Structural Equation Modelling: An Application to Pedestrian Safety in Washington DC and Exploration of the Impact of Variable Scaling Procedures

Lama Al-Hajj Hassan
lama.a.al.haj.hassan@gmail.com
Faculty of Engineering and Architecture
Department of Civil and Environmental Engineering
American University of Beirut
P.P. Box – 11-0236/Department of Civil and Environmental Engineering
Riad El-Solh/Beirut 1107 2020, Lebanon

Justin P. Schorr*
justin11@gwmail.gwu.edu
School of Engineering and Applied Science
Department of Civil and Environmental Engineering
Traffic and Networks Research Laboratory
The George Washington University
Exploration Hall, Room 201-I
20101 Academic Way
Ashburn, Virginia 20147, USA

Samer H. Hamdar
hamdar@gwu.edu
School of Engineering and Applied Science
Department of Civil and Environmental Engineering
Traffic and Networks Research Laboratory
The George Washington University
Science and Engineering Hall, #3810
800 22nd Street, NW
Washington, DC 20052, USA

Stephen Arhin
saarhin@howard.edu
College of Engineering, Architecture and Computer Science
Department of Civil and Environmental Engineering
Transportation Research Center
Howard University
2300 6th Street NW, Suite 2121
Washington, DC 20059, USA

Word Count:

5422 + 1750 (7 Tables/Figures) = 7172

August 1st, 2016

Submitted for presentation at the 96th Annual Meeting of the Transportation Research Board, and publication in Transportation Research Record

Structural Equation Modelling: An Application to Pedestrian Safety in Washington DC and Exploration of the Impact of Variable Scaling Procedures

Al Hajj Hassan, Schorr*, Hamdar and Arhin

ABSTRACT

In a commercially vibrant city like Washington D.C., pedestrian–vehicle collisions remain a constant concern. Among the factors impacting pedestrian safety are environmental, traffic, and roadway geometrical factors. The objective of this paper is not only to identify these factors and their impact on safety using data from the MS2-Howard crash data base. A structural equation modeling approach is applied to establish the relationship between the descriptive exogenous variables and the severity related endogenous variables. This paper also explores the changes in model estimation results based on different variable scaling procedures. Results indicate that uncontrolled roadway segments and pedestrians crossing at non-intersections create the most hazardous situations for pedestrians. Furthermore, differences in coefficient valued based on various scaling procedures create changes in coefficients throughout the entire model. Thus indicating that in order for the structural approach to be applied in a thorough and proper manner, various structures and scales should be tested and results should be interpreted contextually using a variety of different measures.

Keywords: Pedestrians, Safety, Scaling, Structural Equation Modeling (SEM), Washington DC

1.0 INTRODUCTION

A main concern in any metropolitan city is the constant interaction between the many different users of the transportation network. Increased volumes for all types for traffic (pedestrian, motor vehicles, non-motor vehicles, etc.) creates an increase in exposure to hazardous events – and the safety of all users requires examination. In United States (US) cities, high pedestrian volumes are of specific concern as 4,735 fatal and 66,000 non-fatal pedestrian collisions were documented in 2013 (*1*). In a 2015 report the Analysis Group Road Safety Annual Report revealed (2000 – 2013 data) that the 54% reduction in car occupant deaths was not successfully met for vulnerable road users – as although the drop was 36% for pedestrians, 35% for cyclists, and 22% for motorcyclists, these percentages have leveled off since 2010 (*2*).

In today's metropolitan cities, transportation infrastructure is comprised of many different of roadways featuring varying geometric and traffic flow characteristics. Recently, The Federal Highway Administration (FHWA) has been stressing the incorporation of pedestrian safety measures into existing infrastructure. Cities take measures in order to ensure infrastructure is safe for pedestrians, some examples are separating sidewalks from roads, reducing speed limits, making sidewalks accessible for the disabled, adding curb extensions and refuge islands, utilizing pedestrian signals and traffic calming measures (*3*). In order to ensure the safety of pedestrians (and all users of the transportation network) the transportation research community is constantly exploring new data driven approaches. As computing capabilities and data collection methods continue to evolve, datasets containing thousands of microscopic entries can be quickly and accurately analyzed in order to provide insights on the macroscopic level of a city or geographic region.

Research in this study focuses on pedestrian safety in Washington, D.C., where the number of pedestrian fatalities ranks first in the U.S at a rate triples the national average of 14.5%(4). The dataset analyzed in is specific to the region and is gathered from the Traffic Crash Location System (TCLS) in Midwestern Software Solutions (MS2). Analysis is conducted using a Structural Equation Modeling (SEM) approach (*5*), an approach which begins with the identification of pertinent variables which are grouped into dimensions, and ends with the quantification of safety into a single valued "index" in order to assess safety on a macroscopic level.

1.1 Objectives

The application of the SEM approach to transportation safety is a recent development, and to this point no study has examined the impact of variable scales and coding procedures on model results. Thus, in addition to gaining insights into some of the factors impacting pedestrian safety in and urban area (Washington, D.C), one of the main objective of this study is to understand the impact that different coding and scaling procedures have on structural model results.

Specific objectives of this study are organized as follows:

- *Identify variables impacting pedestrian safety.* Variables considered include vehicle and pedestrian characteristics, environmental, infrastructure and traffic flow characteristics, and collision severity metrics such as the number of injuries and fatalities.
- *Group variables into relevant dimensions (environmental, traffic related, infrastructure, etc.) through a factor analysis and pattern examination.* Here, the data set is filtered to exclude missing and incomplete data. Different factor structures are postulated and tested using the SAS Software (Statistical Analysis Software). Tests are conducted using different scaling procedures for exogenous variables.

- *Apply the SEM approach to establish a relationship with the endogenous, severity variables.* To establish these relationships different structural models are tested using the LISREL Software. Multiple model structures are tested, again using different scaling procedures for exogenous variables.
- *Analyze model results to gain insights into factors that impact pedestrian safety.* Results for statistically significant models are analyzed and explained within the context of the modelling technique and scaling procedure. Furthermore, a comparative analysis is conducted – analyzing and identifying the impact that scaling procedures have on model estimation.

The rest of this study is organized as follows. In Section 2 (Background and Motivation) a short overview of previous studies on pedestrian-vehicle collisions is provided, as well as a brief discussion of the application of the SEM approach to transportation safety analysis. This is followed by a section describing the dataset (along with the major limitations) and the methodologies and modelling techniques used in this study. Model results are presented in Section 4, which also features an in depth analysis of the structural model results as well as a comparative analysis on the impact and scaling procedures. Finally, conclusions and suggestions for future works are provided in Section 5.

2.0 BACKGROUND AND MOTIVATION

Pedestrian movement within a city presents an additional safety consideration (as opposed to a highway network) for transportation researchers. Pedestrian-vehicle collisions cause both social and psychological trauma – especially considering the average age of pedestrian fatalities is 46 while the average age of those who sustain injuries is 36 *(1)*. For this reason and many more, extensive research is conducted on the topic – continuing to branch out to other yet-to-be-explained territories. From the International Traffic Safety Data and Analysis Group's (ITSD's) (2) questioning of reasoning behind the faster drop in fatality rate than injury rate in pedestrian crashes to Griswold's (6) recommendation to further study the correlation between pedestrian walking/crossing activities at certain time periods with crash rates. ITSD (2) recommends expanding crash related data sources beyond police reports as they are usually inadequate to carry out a complete analysis of the collision consequences. Therefore this paper relied on MS2 database that consists of police and standard crash detailing.

While previous studies have examined the impact of environmental, traffic and geometric characteristics on pedestrian safety *(7,8,9,10,11)*, the SEM approach used in this study allows for straightforward safety comparisons through the use of a singular index value. While the application of this model to transportation safety has been explored by other authors (5,12), this study contributes to the research literature in two main ways:

1. The application of the SEM approach to pedestrian safety in an urban setting.
2. The evaluation and validation of the SEM approach itself by comparing models estimated using different scaling procedures.

The cause and effect nature of the SEM approach is well suited for analyzing data where relationships between variables are initially unknown, and must be postulated by the modeler based on previously existing theories. These hypothesis are then either confirmed or rejected based on the validity of the estimated model (13). The popularity of the SEM approach (not limited to transportation) stems partially from its innate ability to capture the effect of a large number of exogenous variables, all of which can be included and grouped for analysis (5). However, to this point no transportation related study has examined the impact that the scales of the exogenous variables themselves have on model results – specifically the statistical significance of the model and the paths within and the sign and magnitude of the dimensional coefficients and individual endogenous and exogenous variables. Thus, in addition to examining pedestrian safety in an urban setting on a macroscopic level, this study is highly motivated by the desire to explore some of the nuances in application of the SEM approach for analyzing transportation safety.

3.0 DATA DESCRIPTION AND METHODOLODGY

3.1 Database Description and Limitations

Data used in this research is taken from MS2, a cloud-based Transportation Data Management System that is maintained by the District Department of Transportation (DDOT) Howard University Traffic Data Center. In this data set, crash standard and police reports can be filtered from the Traffic Crash Location System (TCLS) field and AADT and traffic parameters can be extracted from Traffic Count Database System (TCDS) field. This database allows for the monitoring of all types of collision events across the country, and organizes the data on the state and city levels for location specific analysis and comparisons. The web-based system facilitates data aggregation and any agency with the proper authorization can import geographically coded collision records into the data base, which later undergo quality checks. Unlike Fatality Analysis Reporting System (FARS), data sets like MS2 and National Automotive Sampling System-General Estimates System (NASS-GES) data are not limited to fatal injuries suffered in motor vehicle traffic collisions. In addition, MS2 includes collisions even if they didn't involve motor vehicles, a criterion required for a collision record to be added to the FARS database. The MS2 database also includes collision data that is not reported to police as the company itself collects, manages, and analyzes all traffic related information; and in the case a police report is available, it is included in the respective records. This is an advantage compared to NASS-GES, which restricts its data to police reported collisions despite the fact that half of motor vehicle collisions in U.S. fail to be reported (*1*).

Although the MS2 database is comprehensive and provides unique information not found in other transportation related data sets, it still suffers from a number of limitations. The major issues that arise with a database of this sort are continuity and completeness. While these are issues that often arise in the other National databases, they are enhanced in the MS2 set when considering the various sources that are being aggregated. Obviously including non-police reported incidents in the set is a benefit in terms of comprehensiveness, but these incidents often feature missing information for a number of variables. Furthermore, even with police reports the format and procedure for collecting and reporting collision related information varies between States and Cities within. With that said, the comprehensiveness of the dataset makes it an asset to transportation researchers. MS2 is highly regarded throughout the transportation industry and has many users, including several government and private transportation groups (*14*). Using the MS2 database one can build a custom search that groups specific variables into one report form. If the user knows the location they want to analyze, they are able to draw contours around the area in question using different tools on the GIS map, and the data is restricted to these locations. Furthermore, collision trend analysis is facilitated by the software through generation of crash diagrams and maps.

For this study, collision data from 2008 – 2013 was exported and then filtered to include only pedestrian related collisions. Collisions with missing data in any of the parameter fields were omitted. Additional data on person and vehicle specific characteristics of the collisions was used to enrich the dataset through a merging procedure done based on the Crash Identification Number. A total of 5123 detailed vehicle-pedestrian collisions were aggregated, and Table 1 provides a list of the number of records per year. Only records with complete data were considered for analysis, and this dataset featured 2250 records.

Table 1: Collision Records by Year

Year	Total Records
2008	749
2009	1019
2010	836
2011	874
2012	805
2013	840
Total	5123

3.2 Methodology

As previously discussed, SEM is used to estimate the directional paths and associated coefficients, in this case for pedestrian-vehicle collisions extracted from the DDOT-Howard University Traffic Data Center database. Before this estimation can be conducted, a factor analysis must first be conducted to examine potential dimensional groupings of variables. This factor analysis is conducted using the SAS Software, and multiple factor structures are tested before a combination of statistical results and modeler discretion is used to select the final factor structure (*5*). Once the variables and dimensional groupings have been identified, various structural models are tested using the LISREL software. The model is refined based on the factor analysis and logical inference until a statistically significant structure is achieved, such that results represent a significant relation between the considered parameters. Various goodness of fit statistics (discussed later) are used to assess model validity *(5)*.

One of the main challenges associated with SEM is that the chances of achieving good fit decreases with higher degrees of freedom; so typically tests of multiple variable combinations and model structures are required to achieve a statistically significant model. Furthermore, SEM typically performs best when variables all feature similar ranges (for example survey results that are coded using a Likert Scale) (*5*) – a characteristic that is not typically associated with collision related data. With this in mind, different scaling procedures are used such that a comparative assessment can be conducted and the impact of these different scales can be better understood (all variable codes are provided in Table 2). Table 2 includes only the variables used in the final analysis: in order to conserve the sample size, demographic characteristics and vehicle specifications were omitted, and although a number of additional variables were used for model testing – the following variables (Table 2) constitute those that comprised the final structure.

Table 2: Variable Scales and Descriptions

Intersection	Not at Intersection; Private Property, Other : 0
	At Intersection, Within 100ft of Intersection: 1
Grade / Curve	Curve , Crest, Grade, Ramp: 1
	Bridge, Level, Straight, Underpass: 0
Surface Obstruction	Snow, Ice, Repairing, Sand, Slush, Water, Wet, Other: 1
	Dry: 0
Lighting	Day or Street Lights On:1
	Defective, None, Street Lights off or Night: 0
Precipitation	Blowing Sand, Fog/Mist, Rain, Sleet/Hail, Snow:1
	Clear, Severe Crosswind: 0
Traffic Density 1	Light: 1
	Medium: 2
	Heavy: 3
Traffic Density 2	Light: 1
	Medium: 2.5
	Heavy: 10
Divided	Divided Positive:1
	Not Divided, Divided Unprotected: 0
Traffic Control 1	None: 0
	Turn Restricted, Yield: 1
	Stop Sign, Flashing: 2
	Signal, Officer: 3
Traffic Control 2	None: 0
	Turn Restricted, Yield: 1
	Stop Sign, Flashing: 2.5
	Signal, Officer: 5
Speed Limit	Speed Limit / 10

Note that multiple scales were tested for Lighting but only dummy coding produced converging models. From the table, an illustrative example of different potential scales can be seen for the variable Traffic Density. In the MS2 dataset there are three options, Light, Medium or Heavy. The ambiguity associated with this coding method raises a concern regarding the manner in which this variable should be represented. One method would be to use an ordinal scale, assigning a value of 1 to Light, 2 to Medium and 3 to Heavy. However, when taking the approximate average of the AADT on the roadways in each category, roadways classified as Medium have approximately 2.5 times the AADT as those classified as light, and those classified as Heavy have an AADT that is approximately 10 times higher. As such, the alternative coding structure for Traffic Density is 1 for Light, 2.5 for Medium and 10 for Heavy.

3.2.1 Basic Formulation

The formulation of the structural model follows that of Hamdar et al (*15*) (please refer to for full formulation in 15 if needed). The measurement model for latent variables is expressed as follows:

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

Where: x variables are exogenous indicators of the ξ the latent variables; y variables are endogenous indicators of the η the latent variables; and ε , δ represent the error terms.

The structural model is then expressed as:

$$\eta = \alpha + B\eta + \Gamma\xi + \varsigma \quad (2)$$

Where: η is a vector of endogenous latent variables; ξ is a vector of exogenous latent variables; and ζ is a vector of errors terms in equations with mean zero.

4.0 RESULTS AND ANALYSIS

4.1 Preliminary Analysis – Factor Structure

Using the SAS Software, factor analysis was performed with all exogenous variables available in the data set. After testing multiple factor structures with the different variable codes, it became readily apparent that the magnitude of the factor score is the important metric, not the sign. Furthermore, regardless of the scale, the dimensional groupings suggested by the factor analysis remained constant. Table 3 shows the factor structure used for analysis.

Table 3: Final Factor Pattern

Factor Structure (Correlations)			
	Factor1	Factor2	Factor3
Precipitation	**0.97484**	0.07348	-0.01732
Surface Condition	**0.97330**	0.07155	-0.03164
Traffic Control 2	0.10070	**0.82186**	0.17374
Near Intersection	0.11786	0.75989	**0.05609**
Traffic Density 2	-0.02282	**0.42600**	0.04265
Lighting	**-0.10487**	0.21698	-0.18029
Curve/Grade	0.03544	-0.11839	**0.10796**
Divided	-0.06057	0.11454	**0.72946**
Speed Limit	-0.01185	**0.04886**	0.68824

The final dimensions included nine total variables, constituting three dimensions of three variables each. In addition to considering only the magnitude of the factor score, practical consideration based on the physical meaning of each variable was given in order to establish the following final dimensions: *L1: Roadway Geometric Characteristics* (Divided, Intersection Collision, Curve/Grade), *L2: Traffic Related Characteristics* (Speed Limit, Level of Control, Traffic Density), and *L3: Environmental Conditions* (Lighting, Precipitation, Surface Condition).

4.2 Structural Model

After testing several structures using the aforementioned dimensions a statistically significant converging model was achieved using LISREL software and is presented in Figure 1.

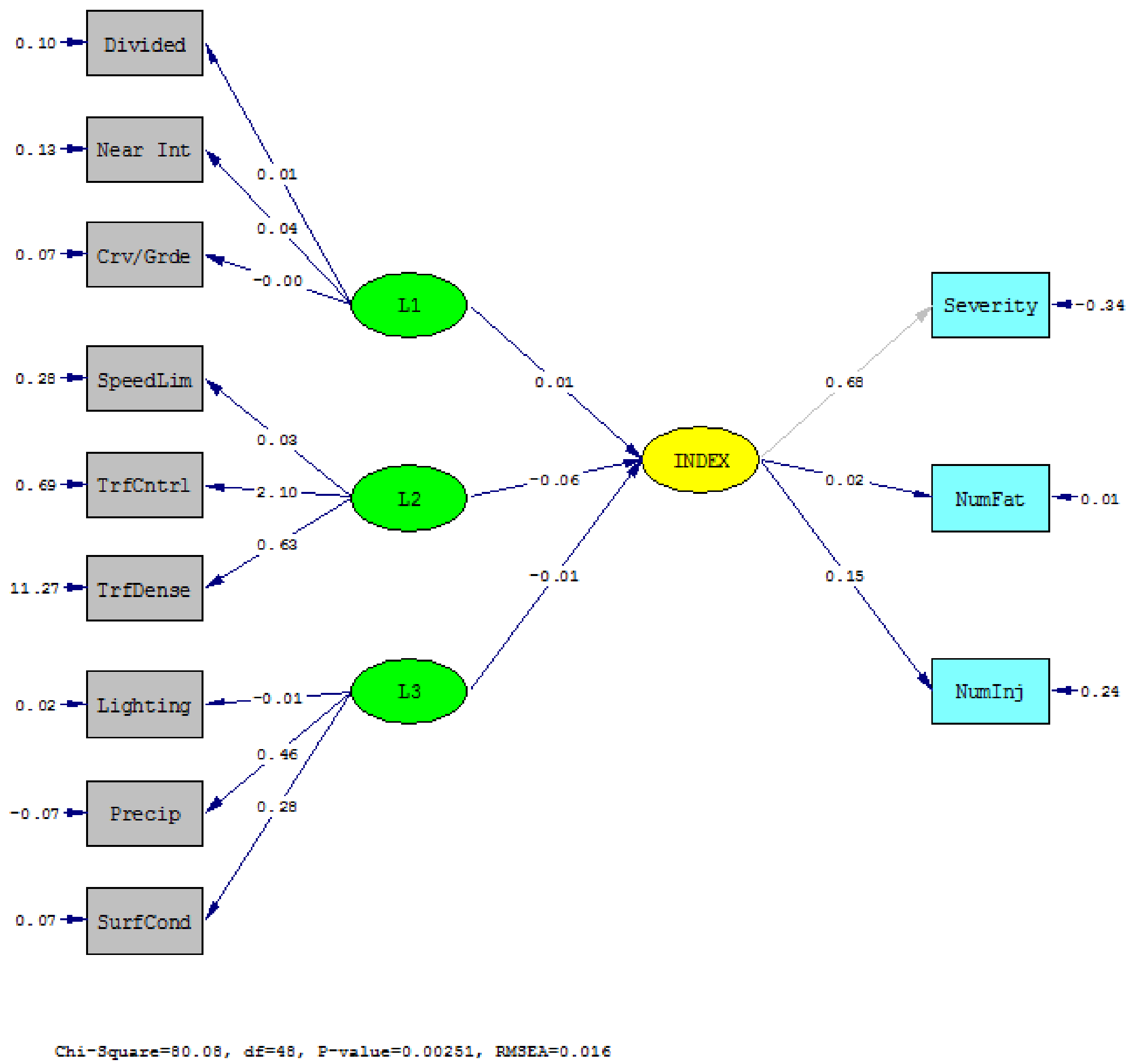

Figure 1: Structural Model

The measurement equations and relevant statistics describing the model are provided in Tables 4A and 4B.

Table 4A: Measurement Equations

Structural Model				
Equation			*Error Variance*	*R^2*
Index=0.01*L1-0.06*L2-0.01*L3			0.95	-0.0034
Endogenous Measurement Model				
Variable	*Equation*		*Error Variance*	*R^2*
Severity	Severity = 0.68*Index		0.34	3.65
Num. Fatalities	Number of Fatalities = 0.02*Index		0.01	0.027
Num. Injuries	Number of Injuries = 0.15*Index		0.24	0.093
Exogenous Measurement Model				
Dimension	*Variable*	*Equation*	*Error Variance*	*R^2*
L1	Near Intersection	Near Intersection = 0.04*L1	0.13	0.050
	Curve / Grade	Curve / Grade = -0.0093*L1	0.07	0.0013
	Divided	Divided = 0.014*L1	0.095	0.0019
L2	Traffic Density	Traffic Density = 0.63*L2	11.27	0.033
	Traffic Control	Traffic Control = 2.13*L2	0.69	0.89
	Speed Limit	Speed Limit = 0.026*L2	0.28	0.0024
L3	Lighting	Lighting = -0.0086*L3	0.021	0.0034
	Precipitation	Precipitation = 0.46*L3	-0.069	1.51
	Surface Condition	Surface Condition = 0.28*L3	0.068	0.55

Table 4B: Model Statistics

T Values		Error Covariance Terms	
Variables	*Value*		
L1/Divided	0.36	*Variables*	*Value*
L1/Near Intersection	0.37	Precipitation/Surface Condition	-2.31
L1/CurveGrade	-0.36	Divided/Traffic Control	-0.04
L2/Speed Limit	2.31	Divided/Speed Limit	0.01
L2/Traffic Control	10.09	Divided/Traffic Density	0.03
L2/Traffic Density	6.89	Speed Limit/Traffic Density	-0.03
L3/Lighting	-3.19	Speed Limit/Traffic Control	-0.02
L3/Precipitation	10.66	Traffic Control/Traffic Density	0.76
L3/Surface Condition	10.39	**Cronbach's Coefficient Alpha**	
L1/Index	0.32		
L2/Index	-2.1	*Dimension*	*Value*
L3/Index	-1.57	L1	0.3296
Index/Num. Fatalities	2.52	L2	0.2077
Index/Num. Injuries	2.68	L3	0.6745

The proposed model is representative of a relatively large sample size of 2550 records therefore Chi-Squared tests often encounter problems (*13*) Therefore, although the p-value associated with the Chi-Squared test (p=0.00251) indicated an excellent statistical fit, statistical significance is evaluated using other metrics as well. As suggested by Golob (*13*) Root Mean Square Error of Approximation (RMSEA) with a 90% confidence interval was examined, as well as the Goodness of Fit Index (GFI), the Adjusted Goodness of Fit Index (AGFI) and the Standardized Root Mean Square Residual (SRMR). The RMSEA has a value of 0.016 and a 90% confidence interval of 0.0096 to 0.022 indicating that the whole interval lies well below the threshold of 0.05 and thus indicating and excellent fit (*13, 16*). This statistical significance is further confirmed by the GFI (0.99) and the

AGFI (0.99), as well as the SRMR (0.022) which is well below the threshold of 0.08 indicating a good fit (*16*). For an alpha value of 0.05, t-values outside the -1.96:1.96 interval are considered significant, and indicate that we can be more confident in the associated path within the structural model. From Table 5B, the Roadway Geometric Characteristics dimension (L1) has t-values that are inside the range described above. This is likely due to the manner in which these variables are coded (dummy values of 1 and 0) – and the remaining dimensions have statistically significant t-values, constituting a statistically significant model on the whole.

4.3 Analysis of Results

From the estimated structural model (Figure 1), ***increases*** in the SPI are indicative of a ***decrease*** in safety (as it indicates an increase in collision severity, the number of injuries and the number of fatalities – all endogenous variables). The signs of the dimensional coefficients further indicate that *increases* in the variables contained in dimension L1 (Roadway Geometric Characteristics) are indicative of a *decrease* in safety, while *increases* in the variables of dimensions L2 (Traffic Related Characteristics) and L3 (Environmental Conditions) are indicative of an *increase* in safety. The main contributor to the index is the Traffic dimension, with an absolute coefficient value of 0.06 which is six times that of Roadway Geometric features and Environmental Conditions factors that both share an equal absolute coefficient value of 0.01. In order to properly analyze model results, additional metrics must be used to make further insight into the impact each variable has on the index. For this reason, the average, standard deviation and coefficient of variation are provided for each variable in the model. Using the average and the standard deviation in combination with the variable coefficient value and the dimensional coefficient, the average contribution each variable makes to the index can be computed, Additionally, the impact on the index from a one standard deviation change in each variable is computed. By understanding the degree to which each coefficient varies (the coefficient of variation) within the context of the contribution that a variation in that variable has on the index (Deviation SPI Contribution) allows for model results to be analyzed in a complete and comprehensive fashion.

The analytical statistics discussed above for all variable in the structural model are presented in Table 5.

Table 5: Analytical Statistics

Variable	Coeff.	Dim.	Mean Value	Standard Deviation	Coefficient Of Variation	Average SPI Contribution	Deviation SPI Contribution
Near Intersection	0.04	0.01	0.8396	0.3670	0.4372	0.2099	0.0918
Curve / Grade	-0.0093	0.01	0.0714	0.2575	3.6080	-0.0767	-0.2769
Divided	0.014	0.01	0.1071	0.3093	2.8886	0.0765	0.2209
Traffic Density	0.63	-0.06	3.5812	3.4157	0.9538	-0.3411	-0.3253
Traffic Control	2.13	-0.06	3.0508	2.2611	0.7412	-0.0859	-0.0637
Speed Limit	0.026	-0.06	2.3496	0.5330	0.2268	-5.4222	-1.2299
Lighting	0.0086	-0.01	0.9780	0.1466	0.1499	-1.1373	-0.1704
Precipitation	0.46	-0.01	0.1616	0.3681	2.2785	-0.0035	-0.0080
Surface Condition	0.28	-0.01	0.1710	0.3766	2.2024	-0.0061	-0.0134

Looking first at the Traffic dimension (L2), positive valued coefficients throughout the dimension demonstrate that increases in all variables within the dimension are indicative of an increase in safety. While the increase in safety associated with an increase in traffic control measures is intuitive and a finding of many studies (*9, 17*), this is not the case for the other variables within this dimension. From the coefficient values of Speed Limit and Traffic Density, model results indicate that higher speed limits and more dense traffic conditions are indicative of an increase in safety. While this result may not be intuitive, several studies (*18, 19*) resulted in similar conclusions relating them to the safety in numbers concept as for the higher speed limit paradox, Zegeer (*20*) explains that pedestrians crossing higher speed roads would cross more carefully and would avoid short-gaps. Moreover, the data set shows that 17.5% of collisions occurred in heavy traffic conditions as compared to 31.3% in light traffic. Furthermore, 98.5% of collisions in the data set occurred on roads with speed limits less than 30 mph – which is most certainly due to the urban area under consideration. The counterintuitive nature of the results in this dimension demonstrate that model results must be interpreted carefully and contextually, as the structural equation model speaks to the macroscopic impact of the variables considered. While higher speeds may often be associated with severe collisions (*21*), data in this study demonstrates that other impactful variables such as traffic control must be considered to fully understand the challenges at hand. Thus, from the analytical statistics in Table 6, the variables for Traffic Control and Traffic Density have relatively average values (as compared to the other variables in the model) for their coefficients of variation as well as their deviation contributions. Speed Limit, however, has the highest deviation contribution in the model – but the second lowest coefficient of variation, indicating that speed limits across Washington, DC are relatively uniform – at least in the areas where pedestrian-vehicle collisions are occurring.

Moving to the Environmental Conditions dimension (L3), positive valued coefficients for precipitation and surface conditions indicate that increases in these variables are indicative of an increase in pedestrian safety, while lighting has a negative coefficient value, indicating the opposite. In other words, situations where there are adverse weather conditions (heavy rain, wet roads) or reduced roadway lighting indicate an increase in pedestrian safety. While potentially counterintuitive, this result can be better interpreted by realizing that there is lighter pedestrian traffic in rainy weather and during the night, reducing the chances of any type of pedestrian-vehicle interaction. Mohamed et al (*17*) adds that in bad weather conditions drivers are inclined to drive more carefully. Similarly, Nance (*22*) found that the risk for child pedestrians increases when the streets were dry with no adverse weather conditions and during day time. While the variables for surface condition and precipitation have high coefficients of variation, their deviation contributions are the lowest in the model, indicating that the impact of these variables on macroscopic pedestrian safety is minimal. While the deviation contribution from lighting is slightly higher, the associated coefficient of variation is the lowest of all variables in the model.

Finally, in the Roadway Geometric Characteristics dimension (L1), positive coefficients for the variables Divided and Near Intersection indicate a decrease in safety on divided roadways (those with a median) and at locations closer to intersections, while a negative valued coefficient for Curve / Grade indicated that there is an increase in safety as this variable increases. Results for the variables Divided and Near Intersection are as expected as more complicated geometric features often associated with intersections and divided roadways increase pedestrian exposure to interactions with vehicles as argued by multiple studies (*7, 21)*. Looking further at the results associated with proximity to intersections, collisions involving vehicles making left or right turns constituted 40% of all collisions in the dataset. One potential explanation for the increase in pedestrian safety on roadways with curves or grades is that drivers may feel an increased level of comfort on straight roadways which may be indicative of higher speeds or less driver attention; consequently increasing pedestrian vulnerability. Furthermore, roadway design in urban setting is conducted with pedestrians in mind – and the exposure of pedestrians on these intuitively more dangerous curved or sloped roadways may be restricted. Variables in this dimension have coefficients of variation that are among the highest in the model, as well as relatively high average and deviation contributions – indicating that changes in these variables have a definite impact on pedestrian safety.

4.4 Analysis of Variable Scales

To examine the impact of variable scales, three alternative model structures were tested based on the most statistically significant model achieved (Figure 1). The scales for the variables Traffic Density and Traffic Control were varied based on Table 2, and the percentage change in coefficients and t-values are presented in Table 6 along with the p-value and RMSEA for the estimation. In the model presented in Figure 1, Traffic Density 2 and Traffic Control 2 are used. It should be noted that a converging model using Traffic Control 1 and Traffic Density 1 could not be achieved – though intermediate solutions suggested more iterations could produce significant results.

Table 6: Percentage Change in Coefficients for Varying Scales

Variable	Control 1 / Density 2		Control 2 / Density 1	
	Coefficient	T-Value	Coefficient	T-Value
Traffic Control	-41.23%	-1.20%	2.37%	19.76%
Speed Limit	-33.33%	-9.87%	-33.33%	0.00%
Traffic Density	-6.35%	-5.60%	-73.02%	20.37%
L1	16.67%	1.94%	0.00%	-1.94%
Near Intersection	80.00%	0.00%	0.00%	-11.11%
Curve/Grade	0.00%	0.00%	0.00%	0.00%
Divided	0.00%	0.00%	0.00%	0.00%
L2	50.00%	2.70%	0.00%	0.00%
Lighting	0.00%	5.09%	0.00%	0.00%
Precipitation	-6.67%	23.24%	0.00%	0.00%
Surface Condition	7.14%	22.48%	0.00%	-0.10%
L3	0.00%	-4.35%	0.00%	1.24%
Severity	0.00%	-----	0.00%	-----
Injuries	6.67%	5.95%	0.00%	6.35%
Fatalities	0.00%	-6.34%	0.00%	-5.97%
p-value	0.0061		0.006	
RMSEA	0.018		0.02	

Here, it can be seen that changing scale for the variable Traffic Control impacts not only the variables within its own dimension – but the variable in the other dimensions as well. This is an interesting result as the alternative scaling for Control was chosen by the authors – and produced the most statistically significant structural models. The scale for Density, conversely, was derived from the actual AADT of the roadways. This result indicated that not only must the modeler use caution when selecting a scale – but they should also test alternative variable scales when utilizing the SEM approach. Importantly, the sign of each exogenous variable remained constant throughout the alternative structures – further supporting the macroscopic analysis conducted in the previous subsection.

5.0 CONCLUSION

This paper examined impactful variables on pedestrian safety. Pedestrian-vehicle collision risk factors in the Washington D.C area were postulated and tested using the structural equation modeling approach. The analysis was based on dataset taken from MS2 software managed by Howard University. After filtering the set 2550 pedestrian-vehicle related records were used and these are distributed over six years from 2008 – 2013. The structure's components and dimensions were identified after running a factor analysis using SAS software and the LISREL software was used to estimate the structural models. Various model structures were tested and results produced a statistically significant model with an extremely good fit. Traffic control measures and intersection proximity were found to be among the most influential variables on pedestrian safety. Additionally, the application of the SEM approach itself was evaluated by testing various variable scales. Results of these alternative model structures demonstrated that changing the scale of a single variable can have a definite impact in terms of magnitude (but not sign) on the other variables and dimensions in the model. Care must be taken when applying the SEM approach to transportation safety, and future studies should explore how other changes in the modeling process impact estimations and analysis. SEM offers a powerful macroscopic analysis tool that quantifies safety into a single index variable – and more exploration of the approach and its application to transportation safety will allow for new insights to be made.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1351647. 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. The authors would also like to thank the District Department of Transportation (DDOT) Howard University Traffic Data Center for providing the data.

6.0 REFERENCES

(1) NASS General Estimates System. National Highway Traffic Safety Administration (NHTSA). http://www.nhtsa.gov/Data/National+Automotive+Sampling+System+(NASS)/NASS+General+Estimates+System. Accessed July 9, 2015.

(2) International Traffic Safety Data and Analysis Group. International Traffic Safety Data and Analysis Group Road Safety Annual Report 2015. Http://internationaltransportforum.org/Pub/pdf/15IRTAD_Summary.pdf. 2015. Accessed June 26, 2015.

(3) Smart Growth America. Dangerous by Design 2014. May 2014. http://www.smartgrowthamerica.org/documents/dangerous-by-design-2014/dangerous-by-design-2014.pdf. Accessed July 22, 2015.

(4) NHTSA. Traffic Safety Facts. February, 2015. http://www-nrd.nhtsa.dot.gov/Pubs/812124.pdf. Accessed June 25,2015.

(5)Schorr, J., S. Hamdar. Safety propensity index for signalized and unsignalized 10 intersections: Exploration and assessment. Accident Analysis and Prevention. October, 2014. http://www.sciencedirect.com/science/article/pii/S0001457514001407?np=y. Accessed July 24, 2015.

(6) Griswold, Julia, B. Fishbain, S. Washington, and D. R. Ragland. Visual assessment of pedestrian crashes. Accident Analysis & Prevention. January, 2011. http://www.sciencedirect.com/science/article/pii/S0001457510002617. Accessed June 26, 2015.

(7) Pulugurtha, Srinivas S., and V. R. Sambhara. Pedestrian crash estimation models for signalized intersections. Accident Analysis & Prevention. January, 2011. http://www.sciencedirect.com/science/article/pii/S0001457510002770. Accessed June 24, 2015.

(8)Cottrill, Caitlin D., and P.V. Thakuriah. Evaluating pedestrian crashes in areas with high low-income or minority populations. Accident Analysis & Prevention. November, 2010. http://www.ncbi.nlm.nih.gov/pubmed/20728622.Accessed June 25, 2015.

(9) Rothman, Linda, A.W. Howard, A. Camden, and C. Macarthur. Pedestrian crossing location influences injury severity in urban areas. Injury Prevention. June, 2012. http://injuryprevention.bmj.com/content/early/2012/06/21/injuryprev-2011-040246.Accessed June 25, 2015.

(10) Scheiner, Joachim, and C. Holz-Rau. A residential location approach to traffic safety: Two case studies from Germany. Accident Analysis & Prevention. January, 2011. http://www.ncbi.nlm.nih.gov/pubmed/21094329. Accessed June 25, 2015.

(11) Rothman, Linda, C. Macarthur, T. To, R. Buliung, and A. Howard. Motor vehicle-pedestrian collisions and walking to school: the role of the built environment. Pediatrics. April, 2014. http://pediatrics.aappublications.org/content/early/2014/04/02/peds.2013-2317.Accessed June 26, 2015.

(12) Hamdar, Samer H., J. Schorr. Interrupted versus uninterrupted flow: A safety propensity index for driver behavior. Accident Analysis & Prevention. June, 2013. http://www.sciencedirect.com/science/article/pii/S0001457513000420. Accessed June 25, 2014.

(13) Golob, T. F., 2001. *Structural Equation Modelling for Travel Behavior Research*. January, 2003. http://www.sciencedirect.com/science/article/pii/S0191261501000467. Accessed June 27, 2015.

(14) MS2 Modern Traffic Analytics. *Client Map and List*. Midwestern Software Solutions. http://www.ms2soft.com/clients/. Accessed July 31, 2015.

(15) Hamdar, S.H., H.S. Mahmassani, R.B. Chen. *Aggressiveness propensity index for 8 driving*. Accident Analysis and Prevention. January, 2008. http://www.sciencedirect.com/science/article/pii/S0001457507001108. Accessed June 25, 2015.

(16) Hu, L. & P. M. Bentler. *Fit Indices in Covariance Structure Modeling: Sensitivity to 852 Underparameterized Model Misspecification*. Psychological Methods.1998. http://people.cehd.tamu.edu/~okwok/epsy651R/Articles/HuBentler1998.pdf. Accessed June 25, 2015.

(17) Mohamed Gomaa, Mohamed, N. Saunier, L. F. Miranda-Moreno, and S.V. Ukkusuri. *A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada*. Safety science. April, 2013. http://www.sciencedirect.com/science/article/pii/S0925753512002664. Accessed July 20, 2015.

(18) Miranda-Moreno, F. Luis, P. Morency, and A. M. El-Geneidy. *The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections.* Accident Analysis & Prevention. September, 2011. http://www.ncbi.nlm.nih.gov/pubmed/21658488.Accessed June 24, 2015.

(19) Ewing, R., R.A. Schieber, C.V. Zegeer. *Urban sprawl as a risk factor in motor vehicle occupant and pedestrian fatalities*. American Journal of Public Health. September, 2003. http://www.minority.unc.edu/sph/minconf/2004/materials/ewing.et.al.pdf. Accessed June 28, 2015.

(20) Zegeer, Charles, J. Stewart, H. Huang, and P. Lagerwey. Safety effects of marked versus unmarked crosswalks at uncontrolled locations: analysis of pedestrian crashes in 30 cities. In *Transportation Research Record*: *Journal of the Transportation Research Board No.1773,* Transportation Research Board of the National Academies, Washington, D.C, 2001, pp. 56-68.

(21) Zegeer, V. Charles, and M. Bushell. Pedestrian crash trends and potential countermeasures from around the world. Accident Analysis & Prevention. January 2012. http://www.sciencedirect.com/science/article/pii/S0001457510003842. Accessed June 23, 2015.

(22) Nance, Michael L., L. A. Hawkins, C. C. Branas, C. V.-O'Neill, and F. K. Winston. *Optimal driving conditions are the most common injury conditions for child pedestrians.* Pediatric emergency care. September, 2004. http://www.ncbi.nlm.nih.gov/pubmed/15599256. Accessed June 23, 2015.

(21) Harwood, D.W., D.J. Torbic, D.K. Gilmore, C.D. Bokenkoger, J.M. Dunn, C.V. Zegeer, S. Srinivasan, D. Carter, C. Raborn, C. Lyon, B. Persaud . Pedestrian Safety Prediction Methodology. National Cooperative Highway Research Program (NCHRP). March, 2008. http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_w129p3.pdf. Accessed June 22, 2015.