Exploration of socio-demographic characteristics and culture as factors in differential safety performance across geography

Final Report

Southeastern Transportation Center

Xuehao Chu, Steve Polzin, and Jodi Godfrey
April 2017

US Department of Transportation grant DTRT13-G-UTC34

DISCLAIMER

1. Report No.

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
Exploration of Socio-Demographic Characteristics and Culture as Factors in Differential Safety Performance across Geography

5. Report Date
April 2017

6. Source Organization Code
N/A

7. Author(s)
Chu, Xuehao; Polzin, Steven; Godfrey, Jodi

8. Source Organization Report No.

9. Performing Organization Name and Address

Southeastern Transportation Center
309 Conference Center Building
Knoxville, Tennessee 37996-4133
865.974.5255

10. Work Unit No. (TRAIS)

11. Contract or Grant No.
DTRT12-G-UTC34

12. Sponsoring Agency Name and Address

US Department of Transportation
Office of the Secretary of Transportation–Research
1200 New Jersey Avenue, SE
Washington, DC 20590

13. Type of Report and Period Covered
Final Report: September 2013 – April 2017

14. Sponsoring Agency Code
USDOT/OST-R/STC

15. Supplementary Notes: None

16. Abstract

Purpose: The objective was to seek a better understanding of the fatality rate (i.e., fatalities per unit of exposure) being historically higher in the Southeastern (SE) region than in non-Southeastern (NSE) regions of the United States.
Methods: This study explored how the exposure measured could affect the relative safety performance between the SE and NSE regions. More important, it explored sources and the degree of the differential safety performance in the overall fatality rate. Further, it explored how SE and NSE differ in the components of the fatality rate, i.e., the risk of involvement and the risk of death once being involved. Finally, it carried out linear regression analysis of general fatality rates and the risk of death for pedestrians.
Results: SE and NSE differences in the overall fatality rate did not vary much across several socio-demographic groups and time of day and day of week periods but did vary significantly across functional classes. In fact, rural non-freeway facilities accounted for almost 80% of the difference in the overall fatality rate. From the analysis of the components of the fatality rate, it was found that the higher fatality rate in SE was largely attributable to its higher risk of death. For pedestrians during 2009, for example, the fatality rate was 78% higher in SE than in NSE and the much higher fatality rate primarily resulted from its 85% higher risk of death. The linear regression analysis of general fatality rates indicate that SE and NSE differences largely disappear once differences in socio-demographic factors or in risk-taking behaviors were controlled for. The linear regression analysis of the pedestrian risk of death indicates that a 10% drop in statewide average speed of vehicle travel could reduce of the risk of pedestrian death by 3.1 percentage points, which is significant relative to the average risk of pedestrian death at 7.99% in the SE region.

17. Key Words
Southeastern Region; States; Fatalities; Pedestrians; Panel Data

18. Distribution Statement

Unrestricted

19. Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
Unclassified

21. No. of Pages
#64

22. Price
N/A

Form DOT F 1700.7 (8-72)

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Consisting of Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee, the Southeastern (SE) region of the United States had been known to have consistently some of the highest number of total fatal crashes and mileage-based fatality rates in the country. It also has many of the states and metropolitan areas that had been considered to be the most dangerous to pedestrians. Prior studies on the differential safety performance for the SE region are few and limited. One single study of inter-regional differences was primarily interested in the role of seat-belt use, vehicle travel by functional classification, and speed limits. Several other studies of traffic safety issues for the SE region focused on intra-regional rather than inter-regional differences. This study contributes to the literature on understanding the differential safety performance for the SE region through a range of exploratory analyses. The following are some highlights from each of these analyses:

- **Variability in regional differences of fatality rates**: SE and NSE differences in the fatality rate based on person hours traveled did not vary much during 2009 across gender and age groups or across gender and race-ethnicity groups, or across time of day and day of week periods. They do vary significantly, however, across functional classifications, particularly between rural versus urban areas and between freeway versus non-freeway facilities in a given area type.
- **Relative role of different travel segments**: Both the fatality rate and its share of exposure for a travel segment determine its contribution to the overall difference in fatality rates between SE and NSE regions. During 2009, male driving during late evening hours contributed one quarter of the overall difference in fatality rates between SE and NSE regions. During 1994-2013, driving on non-freeway facilities in rural areas account for almost 80% of the overall difference in fatality rates between SE and NSE states.
- **Role of risk components for all person types**: The mortality rate of all person types was 57% higher during 2007-2009 in SE than in NSE, resulting from 21% higher per

capita exposure and 30% higher fatality rate. In addition, the 30% higher fatality rate in SE primarily resulted from its 27% higher risk of deaths.

- **Regression analysis of fatality rates for all persons**: Differences in system and weather conditions do not account for the much higher fatality rates in SE. Differences in socio-demographic factors or in general risk-taking behaviors do. All four groups of explanatory variables account for 80% of the state-to-state variation in the fatality rate. When all four groups were accounted for, the fatality rate actually is lower in SE than in NSE regions.
- **Alternative exposure measures for pedestrian fatality rates**: Exposure measures for pedestrian fatality risks should account for both vehicle and pedestrian activities. A reasonable metric appears to be the square root of the product of vehicle and pedestrian activities. For frequent measurement, a reasonable choice would be to use VMT for vehicle activities and the product of resident population and the share for walking for commuting from the American Community Survey (ACS) as pedestrian activities.
- **Decomposition of mortality rates for pedestrians**: The pedestrian mortality rate in 2009 was 44% higher in SE than in NSE regions and this higher mortality rate in SE resulted primarily from 85% higher risk of pedestrian deaths in SE.
- **Regression analysis of the risk of pedestrian deaths**: While socio-demographic factors and other included variables being controlled for, a 10% drop in average speed at the state level could reduce the risk of pedestrian deaths by 3.135 percentage points.

DESCRIPTION OF PROBLEM

The Southeastern (SE) region of the United States had been known to have consistently some of the highest number of total fatal crashes and mileage-based fatality rates in the country (Washington et al., 1999; Stamatiadis and Puccini, 2000; Dixon 2005; Wang, 2006). The SE region also has many of the states and metropolitan areas that had been considered to be the most dangerous to pedestrians (McCann and DeLille, 2000; Ernst and Shoup, 2009). The SE region consists of the eight states of Standard Federal Region 4, one of ten such regions designated by the Office of Management and Budget (1974). Figure 1 shows the eight states of the SE region and all other non-Southeastern (NSE) regions.

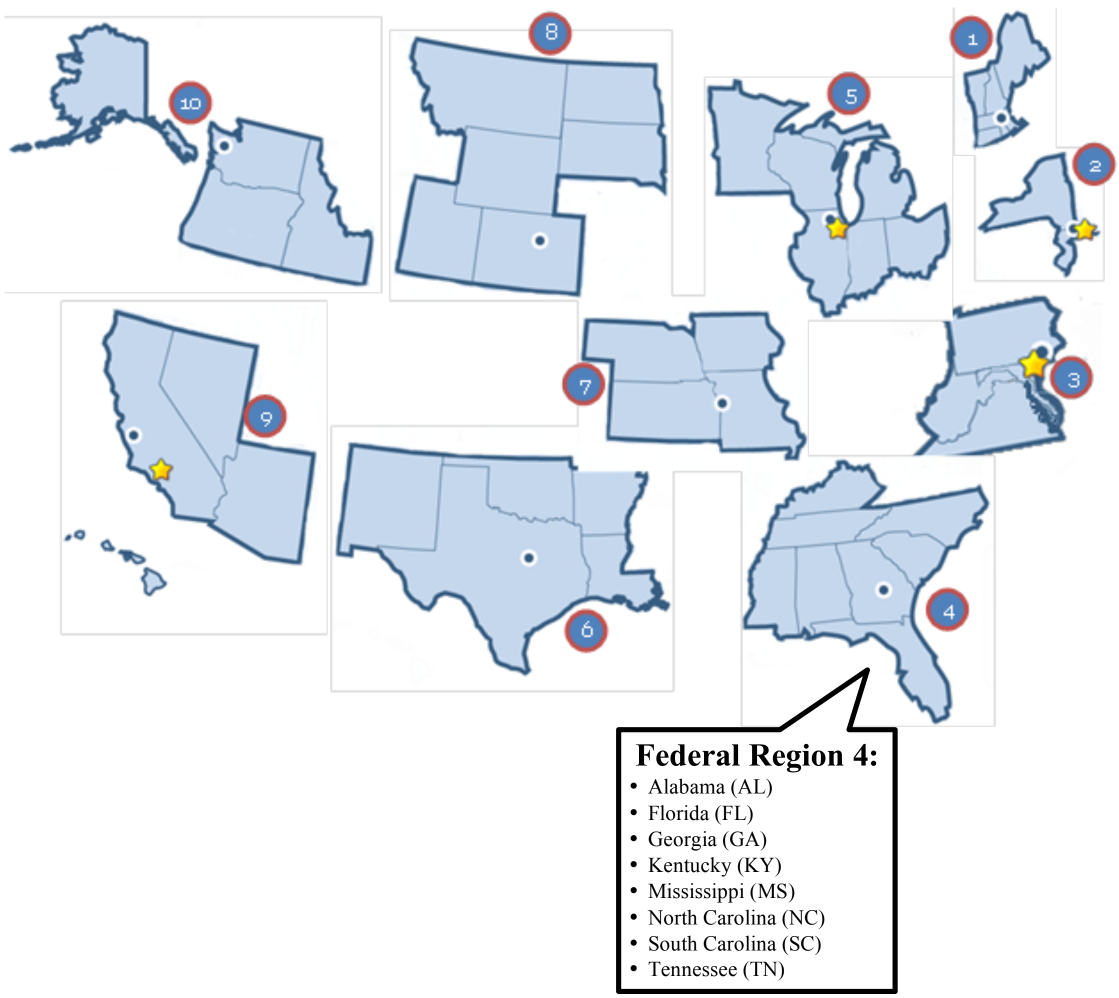


Figure 1. Federal Regions

Prior studies on the differential safety performance for the SE region are few and limited. Washington et al. (1999) tested the statistical significance of the difference in mileage-based rates of fatal crashes between SE and non-Southeastern (NSE) regions and concluded that the elevated rates of fatal crashes in the SE region in fact were statistically significant during 1995. They also suggested that these statistically significant differences are related to seat-belt use, vehicle travel by functional classification, and speed limits. All other four prior studies, Stamatiadis and Puccini (2000), Dixon (2005), Wang (2006), and Lambert and Meyer (2006), were interested in traffic safety issues for the SE region but focused on intra-regional differences rather than inter-regional differences. Stamatiadis and Puccini (2000) were interested in socioeconomic and demographic factors but not cultural factors. Lambert and Meyer (2006) studied the role of land-use patterns and the quality of emergency medical service in the number of fatal crashes per capita across metropolitan areas in the SE region. Both Dixon (2005) and Wang, however, primarily were interested in the role of roadway characteristics such as widening shoulders, enhancing delineation, protecting clear zones, etc.

Beyond the SE and NSE regions, there is a general literature on understanding differential safety performance across geographies. The brief review here focuses on the approach and statistical methodology. The studies in this literature differ in the dependent variable of interest, in the level of geography, in the methodology, etc. The dependent variable can be in the form of rates (Traynor, 2008; Eksler et al., 2008) or in the number of fatalities and injuries (Noland, 2003). The level of geography can be at the county level within a single state (Aguero-Valverde and Jovanis, 2006; Traynor, 2008), at the county level within the entire US (Keeler, 1994; Aguero-Valverde and Jovanis, 2006), at the level of metropolitan areas (Dewey et al., 2003), at the state level (Noland, 2003), and even at the country level (Eksler et al., 2008). The methodology used varies widely across the studies, including descriptive (Washington et al., 1999), linear regression (Grabowski and Morrisey, 2004), count-data models (Noland, 2003), Bayesian ecological regression (Eksler et al., 2008), etc.

This study contributes to the literature on understanding the differential safety performance for the SE region through a range of exploratory analyses. Methodologically, this study contributes to the general literature by presenting a simple approach to attributing overall geographic differences in rates (e.g., fatality rates) to individual travel segments (e.g., gender) and by using the square root of the product of both vehicle and pedestrian activities as a measure of exposure for pedestrian crash and injury risks.

APPROACH AND METHODOLOGY

Introduction

The overall approach and methodology for this study was exploratory. Specifically, this study approached the differential safety performance between the SE region and NSE regions from several angles:

- Variability in regional differences of fatality rates
- Relative role of different travel segments
- Role of risk components for all person types
- Regression analysis of fatality rates for all persons
- Alternative exposure measures for pedestrian fatality rates
- Decomposition of mortality rates for pedestrians
- Regression analysis of the risk of pedestrian deaths
- Use of a panel dataset

These are explored in some detail in individual sub-sections with their own data and methodology used.

Variability in Regional Differences of Fatality Rates

The issue explored was whether and to what degree the differential safety performance between the SE region and NSE regions vary over time, across travel conditions, and among socio-demographic groups. The issue is whether differential safety performance is relatively uniform across these conditions and groups or whether it is particularly different for some of them.

The focus was on fatalities and on those travel conditions and socio-demographic groups that can be defined with variables readily available in the Fatality Analysis Recording System (FARS). Socio-demographic groups were defined by person age, gender, and race and ethnicity. Travel conditions were based on day of week, time of day, and functional classification of roadways.

Safety performance was measured by fatalities per unit of exposure. For variability across travel conditions and socio-demographic groups, it was for a single year of 2009. Exposure was measured by person hours traveled (PHT). PHT data were estimated from the

2009 National Household Travel Survey (NHTS). Because PHT data estimated from the NHTS are for household travel only, fatalities in this case were limited to pedestrians, bicyclists, and occupants of light vehicles. Fatalities of vehicle occupants for non-household travel (commercial vehicles, international tourists, etc.) were included as well due to lack of information for exclusion.

For temporal variability, it was for the 20-year period from 1994 to 2013. The measure of exposure was in terms of vehicle miles traveled (VMT). VMT data by functional classification were directly from the annual *Highway Statistics* series of the Federal Highway Administration (FHWA). Fatalities of all person types were included for this analysis.

With three separate graphs, Figure 2 shows the results on the variability of the difference in PHT-based fatality rates between SE and NSE regions during 2009. The top graph is for both gender and age groups; the middle one is for both gender and race-ethnicity; and the bottom one is for both day of week and time of day.

For either SE or NSE, the difference in the vertical scales across the three graphs reflects the degree of variation in the fatality rate across the time periods and among the socio-demographic groups. Fatality rates vary the most across the time period, peaking just after midnight. Fatality rates vary significantly among gender and age groups, peaking for the very old and for young adults. Fatality rates vary much less among the race and ethnicgroups.

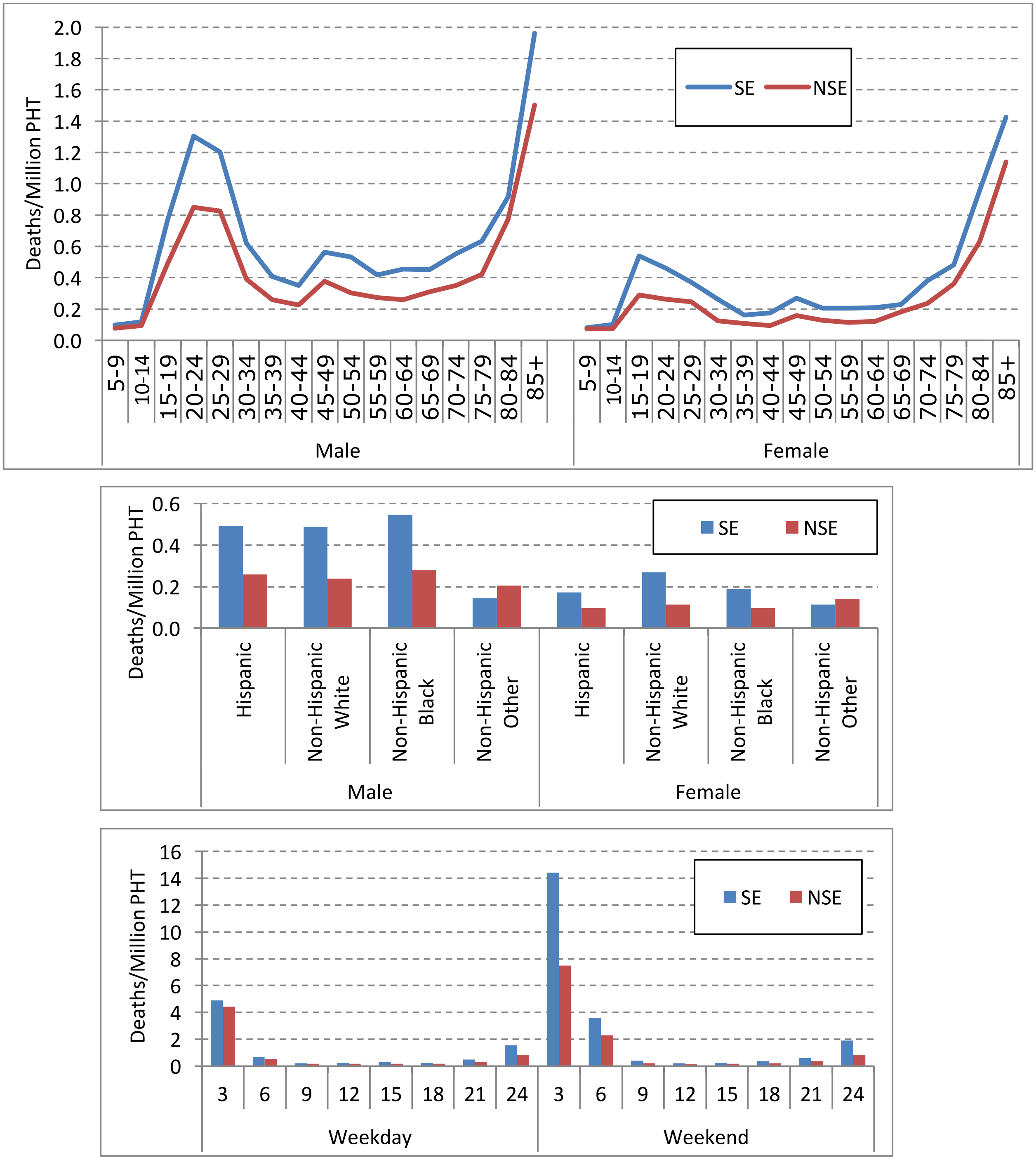


Figure 2. Fatalities per Million Person Hours Traveled (PHT). *Top*: by gender and age; *Middle*: by gender and race and ethnicity; and *Bottom*: by day of week and time of day.

Except for the Female Non-Hispanic Other group, the fatality rate is consistently much higher for SE than for NSE. Excluding Non-Hispanic other races, the PHT-based fatality rate during 2009 was higher in the SE region than in NSE regions as a whole for each gender-age group, each day of week and time of day combination, and every combination of gender and race-ethnicity. That is, the observed differential safety performance between SE and NSE at the aggregate level persists at these less aggregated levels.

Also with three graphs, Figure 3 shows the results on the variability of the difference in VMT-based fatality rates between SE and NSE regions from 1994 to 2013 and across functional classes. The top graph is for urban versus rural areas; the middle one is for freeway versus non-freeway facilities in rural areas; and the bottom one is for freeway versus non-freeway facilities in urban areas.

Variability of the difference in VMT-based fatality rates between SE and NSE regions are different over time versus across functional classes. The difference stayed largely similar over this 20-year period. The difference, however, varied significantly across functional classes, at least for the aggregated classes considered:

- The absolute difference in VMT-based fatality rates is significantly greater in rural areas than in urban areas (top graph).
- For rural areas, the absolute difference is significantly greater for non-freeways than for freeways (middle graph).
- For urban areas, the absolute difference is similar in magnitude between freeways and non-freeways (bottom graph).

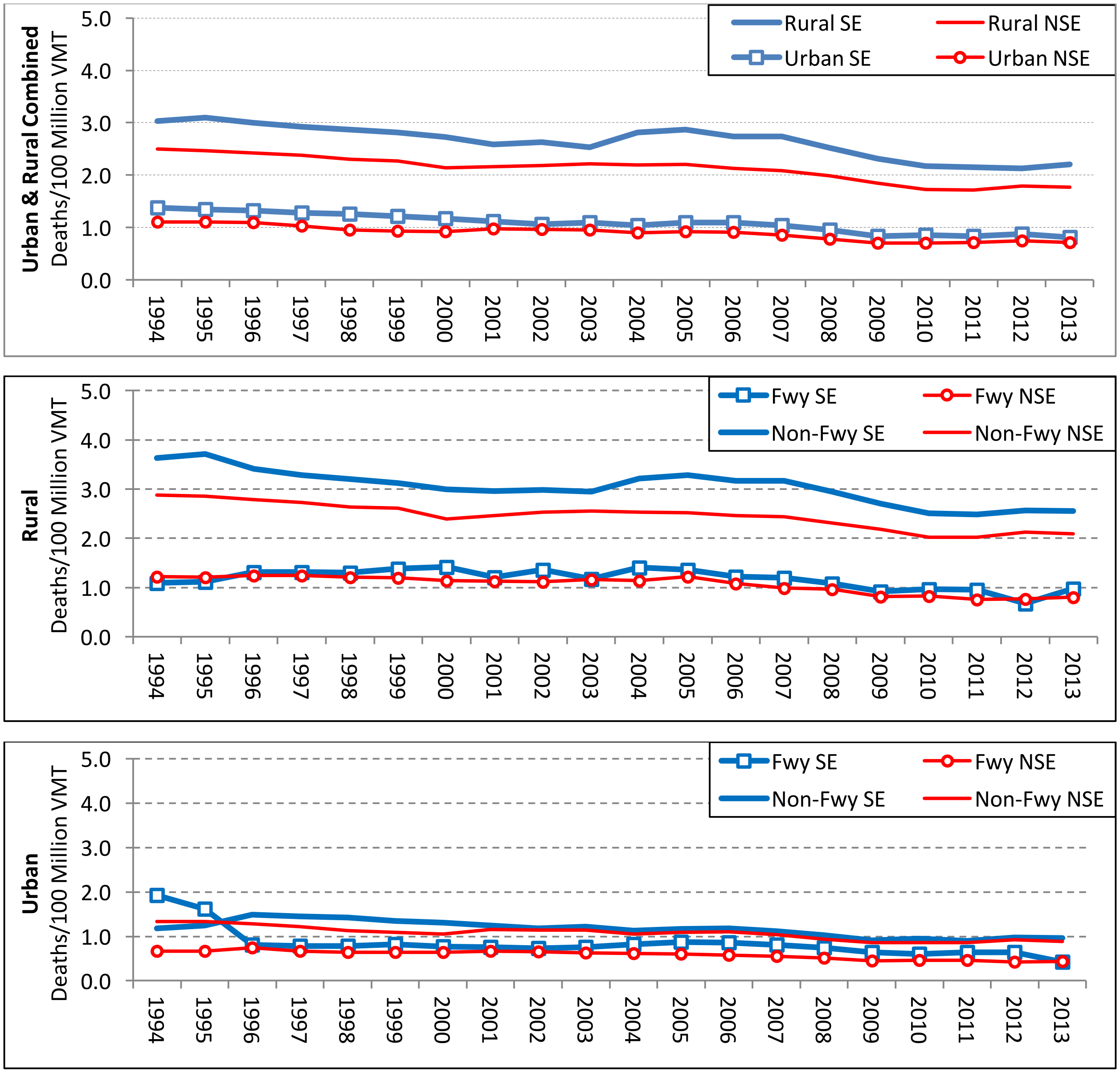


Figure 3. Trends in Fatalities per 100 Million VMT by Functional Class. *Top*: by urban versus rural; *Middle*: by freeway versus non-freeway in rural areas; and *Bottom*: by freeway versus non-freeway in urban areas. The 1994 and 1995 data for Fwy SE probably were due to some unknown errors.

Relative Roles of Different Travel Segments

The issue explored was the relative roles of different travel segments and their contribution to the overall difference in the fatality rate between SE and NSE regions. From the previous section on the variability of the SE and NSE difference in fatality rates, the absolute difference is much higher for some groups or travel conditions than for others. The question was whether those groups or travel conditions with a much higher absolute difference in fatality rates necessarily contribute more to the overall SE and NSE difference in fatality rates. Understanding this relative contribution of various groups and travel conditions is important in developing effective countermeasures for reducing fatalities.

Consider the bottom graph in Figure 2 as an example. For either SE or NSE regions, the fatality rate is dramatically higher during late evening and early morning hours. Further, the absolute SE and NSE difference in the fatality rate is also significantly higher during those hours. Do these hours also contribute the most to the overall difference in the fatality rate between SE and NSE regions?

Taken from Chu (1999), the exploration started with a mathematical relationship that links the overall difference in fatality rates between regions with the differences in fatality rates for individual travel segments. In fact, the overall fatality rate for a given region, R, can be expressed as follows:

$$R = \frac{F}{E} = \frac{\sum_i F_i}{\sum_i E_i} = \sum_i \frac{F_i}{\sum_i E_i} = \sum_i \frac{F_i}{\sum_i E_i}\left(\frac{E_i}{E_i}\right) = \sum_i \left(\frac{F_i}{E_i}\right)\left(\frac{E_i}{\sum_i E_i}\right) = \sum_i R_i S_i$$

with

F = overall number of fatalities

E = overall amount of exposure

i = segment i

F_i = fatalities for segment i

E_i = exposure for segment i

R_i = fatality rate for segment i

S_i = share of exposure for segment i

While Chu (1999) applies this relationship to temporal changes, the current analysis applies to regional differences. Specifically, the relative contribution of segment i to the overall difference in fatality rates between SE and NSE regions, $R^{SE} - R^{NSE}$, can be determined with the following:

$$\frac{R_i^{SE} S_i^{SE} - R_i^{NSE} S_i^{NSE}}{R^{SE} - R^{NSE}} 100 .$$

It is clear from this formula that both the fatality rate for a specific segment and its share of exposure determine the relative contribution of this segment on the overall difference. This formula was then applied to the 2009 data for the same socio-demographic groups and travel conditions.

Figure 4 shows the results for the various combinations of gender and age groups. Showing the fatality rates for each of these travel segments in each region, the bottom graph repeats the top graph in Figure 2. The middle graph shows the distribution of exposure across the same gender and age travel segments. The top graph shows the relative role of these travel segments to the overall difference in fatality rates between SE and NSE regions. Males aged from 20 to 24 contributed 9.2% to the overall SE and NSE difference in fatality rates. In contrast, females aged from 20 to 24 contributed just over 3.4%. Though having the highest fatality rate among all age groups of a given gender, the very old population, 85+ for both males and females, contribute little to the overall SE and NSE difference in fatality rates. This is because they just do not drive enough to make a large impact on the overall SE and NSE difference in fatality rates.

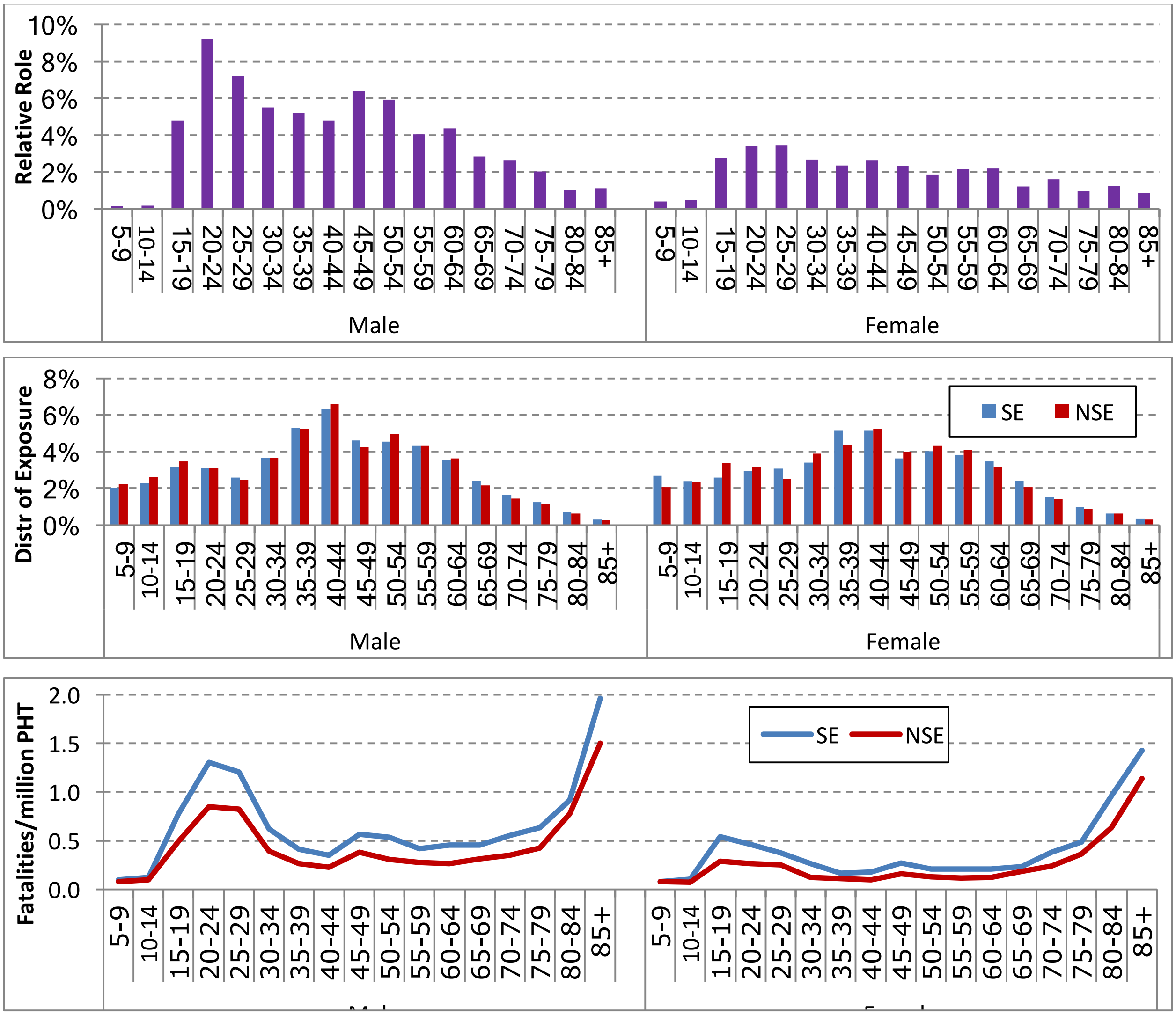


Figure 4. Relative Roles of Gender and Age Groups. *Bottom*: fatalities per million person hours traveled (PHT) by gender and age travel segments; *Middle*: distribution of exposure across the same gender and age travel segments; *Top*: the relative role of these travel segments to the overall absolute difference in fatality rates between SE and NSE regions.

Figure 5 shows the results for the aggregated functional classes used before for the 20-year period from 1994 to 2013. As discussed with respect to Figure 3, non-freeways in rural areas are likely to play a significant role in the overall difference in fatality rates between SE and NSE regions. While not quantified then, this significant role of non-freeways in rural areas is quantified here. Almost 80% of the overall SE and NSE difference in the fatality rate is attributable to rural non-freeway facilities.

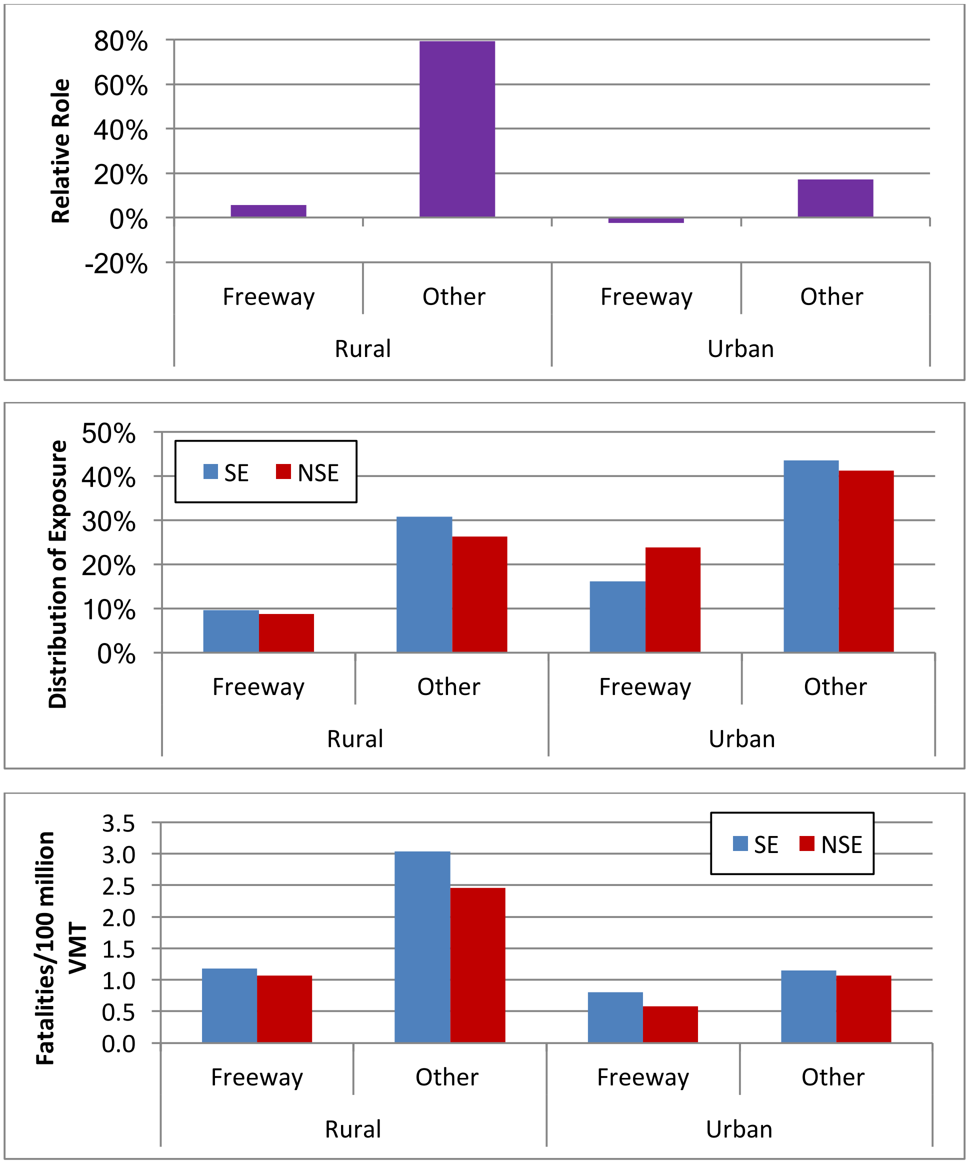


Figure 5. Relative Roles of Functional Classes. *Bottom*: fatalities per 100 million VMT by functional class; *Middle*: distribution of VMT across the same functional classes; *Top*: the relative role of these functional classes to the overall absolute difference in fatality rates between SE and NSE regions.

Figure 6 shows the results for the various combinations of gender and time-of-day segments. Similar to the earlier graphs, they show from the top to the bottom the relative roles of travel segments defined by gender and time of day, the relative distributions of exposure, and fatality rates for 2009. Traveling between midnight and 3 PM had the highest fatality rates. Because of the little account of traveling during these early morning hours, however, traveling in this period did not have the highest contribution to the overall SE and NSE difference in fatality rates. Males traveling between 9 PM and midnight did. In fact, they contributed more than 25% to the overall SE and NSE difference in fatality rates.

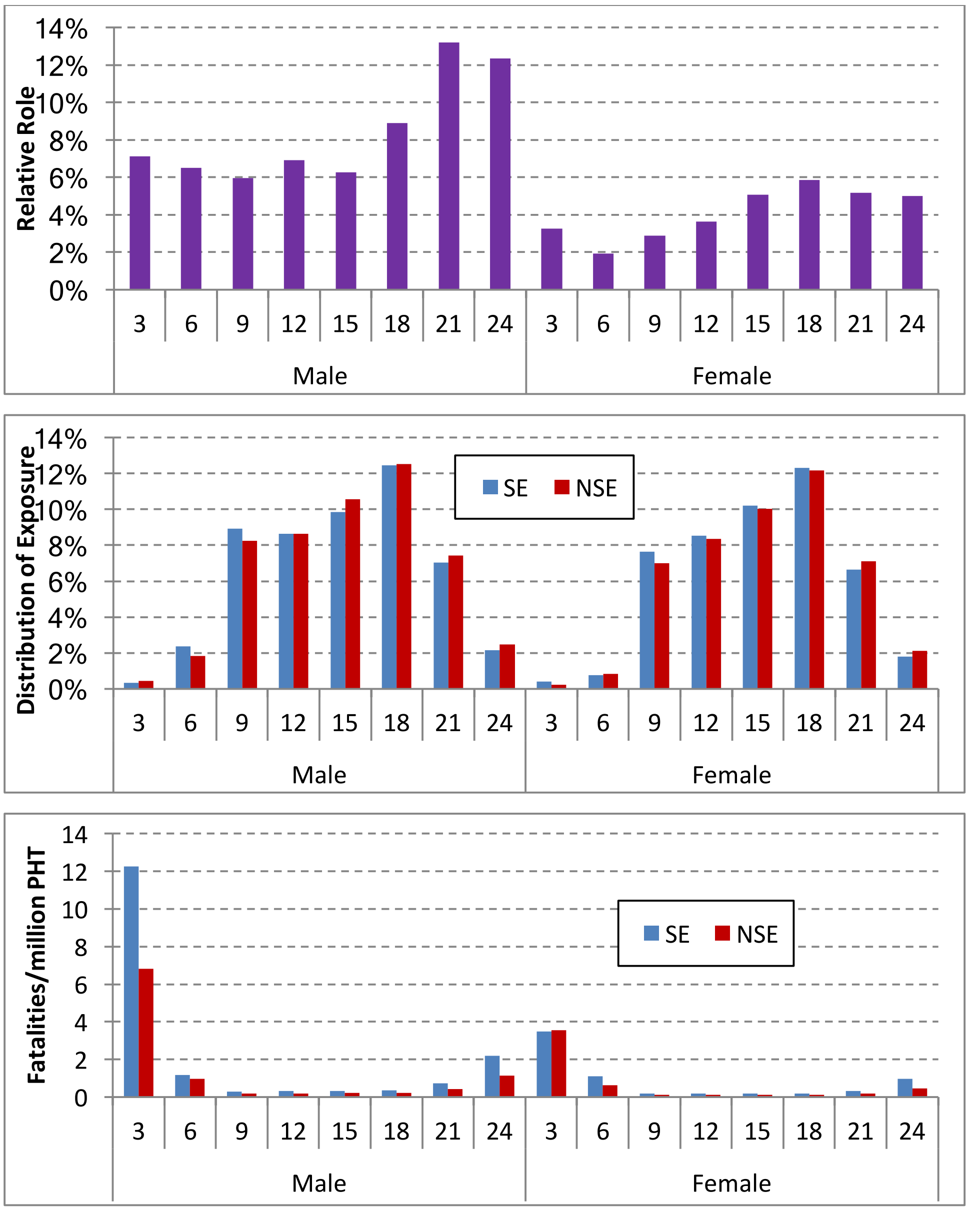


Figure 6. Relative Roles of Gender and Time of Day. *Bottom*: fatalities per million person hours traveled by gender and time-of-day segments; *Middle*: distribution of exposure across the same travel segments; *Top*: the relative role of these travel segments to the overall absolute difference in fatality rates between SE and NSE regions.

Role of Risk Components for All Person Types

The issue explored here was how the SE region is different from NSE regions in the risks for the different components of the crash and fatality process. Instead of directly comparing mortality rates (fatalities per capita) or fatality rates (fatalities per unit of exposure), further insights may be gained from decomposing these rates and comparing each component.

The method of decomposing mortality rates or fatality rates is frequently used in the literature. Baker and Li (1996) use it to explore the gender discrepancy in deaths rates from bicycling injuries. Dellinger et al. (2002) investigate the potential benefit of crash prevention interventions aimed at different components of the fatal crash involvement rate for older drivers. Goldstein et al. (2011) use it to explain regional disparities in traffic mortality. Zhu et al. (2013) use it to explore the gender discrepancy in per capita pedestrian deaths.

For the current analysis, the mortality rate was decomposed into the product of the fatality rate and per capita VMT:

$$\text{Mortality Rate} = \underbrace{\frac{\text{Fatalities}}{\text{VMT}}}_{\boxed{\text{Fatality Rate}}} \cdot \underbrace{\frac{\text{VMT}}{\text{Population}}}_{\boxed{\text{Per Capita VMT}}}$$

The fatality rate was further decomposed into the following three multiplicative components:

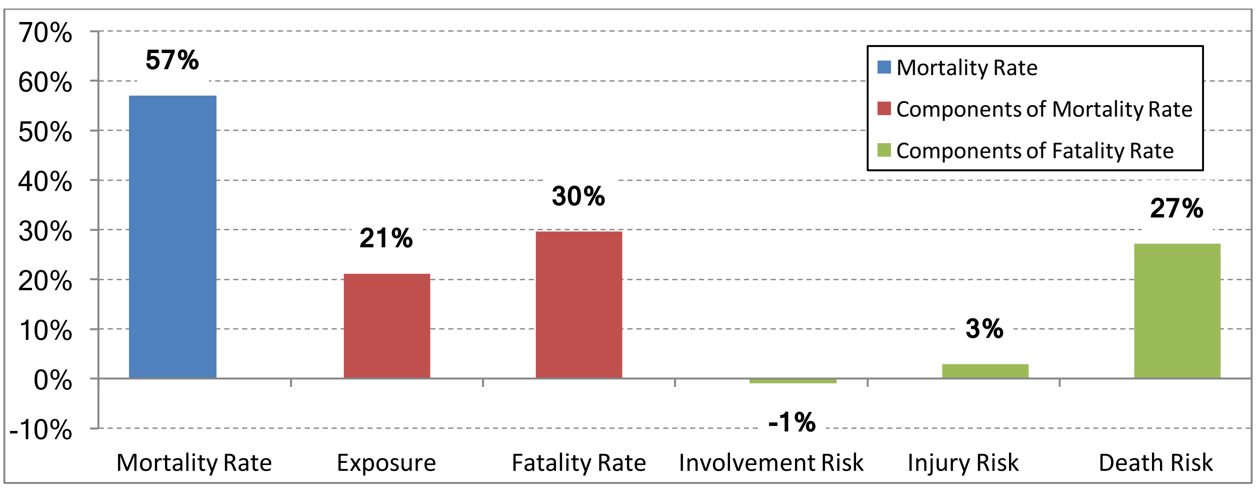
$$\text{Fatality Rate} = \underbrace{\frac{\text{Injury Crashes}}{\text{VMT}}}_{\boxed{\text{Involvement Risk}}} \cdot \underbrace{\frac{\text{Injuries}}{\text{Injury Crashes}}}_{\boxed{\text{Injury Risk}}} \cdot \underbrace{\frac{\text{Fatalities}}{\text{Injuries}}}_{\boxed{\text{Death Risk}}}$$

where injury crashes and injuries include all severities and all person types, including vehicle occupants, pedestrians, bicyclists, etc.

This approach was also applied to SE and NSE regions for the three-year period from 2007 to 2009. Three years were used to reduce the randomness in the safety data, particularly for smaller states because initially the analysis was done at the state level. These three particular years were used primarily because of the difficulty of getting data on non-fatal crashes and non-fatal injuries. Data were obtained at the state level for 44 states, including all eight SE states. Data were not available on nonfatal crashes and nonfatal injuries at the state level for seven states: Colorado, Hawaii, Maine, Massachusetts, Rhode Island, Utah, and West Virginia. Data on nonfatal crashes and injuries for most of the 44 states were from what

NHTSA had summarized from its State Data Program for the ten-year period from 2000 to 2009: *State Data System Crash Data Report: 2000-2009*, 2014. Data on nonfatal crashes and injuries for the other states were obtained from summary statistics that were available online. Fatality data were from NHTSA's annual *Traffic Safety Facts*. Population data were from Table 1, Intercensal Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2000 to July 2010 (ST-EST00INT-01), U.S. Census Bureau, Population Division, September 2011. Finally, VMT data were from FHWA's annual *Highway Statistics* series.

Figure 7 shows the results in terms of how SE was different from NSE for the various risks and component risks. The SE mortality rate was 57% higher during the 2007-2009 period, resulting from its 21% higher per capita exposure and 30% higher fatality rate than NSE regions. That is, SE residents not only drove more on a per capita basis but also were far more likely to be fatally injured in vehicle crashes.



Notes:

- Mortality rate = fatalities per capital
- Exposure = VMT per capita
- Fatality rate = fatalities per VMT
- Involvement risk = injury crashes of any severity per VMT
- Injury risk = injuries of any severity per injury crash
- Death risk = fatal injuries per injury of any severity

Figure 7. SE and NSE Difference in Risks and Component Risks, 2007-2009. The differences are measured as 100*(SE – NSE)/SE.

In addition, SE's 30% higher fatality rate resulted from its 1% lower involvement risk, 3% higher injury risk, and 27% higher risk of death. That is, the fatality rate was much higher in SE primarily because people injured in vehicle crashes in the SE region were much more likely to die from them. It is noted that differences in the injury risk largely depend on differences in average vehicle occupancies. On the other hand, differences in factors within a car that may distract the driver also influence differences in the involvement risk. One example is a car driven by a youth driver and how this driver may be distracted by the presence of youth passengers.

Regression Analysis of Fatality Rates for All Person Types

This explored the role of general risk-taking behaviors in the fatality rates at the state level. General risk-taking behaviors are any risk-taking behavior towards the general life of a person, including financial, health, etc. This exploration expanded the regression analysis in the thesis completed in 2015 by Jodi Godfrey as part of the same research project: *Risk-Taking Characteristics as Explanatory Variables in Variations of Fatality Rates in the Southeastern United States*. Similar to the thesis work, linear regression was done on one cross-section of all individual states for a combined 3-year period (2007-2009). Fatalities included all person types: vehicle occupants, bicyclists, pedestrians, etc. In addition, fatality rates were based on VMT as exposure.

The current work differed from the thesis work in several important ways:

- A dummy variable for the SE region was included as an explanatory variable to measure the differential safety performance between it and other regions.
- A broader set of explanatory variables were considered with their sources of data described in the appendix:
  - Average population density (resident population/square miles)
  - Vehicle traffic density (vehicle miles traveled over lane miles)
  - Weather norms (average January temperature, annual rainfall)
  - Race and ethnicity of resident population
  - Gender and age of resident population
  - Per capita personal income in constant dollars

- Unemployment rate
- Sales of accommodation and food services (as exposure for visitors)
- Credit scores
- Violent crimes per capita (violent crimes/resident population)
- Education attainment (without high school diploma)
- Alcohol consumption per capita for 21+
- Smoking for 18+ (ever smoked 100+ cigarettes and currently smoking)
- Drug use for 18+ (use of any illicit drug past month)
- Obesity for 18+
- Seat belt use by drivers and front passengers

- These potential explanatory variables were grouped into several categories:
  - Land use and system conditions: population density and traffic density
  - Weather: rain, January temperature
  - Socio-demographic: age, gender, race, ethnicity, income
  - Risk-taking behavior
- These categories of explanatory variables were sequentially added to a base model, which included a constant and the SE dummy but excluded all of these categories. The interest was in how the coefficient of the SE dummy would change from adding each category separately and sequentially. Of particular interest was the category on risk behavior. The hope was that there is still a considerable amount of difference after controlling for the other categories but this difference would disappear once risk behavior is also controlled for.

Regression analyses are common in safety research for exploring the role of variables of interest while controlling for other potential contributing factors. Sometimes, the number of fatalities is directly used as the dependent variable, in which case count-data models are typically used. When the dependent variable of interest is some form of fatality risks, however, linear regression is used instead. One primary argument against using linear regression is that the estimated model from linear regression assumes a linear relationship between the number of fatalities and exposure. The current exploration used linear regression. One reason is that the primary interest was on geographic differences in fatality risks rather than the number of fatalities. More important, the estimated models are not intended for prediction as part of this research. Rather they help explore the statistical and quantitative roles of risk-taking factors in geographic differences in fatality rates.

Table 1 shows the results from this regression analysis. The dependent variable was in annual fatalities per 100 million VMT. The top half of the table has the results from adding each category of explanatory variables separately and the bottom half has the results from adding each category sequentially. For each variable, both its coefficient and the t-statistic are shown. For each regression, the R^2 adjusted for the number of explanatory variables included (the larger the better) as well as the standard error of the regression (the smaller the better).

When each category was added separately, the coefficient for the SE dummy stayed significant both in its magnitude and its statistical significance when system features or weather features were controlled. The magnitude and statistical significance of the dummy, however, essentially disappeared when either the socio-demographic factors or the risk-taking behavior variables were controlled. This is important as it indicates that various variables about people, both their socio-demographic factors and risk-taking behaviors, can explain the differences and there is nothing inherent about the weather and physical environment of the SE region.

Table 1. Linear Regression of VMT-Based Fatality Rates for All Person Types, 2007-2009.

Variables and Categories		Region		System		Weather		Socio-Demographic		Risk-Taking Behavior	
		Coeff.	t stat	Coeff.	t stat	Coeff.	t stat	Coeff.	t stat	Coeff.	t stat
Constant		0.907	25.36	1.096	1.54	1.296	8.51	5.056	2.45	9.408	1.70
Region	Southeastern dummy	0.367	2.96	0.423	4.10	0.404	2.82	-0.003	-0.02	-0.064	-0.56
System	Ln(VMT/LM)			-0.099	-1.10						
	Ln(population density)			-0.090	-1.98						
Weather	Average Jan temperature					0.007	1.69				
	Annual rain					-0.008	-2.13				
Socio-demographic	Under 18							-0.020	-0.58		
	65+							-0.006	-0.16		
	Black							0.008	2.12		
	Hispanic							0.000	0.12		
	Ln(per capita per. income)							-0.651	-1.52		
	Bachelor+							-0.034	-2.90		
	Per capita sales							0.001	0.94		
Risk behavior	Ln(per capita violent crimes)									-0.109	-1.36
	Adult smokers									0.007	0.38
	Adult drug users									-0.051	-2.24
	Per capita alco. consumption									0.210	2.59
	No health insurance									0.015	1.25
	Without high school									0.010	0.99
	Obese adults									-0.021	-0.95
	Poor general health									0.121	2.71
	Credit score									-0.012	-1.64
Adjusted R^2		0.13		0.41		0.19		0.51		0.55	
Standard Error of the Estimate		0.32		0.27		0.31		0.24		0.23	
Constant		0.907	25.36	1.096	1.54	-0.352	-0.54	3.288	2.05	7.604	1.46
Region	Southeastern dummy	0.367	2.96	0.423	4.10	0.150	1.42	0.007	0.08	-0.161	-1.51
System	Ln(VMT/LM)			-0.099	-1.10	-0.251	-3.14	-0.119	-1.42	-0.091	-1.06
	Ln(population density)			-0.090	-1.98	-0.095	-2.38	-0.101	-2.11	-0.101	-2.17
Weather	Average Jan temperature					0.016	4.64	0.015	3.39	0.011	2.47
	Annual rain (inches)					0.004	1.15	-0.008	-1.73	-0.003	-0.72
Socio-demographic	% Under 18							-0.063	-2.25	-0.038	-1.02
	% 65+							-0.009	-0.30	0.002	0.06
	% Black							0.014	3.62	0.007	1.15
	% Hispanic							-0.003	-0.75	-0.013	-2.46
	Ln(per capita per income)							-0.003	-0.01	-0.297	-0.67
	% Bachelor+							-0.029	-2.73	-0.008	-0.52
	Per capita sales							0.000	-0.80	-0.001	-0.79
Risk behavior	Ln(per capita violent crimes)									0.076	0.94
	% Adult smokers									-0.010	-0.56
	% Adult drug users									-0.037	-1.55
	Per capita alco. consumption									0.138	1.79
	% No health insurance									0.020	1.47
	% Without high school									-0.001	-0.07
	% Obese adults									-0.043	-1.94
	% Poor general health									0.095	2.66
	Credit score									-0.006	-0.87
Adjusted R^2		0.13		0.41		0.61		0.75		0.80	
Standard Error of the Regression		0.32		0.27		0.22		0.17		0.16	

Alternative Exposure Measures for Pedestrian Fatality Risks

As already mentioned in the introduction section, many of the states and metropolitan areas ranked the most dangerous to pedestrians are in the SE region. Partly as a result of this reason, pedestrian safety and pedestrian fatality risks had become a far more significant transportation safety issue in the SE region than transportation safety in general. By fatality risks we mean ratios of the number of fatalities over the amount of exposure during a given period within a given geography. Such rankings of geographic areas on pedestrian fatality risks, however, are frequently being criticized for the specific measure of exposure used. One measure, population for example, may be criticized for not accounting for the amount of travel. Some measures may be criticized for accounting for vehicle activities but not for pedestrian activities. Still other measures may be criticized for accounting for pedestrian activities but not for vehicle activities. The issue explored here was how different measures of exposure for measuring pedestrian fatality risks at the state level relate to each other and how they affect the difference in pedestrian fatality risks between SE and NSE regions.

This exploration focused on a single year of 2009 because some of the measures of exposure were estimated from the 2009 NHTS. The measures of exposure explored are summarized in Table 2 and discussed below:

- Using resident population is common in assessing pedestrian fatality risks at aggregate levels and the resulting risk is often referred to as the morality rate because it does not account for travel activities.
- The product of resident population and the walk share of workers for commuting to work from the American Community Survey (ACS) has been used by reports that rank states and metropolitan areas on their pedestrian fatality risks, including McCann and DeLille (2000) and Ernst and Shoup (2009).
- The product of resident population and the walk share of person trips for all purposes likely represents an improvement over the similar product with the walk share for commuting only.
- It is not uncommon to use measures of vehicle activities alone as exposure for pedestrian fatality risks. Three were considered that measure vehicle activities in miles, trips, and travel time, respectively.

- It is not uncommon either to use measures of pedestrian activities alone as exposure for pedestrian fatality risks. Two were considered, including walk trips for all purposes and person hours walked for all purposes. The latter also included time for access to and egress from transit.
- These measures of exposure are not satisfactory because they either do not account for travel activities at all or account for just vehicle activities or pedestrian activities but not both. It takes both vehicle and pedestrian activities to have vehicle and pedestrian crashes. Traditionally, single vehicle crashes are captured but pedestrian events without vehicle involvement are not captured by crash databases. As a result, it is critical for a good measure of exposure to account for both vehicle and pedestrian activities.

Table 2. Alternative Measures of Exposure

<table>
<tr><td>Name</td><td>Description</td><td>Explanation</td></tr>
<tr><td>Pop</td><td>Resident population</td><td>No travel activities</td></tr>
<tr><td>PopACS</td><td>Resident population times the walk share of workers commuting to work from the American Community Survey (ACS) for 2009</td><td rowspan="2">Used as indicator for walking population or walking activities. PopACS is used by Mean Streets and Dangerous by Design reports.</td></tr>
<tr><td>PopNHTS</td><td>Resident population times the walk share of person trips for all purposes from 2009 NHTS</td></tr>
<tr><td>VMT</td><td>Total vehicle miles traveled (VMT) for all functional classes from the Highway Statistics series for 2009</td><td rowspan="3">Vehicle activities but no pedestrian activities</td></tr>
<tr><td>VT</td><td>Vehicle trips for all purposes of household travel from 2009 NHTS</td></tr>
<tr><td>VHT</td><td>Vehicle hours traveled for all purposes of household travel from 2009 NHTS</td></tr>
<tr><td>WT</td><td>Walk trips for all purposes from 2009 NHTS</td><td rowspan="2">Pedestrian activities but no vehicle activities</td></tr>
<tr><td>PHW</td><td>Person hours walked for all purposes from 2009 NHTS</td></tr>
<tr><td>VTtWT</td><td>Square root of VT times WT</td><td rowspan="3">Both vehicle and pedestrian activities with direct measure of pedestrian activities</td></tr>
<tr><td>VHTtPHW</td><td>Square root of VHT times PHW</td></tr>
<tr><td>VMTtPHW</td><td>Square root of non-freeway VMT times PHW</td></tr>
<tr><td>VMTtACS</td><td>Square root of total VMT times PopACS</td><td rowspan="3">Both vehicle and pedestrian activities with indirect measure of pedestrian activities</td></tr>
<tr><td>VTtACS</td><td>Square root of VT times PopACS</td></tr>
<tr><td>VTtNHTS</td><td>Square root of VT times PopNHTS</td></tr>
</table>

The literature provides little guidance on accounting for both vehicle and pedestrian activities. Using the product of vehicle volumes and pedestrian crossing activities at disaggregated levels such as intersections has appeared in the literature. Papadimitrious et al. (2012) use the product of vehicle volume and pedestrian crossing time. Knoblauch et al. (1996) suggest a similar product. Older and Grayson (1976) and Cameron (1982) used the product of volumes of crossing pedestrians and vehicles for roadway sections. Javid and Seneviratne (1991) use the product of volumes of crossing pedestrians and vehicles at intersections.

At more aggregated levels, there is no known prior work that accounts for both vehicle and pedestrian activities in measuring exposure for pedestrian risks. Based on data from the National Travel Survey of the U.K. during 1972-1973, however, Goodwin and Hutchinson (1977) suggested that the average number of pedestrian crashes is approximately proportional to the product of pedestrian hours walked and VMT.

Initially, directly using the product of vehicle and pedestrian activities as exposure was tried at the state level. It turned out that such products were clearly inappropriate as measures of exposure because they favor larger geographies. To illustrate, consider a simple example. State A has 10 times the number of pedestrian fatalities as does State B. In addition, State A also has 10 times the amount of vehicle activities and 10 times the amount of pedestrian activities as does State B. Intuitively, the pedestrian fatality risk should be similar between the two states in this example. With exposure measured as the direct product of vehicle and pedestrian activities, however, the pedestrian fatality risk for State A (i.e., the larger one) would be just one-tenth of that for State B (i.e., the smaller one).

To avoid the problem shown in this example but to still be able to account for both vehicle and pedestrian activities, the square root of the product of vehicle and pedestrian activities was explored. There is no known study that uses the square root of the product of pedestrian and vehicle activities as a measure of exposure for pedestrian risks. At the disaggregated level, however, using the square root of crossing flows has appeared in the literature. Gårder (1989), for example, uses the square root of the product of hourly crossing pedestrian and vehicle volumes at signalized intersections. More generally, Chapman (1973) suggested using the square root of the product of conflicting vehicle flows at intersections as exposure of vehicle crashes.

As one would expect, this square-root approach did avoid the problem illustrated with the example. Six such measures were considered. Three of them accounted for pedestrian activities directly by walk trips or person hours walked. The other three accounted for pedestrian activities indirectly by the product of resident population and walk shares.

Table 3 shows how these various measures of exposure relate to each other by showing the correlation coefficients (cc) among them for 2009. The abbreviated names for the measures in the columns and rows follow the same order as in Table 2. This table does reveal some interesting relationships among some of these measures of exposure.

Table 3. Correlation Coefficients among Alternative Measures of Exposure

	Pop	PopACS	PopNHTS	VMT	VT	VHT	WT	PHW	*VTtWT*	*VHTtPHW*	*VMTtPHW*	*VMTtACS*	*VTtACS*
PopACS	0.760												
PopNHTS	0.733	0.889											
VMT	0.893	0.456	0.430										
VT	0.877	0.485	0.448	0.945									
VHT	0.877	0.478	0.452	0.954	0.986								
WT	0.725	0.898	0.993	0.431	0.458	0.456							
PHW	0.662	0.853	0.948	0.343	0.380	0.375	0.948						
VTtWT	0.937	0.879	0.905	0.745	0.767	0.761	0.911	0.841					
VHTtPHW	**0.890**	**0.886**	**0.928**	**0.657**	**0.669**	**0.679**	**0.924**	**0.920**	**0.969**				
VMTtPHW	0.906	0.850	0.887	0.715	0.661	0.668	0.882	0.864	0.950	**0.969**			
VMTtACS	0.923	0.933	0.841	0.725	0.698	0.697	0.844	0.771	0.947	**0.921**	0.937		
VTtACS	0.911	0.953	0.863	0.688	0.700	0.704	0.867	0.801	0.956	**0.938**	0.919	0.989	
VTtNHTS	0.931	0.872	0.918	0.734	0.739	0.753	0.912	0.844	0.990	**0.976**	0.954	0.941	0.953

Note: Exposure measures in italic represent taking the square root of the project involved.

Notions:

Pop	Resident population
ACS	walk share of workers commuting to work from ACS
NHTS	walk share of person trips for all purposes from 2009 NHTS
VMT	Total vehicle miles traveled
VT	Vehicle trips for all purposes of household travel from 2009 NHTS
VHT	Vehicle hours traveled for all purposes of household travel from 2009 NHTS
WT	Walk trips for all purposes from 2009 NHTS
PHW	Person hours walked for all purposes from 2009 NHTS

Population (i.e., Pop) correlates well with vehicle activities alone and even better with the measures that account for both vehicle and pedestrian activities. Population, however, does not correlate well with pedestrian activities alone, particularly with the amount of time spent walking (cc = 0.662).

The product of resident population and ACS walk share for commuting (i.e., PopACS) correlates poorly with vehicle activities alone (cc < 0.5). It does much better with pedestrian activities alone. In fact, its correlation with pedestrian activities is much higher than the correlation of population with pedestrian activities. For example, cc = 0.662 between Pop and PHW but cc = 0.853 between PopACS and PHW. As a result, the primary objective of multiplying the walk share for commuters in this measure is accomplished to some degree. However, this ACS-based product of exposure correlates less well with the three exposure measures that account for both vehicle activities and directly-measured pedestrian activities (i.e., WT or PHW) than population alone with these three measures. For example, cc = 0.937 between Pop and VTtWT but cc = 0.879 between PopACS and VTtWT.

The product of resident population and NHTS walk share for all purposes (i.e., PopNHTS) correlates much better with pedestrian activities than the product of population with ACS walk share for commuting. For example, cc = 0.853 between PopACS and PHW but cc = 0.948 between PopNHTS and PHW. That is, it is more effective to use a walk share for all purposes than to use a walk share for commuting only to approximate pedestrian activities.

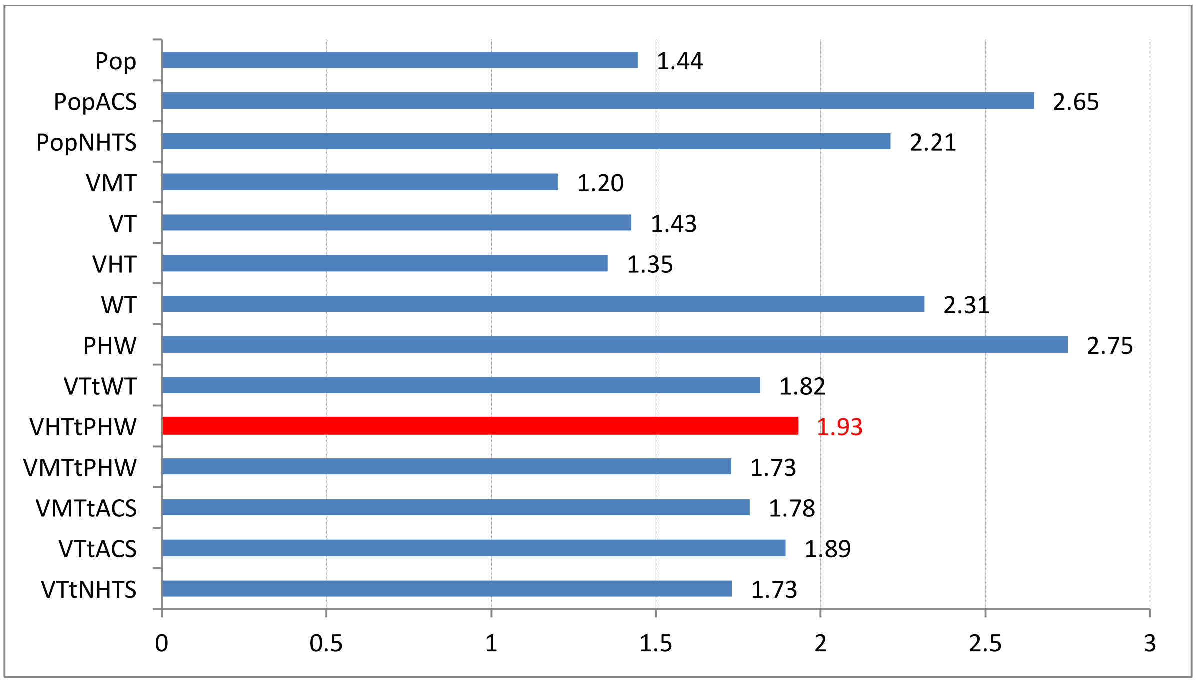
Vehicle activities correlate poorly with pedestrian activities, particularly when pedestrian activities are in terms of time spent (cc < 0.4). This is another reason for exposure measures for pedestrian safety to not account for either alone but to account for both simultaneously.

Besides other measures of vehicle activities alone, VMT correlates well with population (cc = 0.893) but not with any other measure considered. In this sense, it may not be a much better measure of exposure for pedestrian fatality risks, at least at the state level. The measures that account for both vehicle and pedestrian activities correlate much better with pedestrian activities alone than with vehicle activities alone. For example, cc = 0.767 between VT and VTtWT but cc = 0.911 between WT and VTtWT.

It might be helpful to select a single measure as the basis of comparison and compare all other measures with it. The measure that accounts for both vehicle time and pedestrian time, VHTtPHW, was selected as the base of comparison. Its correlation coefficients with all other measures are highlighted in bold in Table 3. Those measures in terms of vehicle activities do not correlate well with this base measure (cc < 0.7) and may not be proper choices as exposure for pedestrian fatality risks. Population or population adjusted for walk share for commuting may do much better than vehicle activities alone. All other measures are highly correlated with this base measure (cc > 0.9) and may serve as reasonable alternatives of exposure for pedestrian fatality risks. The two measures that account for pedestrian activities alone are among these other measures. Based on this comparison, pedestrian activities alone may not be bad measures of exposure after all for pedestrian fatality risks. Population adjusted for walk share for all purposes, PopNHTS, is also among these other measures and it is a surprise that it does so well relative to this base measure.

As Figure 8 shows, the pedestrian fatality rate is higher in SE than in NSE regions across all measures of exposure explored. This is somewhat comforting. However, the relative difference in pedestrian fatality rates between SE and NSE regions vary significantly across those measures that either do not account for either vehicle or pedestrian activities, those measures that try to account for pedestrian activities indirectly, or those that account for only vehicle or pedestrian activities. This lack of robustness in the relative difference from using these measures of exposure may indicate that these are not proper exposure for measuring pedestrian fatality rates, especially when for comparison over aggregated geographies. The good news is that the relative difference is highly robust across all those measures that account for both vehicle and pedestrian activities. In fact, the SE over NSE ratio of pedestrian fatality rates falls in a narrow range between 1.73 and 1.93. The measure chosen earlier as the base of comparison resulted in an estimate of the relative difference at 1.93, which is the high end of this range.

The conclusions on whether a measure of exposure may be appropriate for measuring pedestrian fatality risks differ significantly between those from comparing the degree of correlation among these measures and those from comparing the relative difference between SE and NSE regions across these measures. It appears that close correlation does not necessarily lead to robust outcomes in the relative difference in pedestrian fatality risks. The overall conclusion should be based on the comparison of relative differences rather than the comparison of correlation levels. Consequently, it is best to use measures of exposure for pedestrian fatality risks that account for both vehicle and pedestrian activities. When no better alternative is available for combining vehicle and pedestrian activities in a single measure, use the square root of their product.



Notes:

Pop	Resident population
ACS	walk share of workers commuting to work from 2009 ACS
NHTS	walk share of person trips for all purposes from 2009 NHTS
VMT	Total vehicle miles traveled
VT	Vehicle trips for all purposes of household travel from 2009 NHTS
VHT	Vehicle hours traveled for all purposes of household travel from 2009 NHTS
WT	Walk trips for all purposes from 2009 NHTS
PHW	Person hours walked for all purposes from 2009 NHTS

Figure 8. Comparison of SE over NSE Ratio of Pedestrian Fatality Rates across Exposure Measures

This overall conclusion also has important implications for frequent measurement of exposure for pedestrians. Given the low frequency of the NHTS, frequent measurement of pedestrian exposure should rely on measures that account for both vehicle and pedestrian activities and use data that are readily available. Among those considered above, only VMTtACS, the square root of VMT times the ACS share for walking, satisfies both requirements. Instead of using PoptACS, the widely cited national rankings of states should be based on VMTtACS instead. Based on VMTtACS being the chosen exposure measure, Figure 8 shows that the pedestrian fatality rate was 78% higher in SE than in NSE regions.

Decomposition of Mortality Rates for Pedestrians

This exploration focused on decomposition analyses of mortality rates for pedestrians rather than for all person types. Three components were used for pedestrians:

1. exposure per capita
2. involvement risk (= pedestrians involved in vehicle crashes per unit of exposure), and
3. death risk (= pedestrians died once involved in vehicle crashes).

The analyses were limited to 2009 because of the need to estimate pedestrian activities from the 2009 NHTS and because of the difficulty of getting data on the number of pedestrians involved in vehicle crashes for more than one year. As before, data on pedestrian fatalities were from FARS. There are no readily accessible databases for nationwide data on pedestrians involved in vehicle crashes. Instead, involvement data were assembled from several sources, including Table 9 of NHTSA 2014 report, *State Data System Crash Data Report: 2000-2009*, DOT-HS-812-052; published summary data by individual states, and data tabulated directly from traffic crash databases of individual states. Even after much effort, involvement data were not available for three states in NSE regions: Massachusetts, New Hampshire, and Rhode Island.

Two separate decomposition analyses were carried out with two different measures of exposure: person hours walked and the square root of the product of vehicle hours traveled and person hours walked. Both vehicle and pedestrian activities for these two measures were estimated from the 2009 NHTS. Table 4 shows the results from these two analyses.

The SE and NSE differences are far more dramatic for pedestrians than those for all person types. While residents in SE walk less, but both of their risk of involvement and risk of death once involved are higher in SE. In 2009, pedestrian fatalities per capita were 44% higher in SE than in NSE regions. This much higher pedestrian mortality rate in SE resulted from 85% higher risks of pedestrians dying once being involved in vehicle crashes as well as from differences in the other two components. The differences in the other two components, however, depend on the exposure measure used. When exposure was measured with person hours walked alone, the higher pedestrian mortality rate in SE also resulted from 47% fewer hours walked per capita and from 49% higher risks of pedestrians being involved in vehicle crashes per hour walked. When both vehicle and pedestrian activities were accounted for in the measure of exposure, however, the overall 44% higher pedestrian mortality rate in SE also resulted from 25% lower exposure per capita and from only 4% higher risks of pedestrians being involved in vehicle crashes per unit of exposure.

Table 4. Results from Decomposition Analyses of Pedestrian Mortality Rates

Components	Alternative Measures of Exposure	
	PHW (Person Hours Walked)	VHTtPHW (Square Root of Vehicle Hours Traveled times PHW)
Mortality Rate	+44%	+44%
Exposure per Capita	-47%	-25%
Risk of Involvement	+49%	+4%
Risk of Death	+85%	+85%

Notes:

- Mortality Rate = pedestrian fatalities per unit of exposure
- Risk of Involvement = pedestrians involved in vehicle crashes per unit of exposure
- Risk of Death = pedestrian deaths per pedestrian involved in vehicle crashes

These results show the importance of using appropriate measures of exposure for risk comparison and analyses for pedestrians. In addition, it is clear that the much higher pedestrian mortality rate in SE is largely due to the significantly higher risk of pedestrians dying once being involved in vehicle crashes. This is an important finding because effective countermeasures to reduce the injury severity of involved pedestrians are not necessarily the same as those to reduce the risk of pedestrians being involved in crashes.

The natural question is why the risk of death for pedestrians is so much higher in the SE region than in NSE regions. One potential reason is the existence of underreporting of pedestrian crashes with no injury or minor injury to pedestrians. Underreporting of pedestrian injuries has been reported for some locations in the US (Sciortino et al., 2005; Stutts and Hunter, 1998). If the degree of underreporting of pedestrian crashes with no or minor injury were much higher in the SE region than in the NSE regions, this data problem would have led to an artificially lower risk of involvement but higher risk of death for the SE region.

While this is a possibility, there is no obvious reason as to why pedestrian crashes would be underreported more in the SE than in NSE regions. There is no empirical evidence to support this possibility either.

Besides the possibility of data problems, the injury severity of pedestrians involved in vehicle crashes is determined largely by three sets of direct determinants: impact speed, impact configuration, and pedestrian attributes. While the mass of an involved vehicle is an important determinant of injury severities to both its own occupants and the occupants of the other vehicles involved, the mass of the vehicle is unlikely to be a significant factor in determining the injury severity of a pedestrian. One direct determinant of the risk of pedestrian deaths once involved in vehicle crashes is the access and quality of emergency care. More generally about pedestrian injury severity, Siddiqui and Chu (2006) summarize these direct determinants as follows:

- **Impact Speed**—The most important of these is the impact speed, which is the speed of the vehicle upon striking the pedestrian (Lee and Abdel-Aty, 2005; Sullivan and Flannagan, 2002; Jensen, 1999; Garder, 2004; and Pitt et al.,1990). The chance of survival by the pedestrian drops quickly between an impact speed of 20 mph and an impact speed of 40 mph (NHTSA, 1999).
- **Impact Configuration**—Besides impact speed, one set of determinants relates to impact configuration between the pedestrian and the vehicle (Yang, 2002). This impact configuration includes several aspects, including the angle at which the vehicle strikes the pedestrian (e.g., frontal versus side), the angle at which the pedestrian is struck (i.e., front, back, side), and the height of the impact on the pedestrian.
- **Pedestrian Attributes**—The final set of determinants relate to the characteristics of the pedestrian. The very young (Jensen, 1999; LaScala et al., 2001; Al-Ghamdi, 2002; and Fontaine and Gourlet, 1997) and the very old (Lee and Abdel-Aty, 2005; Jensen, 1999; LaScala et al., 2001; Al-Ghamdi, 2002; and Fontaine and Gourlet, 1997; and Zajac and Ivan, 2003)) are most vulnerable to suffering from severe injuries. Male pedestrians, being physically stronger and bigger on average than their female counterparts, may be less likely to sustain severe injuries.

Policy analysis of pedestrian safety, however, often requires an understanding of indirect determinants of pedestrian injury severity that go beyond the direct determinants. Adapted from Figure 2 of Siddiqui and Chu (2006), Figure 9 shows these indirect determinants and how they play a role in pedestrian injury severity through their effects on the direct determinants.

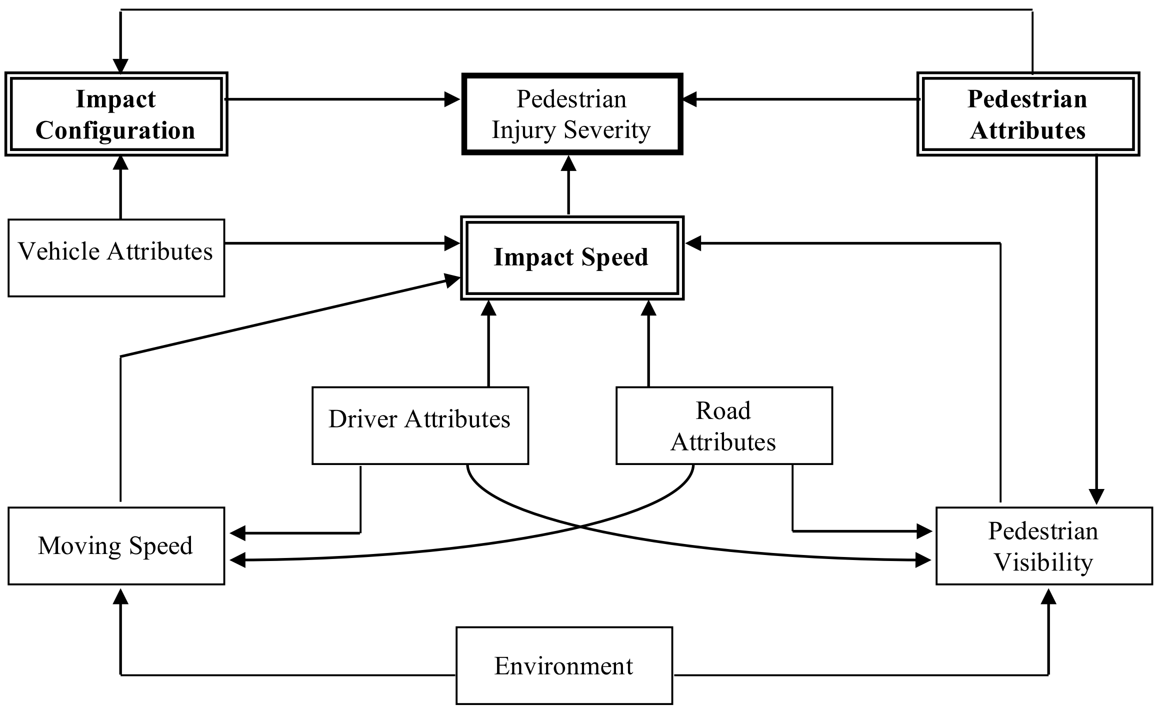


Figure 9. A Framework on the Determinants of Pedestrian Injury Severity.

- Vehicle attributes may affect both impact configuration and impact speed. High profile vehicles, such as SUVs, are more likely to increase the height of the impact on a pedestrian. Holding other factors constant, on the other hand, heavy vehicles are harder to stop, resulting in a higher impact speed.
- In addition to vehicle attributes, several sets of other factors affect the impact speed of a vehicle. These include the moving speed of the vehicle, driver attributes, road attributes, and pedestrian visibility to the driver.
- Furthermore, both driver attributes and road attributes affect the moving speed and pedestrian visibility to the driver.

- Pedestrian attributes, such as whether they wear reflective clothing at night, affect pedestrian visibility to the driver.
- Finally, the environment in terms of weather and light conditions can affect both the moving speed of the vehicle and pedestrian visibility to the driver.

Regression Analysis of the Risk of Pedestrian Deaths

Given the significance of the risk of pedestrian deaths once being involved in vehicle crashes, this explored the role of traffic speed in the variation of this risk of pedestrian deaths across states during 2009. With no direct measurement of traffic speeds available for this exploration, traffic speeds were represented in two ways. In one way, proxy variables were used, including the distribution of population by population density, traffic density on urban arterials (ratio of vehicle miles traveled over lane miles), and average width of arterials in terms of number of lanes. In another way, average speeds of vehicle travel were estimated from the stated trip distance and duration from the 2009 NHTS. In both cases, additional explanatory variables were used to control for differences in socio-economic and demographic characteristics, weather, presence of hospitals, and the share of light trucks in the fleet. Details about the source of data for these variables are provided in Part II and are not repeated here.

Table 5 shows the linear regression results for some of the specified models. The dependent variable is the risk of pedestrian deaths in terms of pedestrian deaths per 100 pedestrians involved in vehicle crashes during 2009. Similar to the regression analysis for fatalities of all person types, the SE region appears as a dummy variable in every model to see how its coefficient may change with the addition of control variables and speed-related variables. In addition to the estimated coefficient and its t-statistic for each explanatory variable, also shown are the adjusted R^2 (the larger the better) and the standard error of the regression (the smaller the better).

The first mode, labeled as Region Only, essentially shows on average how different SE is from the rest of the states in the risk of pedestrian deaths. The average of this risk was 4.231 across the states in NSE regions but was 7.99 (= 4.231+3.759) across the states in the SE region. That is, this SE average is about 89% higher than the NSE average. It is noted that this is different from the 85% value discussed in the previous section. In the previous section, the 85% value was based on risks measured at the regional level (SE vs. NSE) but the 89% value in this section was based on risks that were measured at the state level and were averaged across the states in SE and NSE.

Table 5. Linear Regression of the Risk of Pedestrian Deaths, 2009

Independent Variables and Categories		Region Only		Control Variables Only		Plus Speed Proxies		Plus Average Speed	
		Coeff.	t stat	Coeff.	t stat	Coeff.	t stat	Coeff.	t stat
Constant		4.231	13.82	4.005	1.31	0.284	0.06	-6.66	-0.89
Region	Southeastern dummy (= 1 for SE)	3.759	5.01	0.922	1.17	0.419	0.59	0.784	1.01
Speed	Ln(average speed in miles per hour)							3.135	1.56
	% population with high density					-0.036	-1.03		
	% of population with low density					0.056	2.43		
	Vehicles per lane for urban arterials					-0.684	-0.91		
	Lanes for principal arterials					0.665	0.83		
Control Variables	% male drivers 75+			1.816	2.01	2.254	2.37	2.528	2.54
	% female drivers 75+			-0.958	-1.40	-1.486	-2.00	-1.298	-1.83
	% Black			0.137	3.44	0.140	3.61	0.134	3.52
	% Hispanic			0.041	1.39	0.061	2.03	0.034	1.16
	Ln(per capita income, 000)			-0.198	-4.38	-0.101	-2.05	-0.197	-4.43
	Hospitals/1000 square miles			0.002	0.10	0.000	0.00	0.019	1.01
	Annual rainfall in inches			0.003	0.13	-0.011	-0.52	-0.007	-0.30
	Annual snowfall in inches			0.006	0.66	-0.005	-0.51	-0.001	-0.13
	% light trucks			0.109	3.20	0.059	1.48	0.096	2.79
Adjusted R^2		0.34		0.65		0.73		0.66	
Standard Error of the Regression		1.94		1.41		1.23		1.38	

Notes:

1. Dependent variable = pedestrian deaths per 100 pedestrians involved in vehicle crashes.
2. % population with high density = percentage of population in tracts with population density at least 10,000 persons per square miles. % population with low density = percentage of population in tracts with population density under 500 persons per square miles.

The second mode, Control Variables, adds a set of control variables including both socio-demographic and other controls. All of these control variables have the expected direction of effects on the risk of pedestrian deaths. Most of the socio-demographic variables are highly significant. Among the other control variables, the only statistically significant one is the percentage of light trucks in the vehicle fleet of each state during 2009. With these factors being controlled for, the SE dummy is no longer statistically significant. If taking the coefficient for the SE dummy at its face value, one would see that the difference in the average of the risk of pedestrian deaths between SE and NSE regions is much lower now. The NSE average was 4.005 and the SE average was 4.927 (=4.005+0.922), which is 23% higher than the NSE average.

The third model, Plus Speed Proxies, uses a set of four variables as proxies for traffic speed. These measure the distribution of population density, traffic density for urban arterials in annual vehicles per lane, and roadway width for principal arterials in number of lanes. While only the percentage of population in low density areas is statistically significant, all four of these proxies have expected signs. The control variables are relatively stable except the weather variables, whose coefficients are still highly insignificant but become negative. While still having a positive coefficient, the SE dummy is highly insignificant now. Both adjusted R^2 and the standard error of the regression improved significantly. It is difficult, however, to collectively interpret these variables as proxies for traffic speed on the risk of pedestrian deaths. For the percentage of population in low density tracts, a 10% drop would lower the risk of pedestrian deaths by 0.56 percentage points.

The last model, Plus Average Speed, replaces the set of four proxy variables for traffic speed by the natural log of average speeds. In miles per hours, the average speed for each state was derived from stated duration and distance of vehicle trips in the 2009 NHTS. The control variables are still relatively stable. Using speed directly did improve both adjusted R^2 and the standard error of the regression but by less than using the set of proxy variables. In addition, the log of average speed is marginally significant at the 13% level. More important, average speed is numerically significant as well. A 10% decrease in average speed (e.g., from 30 mph to 27 mph) would drop the risk of pedestrian deaths by 3.135 percentage points. With the average of the risk of pedestrian deaths across SE states being at 7.99%, such a drop in the risk would be highly effective in reducing the number of pedestrian deaths in the SE region.

Use of a Panel Dataset

This study also considered using a panel dataset at the state level for a more comprehensive regression analysis of the role of socio-demographic factors and risk-taking behaviors in geographical differences in pedestrian safety risks. A panel dataset of states consists of data for all states over multiple time periods (e.g., years). Panel data analysis, particularly at the state level, is common in safety research both because of the advantages discussed below and because state-level data are more readily available.

There were two primary reasons for expanding the regression analyses of a single cross section of state-level data described earlier to regression analyses of a panel dataset. Panel data give the researcher a large number of data points, increasing the degrees of freedom and reducing the possibility and degree of collinearity among explanatory variables, and improving the efficiency of estimates (i.e., estimates are more precise with smaller standard errors.). This is particularly important because there was a relatively large number of potential explanatory variables involved and they reduce the degree of freedom quickly in a single cross section of states.

The other reason is that panel data allow the researcher to control for omitted (unobserved or incorrectly measured) variables that are correlated with explanatory variables. Without proper control, such variables would lead to biased estimates. Due to lack of data and time constraints, data for many potential determinants are not available. Some of these include traffic laws other than seat belt usage, EMS capabilities, etc. In addition, data about the transportation system are typically unavailable beyond lane miles and VMT. Each state has its unique features beyond what are measured by the variables considered. As a result, there is a need for better control of unmeasured unique features of individual states.

Ideally the focus should have been on the pedestrian risk of death because this risk was 85% higher in SE than in NSE regions while the pedestrian involvement rate was about the same, at least during 2009. While measuring this risk of death does not require exposure data, it does require data on pedestrian involvement, i.e., all pedestrians involved in crashes with vehicles regardless of whether they died or were injured. Getting the state-level data on involved pedestrians was already extremely difficult for a single year, getting the data for multiple years was considered infeasible.

The focus instead was on the pedestrian fatality rate with exposure being measured as the square root of the product of both vehicle and pedestrian activities. Vehicle activities were measured by VMT and pedestrian activities were approximated with the product of resident population and the share of workers walking to work. As Figure 8 shows, the pedestrian fatality rate as measured in this way was 78% higher in SE and in NSE regions during 2009. While this difference is not as high as for the risk of death, it is still alarmingly high and requires a better understanding of its contributing factors.

The unit of analyses was at the state level largely because more and better data are available at the state level. Many of the data items used are simply not available at a lower geography level for the entire country. VMT is a good example. While a few states are known to estimate and publish county-level VMT, most do not, at least do not publish them. Data for most of the risk-taking behaviors are readily available at the state level in published tables but not at lower geographic levels. In terms of data quality, the annual number of pedestrian fatalities is small for a large portion of counties and can be subject to a high degree of randomness. Annual estimates of the walk share for commuting from the ACS are subject to large margins of error at the county level and are not even made available by the Bureau of Census for small counties. Also, the data on most risk-taking behaviors are based on surveys and the precision of estimates at lower geographies drops quickly.

The duration of the panel from 2001 through 2014 was chosen because annual data for many of the key variables of interest were not readily available before 2001. These include the walk share for commuting and most of those on risk-taking behaviors. Even after a great effort of getting data for the full duration, data for a number of variables are only available for some years of the duration. The second part of the appendix details the sources of the data for all variables in the panel dataset.

Typically panel data analysis uses one of two statistical models: the random effects model and the fixed effects model. For the random effects mode, the units of observations are drawn randomly from the overall population. One good example would be the households selected from all households in a geography for a panel survey. As a result, the random effects model is not well suited for a panel dataset of states.

The fixed effects model can be written as:

$$y_{it} = \mathbf{X}'_{it}\beta + \alpha_i + \varepsilon_{it}$$

with i denoting states and t denoting year. $\mathbf{X}_{it}$ is the vector of explanatory variables explicitly included. β is the vector of parameters for these explanatory variable. The error term is ε_{it}. The fixed effects formulation implies that unobserved differences across groups can be captured in differences in the constant term α_i. Each α_i is treated as an unknown parameter to be estimated and the estimated value absorbs the effects particular to each state. By estimating these state-specific constants, one estimates the pure effect of each explanatory variable by controlling for the unobserved differences across states. The fixed effects method controls for all stable characteristics of the states in the study. This would mean that fixed effects control for time-invariant characteristics of each state, such as the history of each state before the study period.

The focus was to explore the role of risk-taking behaviors in the differential safety performance between the SE and NSE regions. The approach to model specification was to extend the earlier regression with a single cross section of data to the panel dataset. Specifically, these potential explanatory variables were grouped into several categories:

- Land use and system conditions: population density and traffic density by functional class
- Weather: rain, snowfall, winter temperature
- Socio-demographic: age, gender, race, ethnicity, income, income distribution
- Visitor exposure: employment of leisure and hospitality services
- Medical condition: depression, mental illness
- Risk-taking behaviors

A dummy variable for the SE region was going to be included as an explanatory variable to measure the differential safety performance between it and other regions. These categories of explanatory variables were going to be sequentially added to a base model without any of these categories included. The interest was in how the coefficient of the dummy variable for the SE region would change with adding each category separately and sequentially. Of particular interest is the category on risk-taking behaviors. The hope is that there is still a considerable amount of difference after controlling for the other categories but this difference disappears once risk-taking behaviors is also controlled for.

At the end, however, the planned panel data regression analysis was not carried out because of a significant feature of the fixed effect model that was learned after the panel dataset was already built. A major shortcoming of the fixed effects method is that any time-invariant variables cannot be estimated. Because of this shortcoming, the dummy for the SE region cannot be estimated in a fixed effects model. Related to this shortcoming is that the fixed effects model does not work well with data for which temporal variation is minimal or change slowly over time. The culture of a region may change slowly but is largely fixed. The variables used as indicators of risk-taking behaviors do change but slowly. As a result, they cannot be efficiently estimated. Essentially, the fixed effects model is designed to study the causes of temporal changes within a state but not to study the causes of differences across states. Unfortunately, the primary interest of this study is in the causes of differences in the safety performance across states.

CONCLUSIONS

This study explored the differential safety performance between the Southeastern (SE) region and non-Southeastern (NSE) regions. Safety performance was measured in fatality rates (i.e., fatalities per unit of exposure). Part of the exploration was methodological, i.e., how the exposure measure used could affect relative safety performance for pedestrians between SE and NSE regions. It was argued that exposure measures for pedestrian fatality risks should account for both vehicle and pedestrian activities. It was determined that the square root of the product of vehicle and pedestrian activities is a reasonable measure of exposure. For frequent measurement, a reasonable choice would be to use VMT for vehicle activities and the product of resident population and the share of workers walking to work from the American Community Survey as pedestrian activities.

More significantly, the study explored sources and the degree of the differential safety performance. It was found that SE and NSE differences in the fatality rate based on person hours traveled did not vary much during 2009 across gender and age groups or across gender and race-ethnicity groups, or across time of day and day of week periods. They did vary significantly, however, across functional classifications, particularly between rural versus urban areas and between freeway versus non-freeway facilities in a given area type. During 1994-2013, in fact, driving on non-freeway facilities in rural areas accounted for almost 80% of the overall difference in fatality rates between SE and NSE states.

Beyond exploring differences in the fatality rate directly, this study further explored the components of the fatality rate both for all person types combined and for pedestrians. The fatality rate was decomposed into two multiplicative components: the involvement risk and the risk of death. It turned out that the involvement risk was similar between SE and NSE regions and that the SE region has much higher risk of death than the NSE region. For all person types combined during 2007-2009, it was found that the fatality rate was 30% higher in SE than in NSE and this higher fatality rate in SE primarily resulted from its 27% higher risk of death. For pedestrians during 2009, it was found that the fatality rate was 78% higher in SE than in NSE and this much higher fatality rate for pedestrians in SE primarily resulted from its 85% higher risk of death.

This study also conducted linear regression analyses of the fatality rate for all person types and of the risk of death for pedestrians. For all person types combined, the primary interest was in socio-demographic factors and risk-taking behaviors. Differences in system and weather conditions do not account for the much higher fatality rates in SE. Differences in the fatality rate largely disappears once differences in socio-demographic factors or in general risk-taking behaviors are controlled for. For pedestrians, the primary interest was in the statewide average speed of vehicle travel. It was found that a 10% drop in average speed at the state level could reduce the risk of pedestrian deaths by 3.1 percentage points. This is significant relative to the average risk of pedestrian deaths being 7.99% in the SE region.

Further research is needed to investigate the finding that the risk of death for pedestrians was 85% higher in the SE region than in the NSE regions. The regression analysis did find a significant role of average speed in this higher risk of death. However, this exploratory analysis was limited. First, the regression analysis was aggregated at the state level, suffering from the typical shortcomings of using aggregate data. Second, speed was estimated from stated distance and duration of person trips from household surveys. Third, the regression analysis did not explicitly account for other major causes of pedestrian fatalities. Besides the impact speed, other major determinants of pedestrian fatalities are the access and quality of emergency care systems, pedestrian characteristics (e.g., being older aged), and impact configuration (i.e., where and how the vehicle and pedestrian were in contact). The new research should take a disaggregated approach, such as at the level of individual crashes, use speeds of vehicle travel that are more directly related to the crash conditions, and account for all major determinants of pedestrian fatalities.

ACKNOWLEDGEMENTS

The authors are thankful for support received from the Southeastern Transportation Center (United States Department of Transportation grant number DTRT13-G-UTC34). The authors greatly appreciate the sponsorship of the Office of Secretary of Transportation.

References

Aguero-Valverde, J. and P.P. Jovanis (2006). Spatial Analysis of Fatal and Injury Crashes in Pennsylvania. *Accident Analysis and Prevention* 38: 618-625.

Al-Ghamdi, A.S. (2002). Pedestrian–Vehicle Crashes and Analytical Techniques for Stratified Contingency Table. *Accident Analysis and Prevention* 34: 205–214.

Cameron, M.H. (1982). A Method of Measuring Exposure to Pedestrian Accident Risk. *Accident Analysis and Prevention* 14: 397-405.

Chapman, R. (1973). The Concept of Exposure. *Accident Analysis and Prevention* 5: 95-110.

Chu, X. (1999). Effects of Changes in Travel Patterns on Highway Fatalities. *Accident Analysis and Prevention* 31: 221-227.

Dellinger, A.M., J.A. Langlois, and G. Li (2002). Fatal Crashes among Older Drivers: Decomposition of Rates into Contributing Factors. *American Journal of Epidemiology* 155: 234-241.

Dewey, J.F., D. Denslow, D. Lenze, and L. Holt. 2003. *Pedestrian Safety*. Bureau of Economic and Business Research, University of Florida.

Dixon, K. *Southeastern United States Fatal Crash Study*. Georgia Institute of Technology, 2005.

Eksler, V., S. Lassarre, and I. Thomas (2008). Regional Analysis of Road Mortality in Europe. *Public Health* 122: 826-837.

Ernst, M. and L. Shoup. *Dangerous by Design*. Transportation for America. 2009.

Fontaine, H. and Y. Gourlet (1997). Fatal Pedestrian Accidents in France: A Typological Analysis. *Accident Analysis and Prevention* 29: 303-312.

Gårder, P. (1989). Pedestrian Safety at Traffic Signals: A Study Carrried Out with the Help of a Traffic Conflicts Technique. *Accident Analysis and Prevention* 21: 435-444.

Gårder, P.E. (2004). The Impact of Speed and Other Variables on Pedestrian Safety in Maine. *Accident Analysis and Prevention* 36: 522-542.

Goldstein, G.P., D.E. Clark, L.L. Travis, and A.E. Haskins (2011). Explaining Regional Disparities in Traffic Mortality by Decomposing Conditional Probabilities. *Injury Prevention* 17: 84-90.

Goodwin, P.B., and T.P. Hutchinson (1977). The Risk of Walking. *Transportation* 6: 217-230.

Grabowski, D.C., and M.A. Morrisey (2004). Gasoline Prices and Motor Vehicle Fatalities. *Journal of Policy Analysis and Management* 23: 575-593.

Javid, M., and P.N. Seneviratne (1991). Applying Conflict Technique to Pedestrian Safety Evaluation. *ITS Journal*, March 1991.

Jensen, S.U. (1999). Pedestrian safety in Denmark. *Transportation Research Record* 1674: 61-69.

Keeler, T.E. (1994). Highway Safety, Economic Behavior, and Driving Environment. *American Economic Review* 84: 684:693.

Knoblauch, R.L., M.T. Pietrucha, and M. McLean (1996). Field Studies of Pedestrian Walking Speed and Start-Up time. *Transportation Research Record* 1538: 27-38.

Lambert, T.E., and P.B. Meyer (2006). Ex-Urban Sprawl as a Factor in Traffic Fatalities and EMS Response Times in the Southeastern United States. *Journal of Economic Issues* 40: 941-953.

LaScala, E.A., F.W. Johnson, and P.J. Gruenewald (2001). Neighborhood Characteristics of Alcohol-Related Pedestrian Injury Collisions: A Geostatistical Analysis. *Prevention Science* 2: 123-134.

Lee, C., and M. Abdel-Aty (2005). Comprehensive Analysis of Vehicle–Pedestrian Crashes at Intersections in Florida. *Accident Analysis and Prevention* 37: 775–786.

Li, G. and S.P. Baker (1996). Exploring the Male-Female Discrepancy in Deaths Rates from Bicycling Injury: The Decomposition Method. *Accident Analysis and Prevention* 28: 537-540.

McCann, B. and B. DeLille. *Mean Streets 2000: Pedestrian Safety, Health and Federal Transportation Spending*. Surface Transportation Policy Project. 2000.

National Highway Traffic Safety Administration (1999). *Literature Review on Vehicle Travel Speeds and Pedestrian Injuries*. U.S. Department of Transportation, Washington, D.C.

Noland, R.B. (2003). Traffic Fatalities and Injuries: The Effect of Changes in Infrastructure and Other Trends. *Accident Analysis and Prevention* 35: 599-611.

Older, S.J. and G.B. Grayson (1976). An International Comparison of Pedestrian Risk in Four Cities. In: *Proceedings of the International Conference on Pedestrian Safety*. Michlol, Technion, Haifa, Isreal.

Papadimitriou, E., G. Yannis, and J. Golias (2012). Analysis of Pedestrian Exposure to Risk in Relation to Crossing Behavior. *Transportation Research Record* 2299: 79-90.

Pitt, R., B. Guyer, C. Hsieh, and M. Malek (1990). The Severity of Pedestrian Injuries in Children: An Analysis of the Pedestrian Injury Causation Study. *Accident Analysis and Prevention* 22: 549–559.

Sciortino S., M. Vassar, M. Radetsky, M.M. Knudson (2005). San Francisco Pedestrian Injury Surveillance: Mapping, Under-reporting, and Injury Severity in Police and Hospital Records. *Accident Analysis and Prevention* 37: 1102-1113.

Siddiqui, N.A., and X. Chu, Crossing Locations, Light Conditions, and Pedestrian Injury Severity. *Transportation Research Record* 1982: 141-149 (2006).

Stamatiadis, N. and G. Puccini (2000). Socioeconomic Descriptors of Fatal Crash Rates in the Southeast USA. *Injury Control and Safety Promotion* 7: 165-173.

Stutts, J.C. and W.W. Hunter (1998). Police-reporting of Pedestrians and Bicyclists Treated in Hospital Emergency Rooms. *Transportation Research Record* 1635: 88-92.

Sullivan, J.M., and M.J. Flannagan (2002). The Role of Ambient Light Level in Fatal Crashes: Inferences from Daylight Saving Time Transition. *Accident Analysis and Prevention* 34: 487-498.

Traynor, T.L. (2008). Regional Economic Conditions and Crash Fatality Rates – a Cross-County Analysis. *Journal of Safety Research* 39: 33-29.

Wang, C. *Fatal Crash Trends and Analysis in Southeastern States*. Dissertation, School of Civil and Environmental Engineering, Georgia Institute of Technology, 2006.

Washington, S., J. Metarko, I. Fomunung, R. Ross, F. Julian, and E. Moran (1999). *Accident Analysis and Prevention* 31: 135-146.

Yang, J. (2002). *Review of Injury Biomechanics in Car-Pedestrian Collisions*. Crash Safety Divisions, Chalmers University of Technology, Sweden.

Zhu, M., S. Zhao, J.H. Coben, and G.S. Smith (2013). Why More Male Pedestrians Die in Vehicle-Pedestrian Collisions than Female: A Decomposition Analysis. *Injury Prevention* 19: 227-231.

Zajac S.S., and J.N. Ivan (2003). Factors Influencing Injury Severity of Motor Vehicle-Crossing Pedestrian Crashes in Rural Connecticut. *Accident Analysis and Prevention* 35: 369-379.

Appendix

Publications and Presentations

Godfrey, J.A. *Risk-Taking Characteristics as Explanatory Variables in Variations of Fatality Rates in the Southeastern United States*. Department of Civil and Environmental Engineering, University of South Florida. 2015.

Godfrey, J.A. (2015). Risk-Taking Characteristics as Explanatory Variables in Variations of Fatality Rates in the Southeastern United States. *ITE Journal* 85: 43-47.

Polzin, S., X. Chu, and J. Godfrey. A Decomposition Approach to Comparing Pedestrian Fatality Risks across Geographies. *10th UTC Spotlight Conference on Pedestrian and Bicycle Safety*, December 1-2, 2016, Keck Center, Washington, D.C.

Chu, X., S. Polzin, and J. Godfrey. Difference and Variability in Traffic Fatality Rates between Southeast and Other Regions. *2016 UTC Conference for the Southeastern Region*, March 31-April 1, 2016, Knoxville, TN.

Chu, X., S. Polzin, and J. Godfrey. Comparing Traffic Injury Rates between Southeast and Other Regions: A Decomposition Approach. *2016 UTC Conference for the Southeastern Region*, March 31-April 1, 2016, Knoxville, TN.

Sources of Panel Data

This appendix describes the source of data for every data item in the panel dataset. The data items have been organized into the same categories:

- Crash outcomes
- Exposure
- Infrastructure
- Laws
- Land use and system conditions
- Weather

- Socio-demographic and economic
- Risk-taking behavior

Crash Outcomes

- Fatalities by person type (vehicle occupants, pedestrians, bicyclists)
- Fatal crashes by type (single-vehicle crashes, crashes with non-motorists, etc.)

  The data were obtained from using the FARS Query System at https://www-fars.nhtsa.dot.gov//QueryTool/QuerySection/SelectYear.aspx for each year separately. These were also done separately for fatalities versus for fatal crashes.

Exposure

- Resident population

  2001 to 2010: Table 1, Intercensal Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2000 to July 1, 2010 (ST-EST00INT-01, U.S. Census Bureau, Population Division, September 2011.

  2001 to 2010: Table 1, Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2010 (NST-EST2014-01), U.S. Census Bureau, Population Division, December 2014.

- VMT by functional classification

  Annual *Highway Statistics*, Annual Vehicle-Miles of Travel by Functional System, Table VM-2, Federal Highway Administration.

- Employment for leisure and hospitality services

  U.S. Bureau of Labor Statistics, All Employees: Leisure and Hospitality, retrieved as annual data separately for each state from FRED, Federal Reserve Bank of St. Louis; fred.stlouisfed.org.

  The leisure and hospitality sector is part of the service-providing industries and includes two sub-sectors: NAICS 71, Arts, Entertainment, and Recreation and NAICS 72, Accommodation and Food Services.

- Share of light trucks in statewide fleet

  Annual *Highway Statistics*, Truck and Truck-Tractor Registrations, Table MV-9, Federal Highway Administration.

- Share of workers walking to work

  Data on workers:

  2005-2014: 1-year American Community Survey data, retrieved from American FactFinder.
  2000: 2000 Census, retrieved from American FactFinder.

  Data on shares:

  2005-2014: calculated as the ratio of workers usually walk to work over total workers.
  2001-2004: extrapolated from 2000 and 2005 shares.

Infrastructure

- Lane miles by functional classification

  Annual *Highway Statistics*, Functional System Lane-Length, Table HM-60, Federal Highway Administration.

- Centerline miles by roadway width in lanes

  Annual *Highway Statistics*, Public Roadway Length, Miles by Functional System, Table HM-20, Federal Highway Administration.

- Average roadway width in lanes by functional classification

  Derived as the ratio of lane miles over centerline miles.

- Density of community hospitals per 1,000 square miles

  Hospitals: American Hospital Association Annual Survey, Copyright 2016 by Health Forum, LLC, an affiliate of the American Hospital Association, retrieved from Total Hospitals, State Health Facts, http://kff.org/other/state-indicator/total-hospitals/, The Henry J. Kaiser Family Foundation.
  Community hospitals are all nonfederal, short-term general, and specialty hospitals whose facilities and services are open to the public. They represent 85% of all hospitals. The other hospitals are federal hospitals, long term care hospitals,

psychiatric hospitals, institutions for the mentally retarded, and alcoholism and other chemical dependency hospitals.

Hospital density: ratio of number of community hospitals over land area in square miles.

Laws

- Speed limit on any facility

The maximum speed limit in a speed did not vary much over the period of the panel dataset. Data were compiled from various sources.

Farmer, Charles M., Relationship of Traffic Fatality Rates to Maximum State Speed Limits, Insurance Institute for Highway Safety, Table 1, 2016. Available at http://www.iihs.org/frontend/iihs/documents/masterfiledocs.ashx?id=2117, accessed on May 5, 2016.

FHWA, Freight Facts and Figures 2004, Table 3-7, US DOT. Available at http://www.ops.fhwa.dot.gov/publications/publications.htm#fa, accessed on May 4, 2016.

FHWA, Freight Facts and Figures 2005, Table 3-7, US DOT. Available at http://www.ops.fhwa.dot.gov/publications/publications.htm#fa, accessed on May 4, 2016.

FHWA, Freight Facts and Figures 2006, Table 3-7, US DOT. Available at http://www.ops.fhwa.dot.gov/publications/publications.htm#fa, accessed on May 4, 2016.

FHWA, Freight Facts and Figures 2007, Table 3-7, US DOT. Available at http://www.ops.fhwa.dot.gov/publications/publications.htm#fa, accessed on May 4, 2016.

FHWA, Freight Facts and Figures 2008, Table 3-8, US DOT. Available at http://www.ops.fhwa.dot.gov/publications/publications.htm#fa, accessed on May 4, 2016.

FHWA, Freight Facts and Figures 2009, Table 3-9, US DOT, 2009. Available at http://www.ops.fhwa.dot.gov/publications/publications.htm#fa, accessed on May 4, 2016.

FHWA, Freight Facts and Figures 2010, Table 3-9, US DOT, 2010. Available at http://www.ops.fhwa.dot.gov/publications/publications.htm#fa, accessed on May 4, 2016.

FHWA, Freight Facts and Figures 2011, Table 3-11, US DOT, 2011. Available at http://www.ops.fhwa.dot.gov/publications/publications.htm#fa, accessed on May 4, 2016.

FHWA, Freight Facts and Figures 2012, Table 3-14, US DOT, 2012. Available at http://www.ops.fhwa.dot.gov/publications/publications.htm#fa, accessed on May 4, 2016.

FHWA, Freight Facts and Figures 2013, Table 3-16, US DOT, 2014. Available at http://www.ops.fhwa.dot.gov/publications/publications.htm#fa, accessed on May 4, 2016.

NHTSA and FHWA, *The Effect of Increased Speed Limits in the Post-NMSL Era*, Report to Congress, Appendix A, 1998. Available at http://www-nrd.nhtsa.dot.gov/Pubs/808-637.PDF, accessed on May 5, 2016.

NHTSA, *Summary of State Speed Limit Laws*, Twelfth Edition, DOT-HS-811-769, 2013. Available at http://www.nhtsa.gov/staticfiles/nti/pdf/Summary_State_Speed_Laws_12th_Edition_811769.pdf, accessed on May 5, 2016.

NHTSA, *Summary of State Speed Limit Laws, Eleventh Edition*, DOT-HS-811-457, 2011. Available at http://www.nhtsa.gov/staticfiles/nti/pdf/811457.pdf, accessed on May 5, 2016.

Ohio Insurance Institute, *Speed Limit Laws*. Available at https://www.ohioinsurance.org/speed-limit-laws/, accessed on May 4, 2016.

- Legal limit of blood alcohol concentration at 0.08

  NHTSA, 2008, *Statistical Analysis of Alcohol-Related Driving Trends, 1982-2005*, Table 3: Effective Dates of Alcohol Laws in 50 States and DC, http://www-nrd.nhtsa.dot.gov/Pubs/810942.pdf.

  NHTSA, 2001, *Legislative History of 0.08 Per Se Laws*, Table 1: States/Jurisdictions with 0.08 Per se Laws, http://ntl.bts.gov/lib/26000/26000/26071/DOT-HS-809-286.pdf.

- Seat belt laws:
  - Primary for front seats
  - Primary for rear seats
  - Secondary for front seats
  - Secondary for rear seats

Insurance Institute for Highway Safety, Safety Belts, State Laws, http://www.iihs.org/iihs/topics/laws/safetybeltuse?topicName=safety-belts.

Land Use and System Conditions

- Population density, average

  Land area in square miles, excluding water area: Bureau of the Census, State Area Measurements and Internal Point Coordinates, https://www.census.gov/geo/reference/state-area.html.

  Population density: ratio of resident population over land area in square miles.

- Population density, distribution

  2000: Bureau of the Census, Census 2000 Gazetteer Files, Census Tracts, https://www.census.gov/geo/maps-data/data/gazetteer2000.html.

  2010: Bureau of the Census, Census 2010 Gazetteer Files, Census Tracts, https://www.census.gov/geo/maps-data/data/gazetteer2010.html.

  For a given year, the file has the population and land area, excluding water area for each census tract. These data were used to derive the share of state population who lived in census tracts with fewer than 500 persons per square mile and the share of state population who lived in census tracts with at least 10,000 persons per square mile. These shares for 2000 were used every year from 2001 through 2007. These shares for 2010 were used for the other years.

- Vehicle traffic density by functional class

  Derived as the ratio of VMT over LM by functional classification

Weather

Historical weather data are available for individual weather stations within every state. Three data items are considered: average monthly temperature (main interest is in the winter months), average annual snowfall, and annual total precipitation. Pre-tabulated state-level annual data are not found for a continuous period. Given that the general pattern is consistent across states over time, the plan is to use weather norms for each state and repeat them over the panel period.

- Average monthly temperature
  NOAA National Climatic Data Center, averages based on data collected by weather stations throughout each state from 1971 to 2000. Retrieved

from Current Results: Weather and Science Facts, https://www.currentresults.com/Weather/US/average-annual-state-temperatures.php.

These data do not vary across the years for each state.

- Total precipitation

  NOAA National Climatic Data Center, averages based on data collected by weather stations throughout each state from 1971 to 2000. Retrieved from Current Results: Weather and Science Facts, https://www.currentresults.com/Weather/US/average-annual-state-precipitation.php.

These data do not vary across the years for each state.

- Snowfall in a select city of a state

  National Center for Environmental Information, Local Climatological Data Annual Summary with Comparative Data, annual total snowfall in inches, for each selected city in each state. https://www.ncdc.noaa.gov/IPS/lcd/lcd.html.

Socio-Demographic and Economic

- Race and ethnicity of resident population

  2001-2010: Intercensal Estimates of the Resident Population by Sex, Race and Hispanic Origin for States and the United States: April 1, 2000 to July 1, 2010 (ST-EST00INT-SEXRACEHISP), U.S. Bureau of the Census, Population Division, October 2012.

  2011-2014: One-year data from the American Community Survey, extracted from American FactFinder.

- Gender and age of resident population

  2001-2010: Intercensal Estimates of the Resident Population by Age and Sex for States and the United States: April 1, 2000 to July 1, 2010 (ST-EST00INT-AGESEX), U.S. Bureau of the Census, Population Division, October 2012.

  2011-2014: One-year data from the American Community Survey, extracted from American FactFinder.

- Gender and age of drivers

Annual *Highway Statistics*, Licensed Female Drivers by Age and Licensed Male Drivers by Age, Table DL-22, Federal Highway Administration.

- Median household income

  U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplements, Table H-8, Median Household Income by State: 1984 to 2014, Current Dollars.

  Adjusted to constant 2014 dollars with annual, seasonally adjusted implicit price deflator for U.S. gross domestic product from FRED, Federal Reserve Bank of St. Louis, fred.stlouisfed.org.

- Per capita personal income in constant dollars

  Bureau of Economic Analysis, SA1 Personal Income Summary: Personal Income, Population, Per Capita Personal Income, 1960-2015, current dollars, March 2016.

  Adjusted to constant 2014 dollars with annual, seasonally adjusted implicit price deflator for U.S. gross domestic product from FRED, Federal Reserve Bank of St. Louis, fred.stlouisfed.org.

- Unemployment rate

  Bureau of Labor Statistics, States and Selected Areas: Employment Status of the Civilian Non-institutional Population, 1976 to 2015 Annual Averages, https://www.bls.gov/web/laus/ststdsadata.txt.

- Education attainment

  Frank, Mark W., Annual State-Level Human Capital Measures, 1940-2015, High School and College Attainment, revised January 2016, http://www.shsu.edu/eco_mwf/Frank_Education_2015.xlsx.

- Average price of gasoline

  U.S. Energy Information Administration, *Motor Gasoline Price in the Transportation Sector*, Current Dollars per Million Btu, based on the State Energy Data System, http://www.eia.gov/state/seds/sep_update/pr_all_update.csv.

  Adjusted to constant 2014 dollars with annual, seasonally adjusted implicit price deflator for U.S. gross domestic product from FRED, Federal Reserve Bank of St. Louis, fred.stlouisfed.org.

Risk-Taking Behavior

- Unbanked households (2009, 2011, 2013)

  Federal Deposit Insurance Corporation, *2009 National Survey of Unbanked and Underbanked Households*, Table B-1 Banking Status of Households by State, https://www.fdic.gov/householdsurvey/2009/appendices.pdf.

  Federal Deposit Insurance Corporation, *2011 National Survey of Unbanked and Underbanked Households*, Table H-1 2011 Banking Status of Households by State, https://www.fdic.gov/householdsurvey/2011/2011_unbankedreport_app_h-i.pdf.

  Federal Deposit Insurance Corporation, *2013 National Survey of Unbanked and Underbanked Households*, Table G-1 Banking Status by State, 2013, https://www.fdic.gov/householdsurvey/2009/appendices.pdf.

  This survey was conducted only three times during the panel period: 2009, 2011, and 2013. The limited data on average credit scores were included but not used in the panel data analysis.

- Credit scores (2004,2006,2012)

  Consumer credit scores are either based on Experian's scale from 330 to 830 or Fair Isaac Corporation's (FICO) scale from 330 to 850. The FICO is more widely used, but state average data are rarely available. Experian, on the other hand, produces quarterly state average data from a sample of three million credit profiles. Details from this sample are unknown, but this sample is not representative of a state's entire population.

  2004 and 2006: Average Consumer Credit Score, 4th Quarter for 2004 and 2nd Quarter for 2006, https://www.homeimprovementweb.com/information/finance-money/state-credit-scores.htm. These data were based on Experian's credit scale from 330 to 830.

  2012: Maciag, Average Credit Scores by State, Governing Magazine, April 2012, http://www.governing.com/blogs/by-the-numbers/average-credit-scores-for-states.html.

  The limited data on average credit scores were included but not used in the panel data analysis.

- Alcohol consumption per capita for 21+

National Institute on Alcohol Abuse and Alcoholism, *Surveillance Report #104: Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends, 1977-2014*, March 2016. Data were at https://pubs.niaaa.nih.gov/publications/surveillance104/pcyr19702014.txt.

- Drug use for 12+ (use of any illicit drug past month)

SAMHSA, Office of Applied Studies/Center for Behavioral Health Statistics and Quality, *National Survey on Drug Use and Health*.

2014: 2013-2014 National Survey on Drug Use and Health: Model-Based Prevalence Estimates (50 States and the District of Columbia), Table 1, https://www.samhsa.gov/data/sites/default/files/NSDUHsaePercents2014.pdf.

2013: 2012-2013 National Survey on Drug Use and Health: Model-Based Prevalence Estimates (50 States and the District of Columbia), Table 1, https://www.samhsa.gov/data/sites/default/files/NSDUHStateEst2012-2013-p1/Tables/NSDUHsaePercents2013.pdf.

2012: 2011-2012 National Survey on Drug Use and Health: Model-Based Prevalence Estimates (50 States and the District of Columbia), Table 1, https://www.samhsa.gov/data/sites/default/files/NSDUHStateEst2011-2012/ExcelTabs/Web/NSDUHsaeTables2012.pdf.

2011: 2010-2011 National Survey on Drug Use and Health: Model-Based Prevalence Estimates (50 States and the District of Columbia), Table 1, http://archive.samhsa.gov/data/NSDUH/2k11State/NSDUHsae2011/ExcelTabs/NSDUHsaeTables2011.pdf.

2009-2010: Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality, State Estimates of Substance Use and Mental Disorders from the 2009-2010 National Surveys on Drug Use and Health. Table C.1., **http://archive.samhsa.gov/data/NSDUH/2k10State/NSDUHsae2010/FullReport/NSDUHsaeMainReport2010.htm#tabc1.**

2006: Substance Abuse and Mental Health Services Administration, Office of Applied Studies. *State Estimates of Substance Use from the 2005-2006 National Surveys on Drug Use and Health*. OAS Series #H-33, DHHS Publication No. (SMA) 08-4311, Table C.1.

2003-2005, 2007-2008: Bureau of the Census, Statistical Abstracts of the United States, Table on Estimated Use of Selected Drugs by State. Table number: 204 for 2008, 203 for 2007, 196 for 2005, and 195 for 2003-2004.

2002: State Estimates of Substance Use from the 2002 National Survey on Drug Use and Health, Table A.1, http://archive.samhsa.gov/data/2k2State/html/appA.htm#taba.1.

2001: State Estimates of Substance Use from the 2001 National Household Survey on Drug Abuse: Volume II. Individual State Tables and Technical Appendices, Table A.1, http://archive.samhsa.gov/data/nhsda/2k1State/vol1/appA.htm.

- Serious mental illness for 18+ (% suffered last year) (2009-2014)
- Any mental illness for 18+ (% suffered any mental illness past year) (2009-2014)
- Major depressive episode for 18+ (% had 1+ past year) (2009-2014)

SAMHSA, Center for Behavioral Health Statistics and Quality, *National Survey on Drug Use and Health*. The estimated percentage numbers for 2014 were based on the survey data for both 2013 and 2014. The percentage numbers for the other years were similarly estimated.

- Mentally unhealthy days for 18+ (self-stated, average past month)

Estimated directly from the database for the *Behavioral Risk Factor Surveillance System* (BRFSS) for each of these years. The annual database and related documentation were downloaded at https://www.cdc.gov/brfss/annual_data/annual_data.htm.

- Depressive disorder for 18+ (% ever told a professional they have a depressive disorder) (2011-2014)
- Health status for 18+ (reported fair or poor)
- No health insurance for 18+
- Heavy drinking for 18+ (>2 drinks/day for men; >1 drink/day for women past month)
- Binge drinking for 18+ (5+ drinks/occasion for men; 4+ drinks/occasion for women past month)
- Smoking for 18+ (ever smoked 100+ cigarettes and currently smoking)
- Obesity for 18+

2002-2011

Centers for Disease Control and Prevention, Surveillance Summaries in the Morbidity and Mortality Weekly Report, https://www.cdc.gov/brfss/publications/ssummaries.htm. For a given data item, the particular table number can vary from one year to another.

2002 – MMWR Vol 52 No. 55-05 (July 23, 2005): Surveillance for Certain Health Behaviors Among Selected Local Areas — United States, Behavioral Risk Factor Surveillance System, 2002.

2003 – MMWR Vol 54 No. 55-08 (December 2, 2005): Surveillance for Certain Health Behaviors Among States and Selected Local Areas --- Behavioral Risk Factor Surveillance System, United States, 2003.

2004 – MMWR Vol 55 No 55-07 (July 14, 2006): Surveillance for Certain Health Behaviors Among States and Selected Local Areas --- Behavioral Risk Factor Surveillance System, United States, 2004.

2005 – MMWR Vol 56 No. 55-04 (May 11, 2007): Surveillance of Certain Health Behaviors Among States and Selected Local Areas --- Behavioral Risk Factor Surveillance System, United States, 2005.

2006 – MMWR Vol 57 No. 55-07 (August 15, 2008): Surveillance of Certain Health Behaviors Among States and Selected Local Areas --- Behavioral Risk Factor Surveillance System, United States, 2006.

2007 – MMWR Vol 59 No. 55-01 (February 5, 2010): Surveillance of Certain Health Behaviors and Conditions Among States and Selected Local Areas --- Behavioral Risk Factor Surveillance System, United States, 2007.

2008 – MMWR Vol 59 No. 55-01 (December 10, 2010): Surveillance for Certain Health Behaviors Among States and Selected Local Areas --- United States, 2008

2009 – MMWR Vol 60 No. 9 (August 19, 2011): Surveillance of Certain Health Behaviors and Conditions Among States and Selected Local Areas — Behavioral Risk Factor Surveillance System, United States, 2009.

2010 – MMWR Vol 62 No. 55-01 (May 31, 2013): Surveillance of Certain Health Behaviors and Conditions Among States and Selected Local Areas --- United States, 2010.

2011 - MMWR Vol 63 No. 55-09 (October 24, 2014): Surveillance for Certain Health Behaviors Among States and Selected Local Areas — Behavioral Risk Factor Surveillance System, United States, 2011.

2001, 2012-1014

Estimated directly from the database for the Behavioral Risk Factor Surveillance System (BRFSS) for each of these years. The annual database and related documentation were downloaded at https://www.cdc.gov/brfss/annual_data/annual_data.htm.

- Seat belt use by drivers and front passengers

  2007-2014: NHTSA, Seat Belt Use in 2014—Use Rates in the States and Territories, Traffic Safety Facts, June 2015

  2003-2010: NHTSA, Seat Belt Use in 2011—Use Rates in the States and Territories, Traffic Safety Facts, July 2011

  2002: Glassbrenner, Donna, *Safety Belt Use in 2003 – Use Rates in the States and Territories*, Table 1, Document No. DOT HS 809 713, NHTA, March 2004.

  2001: Glassbrenner, Donna, *Safety Belt Use in 2001 – State Rates*, Document No. DOT HS 809 501, NHTSA, October 2002.

- Violent crimes per capita (violent crimes/resident population)

  Annual number of violent crimes: Federal Bureau of Investigation, Uniform Crime Reporting, State and National Crime Estimates by Year(s), https://www.ucrdatatool.gov/Search/Crime/State/StateCrime.cfm.

  Violent crimes per capita: ratio of annual number of violent crimes over resident population.