

# GIS techniques to analyze factors associated with crash occurrence

Shraddha Sagar<sup>1\*</sup> , Nikiforos Stamatiadis<sup>2</sup> 

<sup>1</sup> Gresham Smith, United States

<sup>2</sup> Department of Civil Engineering, University of Kentucky, United States

\* Corresponding author: [shraddha.sagar@greshamsmith.com](mailto:shraddha.sagar@greshamsmith.com)

Handling editor: Aliaksei Laureshyn, Lund University, Sweden

Reviewers: Lai Zheng, Harbin Institute of Technology, China; Niels Agerholm, The Danish Road Directorate, Denmark

Received: 16 September 2021; Accepted: 17 January 2022; Published: 07 February 2022

**Abstract:** Fatality rates in Kentucky have been higher than the national average for the past decade. Prior research postulated that the region's unique socioeconomic conditions could provide a compelling explanation of the higher crash rates of the southeast in the US. This study examines the relationship between safety and socioeconomic characteristics using an extensive spatial analysis of crashes in Kentucky. Quasi-induced exposure technique was utilized to determine the crash propensity of different driver groups. Through a series of GIS techniques and spatial analysis, the relative accident involvement ratios were calculated for each group of drivers. The findings of the study concur with the previous findings between driving behavior and demographic factors such as age and gender. The study also attempted to explain the regional disparities in crash occurrence across the state in terms of economic status. The study concluded that the drivers residing in the Appalachian regions have a higher propensity to cause a crash, regardless of the age and gender of the driver. The findings of the study can be used to identify high risk counties for the Safety Circuit Rider (SCR) program in Kentucky.

**Keywords:** crash occurrence, relative accident involvement ratio (RAIR), choropleth map

## 1 Introduction

It is well documented that most of the traffic collisions factors contributing to the crash are assigned to driver (*NHTSA, 2018*). There are though underlying factors that affect driver behavior resulting in a crash involvement. The socioeconomic and demographic characteristics of the driver and their environment may influence their driving conduct and eventual crash involvement (*Sagar, 2020*). While crashes are highly dependent on the crash site characteristics, past research has demonstrated the influence of the socioeconomic features of the driver's residence zip code on crash occurrence (*Sagar, 2020; Sagar et al., 2020*). It is also possible that the drivers involved in a crash reside in a different zone than that of the crash location. Hence, examining the socioeconomic and demographic features of the driver's residence zip code would unveil critical information about the underlying factors that influence crash occurrence.

Kentucky, a southeastern state in the US, has observed higher fatality rates in the last decade (*KTC, 2018; 2003*). While Kentucky is considered a rural south-eastern state with more than 50 percent of its counties classified as rural (*Harrah, 2015*), its distinct socioeconomic

conditions could be a significant reason explaining the high fatality rate (*Stamatiadis & Puccini, 1999*). Analyzing such socioeconomic factors might aid in understanding crash occurrence reasons and identify the major causes that could contribute to these high crash trends.

In the US, the overall crash rates in Appalachia are consistently higher than in non-Appalachia (*Zhu et al., 2013*). Kentucky has about 50 percent of the counties falling into the Appalachian region with about 25.4 percent of the population is under poverty level which is the largest among other Kentucky areas (*FAHE, 2019*). The rurality and terrain of the Appalachian region could pose unique safety challenges compared to the rest of the country, and higher traffic-fatality rates compared to the remaining U.S. have been observed (*Sagar, 2020*).

Most of the studies that examined the effect of socioeconomic characteristics utilized various statistical techniques; however, this study adopted a spatial analysis approach using GIS platform. The primary goal of this study was to define the at-risk group of drivers by analyzing historic crash data and census data using a spatial analysis approach with a special emphasis in the Kentucky Appalachia region. The results of the study are represented in the form of choropleth maps to identify target groups of drivers/areas to implement efficient safety programs targeting such groups. The fact that the study used Kentucky data, does not preclude the use of the methodology developed here to be applied to any other state or region. Moreover, states with similar socioeconomic characteristics could utilize the findings of the study and follow a similar approach in determining at-risk groups.

## 2 Literature review

The factors that predominantly influence crash occurrence have been widely investigated. In several such past attempts, demographic factors, socioeconomic features, roadway geometric design, traffic characteristics and roadway characteristics of crash locations, and behavior of the involved drivers are identified as associated factors (*Adanu et al., 2017; Brown, 2016; Hanna et al., 2012; Factor et al., 2008; Agüero-Valverde & Jovanis, 2006; Hasselberg et al., 2005; Noland & Oh, 2004*). Income, education level, poverty percentage, employment, and the rurality of an area are some of the strong socioeconomic contributors of crash occurrence. Several studies have established the high risk of crash involvement for drivers living in a rural/low income neighborhoods (*Sagar, 2020; Brown, 2016; Ivan et al., 2016; Lee et al., 2014; Whitlock et al., 2003; Murray, 1998; Westerling et al., 1996*).

Age and gender of driver also play a major role in crash occurrence (*Brown, 2016; Lee et al., 2014; Factor et al., 2008; Noland & Quddus, 2004; Stamatiadis & Puccini, 1999*). *Cooper et al. (1995)* examined the interactions of various driving exposure and crash characteristics to demonstrate that novice drivers have higher crash propensity compared to the general population. (*Stamatiadis & Puccini, 1999*) showed that older drivers (over 65) have significant crash involvement in multi-unit crashes. Young (under 25) and older (over 65) drivers have a high probability to cause crashes, while male drivers have a higher crash propensity than female drivers (*Adanu et al., 2017; Brown, 2016; Factor et al., 2008; Agüero-Valverde & Jovanis, 2006; Chen et al., 2000*).

Though these factors play a vital role in crash occurrence, they do not provide information on the type of drivers or their background involved in crashes. Hence, it is more reasonable to examine the socioeconomic characteristics of the driver to address this lack of information (*Lee et al., 2014; Chandraratna et al., 2005; Blatt & Furman, 1998; Cooper et al., 1995*). To gain such an insight, the socioeconomic characteristics associated with the residence of the drivers involved in the crash rather than examining only the location characteristics are considered (*Sagar et al., 2021; Chen et al., 2010*). This is more salient since the drivers involved are more likely to come from a different zone than that of the crash location.

A Kentucky study reported that unintentional injury death rates in Appalachia are higher than the western region of the Commonwealth (*Kearney et al., 1990*). Lower socioeconomic status has been associated with the higher fatality rate along with other factors such as regional topography and lack of access to immediate medical care. (*Zhu et al., 2013*) compared the traffic fatality rates in Appalachian and the non-Appalachian regions using Poisson models with age, gender, and county-specific population density levels as controlling variables. They found that the traffic fatality rates in the Appalachian region are 45 percent higher than non-Appalachian area. They attributed these higher rates largely to the higher rural population of the region. They also observed that the rates are higher in urban Appalachia compared to urban counties in the rest of the U.S.

A variety of methods have been used to investigate the factors associated on crash occurrence. Several regression techniques, clustering methods, and comparative grouping have been extensively used for this analysis. Several previous studies have adopted spatial techniques using GIS platforms to facilitate traffic safety studies. GIS can be used for crash analysis to examine spatial characteristics of accident locations. *Shahzad (2020)* proposed GIS techniques to display statistical data in geographic form and identify road crashes hotspots. *Budzyński et al. (2018)* described the use of GIS for visualization of road parameter data to identify high-risk sections.

The literature review indicated that several socioeconomic factors estimated from a driver's residence zip code are associated with crash occurrence. Income, education level, poverty percentage, employment, and the rurality of an area are such socioeconomic factors that are shown correlation with crash occurrence in addition to traditional factors of driver age and gender. Prior research has investigated how these variables affect a driver's propensity to cause a future crash, however, there has been minimal discussion on how the regional disparities between areas of different socioeconomic conditions impact crash occurrence. This study conducted a spatial analysis to investigate crash involvement trends of Kentucky drivers and regional disparities between the Appalachia and non-Appalachia area in the state.

### 3 Data

The current study utilizes crash data and spatial databases for the state of Kentucky. Crash data from 2013 through 2016 was obtained from the Kentucky State Police Collision Database (*KSP, n.d.*), while the shapefile of Kentucky at the zip code level and county level are obtained from web resources provided by the University of Kentucky and U.S. Census Bureau, respectively (*KGS, n.d.*; *US Census Bureau, n.d.*).

In the crash database, driver information such as driver's residence zip code is also included. The police officer responsible for reporting the crash determines the driver performance and human factors contributing to the crash occurrence. The human factors coded for each driver are used to determine their at-fault status. For each crash considered, the driver with a human factor code recoded by the police officer is assumed to be the at-fault driver for the crash occurrence (*Chandraratna & Stamatiadis, 2009*). In the crash database, multiple human factors are recorded (if any) for drivers involved in crashes. After various data processing in Python, the human factors recorded to the same driver are aligned to convert the multiple entries to a single entry. Age and gender of the driver are used as the factors to match the records. The first human factor recorded is used to determine the at-fault status of the driver. For each crash, the driver with the first human factor coded as 'non-detected' is considered the not-at-fault, while the driver with a human factor detected is considered the at-fault driver. The crashes in which a human factor code is recorded for both or neither driver are eliminated from the analysis. This selection criterion avoids multiple at-fault drivers for the same crash in two-unit crashes

(Chandraratna et al., 2005). In single-unit crashes, only drivers with a human factor coded are included in the dataset, and these drivers are coded as at-fault. As single-unit crashes have only one vehicle involved, there is no not-at-fault driver group involved in these crashes. As noted above, the not-at-fault driver group from the two-unit crashes is considered as the representative sample for exposure to a crash and as such it can be used as the denominator in the estimation of the relative accident involvement ratio, RAIR (see section 4.1), for the single-vehicle crashes as well (Chandraratna et al., 2005; Stamatiadis & Deacon, 1997).

The final dataset includes drivers with ages between 15 and 90 years. To analyze the RAIR of drivers in different age groups, a conventional categorization of young (aged <25 years), middle-aged (25-64 years), and old driver (>64 years) is used. Table 1 shows the distribution of age in the crash dataset used for the study. About 85 percent of the crash data is property damage only crashes, while the remaining 15 percent involve some type of injury (Fatal: 0.26 percent, Major Injury: 1.23 percent, Minor Injury: 4.76 percent, and Possible Injury: 8.72 percent).

**Table 1. Driver age and gender distribution of crash data**

Two-unit Crashes				Single unit Crashes			
	Age	At-fault	Not-at-fault		Age	At-fault	Not-at-fault
	<25	35,560	18,845		<25	15,424	5,896
<b>Male</b>	25-64	87,484	97,698	<b>Male</b>	25-64	29,597	30,052
	>64	6,610	3,876		>64	1,098	1,209
	<25	31,601	20,941		<25	9,587	6,482
<b>Female</b>	25-64	74,886	97,106	<b>Female</b>	25-64	18,043	29,992
	>64	5,609	3,284		>64	892	1,010
<b>Total</b>		<b>241,750</b>	<b>241,750</b>	<b>Total</b>		<b>74,641</b>	<b>74,641</b>

According to the 2016 American Community Survey, the median household income of Kentucky is about \$59,023 with an unemployment rate of 6 percent (US Census Bureau, 2016). Table 2 shows the distribution of age and gender of the driver population in Kentucky.

**Table 2. Driver age and gender distribution of driver population in Kentucky**

Age Groups	<25	25-64	>64
<b>Male</b>	308,530	1,147,209	288,975
<b>Female</b>	293,341	1,168,885	369,475
<b>Total</b>	<b>601,870</b>	<b>2,316,095</b>	<b>658,450</b>

## 4 Methodology

The study utilized Python for data management and several tools in ArcMap to develop the choropleth maps for the study. Several spatial tools were used to analyze the data visually.

### 4.1 Crash rate estimation

In traffic safety analysis, crash exposure is an important metric that defines the relative degree of risk on roads. This study utilizes the quasi-induced exposure technique developed by Carr (1969) to define crash exposure. This technique derives exposure estimates from not-at-fault drivers in two-vehicle crashes assuming that the not-at-fault drivers are a representative sample

of the total population in question. The quasi-induced approach makes two assumptions: 1) there is an at-fault and a not-at-fault driver in each two-vehicle crash, and 2) the not-at-fault drivers are randomly selected among the drivers/vehicles on the road at the time of the crash occurrence. With the second assumption, the technique assumes that the not-at-fault drivers are a random, i.e., representative, sample of the total population in question and thus uses them in estimating crash rates, i.e., as the denominator in the crash rate computation. The quasi-induced exposure is proved to be more reliable and applicable while examining safety issues pertained to specific driver groups. This method is widely accepted and the assumptions are tested and validated scientifically (*Zhang et al., 2021; Curry et al., 2016; Méndez & Izquierdo, 2010; Chandraratna & Stamatiadis, 2009; Lyles et al., 1991*). The crash rate measure of exposure is developed in terms of the relative accident involvement ratio (RAIR), which is the ratio of the percentage of at-fault drivers to the percentage of not-at-fault drivers from the same subgroup defined as

$$\text{RAIR} = \frac{\text{proportion of at-fault drivers}}{\text{proportion of not-at-fault drivers}} \quad (1)$$

The study utilizes the quasi-induced exposure technique to assess the relative risk of drivers to be at-fault in a crash. Hence, a driver's relative accident involvement ratio (RAIR) is calculated based on age, gender, and residence zip code. These ratios are calculated from the raw data, and a GIS mapping platform is used for visualizing and interpreting the results. Choropleth maps developed in this study are expected to provide insight into whether the at-risk drivers reside in regions with specific socioeconomic or demographic characteristics.

Using the quasi-induced technique, RAIRs of driver groups are estimated in relation to the socioeconomic factors at the county level and represented in the form of choropleth maps which is used to identify target groups/areas with high crash propensity. This would allow policymakers to focus their efforts to improve safety through targeted efforts and specific road safety campaigns.

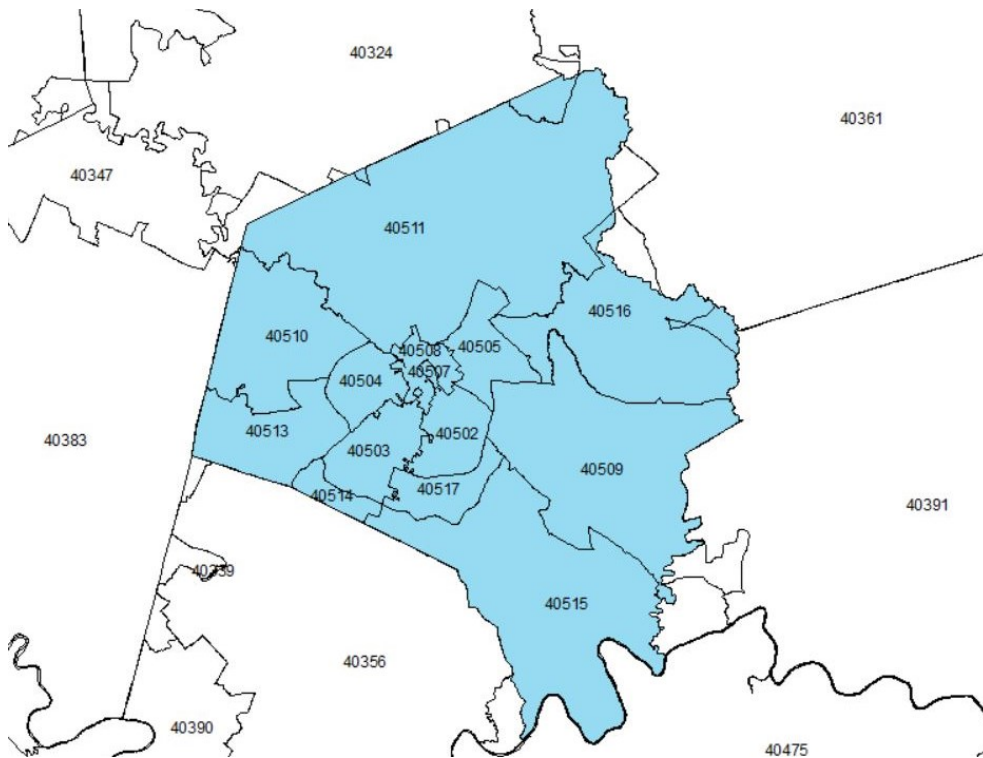
The RAIR were calculated from the crash database for three age categories - young (aged <25 years), middle-aged (25-64 years), and old driver (>64 years); and two gender categories - male and female, for every zip code.

In the next step, the calculated ratios were aggregated at the county level. There are counties that have more than one zip code as well as occasions where several zip codes are split between counties. Therefore, the ratios cannot be directly aggregated at the county level. A series of spatial processes are carried out using ArcMap to proportionally allocate the RAIR among the neighboring counties. This will allow for developing a weighted average based on the county population. As the first step to address this issue, the area of the zip codes that split between counties were calculated. Using GIS techniques, the zip code areas that coincides with the counties were extracted and the proportion of area that split between counties were also calculated. Next, the age and gender distributions of people at the zip code level were collected from the American Community Survey (ACS) database (*US Census Bureau, 2021*). The study requires the age and gender distributions of driver population at the county level, which was not available in a direct format. A surrogate approach was undertaken to address this. This approach was based on utilizing residents in a county over 16 years, since they are legally able to obtain a drivers license. In this case, the population of each county was grouped in the sane age and gender groups to be used as weights in the estimation of the weighted RAIR. This assumption implies that all residents over age 16 have a driver license and it may not hold true at all counties. However, these estimates are used only in the estimation of the weighted average



for the county and it is anticipated to have negligible, if any, impact in the weighted RAIRs and thus the choropleth maps developed.

As some of the zip codes split between counties, the proportion of drivers residing in areas within each portion needs to be calculated. For example, Figure 1 shows the boundaries of Fayette County which has 20 zip codes partially or completely coinciding with its border. About 1.80 percent of zip code 40324 falls into Fayette County while the remaining areas are shared with the neighboring counties. To calculate the weighted RAIR of Fayette County in the next step, the proportion of driver population residing in this portion of zip codes need to be estimated. So, the young population of drivers in zip code 40324 that fall into Fayette County is equal to 1.8 percent of the total young population of drivers in the zip code. Similarly, the age and gender distribution of all zip codes that split between counties were calculated. In this process, it was assumed that the population within each age and gender group follows a similar distribution throughout the zip code.



**Figure 1. Intersection of Zip Codes in Fayette County**

Finally, the weighted RAIR of drivers in each age and gender category for every county was calculated. A weighted RAIR approach was adopted for this purpose where the ratios of all zip codes within the county were weighted to the population, as defined above, in that zip code. Equation (2) is used to calculate the weighted RAIR:

$$\text{Weighted RAIR of } X \text{ county} = \frac{\sum_1^n \text{RAIR}_i \cdot P_i}{P_t}, \quad (2)$$

where  $n$  is the number of zip codes within the county;  
 $\text{RAIR}_i$  is RAIR of any category at the zip code  $i$ ;  
 $P_i$  is population of the category in zip code  $i$  within the county and  
 $P_t$  is total population of the county ( $= \sum P_i$ ).

## 4.2 Choropleth Map Development

Using the weighted RAIR estimated at the county level, choropleth maps were developed in relation with a suitable socioeconomic factor at the county level. Median household income was chosen as it was shown to be a representative of socioeconomic condition of a driver as it is correlated with other socioeconomic variables such as poverty and employment rate (*Sagar, 2020*). The data was collected from the 2016 five-year estimate of the American Community Survey (*US Census Bureau, 2021*). Prior research demonstrated household income to be a better predictor of income (*Lee et al., 2014; Stamatiadis & Puccini, 1999*) than other factors because it better determines a family's overall economic status.

To better analyze the data, a modified ratio was created by normalizing the weighted RAIR for each county by its median household income. That is, the RAIRs are divided by every \$10,000 of household income to create a ratio for each county. The normalized RAIR developed was used to create choropleth maps for the entire state. Normalization ratios are useful when other factors influence the numerical values used to calculate your proportional symbol sizes.

A 4-step process was followed to develop the maps. First, the RAIRs were calculated at the zip code level. Second, the ratios of the zip codes that split between counties were proportioned followed by aggregating the ratios at the county level (as explained in section 4.1). Next, the RAIRs at the county level were normalized by median household income. Lastly, the normalized RAIRs were used to develop the choropleth maps for each age and gender category. Figure 2 is a flow chart showing the major steps followed to develop the choropleth maps.

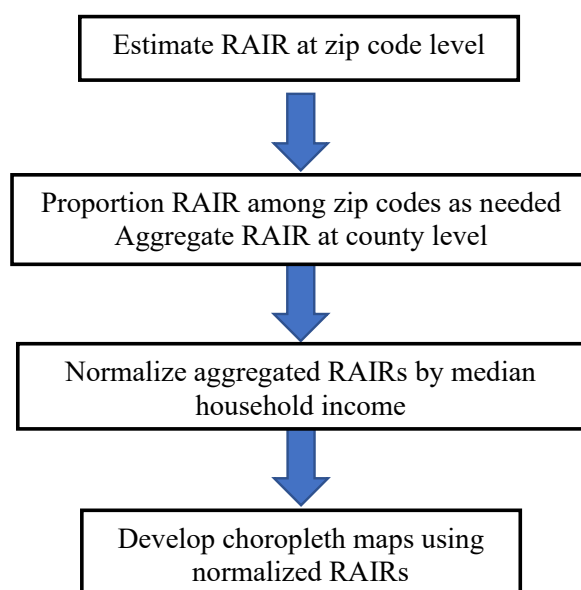


Figure 2. Flow chart of the choropleth maps development process

## 5 Results

Choropleth maps were generated for each county using the normalized RAIRs. In the following maps, counties were shaded to represent the crash involvement risk of drivers in various groups. As it was important to identify if drivers reside in areas where poverty or income could be of concern, the maps also display household income and the Appalachian regions. Median household income was shown on the maps as it is a socioeconomic factor widely recognized as influencing crash occurrences (*Sagar, 2020*). It is correlated with other socioeconomic variables such as poverty and employment rate. Maps were developed for both single- and two-unit crashes.

## 5.1 Two-Unit Crashes

Figure 3 shows the choropleth maps for the three age groups. The normalized RAIR for young drivers (<25 years) range from 0.243 to 0.863, while for middle age (25-64 years) and old age driver (>64 years), the values extend between 0.079 to 0.496 and 0.137 to 0.854. Across the state, the normalized RAIR of young drivers (<25 years) and old drivers (>64 years) are higher compared to the middle age groups (25-64 years). This imply that a young driver or old driver, when involved in a crash, has a greater probability of being the at-fault driver. The counties with relatively higher RAIR are mostly low-income areas in the Appalachia. Overall, the crash propensity of drivers in the Appalachia are higher compared to the drivers residing in the non-Appalachian counties.

Figure 4 shows the choropleth maps developed for male and female drivers. The normalized RAIR for male driver and female drivers ranges from 0.125 to 0.634 and 0.107 to 0.543. While the values for the two gender categories are similar, male drivers have comparatively higher crash propensity. Regardless of the gender of the driver, residents of the counties in the Appalachia have higher RAIR.

## 5.2 Single-Unit crashes

For single-unit crashes, normalized RAIR for young drivers (<25 years) range between 1.00 and 5.15 (Figure 5a), a wider spread than observed for two-unit crashes. Yet, the trend of young drivers diverges from the two-unit crashes. The counties with comparatively lower ratios are in the Appalachia for single-unit crashes. Compared to two-unit crashes, the normalized RAIR is larger for single unit crashes which indicates a greater crash propensity for all drivers. These drivers' inexperience and lack of judgment may explain this phenomenon. Among middle-aged drivers (Figure 5b), RAIR values for single-unit crashes range between 0.67 and 1.05. In Appalachia, risk levels are generally higher than the rest of the state. Many counties with high household median income have lower risk for young drivers compared to middle-aged drivers. For older drivers (Figure 5c), trends for single-unit crashes concur with two-unit crashes. RAIR values are from 0.037 to 0.363, with most counties in the Appalachia falling into high-risk category. While older drivers are less likely to cause a single-unit crash than drivers in the other age groups, they have a higher risk overall. This finding may result from the dataset having a relatively small number of older drivers. No regional or socioeconomic trends are observed for this category.

Figure 6a and 6b shift the focus to the role of gender in single-unit crash risk. The normalized RAIR for male drivers range from 0.152 and 0.622 while for females the range is between 0.085 and 0.48. These maps demonstrate that the likelihood of male drivers causing a single-unit crash is much higher than their propensity to be at fault in a two-unit crash. Female drivers once again have lower RAIRs than males, demonstrating they are less likely to be the cause of a single-unit crash.



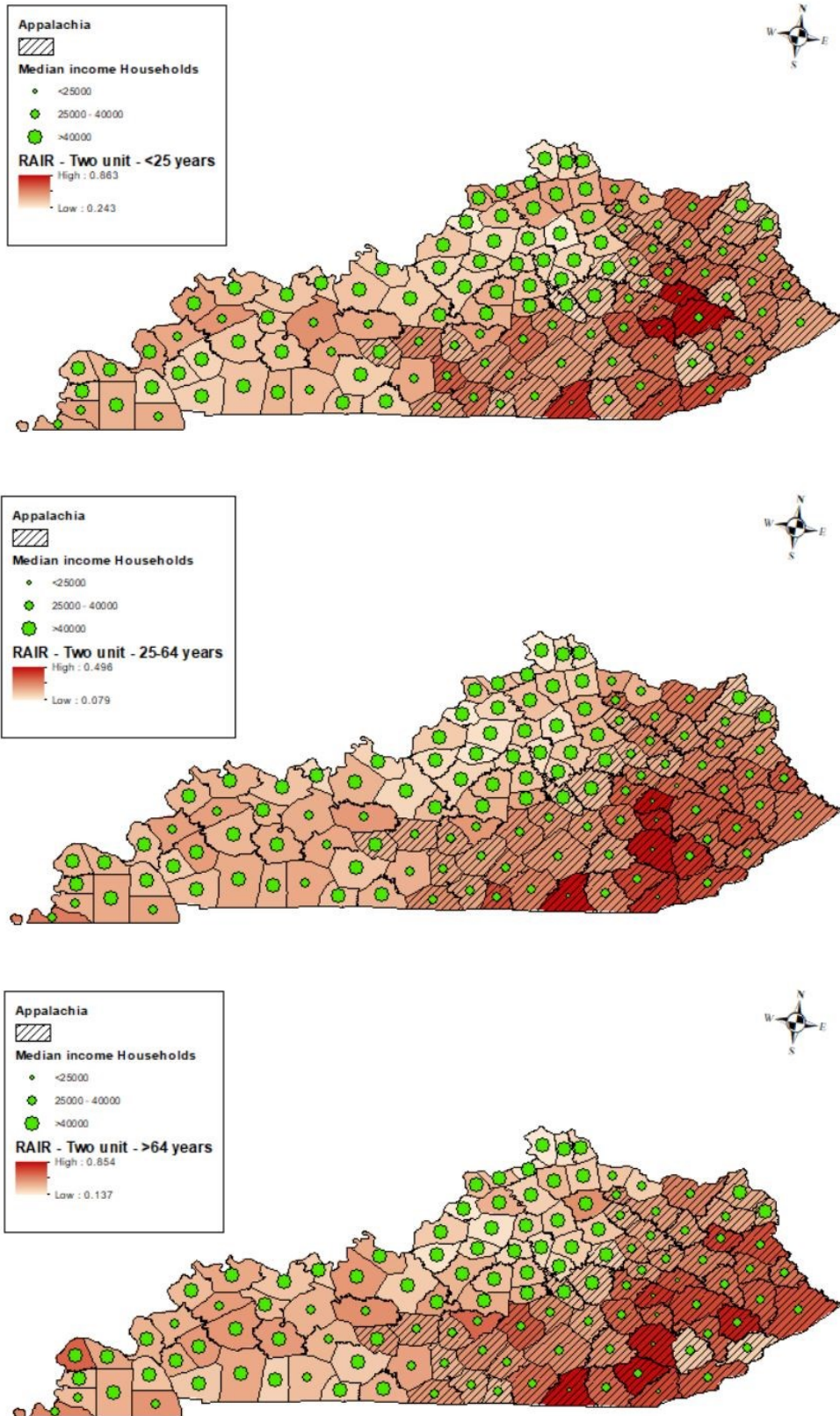


Figure 3. Choropleth map for age (<25, 25-64 and >64 years), two-unit crashes

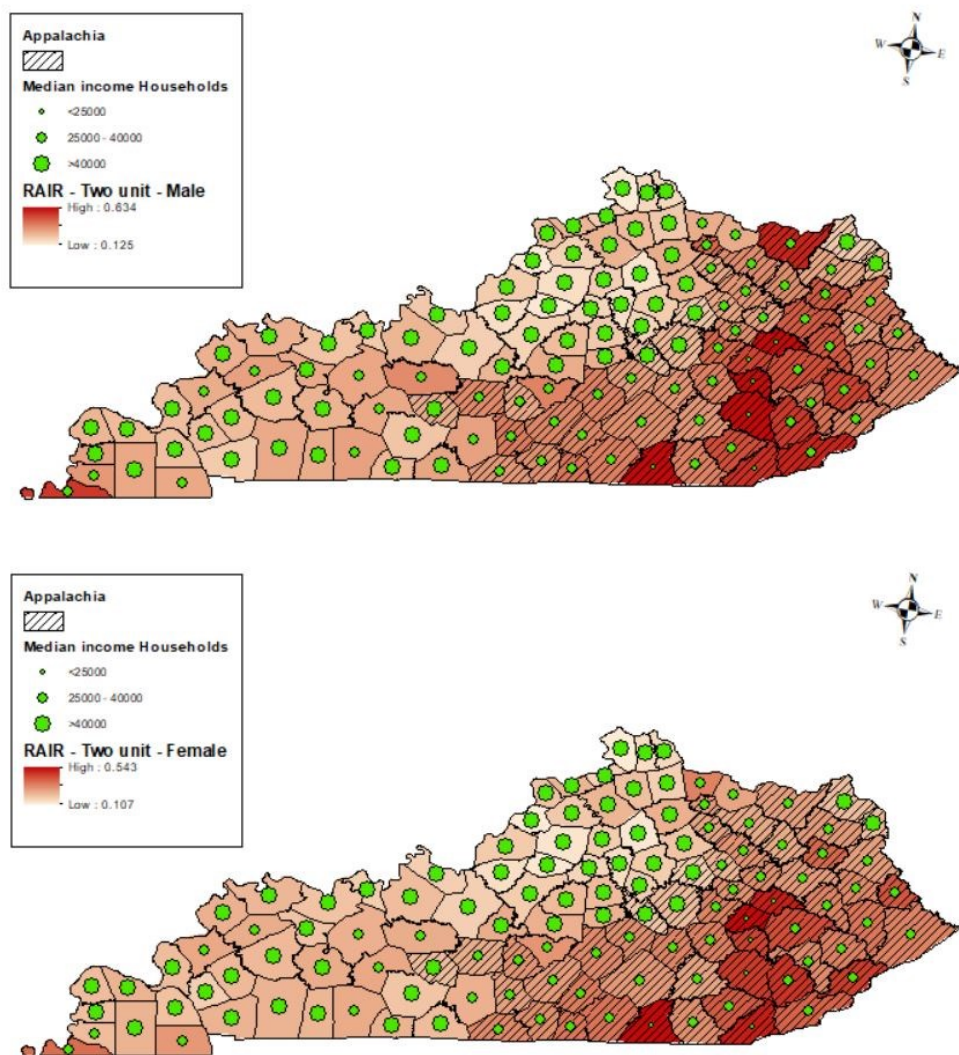


Figure 4. Choropleth map for gender, two-unit crashes

## 6 Application of choropleth maps

The findings of the spatial analysis can be used to identify high risk counties that can be targeted for safety programs. The Safety Circuit Rider (SCR) program of the Federal Highway Administration (FHWA) is a safety program that provides safety-related support to agencies responsible for local road safety with a goal of reducing the frequency and severity of roadway crashes (*Gross et al., 2009*). Kentucky implements the SCR program through the identification of six high risk counties annually and completion of a detailed crash data analysis and road safety audits on the county public roadways (*KTC, n.d.*). The goal of the program is to develop a set of countermeasures to reduce crashes at the identified high-risk areas. The choropleth maps developed from the study can be used to identify high risk counties for both two-unit and single-unit crashes, identified based on its driver's propensity to cause a crash.



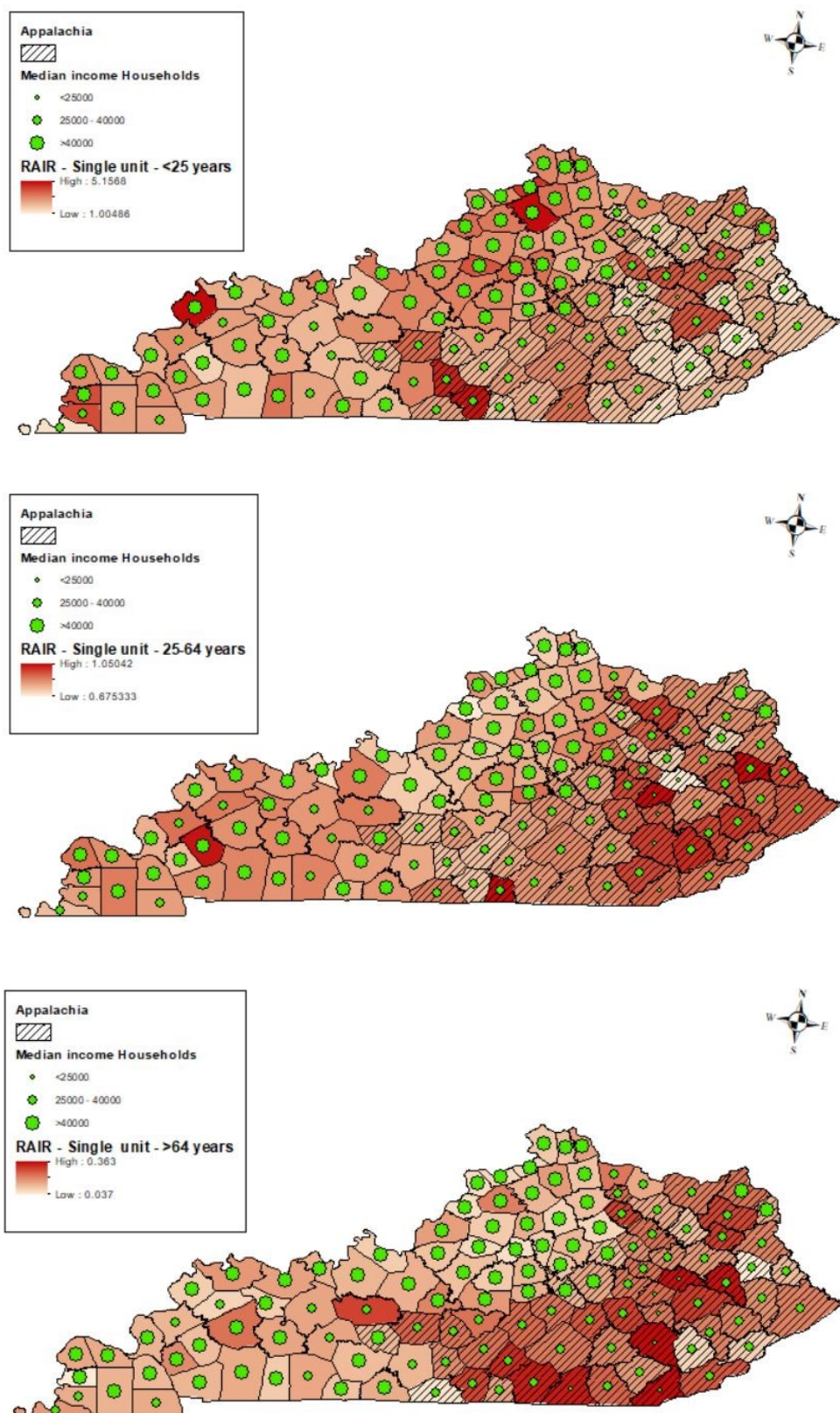


Figure 5. Choropleth map for age (<25, 25-64 and >64 years), single-unit crashes

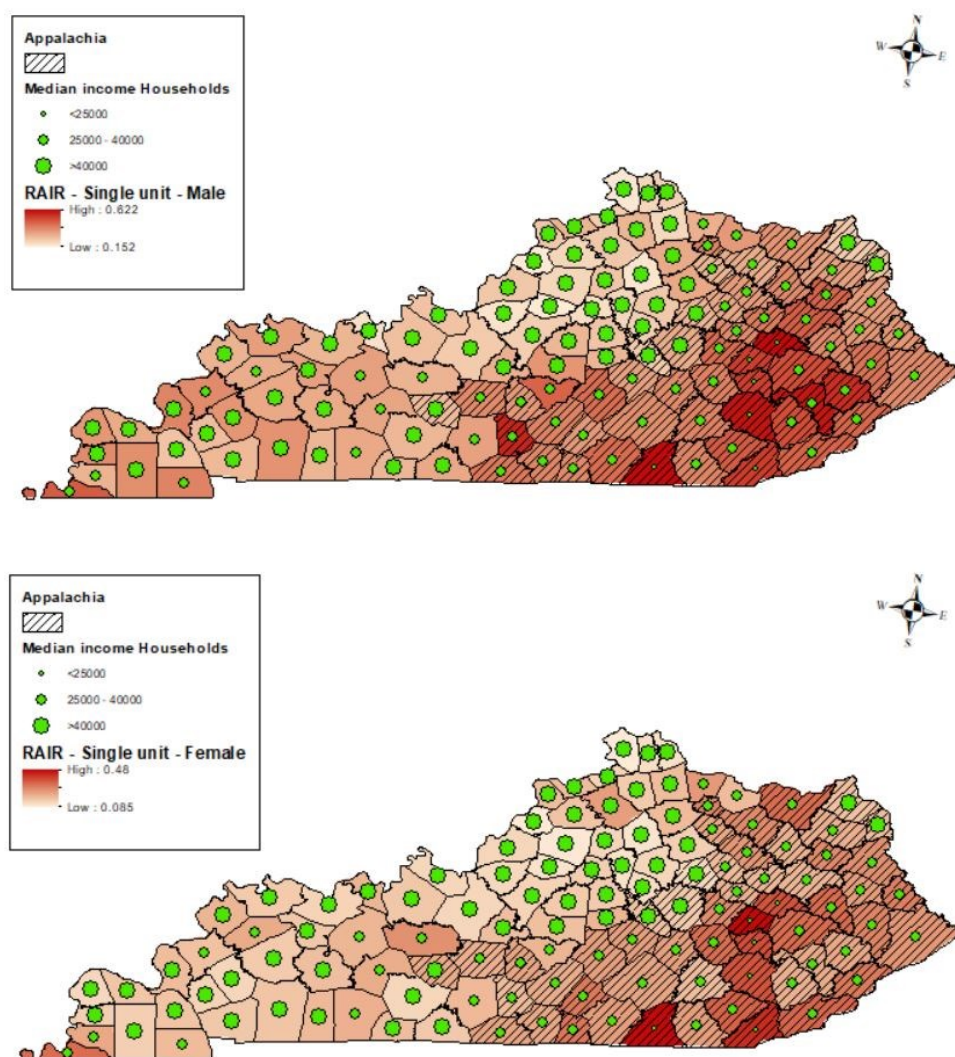


Figure 6. Choropleth map for gender, single-unit crashes

## 7 Discussion

This research aimed to identify the driver groups with higher likelihood to cause a crash, through spatial analysis. The study analyzed crash data using quasi-induced exposure technique and through a series of GIS techniques to develop choropleth maps representing the risk of different driver groups at the county level. A well-structured GIS process was followed for the development of the maps for two- and single-unit crashes.

The findings from the analysis were consistent with previous research on the relationships between driving behavior and other demographic factors such as age and gender (*Stamatiadis & Puccini, 1999; Stamatiadis & Deacon, 1997*). The results indicated that young drivers in two-unit crashes are more prevalent in low-income counties. Young male drivers show a higher risk than females and this could be attributed to the aggressive and risk-taking behavior of male drivers (*Stamatiadis & Puccini, 1999*). Compared to young drivers, middle-aged drivers are less probable to cause two-unit crashes. The crash propensity decreases with age, probably due to betterment in judgment and decision making gained through experience (*Sagar, 2020*). The crash propensity of drivers increases during old age. Above 65 years of age, drivers have a higher probability to be the at-fault driver, indicating the aging-related challenges that affect driving performance (*Staplin et al., 1999*). While the trends of single-unit crashes and two-unit crashes remain similar for middle age and old age drivers, they differ for young drivers. For

single unit crashes, young drivers in higher income counties are more likely to be the at-fault driver in a crash. One explanation for this trend is the relatively smaller dataset with a smaller number of young drivers in these counties. One could also argue that young drivers in higher income counties have a greater ability to obtain a vehicle. However, if this was true, a similar trend for young drivers in these counties would have also been observed for two-vehicle crashes. This does not hold and thus the association between county income and ability to own a car is tenuous. Across the state, female drivers are less likely to be involved in a crash (both single-unit and two-unit) than males. This agrees with prior research (*Stamatiadis & Deacon, 1997*). Additionally, the crash propensity for residents in the Appalachian counties is apparent based on the spatial analysis.

Regardless of age and gender, drivers in low-income counties have higher crash involvement, which concur to the findings the previous research (*Sagar et al., 2021*). Moreover, since most of the Appalachian counties are low income counties, drivers in these areas are more likely to cause crashes. The choropleth maps developed in this study can be used to identify the top at-risk counties in the state that then can be targeted for safety programs such as the Kentucky Circuit Rider program. This can be very helpful for improving safety and it can be considered the main contributions of the study.

It should also be noted that the study does not account multi-unit crashes involving three or more vehicles. The eliminated data accounts for about 9 percent of the overall crash data and its analysis could be potential future research. Also, this study did not consider several important socioeconomic variables, such as education, rurality and employment that are potential factors contributing to crash occurrence. The study utilized historic crash data which has several disadvantages. According to the 2018 reports of NHTSA, about 10 million or more crashes go unreported each year which are not accountable. Also, the police-reported crash data could lead to a bias in safety-related study, which is not preventable.

### **CRedit contribution statement**

**Nikiforos Stamatiadis:** Conceptualization, Funding acquisition, Methodology, Supervision, Validation, Writing—review & editing. **Shraddha Sagar:** Conceptualization, Methodology, Writing—original draft.

### **Declaration of competing interests**

The authors report no competing interests.

### **Acknowledgement**

The authors would like to acknowledge the support of the Kentucky Injury Prevention Center and the Kentucky Transportation Cabinet.

### **References**

- Adanu, E. K., R. Smith, L. Powell, S. Jones (2017) *Multilevel analysis of the role of human factors in regional disparities in crash outcomes*. *Accident Analysis & Prevention* 109, 10-17: <https://doi.org/10.1016/j.aap.2017.09.022>
- Agüero-Valverde, J., P. P. Jovanis (2006) *Spatial analysis of fatal and injury crashes in Pennsylvania*. *Accident Analysis & Prevention* 38 (3), 618-625: <https://doi.org/10.1016/j.aap.2005.12.006>
- Blatt, J., S. M. Furman (1998) *Residence location of drivers involved in fatal crashes*. *Accident Analysis & Prevention* 30 (6), 705-711: [https://doi.org/10.1016/s0001-4575\(98\)00014-1](https://doi.org/10.1016/s0001-4575(98)00014-1)



- Brown, K. T. (2016) *A safety analysis of spatial phenomena about the residences of drivers involved in crashes*. PhD thesis. Clemson University. Graduate School. 1746: [https://tigerprints.clemson.edu/all\\_dissertations/1746](https://tigerprints.clemson.edu/all_dissertations/1746)
- Budzyński, M., W. Kustra, R. Okraszewska, K. Jamroz, J. Pyrchla (2018) *The use of GIS tools for road infrastructure safety management*. E3S Web of Conferences 26, 00009: <https://doi.org/10.1051/e3sconf/20182600009>
- Carr, B. R. (1969) *A statistical analysis of rural Ontario traffic accidents using induced exposure data*. Accident Analysis & Prevention 1 (4), 343-357: [https://doi.org/10.1016/0001-4575\(69\)90080-3](https://doi.org/10.1016/0001-4575(69)90080-3)
- Chandraratna, S., N. Stamatiadis (2009) *Quasi-induced exposure method: Evaluation of not-at-fault assumption*. Accident Analysis & Prevention 41 (2), 308-313: <https://doi.org/10.1016/j.aap.2008.12.005>
- Chandraratna, S., N. Stamatiadis, A. Stromberg (2005) *Potential Crash Involvement of Young Novice Drivers with Previous Crash and Citation Records*. Transportation Research Record 1937 (1), 1-6: <https://doi.org/10.1177/0361198105193700101>
- Chen, H. Y., R. Q. Ivers, A. L. C. Martiniuk, S. Boufous, T. Senserrick, M. Woodward, M. Stevenson, R. Norton (2010) *Socioeconomic status and risk of car crash injury, independent of place of residence and driving exposure: results from the DRIVE Study*. Journal of Epidemiology & Community Health 64 (11), 998-1003: <https://doi.org/10.1136/jech.2009.091496>
- Chen, L.-H., S. P. Baker, E. R. Braver, G. Li (2000) *Carrying Passengers as a Risk Factor for Crashes Fatal to 16- and 17-Year-Old Drivers*. JAMA 283 (12), 1578: <https://doi.org/10.1001/jama.283.12.1578>
- Cooper, P. J., M. Pinili, W. Chen (1995) *An examination of the crash involvement rates of novice drivers aged 16 to 55*. Accident Analysis & Prevention 27 (1), 89-104: [https://doi.org/10.1016/0001-4575\(94\)00052-n](https://doi.org/10.1016/0001-4575(94)00052-n)
- Curry, A. E., M. R. Pfeiffer, M. R. Elliott (2016) *Validation of quasi-induced exposure representativeness assumption among young drivers*. Traffic Injury Prevention 17 (4), 346-351: <https://doi.org/10.1080/15389588.2015.1091072>
- Factor, R., D. Mahalel, G. Yair (2008) *Inter-group differences in road-traffic crash involvement*. Accident Analysis & Prevention 40 (6), 2000-2007: <https://doi.org/10.1016/j.aap.2008.08.022>
- FAHE (2019) *Appalachian Poverty*: <https://fahe.org/appalachian-poverty> (Acc. 4 Feb 2022)
- Gross, F., D. Nabors, R. Eck, M. Hood (2009) *Safety Circuit Rider Programs Best Practices Guide*. Federal Highway Administration Office of Safety, U.S. Department of Transportation. FHWA-SA-09-019: [https://safety.fhwa.dot.gov/local\\_rural/training/fhwas09019/fhwas09019.pdf](https://safety.fhwa.dot.gov/local_rural/training/fhwas09019/fhwas09019.pdf)
- Hanna, C. L., L. Laflamme, C. R. Bingham (2012) *Fatal crash involvement of unlicensed young drivers: County level differences according to material deprivation and urbanicity in the United States*. Accident Analysis & Prevention 45, 291-295: <https://doi.org/10.1016/j.aap.2011.07.014>
- Harrah, J. (2015) *Kentucky Metropolitan Areas Out-Perform Rural and Small Urban Areas*. The Community Research Collaborative Blog: <http://crcblog.typepad.com/crcblog/kentucky-metropolitan-areas-out-perform-rural-and-small-urban-areas.html> (Acc. 4 Feb 2022)
- Hasselberg, M., M. Vaez, L. Lucie (2005) *Socioeconomic aspects of the circumstances and consequences of car crashes among young adults*. Social Science & Medicine 60 (2), 287-295: <https://doi.org/10.1016/j.socscimed.2004.05.006>
- Ivan, J., A. Burnicki, K. Wang, S. Mamun (2016) *Improvements to road safety improvement selection procedures for Connecticut*. Joint Highway Research Advisory Council (JHRAC) of the University of Connecticut. JHR 16-328
- Kearney, P. A., L. Stallones, C. Swartz, D. E. Barker, S. B. Johnson (1990) *Unintentional Injury Death Rates in Rural Appalachia*. The Journal of Trauma: Injury, Infection, and Critical Care 30 (12), 1524-1532: <https://doi.org/10.1097/00005373-199012000-00016>
- KGS (n.d.) *Kentucky Geological Survey*. University of Kentucky: <http://www.uky.edu/KGS/gis/bounds.html> (Acc. 4 Feb 2022)
- KSP (n.d.) *Kentucky Collision Analysis*. Kentucky State Police: <http://crashinformationky.org> (Acc. 4 Feb 2022)
- KTC (2003) *Traffic Collision Facts 2002*. Kentucky Transportation Center, University of Kentucky: <https://doi.org/10.13023/ktc.rr.2002.n01>
- KTC (2018) *Traffic Collision Facts 2017*. Kentucky Transportation Center, University of Kentucky: <https://doi.org/10.13023/ktc.rr.2018.n01>
- KTC (n.d.) *Safety Circuit Rider Program*. Kentucky Transportation Center, University of Kentucky: <http://www.kyt2.com/local-services/safety-circuit-rider-program> (Acc. 4 Feb 2022)
- Lee, J., M. Abdel-Aty, K. Choi (2014) *Analysis of residence characteristics of at-fault drivers in traffic crashes*. Safety Science 68, 6-13: <https://doi.org/10.1016/j.ssci.2014.02.019>

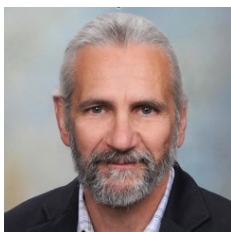


- Lyles, R. W., P. Stamatiadis, D. R. Lighthizer (1991) *Quasi-induced exposure revisited*. Accident Analysis & Prevention 23 (4), 275-285: [https://doi.org/10.1016/0001-4575\(91\)90005-p](https://doi.org/10.1016/0001-4575(91)90005-p)
- Méndez, Á. G., F. A. Izquierdo (2010) *Quasi-induced exposure: The choice of exposure metrics*. Accident Analysis & Prevention 42 (2), 582-588: <https://doi.org/10.1016/j.aap.2009.10.003>
- Murray, Å. (1998) *The home and school background of young drivers involved in traffic accidents*. Accident Analysis & Prevention 30 (2), 169-182: [https://doi.org/10.1016/s0001-4575\(97\)00088-2](https://doi.org/10.1016/s0001-4575(97)00088-2)
- NHTSA (2018) *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*. National Highway Traffic Safety Administration, U.S. Department of Transportation: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812506>
- Noland, R. B., L. Oh (2004) *The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data*. Accident Analysis & Prevention 36 (4), 525-532: [https://doi.org/10.1016/s0001-4575\(03\)00058-7](https://doi.org/10.1016/s0001-4575(03)00058-7)
- Noland, R. B., M. A. Quddus (2004) *A spatially disaggregate analysis of road casualties in England*. Accident Analysis & Prevention 36 (6), 973-984: <https://doi.org/10.1016/j.aap.2003.11.001>
- Sagar, S. (2020) *Effect of Socioeconomic and Demographic Factors of Driver Residence on Crash Occurrence*. University of Kentucky. Civil Engineering: <https://doi.org/10.13023/etd.2020.152>
- Sagar, S., N. Stamatiadis, A. Stromberg (2021) *Effect of Socioeconomic and Demographic Factors on Crash Occurrence*. Transportation Research Record 2675 (12), 80-91: <https://doi.org/10.1177/03611981211027887>
- Sagar, S., N. Stamatiadis, S. Wright, A. Cambron (2020) *Identifying high-risk commercial vehicle drivers using sociodemographic characteristics*. Accident Analysis & Prevention 143, 105582: <https://doi.org/10.1016/j.aap.2020.105582>
- Shahzad, M. (2020) *Review of road accident analysis using GIS technique*. International Journal of Injury Control and Safety Promotion 27 (4), 472-481: <https://doi.org/10.1080/17457300.2020.1811732>
- Stamatiadis, N., J. A. Deacon (1997) *Quasi-induced exposure: Methodology and insight*. Accident Analysis & Prevention 29 (1), 37-52: [https://doi.org/10.1016/s0001-4575\(96\)00060-7](https://doi.org/10.1016/s0001-4575(96)00060-7)
- Stamatiadis, N., G. Puccini (1999) *Fatal Crash Rates in the Southeastern United States: Why Are They Higher?* Transportation Research Record 1665 (1), 118-124: <https://doi.org/10.3141/1665-16>
- Staplin, L., K. H. Lococo, J. Stewart, L. E. Decina (1999) *Safety mobility for older drivers handbook*. National Highway Traffic Safety Administration
- US Census Bureau (2021) *American Community Survey*. United States Census Bureau: <https://www.census.gov/programs-surveys/acs> (Acc. 4 Feb 2022)
- US Census Bureau (n.d.) *Mapping Files*. United States Census Bureau: <https://www.census.gov/geographies/mapping-files.html> (Acc. 4 Feb 2022)
- Westerling, R., A. Gullberg, M. Rosén (1996) *Socioeconomic Differences in 'Avoidable' Mortality in Sweden 1986-1990*. International Journal of Epidemiology 25 (3), 560-567: <https://doi.org/10.1093/ije/25.3.560>
- Whitlock, G., R. Norton, T. Clark, M. Pledger, R. Jackson, S. MacMahon (2003) *Motor vehicle driver injury and socioeconomic status: a cohort study with prospective and retrospective driver injuries*. Journal of epidemiology and community health 57 (7), 512-516: <https://doi.org/10.1136/jech.57.7.512>
- Zhang, G., X. Jiang, X. Qiu, Y. Fan, C. Huang, M. Wei (2021) *Validating the underlying assumption of quasi-induced exposure technique disaggregated by crash injury severity*. Journal of Safety Research 76, 197-204: <https://doi.org/10.1016/j.jsr.2020.12.007>
- Zhu, M., S. Zhao, K. K. Gurka, S. Kandati, J. H. Coben (2013) *Appalachian versus non-Appalachian U.S. traffic fatalities, 2008-2010*. Annals of Epidemiology 23 (6), 377-380: <https://doi.org/10.1016/j.annepidem.2013.04.001>

## About the authors



Dr. **Shraddha Sagar** is a transportation engineer at Gresham Smith. She works as a consultant for transportation projects for several state DOTs (Department of Transportation). She graduated from University of Kentucky with her doctoral degree in 2020. Her primary area of interest focuses on various aspects of traffic engineering, roadway safety and data analytics. Dr. Sagar has published articles in peer-reviewed journals including *Accident Analysis & Prevention* and *Transportation Research Board Record*. She is an active member of professional organizations such as American Society of Civil Engineers, Institute of Transportation Engineers, and ITS TN.



**Nikiforos Stamatiadis** is a Professor of Civil Engineering at the University of Kentucky (UK) since 1990. He teaches several transportation courses at UK at the graduate and undergraduate level. Dr. Stamatiadis is considered a leader in the field of geometric design in the areas of context sensitive solutions and practical design/solutions concepts with an over 20-year experience in research on these topics. Dr. Stamatiadis has published articles in numerous journals including *Accident Analysis and Prevention*, *Journal of Applied Gerontology*, *Transportation Research Board Record*, and *Transportation Quarterly*. He is a member of several professional organizations and is actively involved as a member of several technical committees in the American Society of Civil Engineers, Institute of Transportation Engineers, and TRB.



All contents are licensed under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).