Hotspots and causes of motor vehicle crashes in Baltimore, Maryland: A geospatial analysis of five years of police crash and census data

Zachary Dezman\textsuperscript{a,b}, Luciano de Andrade\textsuperscript{c}, Joao Ricardo Vissoci\textsuperscript{d,e}, Deena El-Gabri\textsuperscript{e}, Abree Johnson\textsuperscript{b}, Jon Mark Hirshon\textsuperscript{e,a,b}, and Catherine A. Staton\textsuperscript{d,e,f,}\textsuperscript{*}

\textsuperscript{a}Department of Emergency Medicine, University of Maryland School of Medicine, Baltimore, Maryland, United States
\textsuperscript{b}National Study Center for Trauma and Emergency Medical Services, University of Maryland School of Medicine, Baltimore, Maryland, United States
\textsuperscript{c}State University of Maringa/UEM, Maringa, Brazil
\textsuperscript{d}Division of Global Neurosurgery and Neurosciences, Duke University, Durham, North Carolina, United States
\textsuperscript{e}Duke Global Health Institute, Duke University, Durham, North Carolina, United States
\textsuperscript{f}Duke Emergency Medicine, Duke University Medical Center, Durham, North Carolina, United States

Abstract

**Introduction**—Road traffic injuries are a leading killer of youth (aged 15–29) and are projected to be the 7th leading cause of death by 2030. To better understand road traffic crash locations and characteristics in the city of Baltimore, we used police and census data, to describe the epidemiology, hotspots, and modifiable risk factors involved to guide further interventions.

**Materials and methods**—Data on all crashes in Baltimore City from 2009 to 2013 were made available from the Maryland Automated Accident Reporting System. Socioeconomic data collected by the US CENSUS 2010 were obtained. A time series analysis was conducted using an ARIMA model. We analyzed the geographical distribution of traffic crashes and hotspots using exploratory spatial data analysis and spatial autocorrelation. Spatial regression was performed to evaluate the impact of socioeconomic indicators on hotspots.

**Results**—In Baltimore City, between 2009 and 2013, there were a total of 100,110 crashes reported, with 1% of crashes considered severe. Of all crashes, 7% involved vulnerable road users and 12% had elderly or youth involvement. Reasons for crashes included: distracted driving (31%), speeding (6%), and alcohol or drug use (5%). After 2010, we observed an increasing trend in all crashes especially from March to June. Distracted driving then youth and elderly drivers

\textsuperscript{*}Corresponding author at: 310 Trent Hall #304 Durham, NC 27710, United States. catherine.staton@duke.edu (C.A. Staton).

Conflict of interest

There are no conflicts of interests in compiling this paper- no financial or personal relationships with other people or organizations that could inappropriately influence the work.
were consistently the highest risk factors over time. Multivariate spatial regression model including socioeconomic indicators and controlling for age, gender and population size did not show a distinct predictor of crashes explaining only 20% of the road crash variability, indicating crashes are not geographically explained by socioeconomic indicators alone.

**Conclusion**—In Baltimore City, road traffic crashes occurred predominantly in the high density center of the city, involved distracted driving and extremes of age with an increase in crashes from March to June. There was no association between socioeconomic variables where crashes occurred and hotspots. In depth analysis of how modifiable risk factors are impacted by geospatial characteristics and the built environment is warranted in Baltimore to tailor interventions.

**Keywords**

Road traffic crash; Hotspot analysis; Geographic information system

**Introduction**

Road traffic injuries (RTIs) are the leading killer of young people (aged 15–29) globally, and without intervention road traffic crashes are projected to be the 7th leading cause of death by 2030 [1]. In the United States, road injuries are the leading cause of death for young men [2]. This is especially problematic because this subset of the population most affected by motor vehicle crashes (MVCs) are at their peak of economic productivity. While MVC mortality has been declining throughout the past decade in the United States, RTIs are increasing throughout the country [3]. These increasing morbidity and high mortality rates due to motor vehicle crashes have significant social and economic consequences [1]. To assess this global problem, there has been a call to decrease this burden of injury through road traffic interventions [1]. In order to develop the most effective interventions, targeting the most impactful change, location-specific MVC epidemiology must be assessed.

With 2.7 million residents, the Baltimore metropolitan area is one of the larger cities in the United States. Despite covering only 92 of the 12,407 square miles of Maryland (0.7%), the city of Baltimore has the greatest proportion of MVCs (21.8%) and RTI (15.3%) in the state [4]. This high crash density makes Baltimore City an ideal location for interventions designed to decrease the statewide burden of MVCs. Policy, legislature, and city-wide regulation have attempted to lower this disproportionate burden of MVC in Baltimore. The current literature on MVC in Baltimore has been focused on specific locations or settings, but a comprehensive epidemiologic analysis has not been published [5]. For example, there have been evaluations of environmental attributes associated with vehicle crashes near Baltimore schools, and of the impact of laws on young drivers, seatbelts and airbags on crashes [6–8]. But current research does not assess how these laws affect traffic incidences at the locations where most MVC occur [9]. The current MVC literature in Baltimore assesses some traffic incident epidemiology but does not perform an overall epidemiologic profile to determine epidemiology, hotspots and modifiable risk factors in order to highlight most important areas to perform more in depth investigations and further safety interventions. Research in other settings has shown that geographically specific motor vehicle crash epidemiology can be used to put forth tailored intervention measures [10–12].
To better understand the location and characteristics of MVCs within the city of Baltimore, this study describes the epidemiology of MVCs from 2009 to 2013, in order to guide further geographic and epidemiologic focused interventions. Through the use of police and census data, we will describe the time trajectory, identify MVC hotspots, the change of hotspot locations over time, modifiable risk factors involved in crashes, and demographic characteristics to further delineate potential interventions to reduce MVC morbidity and mortality.

Our study was approved by the Institutional Review Board of the University of Maryland, Baltimore (HP-00065167).

Methods

Study design

This is an observational, cross-sectional, ecological study using spatial analysis techniques based on car crashes data from 2009 to 2013 in the city of Baltimore, Maryland.

Study population and location

The city of Baltimore is located in the state of Maryland, US, occupying an area of 238.5 km$^2$, with latitude coordinates 39°17′N and longitude coordinates 76°37′W (Fig. 1) [13]. According to the 2010 US Census, Baltimore city has an entirely urban area with 620,961 inhabitants. The Elderly Index (population above 65 years old) of Baltimore city was 11.7% in 2010, and therefore similar to the US average of 13.0%. Baltimore city presented a Gross Domestic Product (GDP) of 152.9 billion USD, the 19th largest within the United States, and a Human Development Index (HDI) of 5.75. This value is higher than the US average HDI of 5.03 [14].

Data sources and variables

Road traffic crashes data—We selected all crashes in Baltimore City from 2009 to 2013 made available from the Maryland Automated Accident Reporting System (MAARS), as provided by Baltimore City Department of Transportation. Crashes were classified according to type of traffic offense (speeding, sleep, distracted/inattentive, aggressive or impaired driving) or characteristic (elderly, youth, pedestrians, motorcycle or bicycles related). Severity of crash was also extracted through the KABCO severity score of 4 or 5 [15]. The KABCO severity score was established by the National Safety Council to be used by police and law enforcement for classifying injury severity (K fatal, A incapacitating, B non incapacitating, C possible injury, O no injury). Severe injuries for this analysis included the K and A crashes for severe and incapacitating injuries.

Socio economic indicators—Socioeconomic data (education, marital status, income, unemployment, poverty and male sex rate) collected by the US CENSUS 2010 were obtained [16]. Data is publicly available and is geographically distributed in Census tracts units. Census tract were used as common geospatial indicator to merge both data sources and conduct data analysis.
Data analysis

Time series—Time series was conducted using the autoregressive integrated moving average (ARIMA) model, which is widely used in epidemiology surveillance. ARIMA model expresses the current traffic crashes incidence linearly with its previous incidence (autoregressive) and the residual series (moving average). The time series model was built according to the Box-Jenkins methodology consisting of three iterative steps: identification, estimation, and diagnostic checking [17,18]. It is possible that several ARIMA models may be identified, and the selection of an optimum model becomes necessary using the Akaike Information Criterion (AIC) and Schwartz Bayesian Criterion (SBC) (Table 1).

Geospatial analysis—Geospatial data was analyzed by geographic locations (points) to evaluate the geospatial distribution of traffic crashes and areas with higher densities of occurrences (hotspots). We applied exploratory spatial data analysis (ESDA) through the software QGIS and GeoDa™ version 0.9.5-i (Spatial Analysis Laboratory, University of Illinois, Urbana Champaign, IL, USA) to determine measures of global spatial autocorrelation, local spatial autocorrelation, and spatial regression [19,20].

Spatial autocorrelation—To evaluate the existence of spatial autocorrelation, we used a Queen-type matrix that allows for the measurement of non-random association between the value of a variable observed in a given geographical unit and the value of variables observed in neighboring units [20]. Using the (I) Global Moran index we calculated spatial autocorrelation evaluating prevalence proportion of traffic crashes in each Census tract. We calculated univariable association for traffic crashes and bivariable associations with Census bases socioeconomic indicators [20,21]. This index identifies if the value of the proportion of traffic crashes tends to be clustered (positive Moran I) or dispersed (negative Moran I) among geographical areas [20–22].

To graphically depict spatial autocorrelation, we applied the local indicators of spatial association (LISA) clustering method. The LISA maps identify significant spatial clusters throughout Baltimore city, with high or low association values for the proportion of traffic crashes [23]. Clustered areas are categorized according to the pattern of characteristics in adjacent districts. High/high (HH) areas are a set of districts with high proportion of crashes surrounded by other districts with high crashes in univariate analysis. The same sense is applied to low/low (LL) set of districts, where districts with low characteristics are surrounded by districts with low values for analyzed variables. When the inverse occurs, districts with low proportion of crashes are surrounded by districts with high crashes, LISA maps categorize them as low/high (L/H) or high/low (H/L) for the opposite pattern.

Spatial regression—To identify which socioeconomic indicators had a higher geospatial impact on the distribution of traffic crashes we conducted a multivariable spatial regression analysis [23]. Using a spatial autoregressive lag (SAR) model, we regressed our independent variables (socio economic indicators) against the proportion of crashes. SAR modeling is a strong approach to understand high spatial autocorrelated data. Interactions are modeled as a weighted average of the neighboring observations through a spatially-lagged dependent variable. The model is weighted based on the neighborhood interaction matrix to analyze
spatial dependency [23]. Finally, we added the proportion of urban population per district to the model to control for the density and exposure.

**Results**

In Baltimore city, between 2009 and 2013, there were a total of 100,110 crashes registered, with an average of 382 crashes a week and 1688 a month. Only one percent of all crashes were considered severe. From all crashes, 7% involved vulnerable road users (VRU) and 12% were related to extreme age population (elderly and youngsters). Reasons for crashes include: speeding (6%), alcohol and drug use (5%), and distracted driving (31%) (Table 2).

**Time series**

After 2010, we observed an increasing trend in the time series for all crashes (Table 2). Seasonality indicated a significant time dependency, increasing from March to June consistently across different years (Fig. 2). Analyzing the variation structure of the time series, the ARIMA model with best performance showed one month variation dependency, which means that variation of a given month was influenced by the previous month. As for the types of crashes, distracted driving related crashes had the highest consistent frequency over time, compared to other types of crashes (Fig. 3), with an increasing trend. Crashes involving the elderly and youth populations had higher frequencies consistently over time.

**Road traffic crashes hotspots and spatial association**

The spatial point distribution the locations for each road traffic crashes in Baltimore city revealed a high density in the center of the city and surrounding areas, and also a spreading following the patterns of main access roads and avenues (Fig. 4A). This pattern remained when plotted the subset of severe crashes (Fig. 4C). Severe crashes have certain areas which are unique hotspots, not seen when evaluating non-severe crashes.

Similarly, the spatial association analysis highlights a moderate spatial dependence for crashes (Moran’s I 0.26, p = 0.001). The center of Baltimore city appears as an important cluster of high risk areas (red polygons) surrounded by other higher risk for crashes (Fig. 4B and D) while the more distant areas grouped as low risk areas for crashes (blue polygons). On the other hand, looking at the spatial association of areas for all road traffic crashes, we can see that one of the main access roads to Baltimore city is considered a high risk area contrasting by low risk areas surrounding it (light blue mark in Fig. 4B). Also, when looking specifically at the spatial association of severe crashes, the areas north of Baltimore Harbor appears as a high risk cluster.

Stratifying the spatial distribution of road traffic crashes by type of crash (Fig. 5A–E) shows a similar pattern to the distribution of all crashes represented in Fig. 4, also concentrated in the center of the city and surrounding areas, which are also vastly populated areas and are targeted with people in traffic to work. However, the pattern changes slightly when looking into the density of crashes due to speeding with more hotspots following the main roads. Distracted driving, on the other hand, is even more condensed in the center of the city.
The spatial association by type of crash, on the other hand, had important differences when compared to the clusters of all crashes. Vulnerable road users (Moran’s I 0.21, p = 0.01) high risk area is located in the center of the city (Fig. 6A), such as extreme age and distracted driving (Moran’s I 0.17, p = 0.01 and 0.09, p = 0.02, respectively) related crashes (Fig. 6C and E). Distracted driving related crashes only deviates by showing high risk polygons surrounded by low risk areas in the northwestern part of Baltimore city, around important access roads to the center of the city. Not surprisingly, the high risk clusters for crashes related to speeding deviate from the common pattern, showing two clusters in areas far from the center, where drivers maintain higher speeds (Fig. 6B), but also suggesting smaller spatial dependence (Moran’s I 0.21, p = 0.02).

Looking more closely to some of the high risk areas for road traffic crashes (Fig. 7) we notice that in the center of the city, crashes seem to be happening more in intersections where potentially there are more crosswalks, traffic lights and other built environment characteristics that emphasized crashes related to vulnerable road users and older and younger people, distracted driving or substance use. On the other hand, at the northwestern and eastern area of the city highlighted in Fig. 7, the high risk area has most severe crashes are located across the extent of a larger road.

**Spatial association with socioeconomic factors**

We used socioeconomic data from the US Census data to explain the spatial distribution of road traffic crashes in Baltimore city. Education, Marital status, and Income showed a negative bivariate spatial association with crashes (Moran’s I −0.21, −0.17, −0.12), suggesting that geographical patterns of crashes are associated with less rates of this geopolitical patterns for these socioeconomic variables. Similarly, however, although significant (p < 0.05), the spatial autocorrelation coefficient was weak indicating that there are areas in the city with the spatial association characteristic are not frequent or large. Multivariate spatial regression model including these socioeconomic indicators and controlling for age, gender and population size did not show a distinct predictor of crashes and was only able to explain 20% of the road crash variability, indicating that crashes in high income areas are not geographically explained by socioeconomic indicators alone.

**Discussion**

This is the first project to conduct a wide ranging epidemiologic profile of road traffic crashes over a 5 year time period in Baltimore Maryland. Our more broad epidemiologic focus allowed for identification of risk factors for crashes as well as defining the population and regions most at risk for road traffic crashes in this geographically and socioeconomically diverse city. Our data found that: as expected, the predominant number of road traffic crashes occurred in the high density center of the city of Baltimore, that there was no association between socioeconomic variables where crashes occurred and hotspots, and there is a significant seasonality of crashes with predictable patterns.

Baltimore City experiences a high density of MVCs most of which are concentrated in the center of the city. This is typical of cities across the United states, more compact cities and regions of cities are shown to have greater total MVCs [24]. The most severe crashes did not
have a significantly different distribution across the city. This is dissimilar from the usual pattern found in the literature where severe crashes are usually further from town centers where traffic is more sprawled [25]. As crash severity has been linked to higher speeds, the diversity of speeds within the city is likely a cause of this even distribution [26,27]. Known modifiable risk factors for crashes including high speed, distracted driving, and alcohol and drug intoxication were cited for crash causes in 5% of the crashes. Our geographic evaluation of the crashes found several census tracts where there were significantly more of these modifiable risk factors.

Census tract 1101 as shown in Fig. 7, had a significant number of crashes due to speeding. This tract contains two major features: 1) one-way inner-city streets that act as a major northbound thoroughfares out of the inner city which run through 2) an increasing pedestrian trafficked urban and university area. This tract also has a major interstate highway with limited speed enforcement. Census tract 2607 as seen in Fig. 7 was a focus for MVCs associated with speed, distracted driving, alcohol, and elderly drivers. This tract contains a major thoroughfare and complex series of intersections: this thoroughfare quickly narrows to two lanes before crossing a combined residential and commercial area with pedestrian traffic. In both cases, there are very obvious combinations of high speed roadways and heavily trafficked pedestrian areas which could be a cause of the hotspots similar to patterns seen in other locations [11,28,29]. Both of these census tracts need to have further evaluation through in depth built environment analysis to inform further safety interventions. Lastly, our analysis determined that census tract 280.01 was a locus for alcohol and drug related crashes surrounded by tracts with very few MVCs. This tract is mostly residential with one large mall which has multiple large thoroughfares and potentially a high density of alcohol outlets. Given alcohol outlet type has been shown to have an impact on the risk of alcohol related road traffic crash in the region and surrounding regions, further research on alcohol outlets in Baltimore is needed [30].

Our data showed that road traffic crash locations are independent (accounting for only 20% of the variance) from the socioeconomic factors of the surrounding census tract. Since Baltimore is a city with a large mobile workforce that commutes, (40.1%) a large proportion of road traffic crash victims are likely to live in different areas of the city than where the crash occurs [16]. Similarly, as Baltimore has an influx of tourists during summer months, this influx might explain the disparity of socioeconomic status between those who suffer crashes and crash location average socioeconomic status. As a result, modifiable risk factors rather than focusing on certain socioeconomic status population would likely be a better area for further prevention initiatives.

Modeling the time series of road traffic crashes showed an annual variation in the number of MVCs and RTIs in Baltimore, with the greatest number of crashes occurring between March and June. Possible explanations for this increase are improved weather and travel conditions which might increase road use, higher tourism rates at this time of year and the peak of the festival season in Baltimore [31,32]. Overall, there was an increasing trend in the number of crashes was most notable amongst those crashes caused by distracted driving [33].
When comparing our time series analysis to Baltimore events, we are able to uncover possible explanations for the variations. For instance, in 2010 there was an increase in the MVCs related to distracted driving (Fig. 3). This is the same year the Maryland State Highway allowed for more factors to be associated with a MVCs (four instead of two) (Personal communication ZDWD). The definition of distracted driving was also widened in 2010, though this change was retroactively applied to all data collected as far back as 2005 (which includes the data our study is based upon) [33]. There have been a number of interventions aimed at decreasing distracted driving in Maryland. In October of 2013, Maryland joined the District of Columbia and Virginia in banning cell phones while driving; this law allowed police to fine drivers for driving while on a phone [34]. As a part of their “Towards Zero Deaths” campaign, Maryland initiated a public education program called “Park the Phone before You Drive.” [35] The US Department of Transportation also launched a similar awareness campaign called “One Text or Call Could Wreck It All.” [36]

Despite these efforts, our analysis shows that the incidence of distracted driving continues to increase (Fig. 3). This increase in distracted driving incidents has been seen in cities nationwide. There is evidence that supports distracted driving laws and enforcement as effective prevention, however education programs against distracted driving have not been shown to be effective [37].

**Limitations**

This evaluation of the motor vehicle crashes in Baltimore City is retrospective police data provided by the city, and may be subject to variations in reporting. There was a definitional change in the “Distracted Driving” classification between 2009 and 2010 possibly changing the proportion of distracted driving cases reported. While our analysis shows that crashes were most commonly categorized as distracted driving prior to 2009 and continued to rise after 2010, so this definition likely only had a small effect on this reporting. This analysis does not focus on specific crashes, and provides only summary descriptions for the crashes occurring within census tracts. It is not prospective, and does not provide information about the various interventions aimed at decreasing motor vehicle crashes and road traffic injuries in Baltimore City; that being said, it is the first step at identifying locations which are amenable to a built environment analysis which would identify further interventions to decrease motor vehicle crashes in Baltimore City.

**Conclusions**

Baltimore City experiences a disproportionate burden of motor vehicle crashes within the State of Maryland. There were a predominant number of road traffic crashes that occur in the high density center of the city of Baltimore, yet are not associated with the socioeconomic variables where the crashes occur. Road traffic crash hotspots are associated with certain modifiable risk factors, like distracted driving, drugs and alcohol and speeding, have a seasonal variation. As such, further research on the geospatial relationship modifiable risk factors and road traffic injury hotspots are warranted, as well as build environment analysis of the road traffic crash hotspots we identified to determine appropriate interventions.
Acknowledgments

We would like to acknowledge the Baltimore City Department of Transportation for their assistance acquiring the data in order to conduct this project.

Funding

Dr. Staton would like to acknowledge salary support funding from the US National Institutes of Health Fogarty International Center (Staton, K01 TW010000-01A1). Dr. Hirshon would like to acknowledge support by the US National Institutes of Health Fogarty International Center, grant 5D43TW007296.

References

14. Mapping the Measure of America. Measure of America of the Social Science Research Council. 2013
37. Policy Interventions for Safer, Healthier People and Communities Transportation and Health. 2011
Fig. 1.
Geographical localization of Baltimore city.
Source: OpenStreetMap, 2015.
Fig. 2.
Seasonality of road traffic crashes in Baltimore city from 2009 to 2013 by month.
Fig. 3.
Time series of road traffic crashes in Baltimore city from 2009 to 2013 by type of crash.
Fig. 4.
Hotspots and spatial correlation clusters for (a) all crashes and (b) severe crashes.
Fig. 5.
Hotspots of crashes divided by (a) Vulnerable road users, (b) Speed, (c) Distracted driving, (d) Alcohol & drugs, (e) Elderly and youngsters.
Fig. 6.
Spatial clusters of crashes divided by (a) Vulnerable road users, (b) Speed, (c) Distracted driving, (d) Alcohol & drugs, (e) Elderly and youngsters.
Fig. 7.
Severe crashes points distribution of three significant geospatial clusters of high MVCs density areas.

Injury. Author manuscript; available in PMC 2017 November 01.
### Table 1

Data Sources for analysis.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type of data</th>
<th>Range</th>
<th># data entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltimore City Department of Transportation</td>
<td>Road traffic crashes, classified into: Speeding: Speeding tickets</td>
<td>2009–2013</td>
<td>742,228</td>
</tr>
<tr>
<td></td>
<td>Vulnerable road users: Pedestrians, Motorcycles, Bicycles</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distracted Driving: Sleep, Distracted/Inattentive, Aggressive</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extreme age: Elderly, Youngsters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alcohol &amp; drugs: Impaired driving</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US CENSUS</td>
<td>Socio economic indicators:</td>
<td>2010</td>
<td>5565</td>
</tr>
</tbody>
</table>
Table 2

Characteristics of MVCs in Baltimore city from 2009 to 2013.

<table>
<thead>
<tr>
<th>N (%) of crashes from 2009 to 2013</th>
<th>Prevalence of crashes per 100,000 population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2009</td>
</tr>
<tr>
<td>Severe Crashes 1045 (1.1%)</td>
<td>43.93</td>
</tr>
<tr>
<td>Vulnerable road users 6566 (7%)</td>
<td>198.14</td>
</tr>
<tr>
<td>Extreme ages 12,494 (12%)</td>
<td>434.41</td>
</tr>
<tr>
<td>Speeding 6108 (6%)</td>
<td>244.89</td>
</tr>
<tr>
<td>Alcohol &amp; Drugs 4867 (5%)</td>
<td>169.90</td>
</tr>
<tr>
<td>Distracted driving 31,254 (31%)</td>
<td>822.22</td>
</tr>
<tr>
<td>Total 100,110 (100%)</td>
<td>3105.02</td>
</tr>
</tbody>
</table>