Built Environment Factors Contributing to Pedestrian Collisions: A Structural Equation Modeling Approach

Emily Porter*
emilyporter@gwmail.gwu.edu
Department of Civil and Environmental Engineering
Center for Intelligent Systems Research
Traffic and Networks Research Laboratory
The George Washington University
Exploration Hall
20101 Academic Way
Ashburn, Virginia 20147

Justin Schorr
justin11@gwmail.gwu.edu
Department of Civil and Environmental Engineering
Center for Intelligent Systems Research
Traffic and Networks Research Laboratory
The George Washington University
Exploration Hall
20101 Academic Way
Ashburn, Virginia 20147

Samer H. Hamdar
hamdar@gwu.edu
Department of Civil and Environmental Engineering
Center for Intelligent Systems Research
Traffic and Networks Research Laboratory
The George Washington University
Exploration Hall
20101 Academic Way, Room 201-I
Ashburn, Virginia 20147
Phone: (202) 994 6652
Fax: (202) 994 0127

5316+ 1750(7 Tables) = 7066

August 1st, 2014

Submitted for presentation at the 94th Annual Meeting of the Transportation Research Board, and publication in Transportation Research Record

ABSTRACT

Collisions between vehicles and pedestrians continue to occur at a high rate each year, resulting in several thousand pedestrian injuries and deaths in the United States alone. In addition to the natural environmental factors (i.e. precipitation, lighting, temperature, etc.), factors related to the infrastructure in which pedestrians walk impact the safety performance of a given roadway segment. The objective of this paper is to identify infrastructural elements which contribute to pedestrian-vehicle collisions. For this purpose, data from NASS-GES (National Automotive Sampling System - General Estimates System) was analyzed using structural equation modeling (SEM). The corresponding approach allows grouping multiple exogenous factors contributing to pedestrian collisions into groups (i.e. dimensions) consisting of multiple variables providing a more comprehensive analysis of the safety of roadway features. The findings may allow the avoidance of undesired/dangerous design standards adopted by traffic/transportation engineers through altering the surrounding physical environment (like adding artificial light to dark roadways) and thus providing better protection for pedestrians.

Keywords: Pedestrians, Infrastructure, Safety, Structural Equation Modeling

1. Introduction and Motivation

Pedestrian collisions impact individuals from all age groups and from different geographic locations. Walking is a mode of transportation used by almost everyone at some point in their lives, and many pedestrians use travel paths that conflict with vehicular pathways. Accordingly, the nature of pedestrian travel entails some risk, primarily from the proximity of vehicles. Although most pedestrians never experience collisions with vehicles, these crashes can be life altering, and sometimes fatal. The resulting damages from these collisions are felt by drivers and pedestrians alike, both physically and emotionally. In the United States, where many safety measures have already been put into effect, the NHTSA (National Highway Traffic Safety Administration) has estimated that 4,743 pedestrians were killed and 76,000 pedestrians were injured in 2012 (1). These numbers indicate that continued efforts in the field of traffic safety are needed in order to protect pedestrians from vehicular collisions.

Multiple approaches have been adopted to increase pedestrian safety. Among the most popular of these approaches, there are: pedestrian education, vehicle modification, and infrastructural alterations (2). Vehicle modification has received increasing attention throughout the last decade (2). However, even with vast technological improvements made to new vehicles, there are older "non-modified" vehicles that remain on the road. Pedestrian education, which seems to be the most simple and cost effective approach to improve pedestrian safety, has been shown to be successful with children. However, results are less promising regarding the effectiveness of pedestrian education for adults (3, 4). There are no apparent downfalls to pedestrian education, but more effective solutions must be pursued in order to increase safety for pedestrians of all age groups. Infrastructure alterations can successfully improve pedestrian safety (2). In order to have a positive influence on pedestrian safety, it is essential to identify not only the core infrastructure elements that are associated with pedestrian-vehicle collisions; additional factors contributing to the increased exposure to pedestrian collision risk must be recognized for a more comprehensive safety analysis.

Given the above approaches and the suggested limitations, the objective of this paper is to identify elements of the roadway infrastructure which contribute to pedestrian-vehicle collisions. Data from police reported pedestrian-vehicle collisions that occurred throughout the United States in 2011 and 2012 was analyzed using structural equation modeling (SEM). SEM allows identifying factors that lead to a decrease in pedestrian safety while grouping the corresponding exogenous observable variables into different dimensions. The data set was provided by NASS - GES (National Automotive Sampling System - General Estimates System), and variables relating to infrastructure and flow conditions (such as speed limit, number of lanes, traffic flow, and type of intersection) were utilized. Many studies that have examined the effect of roadway infrastructure on pedestrian safety aggregate their data (5, 6 and 7). In this paper, the data was disaggregated so that factors associated with each collision location were separately considered. Additionally, no demographic information was taken into account. While many studies in the past have looked at the effects of demographic and socioeconomic variables on pedestrian safety, the main focus of this paper is to identify the various infrastructure elements, traffic flow characteristics, and environmental and impairment related variables that play a role in pedestrian safety.

To realize the stated objective, the specific research tasks to be achieved in this study are as follows: 1) to identify relevant variables from police accident reports related to pedestrian-vehicle collisions and to build the corresponding data points; 2) to run factor analysis to group variables; 3) to apply structural equation modeling techniques to the "filtered" data points; 4) to identify variables with the most influence on pedestrian safety. The literature related to pedestrian-vehicle collisions is presented in Section 2, the adopted modeling approach (i.e. SEM) and the corresponding data are offered in Section

3. Section 4 includes the numerical analysis and a discussion of the results. The concluding remarks and the future research needs are presented in Section 5.

2. Background

Multiple studies have been conducted in order to determine the factors contributing to pedestrian-vehicle collisions. Most of these studies focus on particular cities (i.e. New York City, Montreal, San Francisco, and Baltimore) in order to narrow the amount of data involved and to form conclusions that are relevant to the city in question. Such specific results can be used to improve future planning of pedestrian filled areas. Moreover, the majority of these studies group collisions that have occurred within small, well defined parts of the city, such as a census tract, zip code zone, police district, etc. (5, 6 and 8). Such pre-classification scheme is adopted in order to take into account population demographic characteristics and land-use features. This type of area-level study gives a broad overview of collisions that occur within a particular area as opposed to focusing on the specific geometric characteristics of collision sites.

Many commonly accepted factors contributing to pedestrian collisions are identified in the aforementioned area-level studies, while other variables' influence on collision has not yet been totally verified. Some of the most commonly identified variables include: Individual factors, like age, gender, and socioeconomic status (9, 10), and race (10). Environmental factors, like traffic volume (7, 11, 12) vehicle speed (13), street type and design (13, 2), and land use (12).

Most area level studies take into account variables relating to infrastructure as well as population information. For example, (7) considered demographic and environmental correlates in their study of pedestrians in San Francisco. They use a spatial autocorrelation corrected regression model to determine factors associated with pedestrian traffic injuries in 1990. Their study uses census tracts to aggregate data, and they use a geographic information system to map locations of pedestrian injuries. The variables used in their study include: demographic factors, like gender, age, marital status, education, income and unemployment, and environmental features, like high traffic flow, complex roadway systems, greater population densities and alcohol availability. Results of their study show that pedestrian injury rates are associated with traffic flow, population density, age composition of the local population, unemployment, gender and education. Furthermore, availability of alcohol was directly related to pedestrian injury collisions in which the pedestrian had been drinking alcohol.

Similarly, the area-level study conducted by (6), based on census tracts, of pedestrian collisions in San Francisco uses multivariate regression modelling to identify risk variables. They include numerous variables about street characteristics (traffic volume, intersections, residential streets, arterial streets with and without public transit, freeways, and highways), land use characteristics (commercial, industrial, neighborhood commercial, residential, higher density residential, residential neighborhood commercial, and land area), population characteristics (employees, residents, age 65 and older, age 17 and under, living below the poverty level last year, unemployed), and commute behaviors (workers commuting by walking and workers commuting by public transit). They find that traffic volume, arterial streets without public transit, proportions of land area zoned for neighborhood commercial use and residential-neighborhood commercial use, land area, employee population, resident population, proportions of people living in poverty, and proportion of people aged 65 and over are all statistically significant predictors of pedestrian-vehicle collisions.

(5) use negative binomial regression models in their area-level study of new York City. In order to provide some insight into the accuracy of different area levels, they use two different sizes of focus areas, one based on zip code and the other based on census tracts. Variables about socio-demographics,

land use, transit, intersections, and road characteristics were included. They conclude that the proportion of multi-lane roads is positively associated with pedestrian collisions. Additionally, they report that a smaller aggregation level (census tract as opposed to zip code zone) provides "greater explanatory power for variations in accident frequencies."

The area-level approach takes into account a broad range of variables while focusing on a particular geographical area. A more specific approach, such as that used by (14) and (15) focuses on the characteristics of each specific collision site. (14) consider the effect of infrastructure, speed limits, and pedestrian characteristics on the severity of pedestrian collisions in Montreal using ordered logit regression techniques. Most of the data was provided by the city of Montreal and Quebec's Automobile Insurance Board. One difference in this study (compared to others mentioned above) is that it tries to capture the infrastructure and design characteristics by examining different buffer zone sizes around the site of the collision. This allows the authors to consider various infrastructural characteristics that are present in each of the buffer zones. The authors included the following variables: injury severity (no injury, minor injury, major or fatal injury), day of week (week or weekend), median income in census tract, population density in census tract, number of schools in buffer zone, total number of intersections in buffer zone, total number of cul-de-sacs in buffer zones, connectivity (number of intersections ÷ (number of intersections + cul-de-sacs)) in buffer zone, percentage residential/commercial land use, vehicle (automobile, motorcycle/moped, vans/trucks/buses, emergency vehicle), daylight, road condition (poor or other), road type (local, arterial, highway), parks, hospitals, vehicle driving direction (straight, backing up, turning left, turning right), slope of roadway, visibility (object, weather, good), intersection (whether collision occurred at intersection or not), total bus and metro stops, schools within zone. Their results show that the main factors associated with injury severity levels were roadway type, vehicle movement, accident location, vehicle type, environmental conditions, population density, road connectivity, and land use mix.

A similar study that uses disaggregated data conducted by (15) examines the impact of personal and environmental characteristics on the severity of injuries sustained in pedestrian-vehicle collisions. The data (from Maryland motor vehicle accident reports) covers a 4 year period, in Baltimore, Maryland. A generalized ordered probit model is used to determine which variables are positively correlated with injury severity. The level of injury was considered in order to lessen the effect of low-injury collisions that may occur more frequently in pedestrian dense areas, while drawing attention to more severe injuries sustained in crashes that occur in lower density areas. Each crash was geocoded to the nearest intersection, so very specific information regarding collision location was not taken into account. The researchers used a 0.25 mile buffer zone around each location to take environmental variables into account. The following variables were considered (most gathered from accident reports, some gathered from US Census, Maryland land use, and Baltimore City public schools): severity of injury (no injury, injury, fatality), age (0-15 years, 16-64 years, 64+), sex, clothing type (dark, other), signal disobedience, substance present, daylight, weather (inclement or fair), road condition (defects or good), pedestrian location (crosswalk or not), vehicle (automobile, motorcycles/mopeds, emergency, trucks/vans/buses), road facility type (local, arterial/higher order facility), transit access (number of stops within 0.25 mile zone), connectivity (number of intersections ÷ (number of intersections + cul-de-sacs) in 0.25 mile zone), population density (in census block of where collision occurred), median income (in census block of where collision occurred), schools (within 0.25 mile zone), percentage commercial/residential land use (within 0.25 mile zone). The results indicate that substance use and crosswalk usage affect the severity of pedestrian injury. The researchers conclude that more detailed data about crash environments are needed for this type of analysis to guide policy recommendations. The authors suggest including variables such as lighting, line of sight, and other pedestrian safety elements.

All of the previously described studies have considered population variables to some degree. (16) focus more exclusively on the infrastructure. Their study identifies pedestrian-vehicle collision hotspots in Vancouver, Canada. Data was gathered from the Insurance Corporation of British Columbia and the British Columbia Trauma Registry over a 6 year period, from 2000 to 2005. Crash locations were mapped using ArcGIS 9.2 and geo-referenced to either an intersection or a midblock. Locations with 5 or more incidents (including "near misses") were considered hot spots. Elements of the infrastructure within 100m of incident location were recorded by researchers. The researchers took into account the following variables: long block, bus stop, curb parking, crosswalk, visual obstruction, signage, number of lanes, left or right turning bans, bars, retail, schools, median, exclusive turns, and calming measures (like speed bumps, road narrowing, and reduced speed limits). The authors found a high correlation between the presence of bars and pedestrian collisions, which support similar findings made by (7). However, the presence of schools did not seem to influence their results, although different studies have concluded the schools increase the risk of collision (17, 18). The authors hypothesize that many safety measures are effectively used near schools.

The literature indicates that several studies have been conducted to identify environmental factors associated with pedestrian-vehicle collisions. Most of the previously done work uses aggregated data, including geographic and population variables, while also focusing on a particular city. The studies that use disaggregated data are all specific to particular cities as well. In this case, the authors intend to focus on the impact of the infrastructure on pedestrian-vehicle collisions, and the NASS GES data allows for an analysis of collisions that occurred throughout the United States. A description of the statistical model and the available data are described in Section 3, followed by more rigorous statistical analysis of the data, including factor analysis and model results in Section 4. Lastly, concluding remarks are offered in Section 5

3. Statistical Model and Data

Data:

Data was acquired from NASS GES (National Automotive Sampling System General Estimates System), which allowed information from pedestrian-vehicle collisions from across the United States to be used. NASS GES data is taken from a nationally representative probability sample selected from all police accident reports. The NASS GES database contains hundreds of variables relating to several different types of traffic collisions. In order to obtain the specific information relevant to pedestrian-vehicle collisions, the data files were filtered and then combined so that only pertinent variables were included. The variables included in this study were taken from the following NASS GES files: the Accident File (type of intersection, weather and light condition), the Vehicle File (driver drinking in vehicle, traffic-way description, number of lanes and speed limit), the Person File (injury severity), the "Drimpair" File (driver impairment), and the "Nmimpair" File (pedestrian impairment). Additional information unrelated to the details of the collision itself were taken into account, including region of the country in which the collision occurred (Northeast, Midwest, South, West), year of occurrence, and weight. The weight variable indicates how many times a particular collision is likely to have occurred throughout the country. This number was used to increase the number of data points so that the true number of collisions that occurred could be taken into account as opposed to just those accidents that were randomly selected by NASS GES. Although using the weight to determine the total number of pedestrian-vehicle collisions that occur each year is imperfect compared to collecting police accident reports from every single collision, it is as accurate as possible considering the available nation-wide data. The convergence and the statistical significance of the resulting structural model should bring insights into the correctness of the adopted technique in constructing the data set in question. The two most recent years of data (2011 and 2012) were considered for the study, resulting in a total of 52,224 pedestrian-vehicle collisions. This number is lower than the actual number of collisions that occurred because collisions with missing data were entirely omitted from the study so that only those with complete information were taken into account. Table 1 summarizes the data by year, and Table 2 describes each variable that was considered, including exogenous, endogenous, and other variables.

Table 1: NASS GES Total Collisions per Year in 2011 and 2012

Collisions by year		
Year	Total collisions	Total pedestrian injuries/fatalities
2011	21994	16278
2012	30229	29221

Table 2: Description of Model Variables

Exogenous variable description		
Exogenous variable	Description	Details
Infrastructure		
X_1 (Int. Legs)	Type of intersection	1: Not an intersection 2: Traffic circle or roundabout 3: T or Y intersection 4: 4-way intersection 5: Five-point or more
X_2 (Lanes)	Number of lanes	1-7 Lanes
X_3 (Speed Limit)	Speed limit (mph)	Speed Limit Divided by 10
X_4 (Flow)	Flow conditions just prior to collision	1: One-Way Traffic 2: Two-Way traffic
Environmental		
X_5 (Weather)	Dummy variable corresponding to weather	0: Clear 1: Visual impairment (precipitation, fog, smoke, etc.)
X_6 (Lighting)	Dummy variable corresponding to lighting conditions	0: Light 1:Dark
Distraction/Impairment		
X_7 (Drinking)	Dummy variable corresponding to driver drinking in vehicle	0: Not drinking in vehicle 1: Drinking in vehicle
X_8 (Impaired)	Dummy variable corresponding to driver impairment (drugs, fatigued, physical impairment, etc.)	0: Not impaired 1: Impaired
X_{10} (P Impaired)	Dummy variable corresponding to pedestrian impairment (drugs, fatigued, physical impairment, etc.)	0: Not impaired 1: Impaired
Endogenous variable description		
Endogenous variable	Description	Details
Y_1 (MaxP_Inj)	Severity of injury sustained	0: No injury 1: Possible injury/ injury severity unknown 2: Non-incapacitating injury 3: Incapacitating injury 4: Fatal injury
Y_2 (N P Inj)	Number of pedestrian injured	0-3 per collision
Other variable description		
Region	Location of collision by region	1: Northeast *(PA, NJ, NY, NH, VT, RI, MA, ME, CT)* 2: Midwest *(OH, IN, IL, MI, WI, MN, ND, SD, NE, IA, MO, KS)* 3: South *(MD, DE, DC, WV, VA, KY, TN, NC, SC, GA, FL, AL, MS, LA, AR, OK, TX)* 4: West *(MT, ID, WA, OR, CA, NV, NM, AZ, UT, CO, WY, AK, HI)*

Application of structural equation model:

This subsection describes the application of structural equation modeling (SEM) to the previously described pedestrian-vehicle collision data. This approach has been used by (19, 20, and 21) to describe various traffic related aggressiveness and safety indexes. For example, (21) used SEM to analyze attributes of signalized and unsignalized intersections to identify which elements were most influential in terms of intersection safety. Previously, (19) created a similar safety index using SEM to describe driver behavior in interrupted versus uninterrupted flow. The success of structural equation modeling in describing the traffic scenarios mentioned above provides the motivation for applying SEM to pedestrian-vehicle collisions.

Structural equation modeling is particularly useful because it makes use of latent (unobserved) variables (22). Latent variables can be used to describe large data sets without actually having to consider every variable independently (22). Latent variables describe multiple observed variables grouped in one dimension and are therefore capable of reducing the complexity of the data set. Factor analysis is used to statistically group large sets of variables. For this project, a factor analysis was conducted using Statistical Analysis Software (SAS) in order to gain insights as to potential variable groupings (23). The factor analysis resulted in the grouping shown below in Table 3. Bolded values in Table 3 are indicative of refined potential dimensional placements as suggested by the factor scores.

Table 3: Factor Analysis Results

Factor Structure			
	Factor1	**Factor2**	**Factor3**
Intersection Legs	-0.00284	**0.74761**	-0.47938
Relation to Roadway	-0.15944	**0.23479**	0.00849
Lighting	0.00845	-0.25612	**0.46591**
Weather	-0.02453	-0.09225	0.05241
# of Occupants	-0.01724	-0.02293	0.00090
Driver Drinking	**0.75977**	-0.04502	0.07251
Flow Conditions	0.01749	-0.06755	-0.03589
Number of Lanes	-0.03196	0.02265	**0.19259**
Speed Limit	0.00335	-0.21729	**0.45397**
Control Device	-0.00209	**0.73935**	-0.46582
Driver Impaired	**0.77032**	-0.02266	0.00074
Driver Distracted	0.02064	0.13637	-0.18982
Pedestrian Impaired	**0.10441**	-0.12414	0.32598

Ultimately, these suggested groupings were refined based on logical connections and the statistical significance of various structural models that were estimated using the LISREL software (24). The final grouping of the variables is based on the following three categories: 1) Infrastructure variables, 2) Environmental variables, and 3) Impairment variables. The use of these well-defined categories resulted in the most statistically significant structural model. Figure 1, shown below, illustrates the use of LISREL software to identify a converging model using the three previously described latent variables.

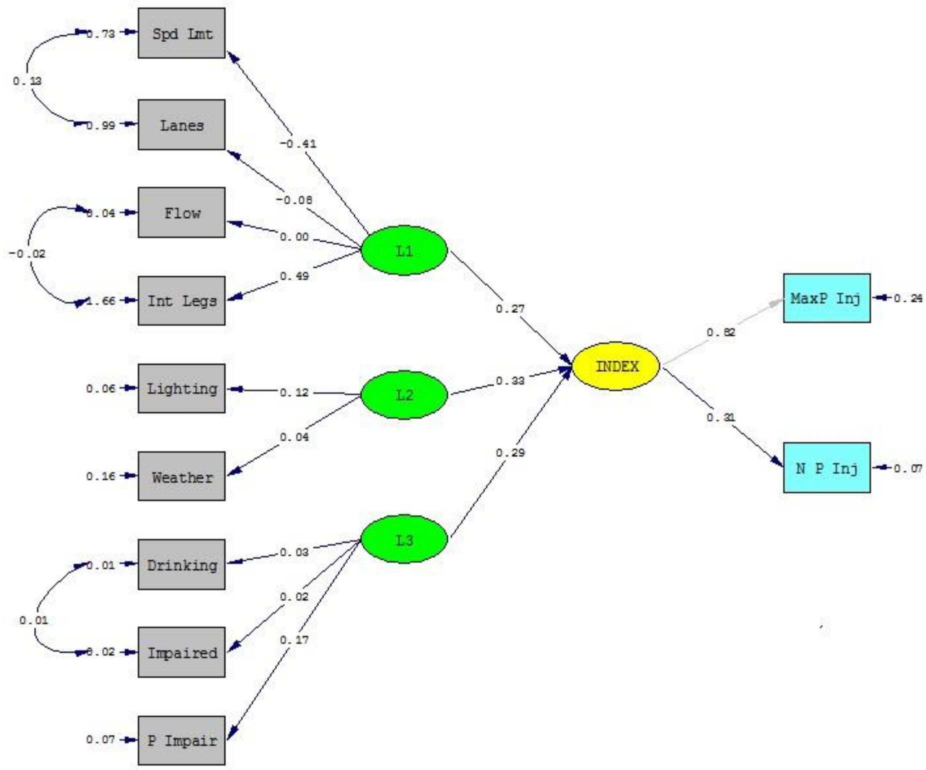

Figure 1: Structural Equation Model Corresponding to the NASS-GES Pedestrian-Vehicle Collisions

The results summarizing the model are shown below in Table 4A. While Chi-squared test experience problems due to the large number of data points (25), other relevant goodness of fit statistics were used to indicate model significance. Goodness of fit was based on the root mean square error of approximation (RMSEA) (25). The converging model resulted in an RMSEA of 0.036 with a 90% confidence interval of 0.034 to 0.037, well below the threshold of 0.05 which indicates statistical significance (25, 26). Furthermore, the standardized root mean square residual of 0.029 was below the accepted value of 0.08 (24). Additionally, the Goodness of Fit Index (0.99) and the Adjusted Goodness of Fit Index (0.98) further support the statistical significance of the model. Considering an alpha value of 0.05, all t-values between -1.96 and 1.96 are considered significant, indicating that each variable's t-value was significant, as shown in the table below in Table 4B.

Table 4A: Details on the Model Measurement Equations

Model measurement equations		
Equation	Errorvar	R^2 Value
Structural model		
Index = 0.27*L1 + 0.33*L2 + 0.29*L3	0.98	0.016
MaxP Inj = 0.82*Index	0.24	0.74
N P Inj = 0.31*Index	0.072	0.56
Exogenous measurement model		
Int Legs = 0.49*L1	1.66	0.13
Flow = 0.0035*L1	0.038	0.00033
Lanes = - 0.079*L1	0.99	0.0063
Spd Lmt = - 0.41*L1	0.73	0.19
Weather = 0.039*L2	0.16	0.0095
Lighting = 0.12*L2	0.062	0.19
Drinking = 0.026*L3	0.014	0.046
Impaired = 0.019*L3	0.021	0.017
P Impaired = 0.17*L3	0.073	0.28

Table 4B: T-Values and Significant Error Covariance Terms

T-values	
Variables	Value
L1/Intersection Legs	56.74
L1/Number of lanes	-14.38
L1/Flow Condition	3.34
L1/Speed Limit	-61.93
L2/Weather	15.10
L2/Lighting	20.07
L3/Driver Drinking	30.30
L3/Driver Impairment	18.54
L3/Pedestrian Impairment	37.04
L1/Index	6.37
L2/Index	8.20
L3/Index	11.79
Index/Number of pedestrian injuries	58.74
Error covariance terms	
Variables	Value
Flow Condition/Intersection Legs	-0.62
Speed limit/Number of lanes	0.73
Driver Impairment/Driver Drinking	0.011

4. RESULTS AND ANALYSIS

General analysis:

Based on the structural model displayed in Figure 1, higher values of the safety index are indicative of a *decrease* in pedestrian safety, as both the number of pedestrians injured and the maximum pedestrian injury severity level increase. High composite contributions to the index indicate variables with greater impact on pedestrian safety. The composite contribution is calculated by multiplying the coefficient of variation (the amount in which a variable changes throughout the data set) by the contribution to the safety index from a one standard deviation change in a variable. This value is paramount for analysis as it considers both how often a variable changes throughout the data set and the contribution to the safety index from a change in that variable.

Table 5: Variable Specific Statistical Measures

Variable	Average	Standard Deviation	Coefficient of Variation	Deviation Contribution	Composite Contribution
L1: Intersection Legs	2.4796	1.3801	0.5566	2.8166	1.5676
L1: Number of Lanes	2.6140	0.9984	0.3819	-12.6376	-4.8267
L1: Flow Condition	1.9604	0.1950	0.0995	55.7055	5.5401
L1: Speed Limit	3.3574	0.9487	0.2826	-2.3138	-0.6538
L2: Lighting	0.0830	0.2759	3.3241	2.2989	7.6418
L2: Weather	0.1998	0.3998	2.0015	10.2517	20.5191
L3: Driver Drinking	0.0148	0.1206	8.1692	4.6386	37.8937
L3: Driver Impaired	0.0213	0.1445	6.7735	7.6045	51.5089
L3: Pedestrian Impaired	0.1148	0.3188	2.7769	1.8751	5.2071

Starting with the infrastructure dimension (L1), positive valued coefficients for intersection legs and flow condition indicate that as intersections become more complex and traffic flow goes from one way to two ways, there is a negative impact on pedestrian safety. The complexity of navigating through traffic for pedestrians in situations involving intersections with numerous intersection legs and multi-directional flow is shown by the model. These findings are supported by both (13) and (2) who found that street type and design are influential in terms of pedestrian safety. The remaining two infrastructure variables, number of lanes and speed limit, both have negative coefficients. The negative values indicate that as these values increase, there is a positive effect on pedestrian safety. The negative influence of the speed limit contradicts the findings made by (13). However, the negative coefficient of these two variables makes sense when considering wide roads with vehicles travelling at high speeds. Situations like this are often discouraging to pedestrians who wish to cross the roadway, which may explain why the coefficients are negative.

The infrastructure dimension is particularly revealing because it takes into account aspects of the physical roadway that have been designed in a certain way. In areas where there is a high amount of pedestrian traffic, this model supports decreasing the number of intersection lanes and limiting vehicular flow conditions. For planning purposes, this model is capable of indicating which factors must be the most carefully considered with regard to their impact on pedestrian safety.

The next dimension (L2) includes environmental characteristics of the roadway, namely lighting condition and weather. Both weather and lighting are coded as dummy variables with 0 meaning light/dry conditions and 1 meaning dark/visual impairment due to precipitation. The composite distribution shown in Table 5 indicates that weather has a more significant impact on pedestrian safety than lighting. The effect of inclement weather on pedestrians has been explored by (27) in the context of increasing signal timing, but not directly related to pedestrian-vehicle collisions. Additionally, (15) recommend paying attention to the effect of lighting conditions on pedestrian-vehicle safety. As shown by the model described here, darkness is a significant contributor to decreased pedestrian safety.

Although the environmental variables are less easy to control during the design stage of roadway construction, there are still some options that can be taken to limit the impact of these variables on pedestrian safety. For example, lighting conditions can be improved by street lights. Visual impairment due to weather conditions is more difficult to change; however it is important for pedestrians to be aware of their increased risk of collision while precipitation is falling.

The last dimension (L3) corresponds to pedestrian and driver impairment. All three of the observed variables in the category, including pedestrian impairment, driver impairment, and driver consuming alcohol in the vehicle are attributed with a decrease in pedestrian safety. Although pedestrian impairment has the least significant composite contribution in this group, it is still influential. The model indicates the driver impairment and alcohol consumption are the most critical with regard to pedestrian safety, with composite contribution values that greatly exceed the contributions from every other variable. The negative impact on pedestrian safety due to alcohol has been reported by many other studies including (7 and 28).

The effect of impaired driving on overall roadway safety is not a new discovery, and many solutions have already been enacted through law. This model illustrates that impaired driving is not only dangerous to others travelling in vehicles, but also that driver impairment has a hugely negative effect on overall pedestrian safety. Driver impairment is more influential than pedestrian impairment in the model because collisions involving impaired drivers are more likely to lead to severe injury than collisions involving non-impaired drivers and impaired pedestrians.

Regional analysis:

One of the main benefits of the SEM approach is that it allows for a ranking system to be developed based on the structural model produced. From the structural model described above, the four regions of the country previously mentioned in Table 2 can be ranked based on the three dimensions considered for analysis as well as the pedestrian safety index as a whole. Table 6, below, displays the average dimensional value for each region as well as the average value of the pedestrian safety index. The table indicates that variable group L1, infrastructure variables, is the most influential in terms of pedestrian safety in region 1, which corresponds to the Northeast of the United States. Variable group L2, environmental variables, is most influential in region 2 of the country, which is the Midwest. Lastly, variable group L3, impairment variables, is the most strongly represented in region 3, which represents the Southern United States. It is interesting to note the different influences of particular variables in the various regions of the country. Although differences exist, it is clear from the table that all three groups of variables are influential in all four quadrants of the country, indicating that the solutions suggested in studies that focus on particular cities may be applicable on a more widespread scale.

Table 6: Regional Specific Latent Variables Values

Region	L1	L2	L3	INDEX
1	144.5770	1.2708	0.3662	146.2141
2	142.8773	2.5115	0.5154	145.9043
3	139.2650	2.0344	0.8083	142.1077
4	143.1763	1.3573	0.7412	145.2749

The above table shows that all four regions of the country have similar safety indexes, although it can be seen that region 1 has the highest safety index value, and therefore it is the least safe region for pedestrians. Since this region corresponds to the Northeastern United States, it is possible that the increase is due to high population density. The relatively small differences in the safety indexes for each region of the country suggest that the observed variables that were taken into account are significant in all four regions.

The convergent model that was produced from NASS GES data indicates that infrastructural and behavioral changes can greatly increase pedestrian safety on roadways. Furthermore, since the data was not specific to a particular city it is likely that widespread solutions are a possibility for improving pedestrian safety in all four regions of the United States.

5. Conclusion

This study made use of structural equation modeling to identify specific parameters that are influential in terms of pedestrian safety, particularly with regard to pedestrian-vehicle collisions. Demographic variables were purposefully excluded from the study so that the findings could be used to identify changes that can be made to the infrastructure as well as driver behaviors that are damaging to pedestrian safety. Since demographic variables cannot be easily changed, their effect on pedestrian-vehicle collisions was not taken into account. The structural equation model identified three main variables groups with large influences over pedestrian safety, including infrastructure variables, environmental variables, and impairment variables. The results produced in this study were created using data from across the country in an effort to make the conclusions as broadly applicable as possible. Future work could identify similar variables in a smaller area in order to verify that the observed variables are as influential on a small scale as they were in this model.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 0927138. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

6. References

1. Traffic Safety Facts 2012. Washington, DC: National Highway Traffic Safety Administration; 2014.
2. Retting, R., Ferguson, S., McCartt, A. A Review of Evidence-Based Traffic Engineering Measures Designed to Reduce Pedestrian-Motor Vehicle Crashes. American Journal of Public Health: September 2003, Vol. 93, No. 9, pp. 1456-1463.
3. Yeaton W. H, Bailey J. S. Teaching pedestrian safety skills to young children: an analysis and one-year follow up. Journal of Applied Behavior Analysis. 1978, 315–329.
4. Wiener, E.L. A. The Elderly Pedestrian: Response to an Enforcement Campaign. In Transportation Research Record: Journal of the Transportation Research Board, Vol. 12, Transportation Research Board of the National Academies, Washington, D.C., 1968, pp. 100–110.
5. Ukkusuri, S., Miranda-Moreno, L., Ramadurai, G., Isa-Tavarez, J. The Role of the Built Environment on Pedestrian Crash Frequency. Safety Science, 50. 2012, 1141-1151.
6. Wier, M., Weintraub, J., Humphreys, E., Seto, E., Bhatia, R. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accident Analysis and Prevention, 41. 2009. 137-145.
7. LaScala, E., Gerber, D., Gruenewal, P. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. Accident Analysis and Prevention, 32. 2000. 651-658.
8. Graham, DJ, Glaister, S. Spatial variation in road pedestrian casualties: The role of urban scale, density and land-use mix, *Urban Studies*, 40. 2003. 1591-1607
9. Laflamme, L., Diderichsen, F. Social differences in traffic injury risks in childhood and you-a literature review and a research agenda. Injury Prevention, 6. 2000. 293-298.
10. Ryb, G., Dischinger, P., Kufera, J., Soderstrom, C. Social, behavioral and driving characteristics of injured pedestrians: A comparison with other unintentional trauma patients. Accident Analysis and Prevention, 39. 2007. 313-318.
11. Lee, C., Abdel-Aty, M. Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. Accident Analysis and Prevention, 37. 2005. 775-786.
12. Loukaitou-Sideris, A., Liggett, R., Sung, H. Death on the Crosswalk: A Study of Pedestrian-Automobile Collisions in Los Angeles. Journal of Planning and Education Research, 26. 2007. 338-351.
13. Ewing, R., Dumbaugh, E. The Built Environment and Traffic Safety: A Review of Empirical Evidence. Journal of Planning Literature, 23. 2009. 347-367.
14. Zahabi, S., Strauss, J., Manaugh, K., Miranda-Moreno, L. Estimating Potential Effect of Speed Limits, Built Environment, and Other Factors on Severity of Pedestrian and Cyclist Injuries in Crashes. . In Transportation Research Record: Journal of the Transportation Research Board, 2247, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 81-90.
15. Clifton, K., Burnier, C., Akar, G. Severity of injury resulting from pedestrian-vehicle crashes: What can we learn from examining the built environment? Transportation Research Part D, 14. 2009. 425-436.

16. Schuurman, N., Cinnamon, J., Crooks, V., Hameed, S. Pedestrian injury and the built environment: an environmental scan of hotspots. *BMC Public Health* 2009.
17. LaScala, E., Gruenewald, P., Johnson, F. An ecological study of the locations of schools and child pedestrian injury collisions. Accident Analysis and Prevention, 36. 2004. 569-576.
18. Newbury, C., Hsiao, K., Dansey, R., Hamill, J. Paediatric pedestrian trauma: The danger after school. Journal of Paediatrics and Child health, 44. 2008. 488-491.
19. Hamdar, S.H., Schorr, J. Interrupted versus uninterrupted flow: a safetypropensity index for driver behavior. Accident Analysis and Prevention, 55. 2013. 22–33.
20. Hamdar, S.H., Mahmassani, H.S., Chen, R.B. Aggressiveness propensity index for driving. Accident Analysis and Prevention, 40. 2008.315–326.
21. Schorr, J., Hamdar, S., Safety propensity index for signalized and unsignalized intersections: Exploration and assessment. Accident Analysis and Prevention, 71. 2014. 93-105.
22. Ullman, J. B. and Bentler, P. M. Structural Equation Modeling. Handbook of Psychology. Four. 2003. 607–634.
23. SAS software, Version 9.3 of the SAS system for Windows. Copyright 2011, SASInstitute Inc. SAS and all other SAS Institute Inc. Product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.
24. LISREL software, Version 8.8 of the LISREL software for Windows. Copyright 2004, Scientific Soft-ware International, Inc. Lincolnwood, IL, USA.
25. Golob, T.F., 2001. Structural Equation Modelling for Travel Behavior Research,Irvine: Institute of Transportation Studies. University of California Irvine, Irvine,California.
26. Hu, L., Bentler, P.M., 1998. Fit indices in covariance structure modeling: sen-sitivity to underparameterized model misspecification. Psychol. Methods,424–453.
27. Perrin, J., Martin, P., Hansen, B. Modifying Signal Timing During Inclement Weather. In Transportation Research Record: Journal of the Transportation Research Board, 1748, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 66–71.
28. Ostrom, M., Eriksson, A. Pedestrian fatalities and alcohol. Accident Analysis and Prevention, 33. 2001. 173-180.

List of Tables:

Table 1: NASS GES Total Collisions per Year in 2011 and 2012
Table 2: Description of Model Variables
Table 3: Factor Analysis Results
Table 4A: Details on the Model Measurement Equations
Table 4B: T-Values and Significant Error Covariance Terms
Table 5: Variable Specific Statistical Measures
Table 6: Regional Specific Latent Variables Values

List of Tables:

Figure 1: Structural Equation Model Corresponding to the NASS-GES Pedestrian-Vehicle Collisions