TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
2019, Vol. 2673(7) 308–318

DOI: 10.1177/0361198119845353
journals.sagepub.com/home/trr
SAGE

Analyzing Pedestrian and Bicyclist Crashes at the Corridor Level: Structural Equation Modeling Approach

Farah J. Al-Mahameed[1], Xiao Qin[1], Robert J. Schneider[2], and Mohammad Razaur Rahman Shaon[1]

Abstract
Pedestrian and bicycle crashes have been increasing at an alarming pace in recent years. Between 2009 and 2016, annual U.S. pedestrian fatalities increased 46%, and bicyclist fatalities increased 34%. Crashes involving pedestrians and bicyclists, or vulnerable road users (VRUs), are negatively correlated with roadway factors, and positively correlated with environmental and socioeconomic factors. However, specific variables representing these factors are often correlated, making it difficult to accurately characterize relationships between individual variables and pedestrian and bicyclist safety. This study used the structural equation model technique to overcome this problem. Pedestrian and bicyclist crash frequency and more than 60 explanatory variables for 200 highway corridors in Wisconsin were collected. The interrelationships between observed "manifest" variables and unobserved "latent" variables were tested. The results suggest that the most important latent variables influencing the crash frequency of VRUs are bicycle/pedestrian-oriented roadway design (e.g., paved shoulders, sidewalks, and bike lanes), exposure (e.g., walking and biking activity, and employment density), and low social status (e.g., educational level, and wage percentage). The benefits of this study may help community planners, transportation researchers, and policymakers with a better understanding of the intricate interrelationship of the influential factors contributing to VRUs road crashes.

Pedestrians and bicyclists are commonly referred to as vulnerable road users (VRUs) because they often suffer the most in road traffic crashes. In addition to not being protected by the body of a vehicle, they are unable to take advantage of many safety features integrated into vehicles such as airbags and seat belts. According to the Fatality Analysis Reporting System, pedestrian fatalities involving motor vehicles increased by 46% (4,109–5,987) and bicyclist fatalities increased by 34% (628–840) between 2009 and 2016 (*1*). VRUs accounted for more than 18% of the 37,461 total U.S. fatalities in 2016, up from a low of 13% in 2003 (*2*).

Walking and bicycling are inexpensive modes of transportation that provide physical activity, use space efficiently, and produce little pollution. However, personal safety is often seen as a barrier to using these sustainable travel modes (*3*). To improve the safety of VRUs, it is important to understand how the characteristics of local environments are associated with pedestrian and bicyclist crashes. Many researchers have explored variables that contribute to pedestrian and bicyclist crashes (*4–9*). Previous studies have identified many variables, or categories of variables, that are related to VRU injuries and fatalities, such as roadway geometry, vehicle characteristics, socioeconomic characteristics, and environmental conditions.

The statistical methods previously used are aimed to construct models that represent the direct relationships between explanatory and dependent variables. However, the causes of crashes often involve intricate relationships among multiple variables, which may not be adequately captured. The interrelationships among explanatory variables may be better understood by applying the structural equation modeling (SEM) technique, as SEM is generally viewed as a combination of factor analysis and path analysis. This highly flexible model structure is capable of representing the complex interrelationships among exogenous and endogenous variables through the inclusion of "unobserved" or latent variables.

[1]Civil and Environmental Engineering Department, University of Wisconsin-Milwaukee, Milwaukee, WI
[2]School of Architecture & Urban Planning, University of Wisconsin-Milwaukee, Milwaukee, WI

Corresponding Author:
Address correspondence to Xiao Qin: qinx@uwm.edu

Specifically, SEM can handle correlations between explanatory variables that represent similar concepts and have overlapped impacts on the dependent variable(s).

This study applies SEM by integrating exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) which is a special case of SEM to establish the relationship between pedestrian and bicyclist crashes and explanatory variables. Specifically, the study explores research questions such as: *What are the most important latent variables associated with the frequency of crashes involving VRUs*? *Among the key latent variables, what combinations of measurable variables provide the most significant representation of these key latent variables*? The analysis uses pedestrian and bicyclist crash data from a unique, detailed dataset describing 200 1-mi corridors along the Wisconsin State Highway System.

Literature Review

Previous studies have shed light on dominant factors related to pedestrian and bicycle crashes. Many have examined roadway geometric characteristics such as number of lanes, median type, speed limits, and speed ratio (i.e., the ratio of speed from crash data over the posted speed limit). Morency et al. confirmed the association between wider roads and higher pedestrian-related crash frequency, as a wider road may encourage drivers to speed and jeopardize pedestrians. Designated right-turn lanes and nearby driveway crossings were associated with higher pedestrian crash risk, whereas median crossing refuges were associated with lower pedestrian crash risk at intersections (*10*). Quiet streets, gentle slopes, and absence of streetcar tracks are some design features associated with lower bicyclist crash risk (*11*). Moreover, cycle tracks (*11*, *12*), traffic diverters, and local streets tend to separate cyclists from the moving traffic, leading to fewer crashes (*11*). Roads with more traffic signals, street parking signs, and automobile trips are associated with more frequent bicycle crashes. Roadways with a speed limit of 35 mph and intersection density are positively related to the likelihood of pedestrian and bicyclist crashes (*11*). In addition, Cai et al. stated that traffic analysis zones (TAZs) with longer sidewalk lengths, more pedestrians, and more employment are more susceptible to have pedestrian crashes, whereas TAZs with longer sidewalk lengths, more employment, and higher population density are more likely to have bicyclist crashes (*13*). Road speed limit was also found to affect bicycle crash frequency, as Siddiqui and colleagues stated that highways with a speed limit greater than 35 mph are more likely to have bicyclist crashes (*24*). Density of signalized intersections, arterial and local road proportion, and sidewalk length, are positively correlated with pedestrian and/or bicyclist–motor-vehicle crashes (*14*). While controlling for exposure variables, several studies have identified specific pedestrian facilities to be negatively associated with pedestrian crashes, including median refuge islands and rectangular rapid flashing beacons (*4*).

Some studies have explored how behaviors are related to pedestrian and bicyclist crashes. Helmet usage, travel programs such as routes to school, wearing reflective clothing, and education related to safety among bicyclists have been associated with fewer bicyclist fatalities. Pedestrians and bicyclists crossing a red-light signal or using mobile devices, and motor-vehicle drivers turning right on red without waiting for other road users to cross are some of the most important safety behaviors studied (*15*). Vehicle speed impact on pedestrian fatality risk reported that car speed positively and strongly affects fatality risk among pedestrians. Above 20 mph, small increases in speed produce relatively large increases in pedestrian injury severity (*16*, *17*).

Other studies have identified exposure as an important variable associated with pedestrian and bicyclist crashes. There are a variety of exposure measures in the literature (i.e., using census journey to work data as an exposure proxy variable (*18*)), but this concept is commonly represented using pedestrian, bicyclist, and automobile counts (*19*, *20*). Several studies showed the relationship between the number of pedestrian and bicyclist crashes and pedestrian and bicyclist activity levels. Results confirm that the relationship is not linear (commonly referred as "safety in numbers" effect): pedestrian or bicycle crash risk (e.g., crashes per crossing or per trip) decreases with the increase in walking or cycling (*19–23*). One challenge for using this important variable in safety analyses is that few jurisdictions have sufficient pedestrian or bicyclist count data, resulting in the use of proxy variables to represent exposure.

Regarding economic, demographic, and social characteristics, Table 1 summarizes factors stated in some previous studies.

However, new research is needed to unravel the true underpinning for pedestrian and bicycle crashes because these minor race groups have relatively low car ownership, live in high-density low-income areas, and tend to walk/bike more (*29*).

Although SEM is common in the academic literature, a relatively small number of traffic safety studies have applied this technique. A study used SEM to examine motor-vehicle crash severity in relation to accessibility, human, vehicle, and roadway-related factors (*30*) which were treated as latent factors. SEM was used to study the frequency of crashes on Korean highways in which several exogenous latent variables were defined, such as driver, road, and environmental factors (*31*). Schorr and Hamdar used SEM to develop a safety propensity index for both signalized and unsignalized intersection (*32*).

Table 1. Summary of Economic, Demographic, and Social Characteristics Influencing Pedestrian and Bicycle Crash Frequency

Author name	Economic, demographic, and social characteristics	Emphasis
Siddiqui et al. (*24*)	Total population, proportion of uneducated population, land use (presence of restaurants and bars), and park coverage	Pedestrians
Nashad et al. (25)	Number of dwelling units, population density, total employment and percentage of households with zero or one car ownership	Pedestrians and bicyclists
Lee and Abdel-Aty (26)	Vehicle-miles traveled (VMT), middle-aged (25–64) and male drivers, neighborhoods with large retail and residential land uses, high vehicular traffic movements, high employment and population density, low-income, and high minority environmental justice areas and races	Pedestrians
Loukaitou-Sideris et al. (*27*)	Hotel room density, number of people walking/biking, population density, school enrollment density, proportion of industrial employment, low-income, and high minority environmental justice areas and races	Pedestrians and bicyclists
Nordback et al. (*28*)	Age <18 years old, neighborhoods with large retail and residential land uses, high vehicular traffic movements, and high employment and population density	Bicyclists
Schneider et al. (*22*)	Percentage of households without access to private vehicles	Pedestrians and bicyclists
Morency et al. (*10*)	Neighborhoods with large retail and residential land uses, high vehicular traffic movements, and high employment and population density	Pedestrians

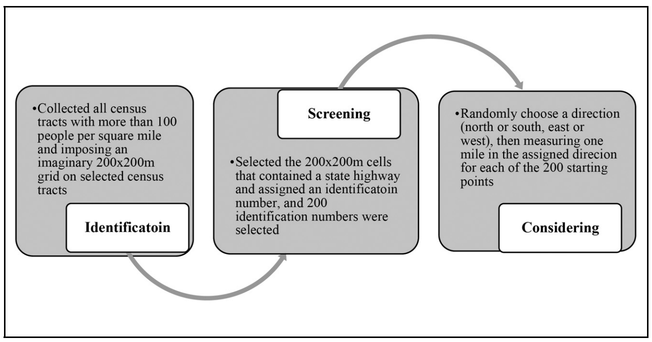

Figure 1. Corridor selection process.

The authors argued that this singular valued index offers an effective method for quantifying and ranking intersection safety as compared with the use of multiple criteria (e.g., number of vehicles involved, total injuries, total fatalities). Wang and Qin used SEM to test single-vehicle crash severity influenced by driver characteristics, highway geometry and roadway conditions, roadside objects, and environmental factors. The authors wanted to understand how the observed variables affect the crash consequence in a direct or indirect manner through collision force, vehicle operating speed before the collision, and severity index, which are unobserved or unmeasurable in most safety studies (*33*). Although the application of the SEM approach in pedestrian and bicyclist crash modeling is rare, other studies have utilized it to model road user traveling behavior and mobility.

Data

This study follows Cai and colleagues' recommendation to further study the common unobserved factors affecting pedestrian and bicyclist crashes (*34*). In contrast to previous studies that primarily focused on predicting pedestrian and bicyclist crashes at specific locations (e.g., intersections), this study focuses on a sample of 200 1-mi-long highway corridors in Wisconsin. Figure 1 illustrates the corridor selection process. The corridors are in the areas with at least 100 residents per square mile, generally including cities, suburbs, and villages but excluding rural areas in Wisconsin. Although the spatial diversity is desirable, the focus was on urbanized areas as these areas tend to have higher volumes of pedestrians and bicyclists and more pedestrian and bicyclist crashes. Among the 200 study corridors, most are located in the Southeast Wisconsin area; 115 had at least one reported pedestrian crash and 67 had at least one reported bicycle crash.

The study examined the frequency of pedestrian and bicyclist crashes reported to police between 2011 and 2015 in each study corridor. These data were gathered from the Wisconsin Department of Transportation (WisDOT) WisTransPortal Database and only included crash records with latitude and longitude coordinates. Explanatory variables were collected from multiple databases including the WisDOT highway inventory, U.S. Environmental Protection Agency's (EPA's) Smart Location Database (SLD), U.S. Census Topologically Integrated Geographic Encoding and Referencing (TIGER/Line) dataset, and Google Maps and Google Street View imagery. Explanatory variables included exposure-related variables (e.g., annualized average daily traffic (AADT)), roadway segment characteristics (e.g., motor-vehicle AADT, average number of through lanes, and posted speed limit), roadway intersection characteristics (e.g., number of residential/nonresidential driveways, number of signalized/un-signalized intersections,

Table 2. Description and Summary Statistics of the Corridor Variables (N = 200)

Notation	Description	Coding	Mean (standard deviation) or percentage
Wisconsin Information System for Local Roads (WISLR)			
Ped_1115	Number of pedestrian crashes (2011–2015)	Continuous	1.91 (3.4)
Bike_1115	Number of Bicyclist crashes (2011–2015)	Continuous	0.85 (1.69)
Google Maps and Google Street View imagery			
High_Spd_Lmt	Posted speed limit higher than 35 mph	1 = Yes 0 = No	1 = 46% 0 = 54%
Pav_Shoulder	Percentage of corridor covered by paved shoulders on both sides (shoulder on only one side for full length = 0.5)	1 = Yes 0 = No	1 = 72 % 0 = 28 %
Bikelane	Percentage of corridor covered by designated bike lanes on both sides (bike lane on only one side for full length = 0.5)	Continuous	0.09 (0.26)
Sidewalk	Percentage of corridor covered by sidewalks on both sides (sidewalk on only one side for full length = 0.5)	Continuous	0.39 (0.44)
Sidepath	Percentage of corridor covered by side paths on both sides (sidepath on only one side for full length = 0.5)	Continuous	0.05 (0.19)
Unsignalized	Unsignalized intersections along corridor	Continuous	5 (4)
Mid_Block	Marked midblock crosswalks across the state highway along the corridor	Continuous	0.04 (0.24)
TWLTL	Percentage of corridor length with a two-way left-turn lane	Continuous	0.06 (0.16)
US Census TIGER/Line dataset			
log_AADT	Natural log of the average of all annualized average daily traffic (AADT) volume counts along the corridor	Continuous	9.359 (0.66)
Walk	Transportation mode used to travel to work (walking)	Continuous	0.027 (0.030)
Bike	Transportation mode used to travel to work (biking)	Continuous	0.006 (0.012)
Employ_Density	Gross employment density (jobs/acre) on unprotected land	Continuous	2.18 (3.81)
Edu_Less_H	Percentage of educational attainment for the population 25 years and over: less than high school	Continuous	0.09 (0.07)
EPA/SLD			
Total_Veh0	Percentage of population with zero car ownership	Continuous	0.07 (0.07)
Low_Wage	Percentage of workers earning $1250/month or less (home location), 2010 decennial census	Continuous	0.28 (0.04)
Poverty	Poverty status in the past 12 months by disability status by employment status for population 20 to 64 years for whom poverty status is determined (percentage)	Continuous	0.12 (0.10)

number of right-turn/left-turn lanes on state highway approaches to all intersections), and socioeconomic data from surrounding census tracts. None of the study corridors had pedestrian or bicyclist counts, so proxy variables were used to represent pedestrian and bicyclist exposure, such as the percentage of workers who regularly walked or bicycled to work, population density, and job density in the surrounding neighborhoods (based on census block groups). Table 2 shows summary statistics and description of the dataset.

Theoretical Framework of SEM

The primary interest of using SEM lies in the test of its theoretical construct which is specified by latent variables and their relationships. As shown in Figure 2 (*31*), an SEM model can be depicted in a path diagram consisting of boxes and circles, which are connected by arrows. Observed variables are usually represented by square or rectangular boxes (e.g., Two-Way Left-Turn Lane (TWLTL)), whereas unobserved or latent variables

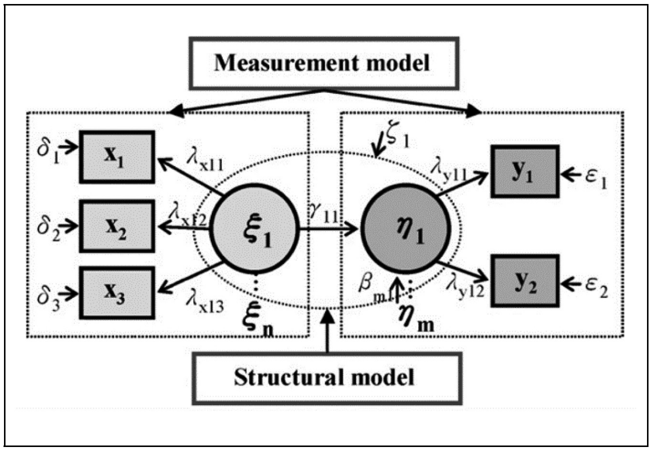

Figure 2. Example of a structural equation model (variable definitions are shown in Table 3).

are usually represented by circles or eclipses (e.g., Low Social Status). A directional arrow (or path) in the model usually indicates a statistical dependence, in which the variable at the tail of the arrow causes the variable at the point. A double-headed arrow does not represent such a statistical dependence, but an indication of correlation between variables.

Through the x-measurement model for exogenous variables, a y-measurement model for endogenous variables, and the structural model between latent variables, SEM is able to differentiate between direct, indirect, and total effects between variables. By combining the structural model with measurement models, SEM expresses the regression effects of exogenous "independent" variables on the endogenous "dependent" ones, as well as, expressing autocorrelation "effects between endogenous variables." For more details see Schumacker and Lomax (*35*).

The formulation of SEM in Equation 1 suggests a structure between the covariances between observed variables (*36*):

$$\begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} A_y & 0 \\ 0 & A_x \end{bmatrix} \begin{bmatrix} \eta \\ \xi \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \delta \end{bmatrix} \tag{1}$$

The model goodness of fit can be measured by the comparative fit index (CFI) (*37*) in Equation 2 and the root mean square error of approximation (RMSEA) (*38*) in Equation 3.

$$\mathrm{CFI} = 1 - \frac{\tau_{\mathrm{est.model}}}{\tau_{\mathrm{indep.model}}} \tag{2}$$

where,

$$\tau_{\mathrm{indep.model}} = \chi^2{}_{\mathrm{indep.model}} - \mathrm{df}_{\mathrm{ndep.model}}$$
$$\tau_{\mathrm{est.model}} = \chi^2{}_{\mathrm{est.model}} - \mathrm{df}_{\mathrm{est.model}}$$

$$\text{Estimated RMSEA} = \sqrt{\frac{\tau_{\mathrm{est.model}}}{\mathrm{Ndf}_{\mathrm{model}}}} \tag{3}$$

where, τ_i: degree of misspecification of the model
χ^2: chi-square statistic
df: degree of freedom
N: sample size

CFI is calculated using χ^2 statistics for two models—the target and the baseline models—and measures how better the model fits with a comparison to the baseline model. The baseline model includes means and variances of the observed variables in addition to the covariances of the observed exogenous variables. Both indices (RMSEA and CFI) assume that the target model is approximately correct, but CFI carries another assumption that the baseline model is also correct. CFI is based on the assumption that all latent variables are uncorrelated and performs well even when the sample size is small (*39*). Values for CFI range between 0 and 1, with a value closer to 1 indicating a better fit. RMSEA—which is a function of a chi-square and degree of freedom—measures the difference between the observed and

Table 3. SEM Elements

Model	Variable	Variable description
Measurement	x	q × 1 column vector of observed exogenous variables
	y	p × 1 column vector of observed endogenous variables
	ξ	n × 1 column vector of latent exogenous variables
	η	m × 1 column vector of latent endogenous variables
	δ	q × 1 column vector of measurement error terms for observed variables x
	ε	p × 1 column vector of measurement error terms for observed variables y
	A_x	The matrix (q × n) of structural coefficients for latent exogenous variables to their observed indicator variables
	A_y	The matrix (q × n) of structural coefficients for latent endogenous variables to their observed indicator variables
Structural	Γ	The matrix (m × n) of regression effects for exogenous latent variables to endogenous latent variables
	β	The coefficient matrix (m × m) of direct effects between endogenous latent variables
	ς	m × 1 column vector of error terms

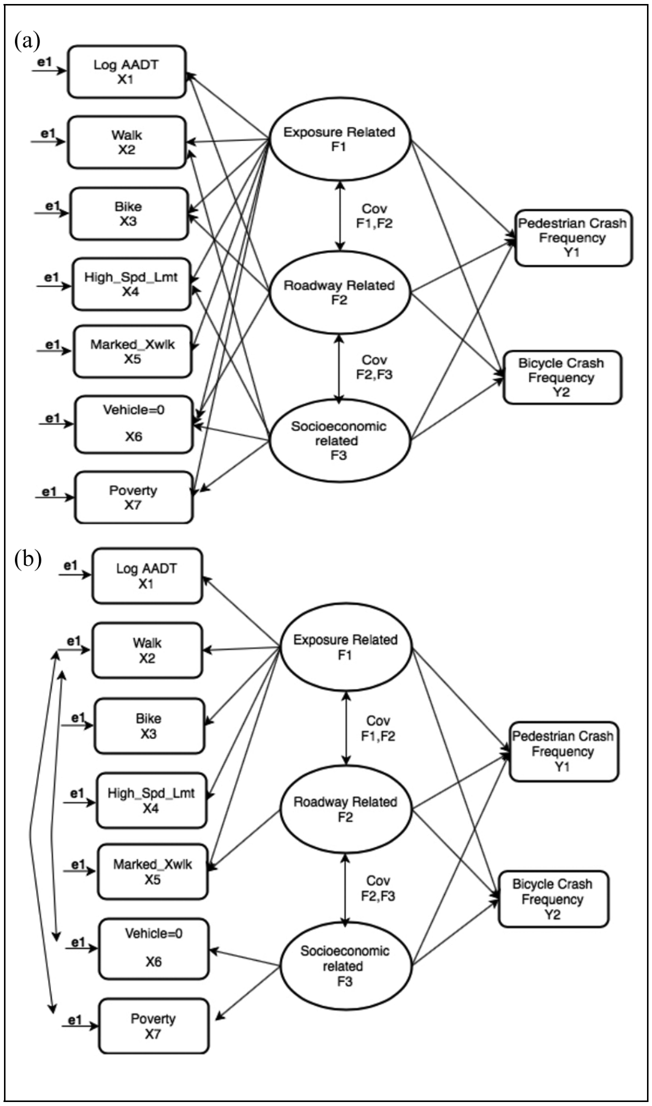

Figure 3. (*a*) Illustration of the conceptual distinction between EFA and (*b*) CFA.

predicted values (*40*). A value of less than 0.08 indicates a good fit model.

SEM Model Specification, Estimation, and Evaluation

Exploratory Factor Analysis

A hypothesized model assessed the relative factors affecting pedestrian and bicyclist crash risk (Figure 3) and indicated that exposure, roadway, and socioeconomic should be used as latent variables that connect exogenous and endogenous variables. The behavioral-related latent is not included as the dataset lacks behavioral input variables. The three latent variables (oval shape) are predictors of the number of crashes that involve either pedestrians or bicyclists (square shape) taking place on the study segments. The latent variables are allowed to correlate. A substantive theoretical model does not exist, so the EFA is used to obtain the empirical factor model and explore the structural portion of SEM. EFA assumes that every observed variable is an indication or a measurement of a latent variable (Figure 3*a*). EFA is usually performed as a precursor to CFA (*41*), which confirms theoretically valid relationships (Figure 3*b*).

Researchers vary in relation to sample size recommendations for factor analysis. A sample size of 200 and the sample-to-variable ratio of 3:1 kept this study within the acceptable ranges for applying factor analysis (*39*). To test the ability to apply factor analysis using the data, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was applied. It is hypothesized that the correlation matrix is an identity matrix, so Bartlett's Test of Sphericity was run to test this hypothesis (*40*). Bartlett's test of sphericity resulted in a P-value of 0.02 (<0.5 is recommended), and a KMO index value of 0.8091 (>0.6 is recommended); both are considered meritorious (*42*).

The strength of the linear relationship between two variables is crucial. Many variables in the study dataset were highly correlated (i.e., the correlation between the percentage of high wage and total vehicles of two or more is 0.79), and these variables were not used together in the same model. A threshold of 0.5 was accepted as a correlation coefficient between the set of variables chosen for the analysis (*39*, *40*).

The number of latent variables can be evaluated using the visual tool called the Scree test. The Scree test showed a clear drop between the third and fourth components, meaning the most suitable number of factors lies between three and four factors. In addition, goodness of fit indices in the three-factor EFA model show acceptable values (RMSEA = 0.000; CFI = 1.064).

The estimation method for factor loading coefficients, which measure the strength between observed and latent variables, relies on data quality. The maximum likelihood (ML) or principal axis factoring method is recommended, depending on whether the data are normally or significantly non-normally distributed (*43*). The ML estimation method was chosen after variables were standardized through the "scale () function" and the rotation method is variance maximizing (varimax) rotation. EFA was conducted using the first part of the sample size of 100 corridors. Table 4 shows the results of the EFA. Despite the cross-loading that appeared in (High_Spd_Lmt) variable, the observed variables that are highly correlated with factors show distinctive characteristics. (High_Spd_Lmt), (Bikelane), (Pav_Shoulder), (Sidewalk), and (Unsignalized) are highly correlated with F1 which can be called pedestrian and bicycle-oriented roadway. (Walk), (Bike), (Employ_Density), and

Table 4. EFA with Varimax Rotation Factor Loadings for the Measurement Model (N = 100)

Variable	Factors[a]		
	Pedestrian and bicycle-oriented roadway (F1)	Exposure (F2)	Low social status (F3)
High_Spd_Lmt	**−0.55**	**−0.64**	−0.28
Bikelane	**0.41**	0.11	0.11
Pav_Shoulder	0.38	−0.11	−0.06
Sidewalk	**0.63**	0.04	0.17
Sidepath	−0.06	0.01	0.01
Unsignalized	**−0.74**	0.14	0.09
Mid_Block_	0.03	0.02	0
TWLTL	−0.25	0.2	0.22
Walk	0.10	**0.69**	−0.02
Bike	0.12	**0.65**	0.17
Employ_Density	0.14	**0.54**	0.26
log_AADT	0.16	**0.48**	0.11
Edu_Less_H	−0.09	0.02	**0.78**
Total_Veh0	0.1	0.16	**0.86**
Low_Wage	0.15	0.27	**0.71**
Poverty	0.14	0.32	**0.90**

[a]Factor loadings in boldface are above (0.4).

(log_AADT) are highly correlated with F2 which can be called exposure. (Edu_Less_H), (Total_Veh0), (Low_Wage), and (Poverty) are highly correlated with F3 which can be called low social status. Therefore, three factors—exposure, social status, and pedestrian and bicycle-oriented roadway—were constructed from the observed variables in the data collection.

Confirmatory Factor Analysis

The CFA model was analyzed using the remaining observations (N = 100). CFA is often used to evaluate a prior theory or hypothesis such as the number of factors, types of factors, whether or not the factors are correlated, and which observed variable are indicators of which factor. Now, given the EFA results, CFA helps cross-validate the structure as well as the factor loadings as EFA is purely data driven. Prior knowledge informs that the presence of ped/bike-friendly facilities, percentage of the working population, AADT, walking/biking, and gross employment density are considered to be related to pedestrian and bicyclist exposure. However, the high score of factor loading in the EFA suggested that high speed limit is also a strong indicator of exposure and thus, the high speed limit was used as an indicator for the latent factor exposure in CFA. By contrast, EFA indicated paved shoulder has a low correlation with pedestrian and bicycle-oriented roadway for any of the three factors but it was kept in CFA because of the prior knowledge.

The EFA result suggests that a CFA is fitted based on three latent variables in the x-measurement model (Figure 3). The fourth latent variable was added following the similar concepts presented in other research work (*32*, *33*). The fourth latent variable in the y-measurement model (Figure 3) is the endogenous latent variable, so-called "Crash Index," which is measured by pedestrian crashes and bicyclist crashes. Figure 4 illustrates the resulting SEM with all latent variables. All variables were significant at the 5% level, and non-significant variables were removed from the final SEM (e.g., Mid_Block and Sidepath).

The overall fit of the model and the significance of some model parameters were evaluated. Both the RMSEA and CFI indices are within the cut-off values of (0.064) and (0.930), respectively. Therefore, the model does fit despite the result of the chi-square test.

Findings and Discussion

The SEM technique enhances safety studies with its ability to build a structure among variables (e.g., pedestrian and bicyclist safety studies at intersections). The ability to include multiple endogenous measures (e.g., pedestrian crashes, bicyclist crashes) is a benefit because it results in a more informative framework. SEM also provides guidance with safety-related data collection, and highlights pertinent variables that can be gathered to represent important latent variables. Several models were tested to identify a statistically significant model.

The final SEM displays standardized parameters for all coefficients. The structural model which can be viewed as a standard regression equation using standardized parameters includes latent exogenous variables

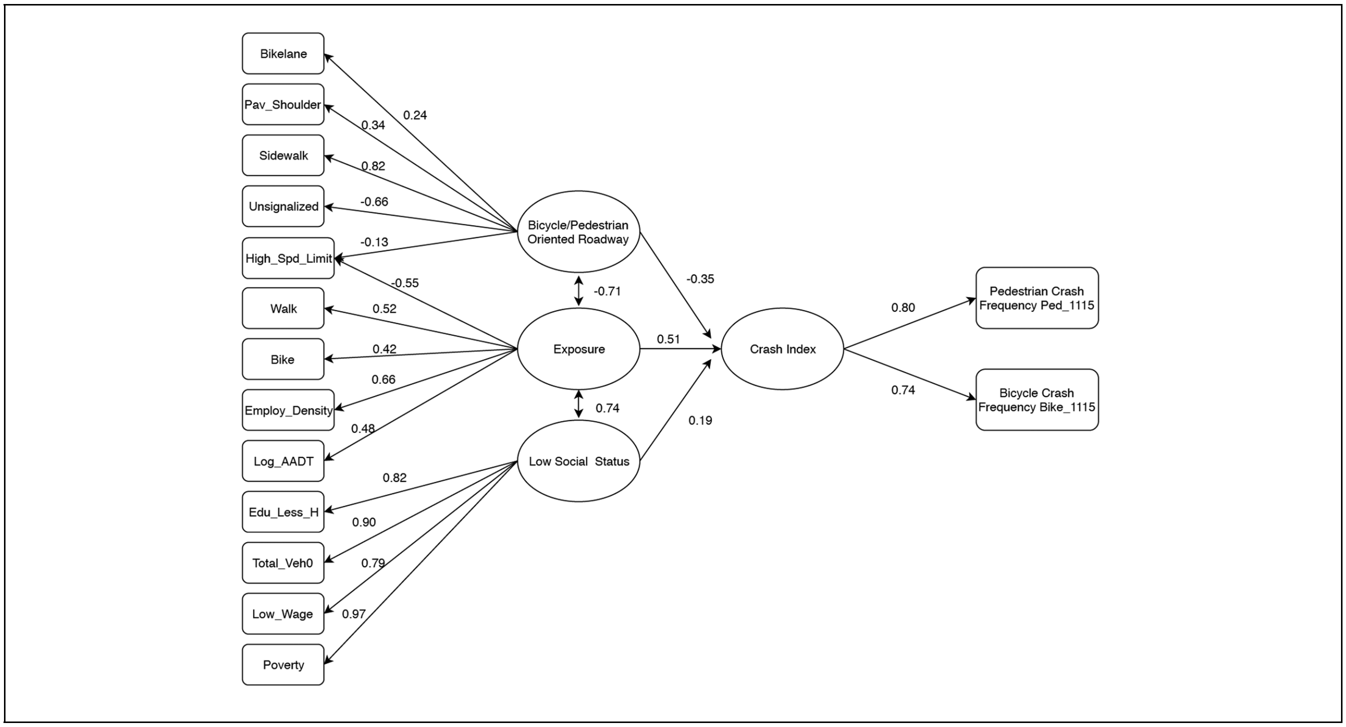

Figure 4. The final structural equation model.

bicycle/pedestrian-oriented roadway, exposure, and low social status, and latent endogenous variable crash index. The regression coefficients show that the crash index is strongly and positively influenced by exposure latent variable (coefficient = 0.51), moderately and negatively affected by bicycle/pedestrian-oriented roadway (coefficient = −0.35), and weakly and positively affected by the low social status (coefficient = 0.19). Regarding the measurement models, the x-measurement model implies that sidewalk coverage along the corridor, bike lane, and paved shoulder coverage are good features of a bicycle/pedestrian-oriented roadway. Bike lanes, paved shoulders, and sidewalks may lead to higher exposure for pedestrians and bicyclists, but they also provide designated space for these VRUs and may decrease the likelihood of crashes. The y-measurement model implies pedestrian and bicycle crash count is strong and positive measures of the crash index.

The low social status latent variable was positively and highly influenced by many variables (e.g., low educational level, and low wage). The results show a positive effect between lower educational level and crash frequency: well-educated residents may have had more driver education training and may be more aware of road safety and the consequences of crashes. People who live in lower-income neighborhoods may travel more by walking and bicycling because of limited resources for automobile travel. Lower rates of car ownership may be positively related to crash frequency through increased pedestrian and bicyclist exposure. It is also possible that areas with higher-income residents may have environments that are more conducive to biking and walking (e.g., more high-quality pedestrian and bicycle infrastructure).

Looking at the exposure latent variable, it can be seen that walking or biking as a transportation mode, as well as employment density and AADT, positively affect pedestrian and bicyclist exposure to traffic, thus leading to more crashes. The high speed variable had dual citizenship, meaning that it was correlated to both the bicycle/pedestrian-oriented roadway variable and the exposure variable in a negative fashion. It is plausible that the locations with a high posted speed limit often indicate fewer pedestrian or bicycle activities, as the roadway design is more vehicle centric. As speed increases, pedestrian and bicycle crashes may be more likely because drivers may not detect pedestrians and bicyclists on the sides of the road, and longer stopping distances are needed to avoid collisions.

These results contain similar conclusions from previous studies. Exceeding the speed limit showed an increase in the probability of being involved in a crash (*32*, *33*). In fact, it was significant in bicycle/pedestrian-oriented roadway (–0.13) and exposure latent variables (–0.55). This underscores the value of SEM, as it is able to clarify complex relationships between variables. A unique conclusion is derived from the correlation between two exogenous latent variables, showing the high positive correlation between low social status and exposure. Owning zero vehicles (shows a low social status) will increase the individual's exposure and therefore increase his/her crash index, leading to more crash involvement. In addition, the presence of paved shoulders tends to decrease the crash index (indirectly) by improving the road design for pedestrians and bicyclists. Paved shoulders provide additional space for pedestrians and bicyclists outside of travel lanes, even the elevation between the shoulder and the roadway, and reduce the presence of gravel or sand that may contribute to bicyclist crashes.

Conclusion

Safety of VRUs is a critical issue affecting the efficiency of our built highways. A reduction in crashes among this particular group of roadway users has always been the top priority of traffic safety researchers and policymakers. Several studies have addressed the issue of pedestrian and bicyclist crash frequency (*44–46*), and many have used standard regression models (e.g., log-linear, multinomial logistic regression, negative binomial). Standard regression models can explain the direct impact of the surrounding factors on pedestrian and bicyclist crash frequencies and injury severities, but they might not explain the underlying complex relationships between variables.

The study applies SEM to develop a conjectured structure that provides a clear portrait between a many highway corridor specific variables and VRU crashes. The structure is featured by the relationship between three exogenous latent variables representing bicyclist and pedestrian-oriented roadway design, exposure, and surrounding social status, and one endogenous variable representing a single value VRU safety quantification for both pedestrian and bicyclist crashes. The relationship between latent and observed variables can also be conveniently established by using the measurement model. Combining both the structural and measurement models in a single modeling process enables the effective distinction between direct, indirect, and synergic effects between variables, and thus more accurately capture the physical underpinning for VRU crashes.

The notable findings from this highway corridor-based VRU study are as follows: the model suggests that bicycle/pedestrian-oriented roadway, exposure, and low social status are strongly related to VRU crash frequency. SEM helps to explain the potentially conflicting information such that an observed variable may affect more than one latent variable in different ways (i.e., High_Spd_Limit), and the results show that high speed limit negatively influences pedestrian and bicyclist exposure to traffic. It is noted that some significant variables in the models were not significant in previous research.

Several limitations existed in this study. The study was limited by not having direct pedestrian and bicyclist volume counts; there may be other land-use variables beyond those considered in this study that contribute to increased pedestrian and bicycle exposure. Future studies should try to use more refined pedestrian and bicyclist exposure data. Pedestrian and bicyclist counts will be helpful to improve the accuracy of latent variable exposure. In addition, crashes were not analyzed if they were not reported to police, or not geocoded in the crash database. The 10-year time period provides more crash data for analysis, but it also increases the chance that a particular corridor had different characteristics when the earliest crashes occurred. The database contained crashes with motor vehicles only, as they appeared to be the most severe, but they have been found to represent only a fraction of total pedestrian and bicycle crashes. Although the sample size of 200 corridors is adequate, more sites are desirable to improve the model fit and significance of the input variables. Further work should also use behavioral data to ensure that these factors are well studied and reduce the potential for omitted variable bias.

Acknowledgments

This work was sponsored by Wisconsin Department of Transportation (WisDOT). The authors thank the students who provided data collection assistance.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: XQ, FJA-M; data collection: FJA-M, MRRS; analysis and interpretation of results: FJA-M, XQ, RJS; draft manuscript preparation: FJA-M, XQ, RJS. All authors reviewed the results and approved the final version of the manuscript.

References

1. *USDOT Releases 2016 Fatal Traffic Crash Data*. NHTSA. https://www.nhtsa.gov/press-releases/usdot-releases-2016-fatal-traffic-crash-data. Accessed July 31, 2018.
2. *FARS Encyclopedia*. https://www-fars.nhtsa.dot.gov/Main/index.aspx. Accessed July 20, 2018.
3. Schneider, R. J. Theory of Routine Mode Choice Decisions: An Operational Framework to Increase Sustainable Transportation. *Transport Policy*, Vol. 25, 2013, pp. 128–137. https://doi.org/10.1016/j.tranpol.2012.10.007.
4. Abdel-Aty, M., S. S. Chundi, and C. Lee. Geo-Spatial and Log-Linear Analysis of Pedestrian and Bicyclist Crashes Involving School-Aged Children. *Journal of Safety Research*, Vol. 38, No. 5, 2007, pp. 571–579.

5. Eluru, N., C. R. Bhat, and D. A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis & Prevention*, Vol. 40, No. 3, 2008, pp. 1033–1054. https://doi.org/10.1016/j.aap.2007.11.010.
6. Sullivan, J. M., and M. J. Flannagan. The Role of Ambient Light Level in Fatal Crashes: Inferences from Daylight Saving Time Transitions. *Accident Analysis & Prevention*, Vol. 34, No. 4, 2002, pp. 487–498.
7. Noland, R., and M. Quddus. Analysis of Pedestrian and Bicycle Casualties with Regional Panel Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2004. 1897: 28–33.
8. Macpherson, A. K., T. M. To, P. C. Parkin, B. Moldofsky, J. G. Wright, M. L. Chipman, and C. Macarthur. Urban/Rural Variation in Children's Bicycle-Related Injuries. *Accident Analysis & Prevention*, Vol. 36, No. 4, 2004, pp. 649–654. https://doi.org/10.1016/S0001-4575(03)00086-1.
9. Kim, J. K., S. Kim, G. F. Ulfarsson, and L. A. Porrello. Bicyclist Injury Severities in Bicycle–Motor Vehicle Accidents. *Accident Analysis & Prevention*, Vol. 39, No. 2, 2007, pp. 238–251. https://doi.org/10.1016/j.aap.2006.07.002.
10. Morency, P., L. Gauvin, C. Plante, M. Fournier, and C. Morency. Neighborhood Social Inequalities in Road Traffic Injuries: The Influence of Traffic Volume and Road Design. *American Journal of Public Health*, Vol. 102, No. 6, 2012, pp. 1112–1119. https://doi.org/10.2105/AJPH.2011.300528.
11. Teschke, K., M. A. Harris, C. C. O. Reynolds, M. Winters, S. Babul, M. Chipman, M. D. Cusimano, J. R. Brubacher, G. Hunte, S. M. Friedman, M. Monro, H. Shen, L. Vernich, and P. A. Cripton. Route Infrastructure and the Risk of Injuries to Bicyclists: A Case-Crossover Study. *American Journal of Public Health*, Vol. 102, No. 12, 2012, pp. 2336–2343. https://doi.org/10.2105/AJPH.2012.300762.
12. Lusk, A. C., P. Morency, L. F. Miranda-Moreno, W. C. Willett, and J. T. Dennerlein. Bicycle Guidelines and Crash Rates on Cycle Tracks in the United States. *American Journal of Public Health*, Vol. 103, No. 7, 2013, pp. 1240–1248. https://doi.org/10.2105/AJPH.2012.301043.
13. Cai, Q., M. Abdel-Aty, and J. Lee. Macro-Level Vulnerable Road Users Crash Analysis: A Bayesian Joint Modeling Approach of Frequency and Proportion. *Accident Analysis & Prevention*, Vol. 107, 2017, pp. 11–19. https://doi.org/10.1016/j.aap.2017.07.020.
14. Chen, P. Built Environment Factors in Explaining the Automobile-Involved Bicycle Crash Frequencies: A Spatial Statistic Approach. *Safety Science*, Vol. 79, 2015, pp. 336–343. https://doi.org/10.1016/j.ssci.2015.06.016.
15. Cooper, J. F., R. J. Schneider, S. Ryan, and S. Co. Documenting Targeted Behaviors Associated with Pedestrian Safety. *Transportation Research Record: Journal of the Transportation Research Board*, 2012. 2299: 1–10.
16. Rosén, E., H. Stigson, and U. Sander. Literature Review of Pedestrian Fatality Risk as a Function of Car Impact Speed. *Accident Analysis & Prevention*, Vol. 43, No. 1, 2011, pp. 25–33. https://doi.org/10.1016/j.aap.2010.04.003.
17. Tefft, B. C. Impact Speed and a Pedestrian's Risk of Severe Injury or Death. *Accident Analysis & Prevention*, Vol. 50, 2013, pp. 871–878. https://doi.org/10.1016/j.aap.2012.07.022.
18. Parkin, J., M. Wardman, and M. Page. Estimation of the Determinants of Bicycle Mode Share for the Journey to Work Using Census Data. *Transportation*, Vol. 35, No. 1, 2007, pp. 93–109. https://doi.org/10.1007/s11116-007-9137-5.
19. Schneider, R. J., M. C. Diogenes, L. S. Arnold, V. Attaset, J. Griswold, and D. R. Ragland. Association between Roadway Intersection Characteristics and Pedestrian Crash Risk in Alameda County, California. *Transportation Research Record: Journal of the Transportation Research Board*, 2010. 2198: 41–51.
20. Nordback, K., W. E. Marshall, and B. N. Janson. Bicyclist Safety Performance Functions for a U.S. City. *Accident Analysis & Prevention*, Vol. 65, 2014, pp. 114–122. https://doi.org/10.1016/j.aap.2013.12.016.
21. Jacobsen, P. L. Safety in Numbers: More Walkers and Bicyclists, Safer Walking and Bicycling. *Injury Prevention*, Vol. 9, No. 3, 2003, pp. 205–209. https://doi.org/10.1136/ip.9.3.205.
22. Schneider, R. J., J. Vargo, and A. Sanatizadeh. Comparison of US Metropolitan Region Pedestrian and Bicyclist Fatality Rates. *Accident Analysis & Prevention*, Vol. 106, 2017, pp. 82–98. https://doi.org/10.1016/j.aap.2017.04.018.
23. Robinson, D. L. Safety in Numbers in Australia: More Walkers and Bicyclists, Safer Walking and Bicycling. *Health Promotion Journal of Australia*, Vol. 16, No. 1, 2005, pp. 47–51. https://doi.org/10.1071/HE05047.
24. Siddiqui, C., M. Abdel-Aty, and K. Choi. Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 382–391. https://doi.org/10.1016/j.aap.2011.08.003.
25. Nashad, T., S. Yasmin, N. Eluru, J. Lee, and M. A. Abdel-Aty. Joint Modeling of Pedestrian and Bicycle Crashes: Copula-Based Approach. *Transportation Research Record: Journal of the Transportation Research Board*, 2016. 2601: 119–127. https://doi.org/10.3141/2601-14.
26. Lee, C., and M. Abdel-Aty. Comprehensive Analysis of Vehicle–Pedestrian Crashes at Intersections in Florida. *Accident Analysis & Prevention*, Vol. 37, No. 4, 2005, pp. 775–786. https://doi.org/10.1016/j.aap.2005.03.019.
27. Loukaitou-Sideris, A., R. Liggett, and H. G. Sung. Death on the Crosswalk: A Study of Pedestrian-Automobile Collisions in Los Angeles. *Journal of Planning Education and Research*, Vol. 26, No. 3, 2007, pp. 338–351. https://doi.org/10.1177/0739456X06297008.
28. Nordback, K., W. E. Marshall, and B. N. Janson. Bicyclist Safety Performance Functions for a U.S. City. *Accident Analysis & Prevention*, Vol. 65, 2014, pp. 114–122. https://doi.org/10.1016/j.aap.2013.12.016.
29. Hilton, J. *Race and Ethnicity in Fatal Motor Vehicle Traffic Crashes 1999 - 2004*. May 2006, National Highway Traffic Safety Administration, U.S. Department of Transportation, DOT HS 809 956.
30. Kim, K., P. Pant, and E. Yamashita. Measuring Influence of Accessibility on Accident Severity with Structural Equation Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2236: 1–10.
31. Lee, J. Y., J. H. Chung, and B. Son. Analysis of Traffic Accident Size for Korean Highway Using Structural Equation

Models. *Accident Analysis & Prevention*, Vol. 40, No. 6, 2008, pp. 1955–1963. https://doi.org/10.1016/j.aap.2008.08.006.
32. Schorr, J. P., and S. H. Hamdar. Safety Propensity Index for Signalized and Unsignalized Intersections: Exploration and Assessment. *Accident Analysis & Prevention*, Vol. 71, 2014, pp. 93–105. https://doi.org/10.1016/j.aap.2014.05.008.
33. Wang, K., and X. Qin. Use of Structural Equation Modeling to Measure Severity of Single-Vehicle Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2014. 2432: 17–25.
34. Cai, Q., J. Lee, N. Eluru, and M. Abdel-Aty. Macro-Level Pedestrian and Bicycle Crash Analysis: Incorporating Spatial Spillover Effects in Dual State Count Models. *Accident Analysis & Prevention*, Vol. 93, 2016, pp. 14–22. https://doi.org/10.1016/j.aap.2016.04.018.
35. Schumacker, R. E., and R. G. Lomax. *A Beginner's Guide to Structural Equation Modeling*. Routledge, New York, 2010.
36. Hamdar, S. H., H. S. Mahmassani, and R. B. Chen. Aggressiveness Propensity Index for Driving Behavior at Signalized Intersections. *Accident Analysis & Prevention*, Vol. 40, No. 1, 2008, pp. 315–326. https://doi.org/10.1016/j.aap.2007.06.013.
37. Bentler, P. M. Fit Indexes, Lagrange Multipliers, Constraint Changes and Incomplete Data in Structural Models. *Multivariate Behavioral Research*, Vol. 25, No. 2, 1990, pp. 163–172. https://doi.org/10.1207/s15327906mbr2502_3.
38. Browne, M. W., and R. Cudeck. Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, Vol. 21, No. 2, 1992, pp. 230–258. https://doi.org/10.1177/0049124192021002005.
39. Tabachnick, B. G., and L. S. Fidell. *Using Multivariate Statistics*, 5th ed. Allyn & Bacon, Inc., Needham Heights, Mass., 2006.
40. Hair, J. F. *Multivariate Data Analysis*. Pearson Education India, 2006.
41. Siddiqui, C., M. Abdel-Aty, and K. Choi. Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 382–391. https://doi.org/10.1016/j.aap.2011.08.003.
42. Pett, M., N. Lackey, and J. Sullivan. *Making Sense of Factor Analysis*. SAGE Publications, Inc., Calif., 2003.
43. Costello, A. B., and J. W. Osborne. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis. *Exploratory Factor Analysis*, Vol. 10, No. 7, p. 9.
44. Wier, M., J. Weintraub, E. H. Humphreys, E. Seto, and R. Bhatia. An Area-Level Model of Vehicle-Pedestrian Injury Collisions with Implications for Land Use and Transportation Planning. *Accident Analysis & Prevention*, Vol. 41, No. 1, 2009, pp. 137–145. https://doi.org/10.1016/j.aap.2008.10.001.
45. Elvik, R. The Non-Linearity of Risk and the Promotion of Environmentally Sustainable Transport. *Accident Analysis & Prevention*, Vol. 41, No. 4, 2009, pp. 849–855. https://doi.org/10.1016/j.aap.2009.04.009.
46. Vilaça, M., E. Macedo, P. Tafidis, and M. C. Coelho. Frequency and Severity of Crashes Involving Vulnerable Road Users - An Integrated Spatial and Temporal Analysis. Presented at 97th Annual Meeting of the Transportation Research Board, Washington, D.C., 2018.

The Standing Committee on Safety Data, Analysis and Evaluation (ANB20) peer-reviewed this paper (19-03034).