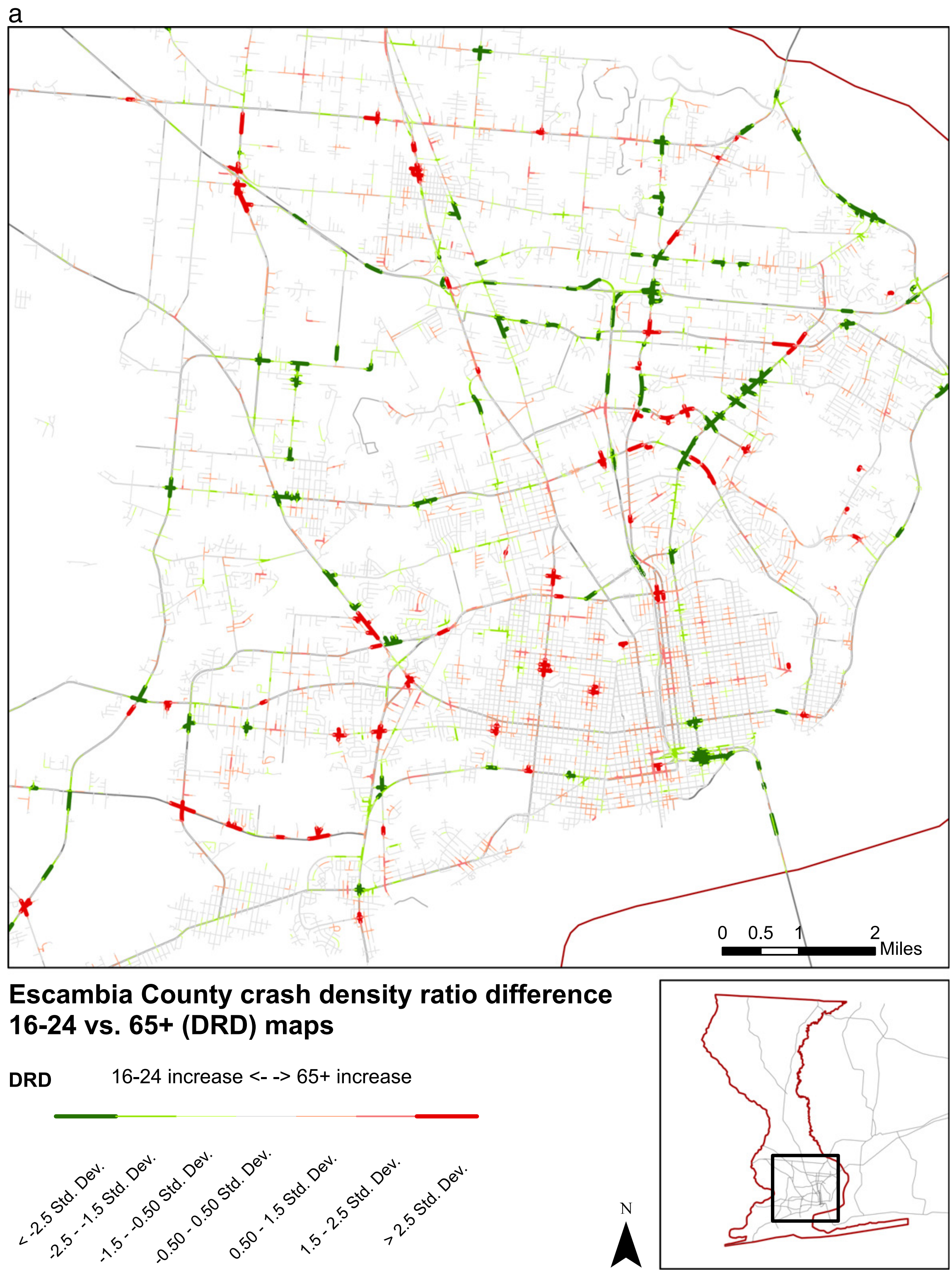

Fig. 7. Crash density ratio difference maps (2008–2012), Escambia County. a) 16–24 versus 65+, and b) 25–64 versus 65+.

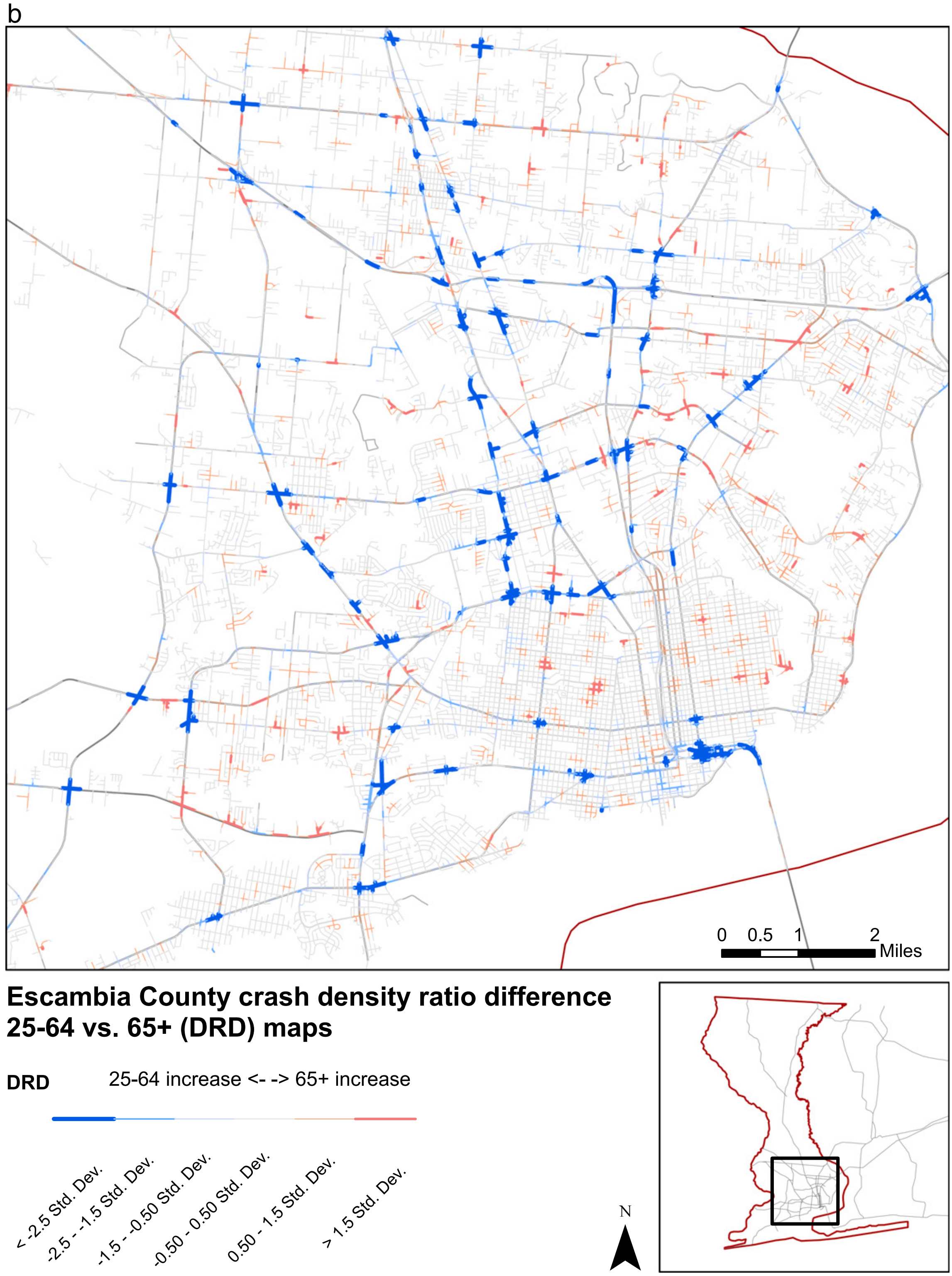

Fig. 7 (*continued*).

older adults. Therefore, a more comprehensive study is needed to successfully extend this analysis to other areas of Florida, and then elsewhere in the U.S. Second, a further differentiation between 65 and 79 and 80 + seniors would be useful, especially since different licensing procedures are employed for 80 + populations. Moreover, in the early periods of older adulthood (close to the age of 65), a substantial number of people may keep working instead of retiring, affecting their driving patterns. Third, this study focuses particularly on 65 + drivers. Therefore, an interesting future direction would be including the other roadway users in the analysis such as 65 + occupants, pedestrians, and cyclists in addition to the drivers. In this study, kernel density estimation is used to investigate the clustering behavior of the crashes.

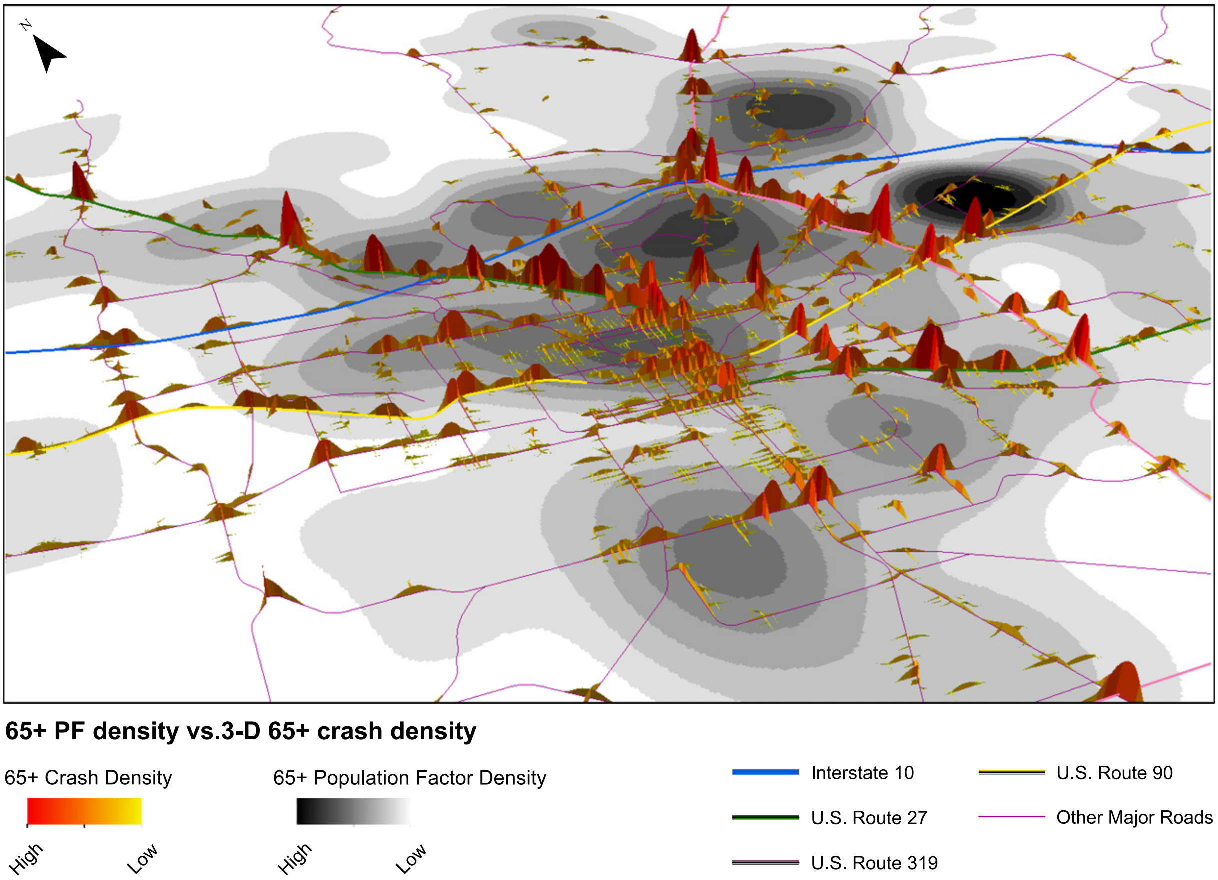

Fig. 8. Comparison of aging population factor density with 3-D density of crashes involving aging drivers, 2008–2012, Leon County.

However, there are other statistically robust methods for detecting the crash clusters such as Geti's Gi* and Local Moran's I, which was used in this study to find samples with no spatial autocorrelation. Extending the use of Local Moran's I for detecting the location of clusters and a comparative analysis of these methods can be a very interesting future work.

Acknowledgments

This project was supported by United States Department of Transportation grant DTRT13-G-UTC42, and administered by the Center for Accessibility and Safety for an Aging Population (ASAP) at the Florida

Table 5
Logistic regression analysis results, a) Leon County, b) Bay County, c) Escambia and Santa Rosa Counties.

Regressors	Leon County				Bay County				Escambia – Santa Rosa County			
	β	SE	p	95%	β	SE	p	95%	β	SE	p	95%
Intercept	−2.84	0.118	≈0	✓	−1.83	0.145	≈0	✓	−1.40	0.086	≈0	✓
Day of week	−0.13	0.060	0.031	✓	−0.23	0.063	≈0	✓	−0.17	0.039	≈0	✓
At peak hour	−0.38	0.054	≈0	✓	−0.35	0.058	≈0	✓	−0.26	0.036	≈0	✓
AADT/10,000	0.02	0.016	0.199	X	−0.00	0.021	0.861	X	0.04	0.012	0.001	✓
Speed limit	0.02	0.002	≈0	✓	0.01	0.003	0.047	✓	0.00	0.002	0.558	X
Alcohol-drug abuse	−1.34	0.230	≈0	✓	−0.62	0.124	≈0	✓	−0.95	0.095	≈0	✓
Intersection presence	0.01	0.051	0.058	X	0.21	0.053	≈0	✓	0.29	0.035	≈0	✓
Traffic control unit presence	0.10	0.051	0.051	X	0.13	0.053	0.016	✓	0.02	0.036	0.629	X
Work zone presence	0.23	0.109	0.033	✓	−0.15	0.178	0.408	X	−0.11	0.100	0.297	X
Weather condition	0.02	0.117	0.890	X	0.19	0.127	0.144	X	−0.04	0.080	0.660	X
Light condition	−0.84	0.062	≈0	✓	−0.93	0.067	≈0	✓	−0.82	0.041	≈0	✓
Road condition	−0.05	0.108	0.675	X	−0.41	0.141	0.004	✓	−0.32	0.126	0.010	✓
Road surface condition	−0.28	0.103	0.006	✓	−0.18	0.109	0.099	X	−0.21	0.069	0.003	✓
Visibility	−0.07	0.078	0.349	X	−0.19	0.082	0.022	✓	−0.08	0.071	0.259	X
Lane departure action	−0.19	0.063	0.002	✓	−0.25	0.072	0.001	✓	−0.48	0.045	≈0	✓
PF(65+)/max [PF(65+)]	1.34	0.219	≈0	✓	0.69	0.132	≈0	✓	0.50	0.126	≈0	✓
	N: 24,020, df: 24,004				N: 14,892, df: 14,876				N: 32,558, df: 32,542			
	$\chi^2 = 752$, $p \approx 0$				$\chi^2 = 668$, $p \approx 0$				$\chi^2 = 1720$, $p \approx 0$			

Generalized linear regression model: $logit(y) \sim 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x14 + x15$. $y \sim$ Probability of a crash to involve an aging driver. Binary variables: "Day of Week" (1:Weekend, 0:Weekday), "At Peak Hour" (1:Yes, 0:No), "Alcohol-Drug Abuse" (1:Yes, 0:No), "Intersection Presence" (1:Yes, 0:No), "Traffic Control Unit Presence" (1:Yes, 0:No), "Work Zone Presence" (1:Yes, 0:No), "Weather Condition" (1:Bad, 0:Clear), "Light Condition" (1:Night, 0:Else), "Road Condition" (1: Defected, 0:Good), "Road Surface Condition" (1:Slippery, 0:Dry), "Visibility" (1:Bad, 0:Clear) and "Lane Departure Action" (1:Yes, 0:No). Continuous Variables: "AADT", "Speed Limit" and "Aging Population Factor Density". Abbreviations: β: estimated coefficient, SE: standard error, p: p value, N: number of observations, df: degrees of freedom, χ^2: Chi^2 statistics vs. constant model, PF(65+): 65+ population factor.

State University (FSU), Florida A&M University (FAMU), and University of North Florida (UNF). We thank the Florida Department of Transportation for providing the data. The opinions, results, and findings expressed in this manuscript are those of the authors and do not necessarily represent the views of the United States Department of Transportation, The Florida Department of Transportation, The Center for Accessibility and Safety for an Aging Population, the Florida State University, the Florida A&M University, or the University of North Florida.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jtrangeo.2016.11.011.

References

Abdel-Aty, M., Abdelwahab, H., 2000. Exploring the relationship between alcohol and the driver characteristics in motor vehicle accidents. Accid. Anal. Prev. 32 (4):473–482. http://dx.doi.org/10.1016/s0001-4575(99)00062-7.

Abdel-Aty, M., Radwan, E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev. 32 (5):633–642. http://dx.doi.org/10.1016/S0001-4575(99)00094-9.

Abdel-Aty, M., Chen, C.L., Schott, J.R., 1998. An assessment of the effect of driver age on traffic accident involvement using log-linear models. Accid. Anal. Prev. 30 (6): 851–861. http://dx.doi.org/10.1016/S0001-4575(98)00038-4.

Abdel-Aty, M., Chen, C.L., Radwan, E., 1999a. Using conditional probability to find driver age effect in crashes. J. Transp. Eng. 125 (6):502–507. http://dx.doi.org/10.1061/(ASCE)0733-947X(1999)125:6(502).

Abdel-Aty, M., Chen, C.L., Radwan, E., Brady, P.A., 1999b. Analysis of crash-involvement trends by drivers' age in Florida. ITE J. 69 (2), 69–74.

Administration on Aging, 2014. A Profile of Older Americans: 2014 [WWW Document]. U.S. Department of Health and Human Services (URL http://www.aoa.acl.gov/aging_statistics/Profile/2014/docs/2014-Profile.pdf accessed 10.1.16).

Alam, B., Spainhour, L., 2008. Contribution of behavioral aspects of older drivers to fatal traffic crashes in Florida. Transp. Res. Rec. 2078:49–56. http://dx.doi.org/10.3141/2078-07.

Alam, B., Spainhour, L., 2014. Logit and case-based analysis of drivers' age as a contributing factor for fatal traffic crashes on highways and state roads in Florida. The Transportation Research Board (TRB) 93rd Annual Meeting (Washington D.C).

Al-Ghamdi, A.S., 2002. Using logistic regression to estimate the influence of accident factors on accident severity. Accid. Anal. Prev. 34 (6):729–741. http://dx.doi.org/10.1016/S0001-4575(01)00073-2.

Baker, T.K., Falb, T., Voas, R., Lacey, J., 2003. Older women drivers: fatal crashes in good conditions. J. Saf. Res. 34 (4):399–405. http://dx.doi.org/10.1016/j.jsr.2003.09.012.

Bayam, E., Liebowitz, J., Agresti, W., 2005. Older drivers and accidents: a meta analysis and data mining application on traffic accident data. Expert Syst. Appl. 29 (3):598–629. http://dx.doi.org/10.1016/j.eswa.2005.04.025.

Bédard, M., Guyatt, G.H., Stones, M.J., Hirdes, J.P., 2002. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accid. Anal. Prev. 34: 717–727. http://dx.doi.org/10.1016/S0001-4575(01)00072-0.

Blazquez, C.A., Celis, M.S., 2013. A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. Accid. Anal. Prev. 50:304–311. http://dx.doi.org/10.1016/j.aap.2012.05.001.

Boyce, T.E., Geller, E.S., 2002. An instrumented vehicle assessment of problem behavior and driving style: do younger males really take more risks. Accid. Anal. Prev. 34 (1):51–64. http://dx.doi.org/10.1016/S0001-4575(00)00102-0.

Brunsdon, C., 1995. Estimating probability surfaces for geographical point data: an adaptive kernel algorithm. Comput. Geosci. 21 (7):877–894. http://dx.doi.org/10.1016/0098-3004(95)00020-9.

Bureau of Economic and Business Research, 2014. Florida Estimates of Population 2014 [WWW Document]. University of Florida (URL http://edr.state.fl.us/Content/population-demographics/data/PopulationEstimates2014.pdf. accessed 10.1.16).

Carr, D., Jackson, T.W., Madden, D.J., Cohen, H.J., 2016. The effect of age on driving skills. J. Am. Geriatr. Soc. 40 (6):567–573. http://dx.doi.org/10.1111/j.1532-5415.1992.tb02104.x.

Chainey, S., Tompson, L., Uhlig, S., 2008. The utility of hotspot mapping for predicting spatial patterns of crime. Secur. J. 21:4–28. http://dx.doi.org/10.1057/palgrave.sj.8350066.

Charness, N., Schaie, K.W., 2003. Impact of Technology on Successful Aging. first ed. Springer Publishing Company, New York.

Collia, D.V., Sharp, J., Giesbrecht, L., 2003. The 2001 National Household Travel Survey: a look into the travel patterns of older Americans. J. Saf. Res. 34 (4):461–470. http://dx.doi.org/10.1016/j.jsr.2003.10.001.

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. West. J. Emerg. Med. 11 (3), 294–301.

Delmelle, E.C., Thill, J.-C., Ha, H.-H., 2011. Spatial epidemiologic analysis of relative collision risk factors among urban bicyclists and pedestrians. Transportation 39 (2): 433–448. http://dx.doi.org/10.1007/s11116-011-9363-8.

Eby, D.W., Molnar, L.J., 2009. Older adult safety and mobility: issues and research needs. Public Work. Manag. Policy 13 (4):288–300. http://dx.doi.org/10.1177/1087724X09334494.

Effati, M., Thill, J.-C., Shabani, S., 2015. Geospatial and machine learning techniques for wicked social science problems: analysis of crash severity on a regional highway corridor. J. Geogr. Syst. 17 (2):107–135. http://dx.doi.org/10.1007/s10109-015-0210-x.

Erdogan, S., Yilmaz, I., Baybura, T., Gullu, M., 2008. Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. Accid. Anal. Prev. 40 (1):174–181. http://dx.doi.org/10.1016/j.aap.2007.05.004.

Evans, L., 2000. Risks older drivers face themselves and threats they pose to other road users. Int. J. Epidemiol. 29 (2):315–322. http://dx.doi.org/10.1093/ije/29.2.315.

FDOT, 2015. Traffic Data. Florida Department of Transportation (URL http://www.dot.state.fl.us/planning/statistics/trafficdata/.accessed 15.1.15).

FHA, 2011. Summary of Travel Trends: 2009 National Household Travel Survey (doi: FHWA-PL-II-022).

Flahaut, B., Mouchart, M., San Martin, E., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones. a comparative approach. Accid. Anal. Prev. 35 (6):991–1004. http://dx.doi.org/10.1016/S0001-4575(02)00107-0.

Getis, A., Ord, J.K., 1992. The analysis of spatial associatiion by use of distance statistics. Geogr. Anal. 24 (3):189–206. http://dx.doi.org/10.1073/pnas.0703993104.

Hellinga, B., Macgregor, C., 1999. Towards modeling the impact of an ageing driver population on intersection design and traffic management. Community Traffic Solutions for the Future Session of the 1999 Annual Conference of the Transportation Association of Canada (Saint John, New Brunswick, Canada. URL http://www.civil.uwaterloo.ca/bhellinga/publications/publications/tac- 1999 modelling older drivers v2.pdf).

Islam, S., Mannering, F.L., 2006. Driver aging and its effect on male and female single-vehicle accident injuries: some additional evidence. J. Saf. Res. 37 (3):267–276. http://dx.doi.org/10.1016/j.jsr.2006.04.003.

Jang, T.Y., 2006. Analysis on reckless driving behavior by log-linear model. KSCE J. Civ. Eng. 10 (4):297–303. http://dx.doi.org/10.1007/BF02830784.

Karacasu, M., Ergul, B., Altin Yavuz, A., 2013. Estimating the causes of traffic accidents using logistic regression and discriminant analysis. Int. J. Inj. Control Saf. Promot. 21 (4):305–312. http://dx.doi.org/10.1080/17,457,300.2013.815632.

Kilamanua, W., Xiaa, J., Caulfieldb, C., 2011. Analysis of spatial and temporal distribution of single and multiple vehicle crash in Western Australia: a comparison study. 19th International Congress on Modelling and Simulation. Perth, Australia:pp. 683–689 (URL http://espace.library.curtin.edu.au/cgi-bin/espace.pdf?file=/2012/02/16/file_1/172,394.).

Koffman, D., Weiner, R., Pfeiffer, A., Chapman, S., 2010. Funding the Public Transportation Needs of an Aging Population, American Public Transportation Association. (URL http://www.apta.com/resources/reportsandpublications/Documents/).

Kononen, D.W., Flannagan, C.A.C., Wang, S.C., 2011. Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. Accid. Anal. Prev. 43 (1):112–122. http://dx.doi.org/10.1016/j.aap.2010.07.018.

Krahe, B., Fenske, I., 2002. Predicting aggressive driving behavior: the role of macho personality, age, and power of car. Aggress. Behav. 28 (1):21–29. http://dx.doi.org/10.1002/ab.90003.

Kvam, P.H., Vidakovic, B., 2007. Nonparametric Statistics with Applications to Science and Engineering. John Wiley & Sons, Hoboken, New Jersey.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. A Policy Pract. 44 (5): 291–305. http://dx.doi.org/10.1016/j.tra.2010.02.001.

Ma, Z., Shao, C., Yue, H., Ma, S., 2009. Analysis of the Logistic Model for Accident Severity on Urban Road Environment. The IEEE Intelligent Vehicles Symposium. Xi'an, China: pp. 983–987 http://dx.doi.org/10.1109/IVS.2009.5164414.

Maddala, G.S., 1986. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge, United Kingdom.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. Anal. Methods Accid. Res. 1:1–22. http://dx.doi.org/10.1016/j.amar.2013.09.001.

Massie, D.L., Campbell, K.L., Williams, A.F., 1995. Traffic accident involvement rates by driver age and gender. Accid. Anal. Prev. 27 (1):73–87. http://dx.doi.org/10.1016/0001-4575(94)00050-V.

Mayhew, D.R., Simpson, H.M., Ferguson, S.A., 2006. Collisions involving senior drivers: high-risk conditions and locations. Traffic Inj. Prev. 7 (2):117–124. http://dx.doi.org/10.1080/15389580600636724.

McGwin, G., Brown, D.B., 1999. Characteristics of traffic crashes among young, middle-aged, and older drivers. Accid. Anal. Prev. 31 (3):181–198. http://dx.doi.org/10.1016/S0001-4575(98)00061-X.

McKenzie, B., Rapino, M., 2011. Commuting in the United States: 2009. U.S. Census Bureau (URL https://www.census.gov/prod/2011pubs/acs-15.pdf).

Merat, N., Anttila, V., Luoma, J., 2005. Comparing the driving performance of average and older drivers: the effect of surrogate in-vehicle information systems. Transport. Res. F: Traffic Psychol. Behav. 8 (2):147–166. http://dx.doi.org/10.1016/S1366-5545(02)00012-1.

Mohaymany, A.S., Shahri, M., Mirbagheri, B., 2013. GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. Geo-spatial Inf. Sci. 16 (2):113–119. http://dx.doi.org/10.1080/10095020.2013.766396.

Mori, Y., Mizohata, M., 1995. Characteristics of older road users and their effect on road safety. Accid. Anal. Prev. 27 (3), 391–404.

NHTSA, 2014. Traffic Safety Facts, 2014 Data [WWW Document]. National Highway Traffic Safety Administration Administration (URL http://www-nrd.nhtsa.dot.gov/Pubs/812,263.pdf. accessed 21.2.15).

NHTSA, 2015. Data-Driven Approaches to Crime and Traffic Safety (DDACTS) [WWW Document]. National Highway Traffic Safety Administration (URL http://www.nhtsa.gov/Driving+Safety/Enforcement+&+Justice+Services/Data-Driven+Approaches+to+Crime+and+Traffic+Safety+(DDACTS). accessed 31.10.15).
Okabe, A., Sugihara, K., 2012. Spatial Analysis along Networks. John Wiley & Sons, Chichester, United Kingdom http://dx.doi.org/10.1002/9781119967101.ch3.
Okabe, A., Satoh, T., Sugihara, K., 2009. A kernel density estimation method for networks, its computational method and a GIS-based tool. Int. J. Geogr. Inf. Sci. 23 (1):7–32. http://dx.doi.org/10.1080/13658810802475491.
Rong, J., Mao, K., Ma, J., 2011. Effects of individual differences on driving behavior and traffic flow characteristics. Transp. Res. Rec. 2248:1–9. http://dx.doi.org/10.3141/2248-01.
Ryan, G.A., Legge, M., Rosman, D., 1998. Age related changes in drivers' crash risk and crash type. Accid. Anal. Prev. 30 (3):379–387. http://dx.doi.org/10.1016/S0001-4575(97)00098-5.
Sabel, C.E., Kingham, S., Nicholson, A., Bartie, P., 2005. Road traffic accident simulation modelling - a kernel estimation approach. The 17th Annual Colloquium of the Spatial Information Research Centre University of Otago (Dunedin, New Zealand. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.624.3867&rep=rep1&type=pdf).
Sandler, M., Nichols, T., Tarburton, R., Allan, B., 2015. Health Care Providers and Older Adult Service Organizations to Assist in the Prevention and Early Recognition of Florida's At-Risk Drivers (doi:BDY17-RFP-DOT-13/14-9032-RC).
Stamatiadis, N., Taylor, W.C., Mckelvey, F.X., 1991. Older drivers and intersection traffic control devices. J. Transp. Eng. 117 (3):311–319. http://dx.doi.org/10.1061/(ASCE)0733-947X(1991)117:3(311).
Staplin, L., Lococo, K.H., Martell, C., Stutts, J., 2012. Taxonomy of Older Driver Behaviors and Crash Risk (doi:DOT HS 811 468A).
Steenberghen, T., Aerts, K., Thomas, I., 2010. Spatial clustering of events on a network. J. Transp. Geogr. 18 (3):411–418. http://dx.doi.org/10.1016/j.jtrangeo.2009.08.005.
U.S. Census Bureau, 2010. 2010 TIGER/Line Shapefile. (URL http://www.census.gov/geo/maps-data/data/tiger.html. accessed 25.7.15).
U.S. Census Bureau, 2012. Households and Families: 2010 (doi:C2010BR-14).
U.S. Census Bureau, 2015. 2010 US Census Blocks in Florida. Florida Geographic Data Library (URL http://www.fgdl.org/metadataexplorer/explorer.jsp. accessed 15.1.15).
Ulak, M.B., Ozguven, E.E., Kocatepe, A., 2015. GIS-based investigation of aging involved crashes: a case study in Northwest Florida. The Road Safety and Simulation International Conference. Orlando, Florida (URL http://www.ce.ucf.edu/asp/cfp_Rss/Cd/docs/FullProceedings.pdf).
Xie, Z., Yan, J., 2008. Kernel density estimation of traffic accidents in a network space. Comput. Environ. Urban. Syst. 32 (5):396–406. http://dx.doi.org/10.1016/j.compenvurbsys.2008.05.001.
Xie, Z., Yan, J., 2013. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. J. Transp. Geogr. 31: 64–71. http://dx.doi.org/10.1016/j.jtrangeo.2013.05.009.
Yamada, I., Thill, J.-C., 2004. Comparison of planar and network K-functions in traffic accident analysis. J. Transp. Geogr. 12 (2):149–158. http://dx.doi.org/10.1016/j.jtrangeo.2003.10.006.