Accident Analysis and Prevention 71 (2014) 93–105

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Safety propensity index for signalized and unsignalized intersections: Exploration and assessment

Justin P. Schorr*, Samer H. Hamdar[1]

Department of Civil and Environmental Engineering, Center of Intelligent Systems Research, Traffic and Networks Research Laboratory, The George Washington University, 20101 Academic Way #201-I, Ashburn, VA 20147, USA

ARTICLE INFO

Article history:
Received 22 December 2013
Received in revised form 9 May 2014
Accepted 13 May 2014
Available online 2 June 2014

Keywords:
Index
Intersections
Safety
Signals
Structural equations
Traffic

ABSTRACT

The objective of this study is to develop a safety propensity index (SPI) for both signalized and unsignalized intersections. Through the use of a structural equation modelling (SEM) approach safety is quantified in terms of multiple endogenous variables and related to various dimensions of exogenous variables. The singular valued SPI allows for identification of relationships between variables and lends itself well to a comparative analysis between models. The data provided by the Highway Safety Information System (HSIS) for the California transportation network was utilized for analysis. In total 22,422 collisions at unsignalized intersections and 20,215 collisions at signalized intersections (occurring between 2006 and 2010) were considered in the final models. The main benefits of the approach and the subsequent development of an SPI are (1) the identification of pertinent variables that effect safety at both intersection types, (2) the identification of similarities and differences at both types of intersections through model comparison, and (3) the quantification of safety in the form of an index such that a ranking system can be developed. If further developed, the adopted methodology may assist in safety related decision making and policy analysis.

1. Introduction

Creating a safer, more efficient transportation network is a main goal of transportation researchers across the world leading to countless studies that analyze different safety aspects as they pertain to transportation. In these studies the quantification of safety (a difficult and ambitious undertaking) is approached in a variety of manners and has a vast array of applications. One such application is the analysis of safety at roadway intersections. Intersection related collisions are of particular concern to transportation researchers as they accounted for 47% of all vehicles involved in collisions and 28% of those involved in fatal collisions on United States roadways in 2010 (NHTSA, 2012). Signalized intersections accounted for 25% of total vehicles in collisions and 8% of vehicles involved in fatal collisions while unsignalized intersections accounted for 22% of total vehicles involved and 20% of vehicles involved in fatal collisions (NHTSA, 2012).

Research in this study is aimed at quantifying and analyzing safety at both signalized and unsignalized intersections by exploring a comprehensive modelling framework that quantifies safety in terms of multiple endogenous variables and considers the combined effects of various exogenous variables on safety as well as on one another. One such framework that allows for the inclusion of these parameters is that of structural equation modelling (SEM). The main benefit of this approach is the quantification of safety in terms of a safety propensity index (SPI). The SPI is a singular value that relates all exogenous variables to all endogenous variables and allows for the development of a ranking system in order to understand and evaluate safety at intersections. Additionally, the estimation of structural models for both signalized and unsignalized conditions creates a powerful comparative framework from which additional analysis can be conducted. This methodology has been utilized by Hamdar et al. (2008) to examine aggressiveness at signalized intersections as well as by Hamdar and Schorr (2013) in order to compare and contrast safety in differing flow scenarios (interrupted and uninterrupted). Within the SEM framework, research in this study will explore further applications of the modelling technique – and a refined analysis approach is proposed to better understand and compare the factors affecting safety in both driving scenarios.

* Corresponding author. +1 2158501886.
E-mail addresses: justin11@gwu.edu (J.P. Schorr), hamdar@gwu.edu (S.H. Hamdar).
[1] Tel.: +1 202 994 6652; fax: +1 202 994 0127.

http://dx.doi.org/10.1016/j.aap.2014.05.008
0001-4575/

Specific research goals to be accomplished throughout this study are as follows: (1) to systematically identify the factors that affect safety propensity at both signalized and un-signalized intersections in a given area; (2) to utilize existing public data repositories (Highway Safety Information System (HSIS)) to study the safety implications of changes in network geometry as an evolving system (in time and space); (3) to validate the formulated structural equation model against alternative model structures estimated using the existing signalized and un-signalized intersection's incident data; and (4) to analyze the validated model to compare the results obtained given the intersection type. The findings may help in understanding how better transportation system performance can be achieved and strategies can be proposed to improve traffic safety and operations.

2. Conceptual framework and background

A great amount of effort has gone into assessing the different factors that contribute to collisions at both signalized and unsignalized intersections. Previous studies have utilized a number of different modelling techniques for both analytical and predictive purposes. Throughout the literature safety is commonly characterized through a single endogenous metric such as collision rate (Vogt and Bared, 1998; Chin and Quddus, 2003; Wang and Abdel-Aty, 2006; Isebrands et al., 2010; Caliendo and Guida, 2012; Wu et al., 2013) or injury/collision severity (Abdel-Aty and Abdelwahab, 2004; Xie et al., 2009; Quddus et al., 2010; Jung et al., 2012) and expansions to this characterization often focus around the categorical grouping of collisions by collision type (Wang et al., 2003; Kim et al., 2006; Bham et al., 2012). The "safety framework" is further defined through the identification and selection of exogenous variables – which are typically related to one of the following categories: environmental conditions, geometric design, traffic characteristics, driver demographics or vehicle characteristics. Many studies examine a combination of two or three variable types such as geometric design/traffic characteristics (Li et al., 1994; Poch and Mannering, 1996; Karlaftis and Golias, 2001; Wang et al., 2003; Wu et al., 2013), driver demographics/environmental conditions (Bham et al., 2012; Jung et al., 2012), or geometric design/traffic characteristics/environmental conditions (Kim et al., 2006; Quddus et al., 2010). While these studies examine the combined effects of these variables on safety, the modelling framework utilized does not allow for examination of the effects individual variables have on one another or for analysis to be conducted on a dimensional level (geometric, environmental, etc.) as opposed to individual variables – all of which are captured through the SEM approach.

An additional area of interest not explored in the studies mentioned above is the comparison and evaluation of safety at both signalized and unsignalized intersections. Comparison of how certain factors (such as lane width, average annual daily traffic (AADT), median width, design speed and the level of actuation/control) have varying effects on safety at both intersection types can provide additional insights for future research and design.

This research separates itself from the aforementioned studies by imploring the analytical SEM approach which lends itself to a greater number and combination of exogenous measures. Furthermore, through the use of SEM formulation endogenous variables are grouped into different dimensions and their complex interactions are formulated. This confirmatory (as opposed to exploratory) approach requires the modeller to postulate the links between variables based on hypotheses and previous empirical results (Golob and Meurs, 1986; Golob, 2001). Once these links have been postulated, the SEM approach establishes causal directional relations between variables and then the model is either accepted or rejected based on its validity (Golob, 2001). When a model is accepted the propensity for safety is captured though a latent scale and index which are related to the observable variables through the SEM formulation. The structure of the model and the SPI itself allows for three major contributions: identification of variables that have a significant effect on safety at both signalized and unsignalized intersections; presenting the manner in which these variables and their effects vary with intersection type; and assessment of the relative importance of different determinants (Hamdar and Schorr, 2013).

In contrast to explicitly simulating drivers' behaviours (Hamdar et al., 2008; Paleti et al., 2010), this paper will feature an empirical data-driven approach (Hamdar and Schorr, 2013) through the HSIS data system. Although the SEM approach has been used to examine safety under different interrupted and uninterrupted flow conditions (Hamdar and Schorr, 2013) as well as at international unsignalized intersections alone (Lee et al., 2008), a model that compares different types of intersections will demonstrate the benefits of the approach itself and provide a new perspective on intersection safety. Comparing results with previous studies will lead to additional insights into potential effects of certain variables – though it is important to keep in mind that the research is fundamentally different and conducted in totally different geographic locations.

The formulation of the structural model given the data availability and the corresponding limitations can be found in Section 3. Section 4 provides the results of the factor analysis and the numerical simulation as well as the model statistics. These numerical results are discussed and the obtained models (signalized versus unsignalized) are compared in Section 5. The concluding remarks and the future research needs are provided in Section 6.

3. Statistical model

3.1. Available data and limitations

Data available through the HSIS for the California transportation network was utilized in this study. In order to build a comprehensive and inclusive data set, all collisions (regardless of collision type or severity) occurring at all intersections (regardless of the level of control) between 2006 and 2010 were considered for analysis. HSIS data was provided on a yearly basis in three separate data sets; the Accident File (containing the variables for lighting, precipitation, collision severity, total injuries, total fatalities and the number of vehicles), the Intersection File (containing the variables for AADT, channelization, intersection legs, control/actuation level and the number of lanes), and the Roadway File (containing the variables for lane width, shoulder width, surface width, median width, divided/undivided and design speed). Also contained in the Accident File was a variable stating where the collision had occurred in relation to an intersection (regardless of the type or level of control). Only collisions occurring in intersections or within 250 feet of an intersection were considered for analysis. Data sets were merged on a year by year basis first by merging the intersection and roadway files through variables for route name and milepost and then by merging this composite file with the accident file through the use of the same variables. For cases where the collision occurred within 250 feet of the intersection, the collision was assigned to the intersection where the absolute value of the difference between the collision milepost and the intersection milepost was at a minimum. Merged data sets for each year were then combined to yield the final set for analysis.

Once the master data set was compiled collisions were classified as having occurred at either an unsignalized or signalized intersection using the traffic control variable from the intersection

Table 1
Total collisions by year.

Year	Unsignalized	Signalized	Total	Total Vehicles	Total Injuries	Total Fatalities
2006	5495	5262	10,757	21,767	7255	103
2007	5161	4870	10,031	20,183	6791	101
2008	4694	4186	8880	17,765	6297	97
2009	4148	3813	7961	16,006	5543	96
2010	3999	3581	7580	15,203	5349	66
Total	23,497	21,712	45,209	90,924	31,235	463
Missing data	1075	1497	2572	–	–	–
Collisions for analysis	22,422	20,215	42,637	–	–	–

file. Each data set was vetted to ensure that no observation where one or more variables were absent or miscoded was considered for analysis. In total there were 22,422 collisions at unsignalized intersections and 20,215 collisions at signalized intersections where all pertinent data was provided (instances where one or more variable was absent are referred to as "missing data" in Table 1). A breakdown of the number of collisions by year at both types of intersections is provided in Table 1 along with the total number of vehicles, injuries and fatalities associated with the collisions considered for analysis from each year.

The following Sections 3.2 and 3.3 present the formulation of the SEM approach as previously explored by Hamdar et al. (2008) and Hamdar and Schorr (2013).

3.2. Measurement models

Measurements models are specified in two sets of equations. The first set (the exogenous measurement model) is represented as follows:

$$\mathbf{X} = \mathbf{\Lambda X}(\boldsymbol{\gamma}) + \boldsymbol{\omega} \quad (1)$$

where $\mathbf{X}$ = vector of exogenous variables; $\mathbf{\Lambda X}$ = matrix of structural coefficients for relating latent exogenous variables to their observed indicator variables; $\boldsymbol{\gamma}$ = vector of latent exogenous constructs; γ_1 = safety propensity associated with "main street characteristics dimension"; γ_2 = safety propensity associated with "intersection characteristics dimension"; γ_3 = safety propensity associated with "environmental dimension"; γ_4 = safety propensity associated with "traffic dimension"; $\boldsymbol{\omega}$ = vector of measurement error terms for observed variables.

Exogenous variables are described in Table 2A, including how they are measured, and the associated variable name by which they will be designated in the next section.

The second set (endogenous measurement model) of equations is summarized in Eq. (2):

$$\mathbf{Y} = \mathbf{\Lambda Y}(\boldsymbol{\eta}) + \boldsymbol{\tau} \quad (2)$$

where $\mathbf{Y}$ = vector of observed endogenous variables; $\mathbf{\Lambda Y}$ = matrix of structural coefficients for latent endogenous variables to their observed indicator variables; $\boldsymbol{\eta}$ = vector of latent endogenous variable; $\boldsymbol{\tau}$ = vector of measurement error terms for observed endogenous variables. Endogenous variables are further described in Table 2B.

The endogenous variables presented in Table 2B represent the manner in which safety is quantified within the context of this study based on the available data. Collisions which lead to a greater number of injuries or fatalities as well as those that have a higher police reported severity level or greater number of vehicles are inherently indicative of a less safe driving environment. The use of multiple endogenous measures creates a more comprehensive safety framework and is one of the benefits of the SEM approach. Drawbacks to this quantification focus around exposure, occupancy, and land use issues and are discussed in the results section.

Table 2A
Exogenous variable description.

Exogenous variable	Description
Proposed main street characteristics dimension variables	
X_1 (ML lane width)	Width of the main street lane (ft)
X_2 (right shoulder width)	Width of main street right shoulder (ft)
X_3 (left shoulder width)	Width of main street left shoulder (ft)
X_4 (surface width)	Width of entire main street surface (ft)
X_5 (median width)	Width of the main street median (ft)
X_6 (divided)	Dummy variable corresponding to a divided or undivided main street Undivided $X_6 = 0$ Divided $X_6 = 1$
Proposed intersection characteristics dimension variables	
X_7 (XS lanes)	Number of lanes on the cross street
X_8 (ML lanes)	Number of lanes on the main street
X_9 (number of intersection legs)	Number of Intersecting legs
X_{10} (ML channelization)	Dummy variable corresponding to the presence of either a left or right turn channelization on the main street No channelization, $X_{10} = 0$ Channelization, $X_{10} = 1$
X_{11} (XS channelization)	Dummy variable corresponding to the presence of either a left or right turn channelization on the cross street No channelization, $X_{11} = 0$ Channelization, $X_{11} = 1$
Proposed environmental characteristics dimension variables	
X_{12} (light)	Dummy variable corresponding to the lighting at time of collision Natural light or function street lamp, $X_{12} = 0$ No lighting, $X_{12} = 1$
X_{13} (precipitation)	Dummy variable corresponding to the precipitation at time of collision No precipitation, $X_{13} = 0$ Precipitation, $X_{13} = 1$
Proposed traffic related characteristics dimension variables	
X_{14} (design speed)	Main street design speed (mph)
X_{15} (XS AADT)	Average annual daily traffic on cross street. (Thousands)
X_{16} (ML AADT)	Average annual daily traffic on main street. (Thousands)
X_{17} (level of control or actuation)	For un-signalized intersections – level of stop control/impedance No control, $X_{17} = 0$ Yield sign, $X_{17} = 1$ Two way stop, $X_{17} = 2$ Four way stop, $X_{17} = 3$ Flashers, $X_{17} = 4$ For signalized intersections – level of actuation Pre-timed, $X_{17} = 1$ Semi-actuated, $X_{17} = 2$ Fully actuated, $X_{17} = 3$

3.3. Structural model

A structural model relating the endogenous latent variables $\boldsymbol{\eta}$ to the latent exogenous variables $\boldsymbol{\gamma}$ can be expressed as:

$$\eta = \mathbf{\Delta}\boldsymbol{\gamma} + \boldsymbol{\xi} \quad (3)$$

Table 2B
Endogenous variable descriptions.

Endogenous variable	Description
Y_1 (total injuries)	Number of injuries per collision
Y_2 (total fatalities)	Number of fatalities per collision
Y_3 (number of vehicles)	Number of vehicles per collision
Y_4 (severity)	Police reported severity level (recoded)
	Severity level 0: $Y_4 = 0$
	Severity level 2: $Y_4 = 1$
	Severity level 3: $Y_4 = 2$
	Severity level 4: $Y_4 = 3$
	Severity level 1 (corresponding to a fatality): $Y_4 = 4$

where $\boldsymbol{\eta}$ = vector of latent endogenous variables; $\boldsymbol{\Delta}$ = matrix of structural coefficients for relating latent exogenous variables to latent endogenous variables; $\boldsymbol{\gamma}$ = vector of latent exogenous constructs ($\gamma_1, \ldots, \gamma_4$ are as previously defined); $\boldsymbol{\xi}$ = vector of measurement error terms for latent endogenous variables.

In the context of this study the objective is to estimate a singular latent endogenous variable, the safety propensity index. As is such the vector term $\boldsymbol{\eta}$ which represents all latent endogenous variables can be defined by a singular value, η_1 – the safety propensity index. Eq. (3) is now presented in matrix form:

$$[\eta_1] = [\,\delta_{11} \quad \delta_{12} \quad \delta_{13} \quad \delta_{14}\,] * \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} + [\xi_1] \tag{4}$$

Note that in Eq. (4) $\boldsymbol{\eta}$ is now represented by the singular value η_1 and the matrix $\boldsymbol{\Delta}$ takes the form $[\,\delta_{11} \quad \delta_{12} \quad \delta_{13} \quad \delta_{14}\,]$. Similarly the measurement equations, Eqs. (1) and (2), are expressed in matrix form in Eqs. (5) and (6) respectively:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \\ X_9 \\ X_{10} \\ X_{11} \\ X_{12} \\ X_{13} \\ X_{14} \\ X_{15} \\ X_{16} \\ X_{17} \end{bmatrix} = \begin{bmatrix} \Omega_{1,1} & 0 & 0 & 0 \\ \Omega_{2,1} & 0 & 0 & 0 \\ \Omega_{3,1} & 0 & 0 & 0 \\ \Omega_{4,1} & 0 & 0 & 0 \\ \Omega_{5,1} & 0 & 0 & 0 \\ \Omega_{6,1} & 0 & 0 & 0 \\ 0 & \Omega_{7,2} & 0 & 0 \\ 0 & \Omega_{8,2} & 0 & 0 \\ 0 & \Omega_{9,2} & 0 & 0 \\ 0 & \Omega_{10,2} & 0 & 0 \\ 0 & \Omega_{11,2} & 0 & 0 \\ 0 & 0 & \Omega_{12,3} & 0 \\ 0 & 0 & \Omega_{13,3} & 0 \\ 0 & 0 & 0 & \Omega_{14,4} \\ 0 & 0 & 0 & \Omega_{15,4} \\ 0 & 0 & 0 & \Omega_{16,4} \\ 0 & 0 & 0 & \Omega_{17,4} \end{bmatrix} * \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} + \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \\ \omega_6 \\ \omega_7 \\ \omega_8 \\ \omega_9 \\ \omega_{10} \\ \omega_{11} \\ \omega_{12} \\ \omega_{13} \\ \omega_{14} \\ \omega_{15} \\ \omega_{16} \\ \omega_{17} \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \lambda_{12} \\ \lambda_{13} \\ \lambda_{14} \end{bmatrix} * [\eta_1] + \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{bmatrix} \tag{6}$$

In addition to the three structural matrices $\boldsymbol{\Lambda X}$ (the matrix of Ω values in Eq. (5)), $\boldsymbol{\Lambda Y}$ (the matrix of λ values in Eq. (6)), and $\boldsymbol{\Delta}$; the following four variance/covariance matrices need to be specified to determine a general structural equation model:

(1) a Variance/covariance (VC)-matrix of latent exogenous variables (Φ)
(2) a VC-matrix of error terms associated with model implied structural equations (Ψ)
(3) a VC-matrix of measurement errors or observed exogenous variables ($\theta\omega$)
(4) a VC-matrix of measurement error terms associated with the observed endogenous variables ($\theta\tau$)

Now that the model has been formulated and pertinent variables have been identified, the model can be applied to the data for the state of California.

4. Application to the data set

Factor analysis was performed using 17 exogenous variables using the Statistical Analysis System (SAS) Software. This type of analysis establishes relationships based on a mathematical function $\boldsymbol{q}(\mathbf{W},\mathbf{Z})$ connecting a variable $\mathbf{X}$ with the set of variables $\mathbf{W}$ and $\mathbf{Z}$ (Hamdar et al., 2008). The measurable values of $\mathbf{Y}$ are known; however the type of function $\boldsymbol{q}(.)$ that should be used and the variables to be included in this function are unknown. Accordingly, we assume that a set of $\mathbf{Y}$ variables are related to a number of functions that operate linearly:

$$X_n = \alpha_{n1}F_1 + \alpha_{n2}F_2 + \cdots + \alpha_{nm}F_m \tag{7}$$

where X is a variable with known data, α is a constant that represents the loading, and F_j is a function $q_j(.)$ of some unknown variables where $j = 1, \ldots, m$; F_j is also referred to as a factor.

The output derived from this analysis is useful in the following manner:

1. *Un-rotated matrix*: deals solely with uncorrelated patterns. Each pattern could potentially involve all (or nearly all) the variables, and therefore may lead to high loadings for several factor patterns.
2. *Pre-rotated matrix*: deals solely with correlated patterns.
3. *Rotated factor matrix*: here the factor matrix covers both correlated and uncorrelated patterns. From this particular patterns can be hypothesized and patterns are easier to uncover and will not include most of the variables.

The following table (Table 3) represents the rotated factor analysis for exogenous variables for unsignalized intersections. It is important to note that the number of factors was reduced from four to three based on consistently low factor scores for environmental characteristics.

Factors scores approximately on the order of 0.1 were considered for analysis (Hamdar et al., 2008) and refined potential dimensional placements are suggested by the factor scores displayed in bold in Table 3. Initial framework was revised based on the findings of the factor analysis and consideration of the physical relevance of the variables to the dimensions they are being associated with: the Environmental Characteristics dimension was

Table 3
Rotated factor pattern – unsignalized intersections.

Rotated pattern	Factor 1	Factor 2	Factor 3
Proposed main street characteristics dimension			
Divided	**0.89409**	−0.32155	−0.00480
Median width	**0.50038**	−0.03742	**0.15781**
L shoulder width	−0.36566	**0.17946**	**0.64001**
R shoulder width	**0.30016**	−0.13260	**0.71369**
Surface width	**0.14481**	**0.90701**	−0.02360
Lane width	−0.04123	**0.54832**	−0.00981
Proposed intersection characteristics dimension			
Intersection legs	0.00945	−0.01524	**0.24839**
ML lanes	**0.89016**	**0.35709**	−0.00026
XST lanes	**0.07189**	0.01202	0.06498
ML channel	**0.41456**	−0.18931	**0.15056**
XST channel	**0.07567**	−0.04361	**0.11611**
Proposed environmental characteristics dimension			
Precipitation	−0.03875	−0.01690	−0.04603
Light	−0.13139	−0.04732	−0.00650
Proposed traffic related characteristics dimension			
Stop control	−0.03068	−0.02644	**0.18669**
ML AADT	**0.67156**	**0.13409**	0.02911
XST AADT	**0.08227**	0.03385	**0.24277**
Design speed	0.03531	−0.01615	**0.23714**

dropped based on the consistently low or negative factor scores, the variable surface width was dropped because it is redundant (shoulder width + lane width), and the variables for shoulder width were dropped since they could not be considered in a relevant dimension. Main line channelization and median width were considered with the other main line characteristics based on the physical relevance of the variables. The resulting final dimensions are L1 – intersections characteristics (design speed, intersection legs, stop control), L2 – main line characteristics (ML AADT, ML lanes, ML channel, median width), and L3 – cross street characteristics (XST AADT, XST Lanes, XST channel).

Several structures were then tested based on these new dimensions ultimately leading to the statistically significant converging model (computed using the LISREL software) displayed in Fig. 1.

The results summarizing the model are presented in Tables 4A and 4B.

As suggested by Golob (2001), for models with large sample sizes (such as this, $N = 22{,}422$) Chi-squared tests often encounter problems. For this reason, the goodness of fit was assessed based on the root mean square error of approximation (RMSEA) (Golob, 2001). For the model above, the RMSEA was 0.052 and the 90%

Table 4A
Unsignalized model measurement equations.

Equation	Errorvar	R^2 value
Structural model		
INDEX = 0.018 × L1 + 0.052 × L2 − 0.023 × L3	1.00	0.0026
Endogenous measurement model		
Severity = 1.54 × INDEX	−0.60	1.33
Number of vehicles = 0.049 × INDEX	0.21	0.011
Total injuries = 0.61 × INDEX	0.83	0.31
Total fatalities = 0.027 × INDEX	0.017	0.043
Exogenous measurement model		
Design speed = 1.26 × L1	121.01	0.013
Stop control = 0.33 × L1	0.15	0.41
Intersection legs = 0.16 × L1	0.24	0.095
ML AADT = 10.23 × L2	99.56	0.51
ML lanes = 1.08 × L2	0.30	0.80
ML channelization = 0.14 × L2	0.23	0.074
Median width = 7.73 × L2	276.63	0.18
XST AADT = 1.42 × L3	3.56	0.36
XST lanes = 0.086 × L3	0.066	0.10
XST channelization = 0.10 × L3	0.060	0.15

Table 4B
Unsignalized model statistics.

t-Values

Variables	Value
L1/design speed	12.15
L1/stop control	38.18
L1/intersection legs	30.26
L2/ML AADT	95.90
L2/ML lanes	113.89
L2/ML channelization	37.78
L2/Median width	59.31
L3/XST AADT	48.81
L3/XST lanes	32.90
L3/XST channelization	38.81
L1/INDEX	1.33
L2/INDEX	7.57
L3/INDEX	−1.71
INDEX/severity	0.00
INDEX/number of vehicles	14.59
INDEX/total injuries	25.34
INDEX/total fatalities	19.80

Error covariance terms

Variables	Value
Design speed/ML AADT	5.60
Design speed/median width	1.23
Design speed/XST AADT	−0.13
Design speed/intersection legs	0.15
Stop control/XST AADT	0.02
Stop control/ML AADT	0.33
Intersection legs/stop control	−0.02
ML AADT/ML channelization	1.15
XST AADT/XST channelization	0.07

Cronbach's alpha

Set	Value
Entire data set	0.509
Dimension L1	0.269
Dimension L2	0.710
Dimension L3	0.408

confidence interval was 0.051; 0.053. The entire confidence interval is around the threshold of 0.05 indicating that the model is statistically significant and has a good fit (Golob, 2001; Hu and Bentler, 1998). For further support, the standardized root mean square residual (SRMR) has a value of 0.040; values less than 0.08 are generally considered to indicate a good fit (Hu and Bentler, 1998). Additional fit statistics such as Goodness of Fit Index (GFI = 0.97) and Adjusted Goodness of Fit Index (AGFI = 0.96) indicated that the model was statistically significant as well. For an alpha of 0.05, *t*-values greater than 1.96 or less than −1.96 are considered significant. *t*-Values indicate that we can be more confident in some paths than others, and nearly all *t*-values in the model were significant.

The following table (Table 5) represents the rotated factor analysis for exogenous variables for signalized intersections. It is important to note that once again the number of factors was reduced from four to three based on consistently low factor scores for environmental characteristics.

Once again, bold values in Table 5 are indicative of refined potential dimensional placements as suggested by the factor scores. Keeping with the same procedure outlined above (utilizing factor scores as well as relevance to sort the dimensions), the final dimensions for signalized intersections are as follows: L1 – intersections characteristics (design speed, intersection legs, level of actuation), L2 – main line characteristics (ML AADT, ML lanes, ML channel, Median Width), and L3 – cross street characteristics (XST AADT, XST lanes, XST channel).

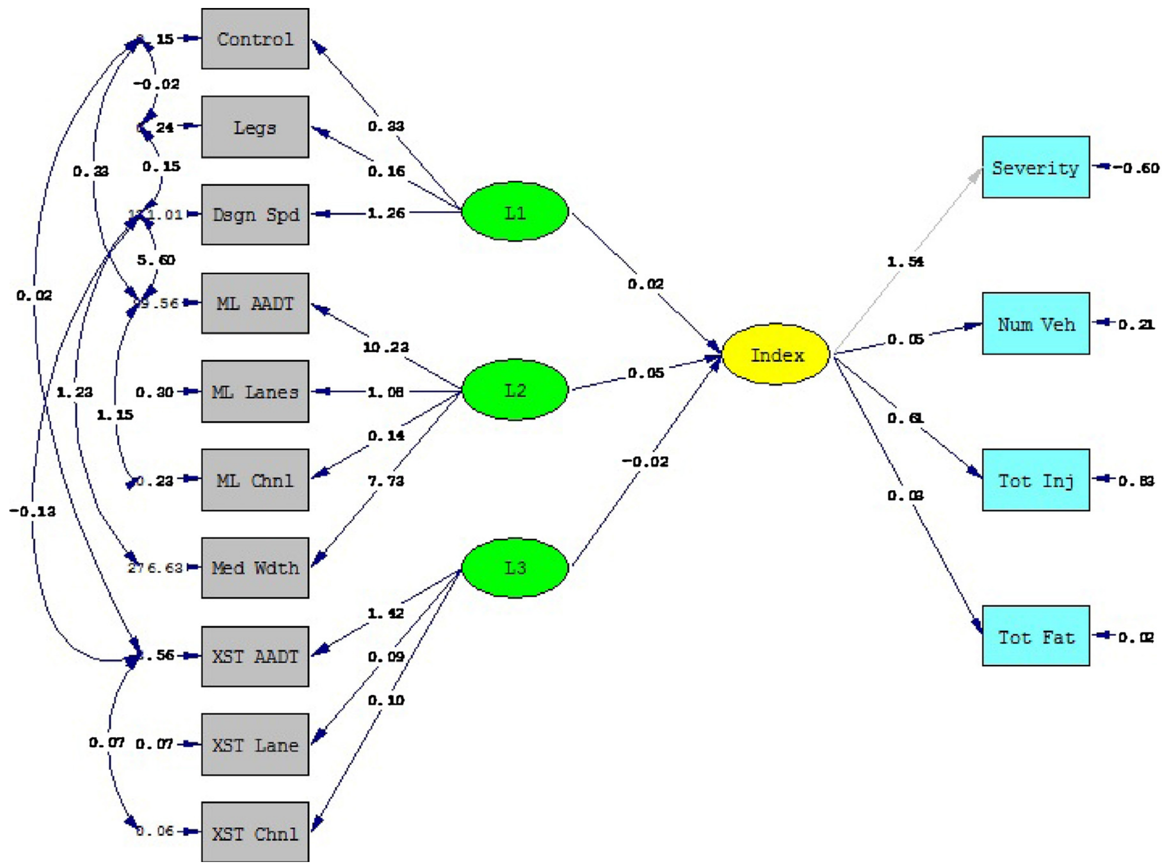

Fig. 1. Structural model – unsignalized intersections.

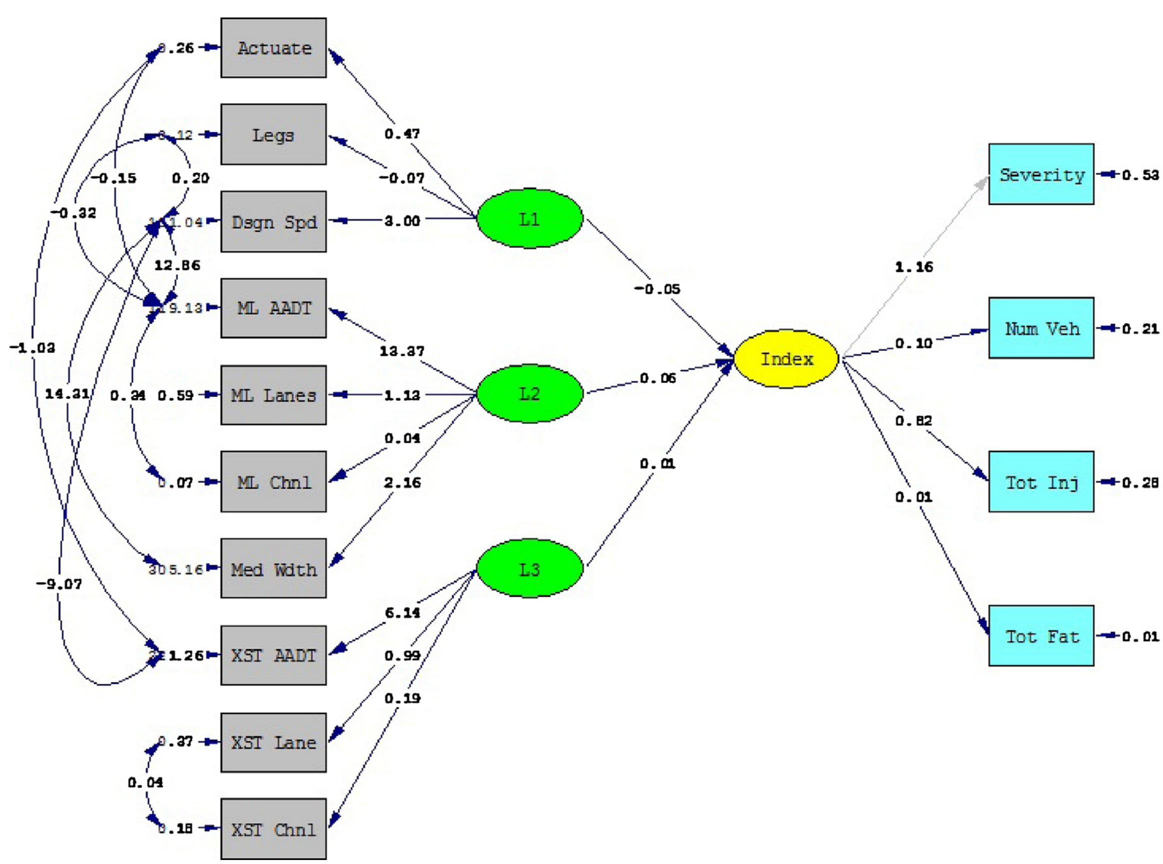

Fig. 2. Structural model – signalized intersections.

Table 5
Rotated factor pattern – signalized intersections.

Rotated pattern	Factor 1	Factor 2	Factor 3
Proposed main street characteristics dimension			
Divided	**0.78651**	−0.48282	**0.09623**
Median width	**0.25463**	−0.31381	−0.23676
L shoulder width	−0.66229	**0.14615**	0.00795
R shoulder width	−0.16323	−0.31720	**0.08584**
Surface width	**0.14925**	**0.85716**	−0.17744
Lane width	−0.07229	**0.35679**	**0.11549**
Proposed intersection characteristics dimension			
Intersection legs	**0.08617**	**0.07984**	−0.03446
ML lanes	**0.86951**	0.34311	−0.07851
XST lanes	**0.32422**	**0.27954**	**0.23967**
ML channel	0.02274	0.05842	**0.59873**
XST channel	**0.10893**	**0.18444**	**0.45461**
Proposed environmental characteristics dimension			
Precipitation	−0.04339	0.00303	0.00316
Light	−0.08473	0.00465	0.04136
Proposed traffic related characteristics dimension			
Actuation	−0.05598	−0.06580	**0.49270**
ML AADT	**0.59755**	**0.19114**	**0.11352**
XST AADT	**0.18146**	**0.16931**	**0.08626**
Design speed	0.00251	−0.09647	**0.20300**

Several structures were again tested based on these new dimensions ultimately leading to the statistically significant converging model (computed using the LISREL software) displayed in Fig. 2.

The results summarizing the model are presented in Tables 6A and 6B.

Again, for models with large sample sizes (such as this, $N = 20{,}215$) Chi-squared tests often encounter problems and goodness of fit was assessed based on the Root Mean Square Error of Approximation (RMSEA) (Golob, 2001). For the model above, the RMSEA was 0.072 and the 90% confidence interval was 0.070; 0.073. The entire confidence interval is under the threshold of 0.08 indicating that the model is statistically significant and has a fair fit (Hu and Bentler, 1998). Moreover, the Standardized Root Mean Square Residual (SRMR) has a value of 0.057 that is also less than 0.08 indicating a good fit (Hu and Bentler, 1998). The Goodness of Fit Index (GFI = 0.95) and the Adjusted Goodness of Fit Index (AGFI = 0.93) confirmed the statistical significance of the model. *t*-Values were assessed in the same manner and as was the case for unsignalized intersections, nearly all *t*-value in this model were significant.

As mentioned by Hamdar and Schorr (2013), the exclusion of certain dimensions gives way to two main biases: over or under estimation of certain dimensions impact on the safety index and over or under estimation of covariances between variables. While this is not ideal, it is important to keep in mind that there are biases associated with any type of statistical analysis. The utilization of factor scores as well as the statistical significance of the model and *t*-values works to alleviate the drawbacks inherent in the approach. Furthermore, it can be noticed that R^2 values are rather low for a selection of the variables in both models. The authors caution the interpretation of these values based upon the research of Kvalseth (1985) and Helland (1987). Calculation and interpretation of R^2 values is straightforward when dealing with linear least squares regression models with an intercept term, but this is far more difficult for other modelling techniques (such as SEM) (Kvalseth, 1985). While Helland (1987) points out that some statisticians take the extreme notion that R^2 values should be ignored completely, Kvalseth (1985) offers the more practical suggestion that these values be considered along with other goodness of fit measures such as the RMSR or the standard error of prediction. For these reasons the authors have investigated the various fit statistics discussed thus far, and in an additional attempt to confirm the results of the factor analysis and check the internal consistency of the model, Cronbach's alpha was utilized (Cortina, 1993). The cut-off for statistical significance is 0.7 (Reynaldo and Santos, 1999), a threshold that all values are under indicating that the model may be unstable. Sijtsma (2009) suggests that alpha is typically calculated when variables are either dichotomous or have the same scale (one popular application is using alpha for surveys involving a

Table 6A
Signalized model measurement equations.

Equation	Errorvar	R^2 value
Structural model		
INDEX = −0.045 × L1 + 0.059 × L2 + 0.0075 × L3	0.99	0.0057
Endogenous measurement model		
Severity = 1.16 × INDEX	0.53	0.72
Number of vehicles = 0.10 × INDEX	0.21	0.046
Total injuries = 0.82 × INDEX	0.28	0.70
Total fatalities = 0.0100 × INDEX	0.0071	0.014
Exogenous measurement model		
Design speed = 3.00 × L1	101.04	0.082
Actuation = 0.47 × L1	0.26	0.46
Intersection legs = −0.068 × L1	0.12	0.036
ML AADT = 13.37 × L2	119.13	0.60
ML lanes = 1.13 × L2	0.59	0.69
ML channelization = 0.040 × L2	0.071	0.022
Median width = 2.16 × L2	305.16	0.015
XST AADT = 6.14 × L3	221.26	0.15
XST lanes = 0.99 × L3	0.37	0.73
XST channelization = 0.19 × L3	0.18	0.17

Table 6B
Signalized model statistics.

t-Values

Variables	Value
L1/design speed	15.76
L1/actuation	16.92
L1/intersection legs	−14.14
L2/ML AADT	91.31
L2/ML lanes	95.54
L2/ML channelization	19.03
L2/median width	15.60
L3/XST AADT	45.00
L3/XST lanes	73.25
L3/XST channelization	47.88
L1/INDEX	−4.04
L2/INDEX	5.63
L3/INDEX	0.71
INDEX/severity	0.00
INDEX/number of vehicles	26.53
INDEX/total injuries	39.18
INDEX/total fatalities	14.98

Error covariance terms

Variables	Value
Design speed/intersection legs	0.20
Design speed/ML AADT	12.86
Design speed/median width	14.31
Design speed/XST AADT	−9.07
Actuation/ML AADT	−0.15
Actuation/XST AADT	−1.03
Intersection legs/ML AADT	−0.32
ML AADT/ML channelization	0.34
XST lanes/XST channelization	0.04

Cronbach's alpha

Set	Value
Entire data set	0.585
Dimension L1	0.020
Dimension L2	0.564
Dimension L3	0.532

Likert scale) and even then only assesses the degree to which variables are inter-related. Given the high amount of variation in the scales of the variables (for example design speed has a range between 25 and 65 while variables such as channelization are either equal to 0 or 1 and those such as stop control vary between 0 and 4), it is not surprising that some alpha values are below the well below the cutoff (specifically in terms of the dimensions that contain the design speed variable) while others meet the requirements or are just slightly under them. Through the testing of alternative factor and model structures, the analysis of fit criteria and the observation of *t*-values, error covariances and R^2 values the models presented above are determined to be consistent. It is important to keep in mind that the SEM approach lends itself to analytical models rather than predictive models and the results obtained above are specific to the area of analysis and cannot be applied universally across all intersections.

5. Analysis of results

For both models, *high values for the SPI are indicative of a decrease in safety*. Observation of the endogenous equations in both models demonstrates that as the SPI increases, so do the values of all endogenous variable. This result works to enhance the validity of the safety framework proposed in this study as the "safer driving environment" features a lower severity level as well as fewer injuries, fatalities and vehicles per collision. Consideration must be given to questions of exposure, occupancy, and land use. In terms of exposure; since individual collisions are the observations in the data set, intersections with high collision rates are represented by repeated data lines with very similar exogenous variables. These repeated entries for intersections with increased collision rates imply that intersections that feature frequent severe collisions will have a higher safety index as compared to intersections with as frequent less severe collisions as well as those with less frequent as severe collisions. Additionally questions may arise when considering that certain geographic areas feature higher vehicle occupancy than others. The implication here is that similarly severe collisions occurring in these areas may have increased fatalities and injuries due to occupancy. Thinking of safety in absolute terms means that the main goal is the reduction in the total injuries and fatalities that occur in collisions, and as is such any slight bias of the model towards intersections featuring high vehicle occupancy does not necessarily constitute a deficiency. Still, consideration should be given to the rare multiple fatality collision that may occur in an area where occupancy is not typically as high. While these types of collisions could potentially skew model results, the robustness of the data set along with the 4 year timeframe over which the data was collected both work to combat this issue. One final consideration needs to be made in terms of the variable for design speed. It is important to make the distinction between design speed, posted speed limits and operating speeds; and while relationships between posted speed limits and operating speeds can be defined with a level of confidence, the same does not hold true when comparing either with design speed (Fitzpatrick et al., 2003). While this is less of an issue in rural settings where posted speed limits are generally higher; issues arise when considering urban settings that require significant engineering judgments as well as social and political considerations when determining the speed limit (Forbes et al., 2012). This discussion of design speed is continued further in the section on the signalized model.

In the following subsections analysis is conducted for signalized and unsignalized intersections independently and then the models are compared to one another.

5.1. Unsignalized intersections

Descriptive statistics for endogenous variables in the unsignalized structural model are provided in Table 7.

Endogenous variables influence the model both in magnitude and variability and both must be examined in order to correctly interpret the model. In terms of magnitude, endogenous variables were averaged and this value was input into the measurement equations (Table 4A) to yield the average contribution that each variable gives to the SPI (Table 7). Although high magnitudes may indicate increased influence on the SPI, if the variable remains constant throughout the majority of the data set than that influence is somewhat reduced. To account for this, the standard deviation is calculated and input into the equation to yield the change in SPI from a one standard deviation change in each variable (Table 7). Furthermore, to understand how frequently these changes occur the coefficient of variation was also calculated (Table 7). By examining both the magnitude that a change in each variable (deviation contribution) as well as how frequently these changes occur (coefficient of variation) the relative influence of each variable can be better understood. Finally, covariance terms throughout the model are explicitly stated in Table 7. While the signs of these terms speak to how one variable "moves" with another in the dataset, the magnitude of covariance terms cannot be interpreted in the same manner. To interpret the magnitude, another measure (such as the Pearson's correlation coefficient) could be used, but due to the sample size and the vastly different scales featured by the variables; such a measure does not allow for additional concrete insights.

Observation of the structural model for unsignalized intersections demonstrates that the model is characterized by positive valued influences from the Main Street and Intersection Characteristics dimensions and a negative valued influence from the Cross Street Characteristics. Taking the absolute value of coefficients (from Fig. 1) shows the largest influence on the safety index comes from the main street characteristics dimension (0.052), followed by cross street characteristics (−0.023) and main street characteristics (0.018).

Starting with the intersection characteristics dimension, positive valued coefficients throughout the dimension indicate that an increase in any of the variables will have an *adverse* effect on safety. The diminishing effects of increased speed on safety have been demonstrated by Haleem and Abdel-Aty (2010) and the NCHRP (2009) and are a somewhat expected result as high speeds crashes are inherently severe. Additionally, Greibe (2003) and Abdel-Aty and Haleem (2011) have found that an increase in the number of intersection legs leads to decreased safety. Once again this result is consistent with expectation as increased complexity forces the driver to make more decisions as they enter the intersection. Finally, the only seemingly counterintuitive result in this dimension is the decrease in safety that is associated with an increased level of control. Greibe (2003) had similar results when he found that speed reducing measures have a negative effect on safety. Further discussion of this finding is provided in subsection on model comparison. High contribution values in Table 7 for design speed and intersection legs demonstrate that they have the highest average contributions and are among the highest deviation contributions of any variable within the model. While both have relatively low coefficients of variation, they remain among the most influential variables in the model.

Moving to the cross street characteristics dimension, positive valued coefficients throughout the dimension indicate that an increase in any variable within the dimension will have a *positive* effect on safety. While the increase in safety with the increase in cross street traffic may be counterintuitive, the deviation contribution from cross street AADT is among the smallest in the model. Additionally, observations of covariance terms for cross

Table 7
Analytical statistics for exogenous variables – unsignalized intersections.

Variable	Average	Standard deviation	Coefficient of variation	Average SPI contribution	Deviation SPI contribution
Intersection characteristics dimension					
Stop control	2.0096	0.5091	0.2533	0.1096	0.0278
Intersection legs	3.4938	0.5147	0.1473	0.3930	0.0579
Design speed	53.7541	11.0728	0.2060	0.7679	0.1582
Main street characteristics dimension					
ML AADT	18.9018	14.2897	0.7560	0.0961	0.0726
ML lanes	2.9541	1.2094	0.4094	0.1422	0.0582
ML channel	0.5488	0.4976	0.9067	0.2039	0.1848
Median width	7.3117	12.6242	1.7266	0.0492	0.0849
Cross street characteristics dimension					
XST AADT	1.4997	2.3594	1.5733	−0.0243	−0.0382
XST lanes	2.0056	0.2707	0.1350	−0.5364	−0.0724
XST channel	0.0764	0.2656	3.4781	−0.0176	−0.0611

Covariance terms

Variable 1	Variable 2	Covariance	Variable 1	Variable 2	Covariance
Intersection legs	Stop control	0.05	ML AADT	Design speed	2.21
Intersection legs	ML AADT	−0.66	XST AADT	ML lanes	−0.01
Intersection legs	XST AADT	0.17	XST AADT	XST lanes	0.1
Intersection legs	ML lanes	−0.04	XST AADT	ML channel	0.12
Intersection legs	XST lanes	0.01	XST AADT	XST channel	0.15
Intersection legs	Median width	0.18	XST AADT	Median width	1.77
Intersection legs	Design speed	0.33	XST AADT	Design speed	0.85
Stop control	ML AADT	−0.39	Design speed	ML lanes	−0.5
Stop control	XST AADT	0.28	Design speed	XST lanes	−0.09
Stop control	ML lanes	−0.07	Design speed	ML channel	0.47
Stop control	XST lanes	0.02	Design speed	XST channel	0.11
Stop control	ML channel	0.01	Design speed	Median width	−1.37
Stop control	XST channel	0.01	ML lanes	XST lanes	0.01
Stop control	Median width	−1.12	ML lanes	ML channel	0.13
Stop control	Design speed	0.41	ML lanes	Median width	8.72
ML AADT	XST AADT	−1.03	XST lanes	ML channel	0.01
ML AADT	ML lanes	11.05	XST lanes	XST channel	0.01
ML AADT	XST lanes	0.37	XST lanes	Median width	−0.03
ML AADT	ML channel	2.1	Median width	ML channel	0.82
ML AADT	XST channel	0.1	Median width	XST channel	0.01
ML AADT	Median width	65.72	ML channel	XST channel	0.02

street AADT indicate that as cross street AADT increases, so does the complexity of the intersection (number of intersection legs and both main and cross street channelization) as well as the design speed and stop control. Coupled with the low deviation contributions across the dimension, these covariance terms reinforce the notion that model results are dominated by variables contained within the other dimensions.

The final dimension is the main street characteristics dimension where positive valued coefficients throughout indicate that an increase in any variable within the dimension will have an *adverse* effect on safety. A number of studies have identified increasing AADT as a contributing factor to decreasing safety, and this is to be expected as increased exposure will often lead to a decrease in safety (Pirdavani et al., 2010; Greibe, 2003; Abdel-Aty and Haleem, 2011). The negative effect of increasing median width on safety has also been identified in previous research by Haleem and Abdel-Aty (2010) where they found that larger median widths are associated with increased accident severity. The results for number of lanes and channelization reinforce the analysis above postulating that increased intersection complexity had an adverse effect on safety.

Coefficient values are used to identify some of the safest (Fig. 3 – Third Street and Alameda Blvd in Coronado) and least safe (Fig. 3 – El Campo Road and US 101 in Arroyo Grande) unsignalized intersection in the data set. There were 7 collisions that occurred in the timeframe of this study at the safer intersection (involving an average of 2 vehicles/collision, resulting in 0.29 injuries/collision

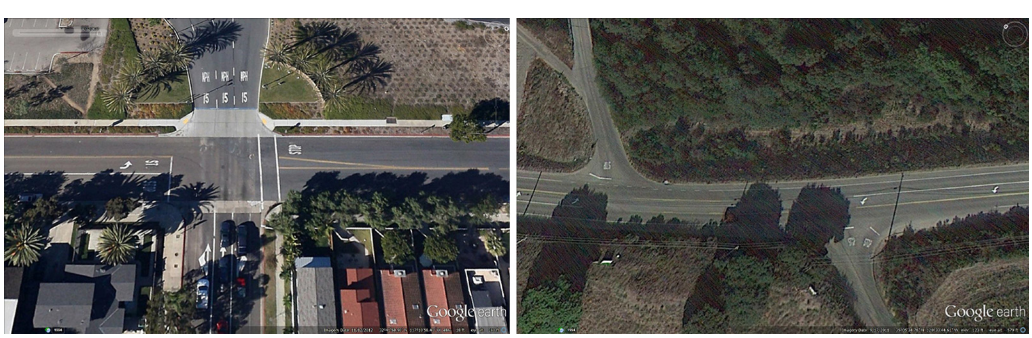

Fig. 3. Safer (left) and less safe (right) unsignalized intersections.

Courtesy: Google Earth.

Table 8
Analytical statistics for exogenous variables – signalized intersections.

Variable	Average	Standard deviation	Coefficient of variation	Average SPI contribution	Deviation SPI contribution
Intersection characteristics dimension					
Actuation	2.6921	0.6909	0.2566	−0.2578	−0.0662
Intersection legs	3.8903	0.3600	0.0925	2.5745	0.2382
Design speed	51.2654	10.4901	0.2046	−0.7690	−0.1574
Main street characteristics dimension					
ML AADT	33.5248	17.2571	0.5148	0.1479	0.0762
ML lanes	4.2447	1.3704	0.3229	0.2216	0.0716
ML channel	0.9215	0.2689	0.2918	1.3593	0.3966
Median width	14.2738	13.7499	0.9633	0.3899	0.3756
Cross street characteristics dimension					
XST AADT	11.0601	16.0912	1.4549	0.0135	0.0197
XST lanes	3.0184	1.1616	0.3848	0.0229	0.0088
XST channel	0.6792	0.4668	0.6873	0.0268	0.0184

Covariance terms

Variable 1	Variable 2	Covariance	Variable 1	Variable 2	Covariance
Intersection legs	Actuation	−0.03	XST AADT	ML lanes	3.45
Intersection legs	ML AADT	0.02	XST AADT	XST lanes	6.05
Intersection legs	XST AADT	0.52	XST AADT	ML channel	0.14
Intersection legs	ML lanes	0.04	XST AADT	XST channel	1.11
Intersection legs	XST lanes	0.05	XST AADT	Median width	2.09
Intersection legs	Median width	0.32	XST AADT	Design speed	−12.44
Intersection legs	Design speed	−0.16	Design speed	ML lanes	0.07
Actuation	ML AADT	0.57	Design speed	XST lanes	−0.03
Actuation	XST AADT	−1.16	Design speed	ML channel	0.36
Actuation	ML lanes	0.02	Design speed	XST channel	0.19
Actuation	ML channel	0.07	Design speed	Median width	14.07
Actuation	XST channel	0.06	ML lanes	XST lanes	0.54
Actuation	Median width	−0.3	ML lanes	ML channel	0.03
Actuation	Design speed	1.4	ML lanes	XST channel	0.08
ML AADT	XST AADT	40.48	ML lanes	Median width	3.2
ML AADT	ML lanes	15.18	XST lanes	ML channel	0.03
ML AADT	XST lanes	5.73	XST lanes	XST channel	0.19
ML AADT	ML channel	0.73	XST lanes	Median width	0.27
ML AADT	XST channel	1.39	Median width	ML channel	−0.53
ML AADT	Median width	22.36	Median width	XST channel	−0.38
ML AADT	Design speed	16.02	ML channel	XST channel	0.03

and having an average police reported severity level of 1.14) while 23 collisions occurred at the less safe intersection (involving an average of 1.96 vehicles/collision, resulting in an average 1.35 injuries/collision and having an average police reported severity level of 2.09). Looking at the characteristics of the intersections, the safer intersection has a lower design speed (25 mph as opposed to 65 mph), less main street AADT (17,900 as opposed to 55,000), and fewer main street lanes (3 opposed to 4). The safer intersection additionally features much higher cross street traffic (9300 as opposed to 500); which as discussed above was a counter intuitive finding of the model. Clearly, the high design speed and complicated geometry of the intersection on the right work to create an unsafe driving environment.

5.2. Signalized intersections

Descriptive statistics for endogenous variables in the signalized structural model are provided in Table 8.

Observation of the structural model for signalized intersections demonstrates that the model is characterized by positive valued influences from the main street and cross street characteristics dimensions and a negative valued influence from the intersection characteristics dimension. Taking the absolute value of coefficients (from Fig. 2) shows that the strongest influence is from the main street characteristics (0.059) followed by intersection characteristics (−0.045) and then cross street characteristics (0.0075).

Looking first at the main street characteristics dimension, positive valued coefficients for all variables indicate that increasing values within the dimension will have an *adverse* effect on safety. Chin and Quddus (2003) found that AADT had the largest influence on the number of collisions at signalized intersections and also found an increase in accident frequency for median widths greater than two feet. Greibe (2003) also identified AADT as the most important factor on safety; Abdel-Aty and Keller (2005) also found AADT to be a major contributing factor on injury severity for a number of different collision types occurring at signalized intersections.

Moving to the intersection characteristics dimension, the signs of the coefficient values indicate that increasing the design speed or the level of actuation will have a *positive* effect on safety while increasing the number of intersection legs will have an *adverse* effect. The increase in safety for an increased level of actuation is consistent with the findings of Chin and Quddus (2003) who found a 5% reduction in collisions when comparing adaptive to pre-timed signals. Coupled with the main street variables for number of lanes and channelization, the influence of the number of intersection legs once again indicates that there is a decrease in safety with an increase in intersection complexity. The variable for design speed presents the only counterintuitive result as high speeds are generally associated with decreased safety, but the opposite appears to be suggested by the model. This result is likely a function of a large number of urban intersections in the dataset. Urban intersections often feature complicated geometry and higher AADT as

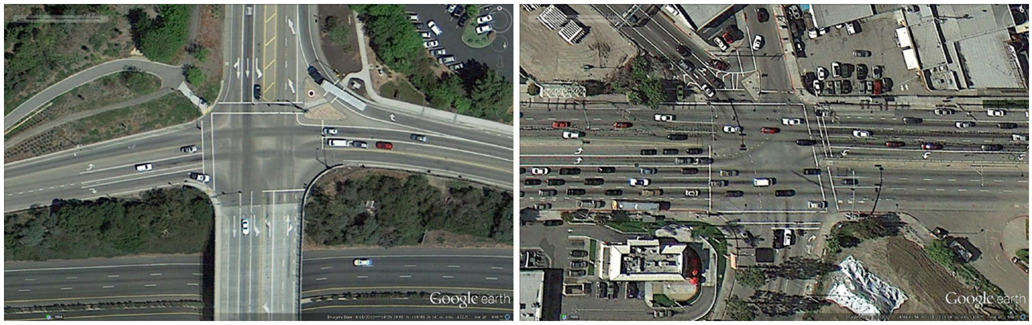

Fig. 4. Safer (left) and less safe (right) signalized intersections.

Courtesy: Google Earth.

compared to their rural counterparts that are often less travelled and contain simple geometry. Furthermore, as discussed previously design speed does not necessarily correlate with the posted speed limit and more importantly, with the operating speeds of the vehicles – specifically in urban settings (Fitzpatrick et al., 2003; Forbes et al., 2012). Additionally, Fitzpatrick et al. (2003) notes that while between 37% and 64% of vehicles on rural roadways were at or below the posted speed limit, this range dropped to 23–52% for an urban/suburban setting. Thus, one potential explanation for this finding is that when considering the large number of urban intersections along with the manner in which posted speed limits are determined in these settings, intersections with higher design speeds are better able to accommodate the tendency of drivers to exceed these posted speeds. Wang et al. (2003) had similar findings in terms of the effects of increasing speed, but also offers caution as to interpreting that result. As was the case for the unsignalized model, the deviation contribution from intersection legs is among the highest in the model; again with a low coefficient of variation. Observation of covariance terms between both design speed and intersection legs and actuation and intersection legs indicates a negative valued relationship; and positive valued relationship is observed in the covariance between design speed and actuation. These relationships offer additional insights into the counterintuitive effects of design speed on safety – as it further reinforces the notion that the dimension is dominated by the variable for intersection legs which has significantly higher average and deviation contributions than both design speed and actuation.

The final dimension is the cross street characteristics dimension where positive valued coefficients indicate that increasing the value of variables within this dimension will have an *adverse* effect on safety. Low deviation contributions from variables within the dimension indicate that these variables are among the least influential in the model.

Coefficient values were used to identify some of the safest (Fig. 4 – Chumash Hwy and State Street Route in Santa Barbara) and least safe (Fig. 4 – Venice Blvd and S Robertson Blvd in Los Angeles) signalized intersections. The safer intersection had 5 collisions within the timeframe of this study (involving an average of 2.4 vehicles/collision, resulting in 0.4 injuries/collision and having an average police reported severity level of 1) while the less safe intersection had 14 collisions (involving an average of 2.14 vehicles/collision, resulting in 1.0 injuries/collision and having an average police reported severity level of 2.36). Looking at the characteristics of the intersections, the safer intersection features less legs (4 as opposed to 5), less main street lanes (4 opposed to 6) and less main street AADT (18,000 opposed to 62,804). Additionally, the safer intersection has a higher design speed (65 mph) than the less safe intersection (40 mph). From this, it appears the complicated geometry of the highly travelled intersection on the right works to create the unsafe driving environment.

5.3. Model comparison

The first interesting observation is the fact that the best models for both signalized and un-signalized intersections contained the same variables and have the same structure. Hamdar and Schorr (2013) performed a similar analysis for interrupted and uninterrupted flow segments and determined that the models for differing flow conditions had different structures. Consistent structure between the two models indicates that while the factors effecting safety vary with flow condition (Hamdar and Schorr, 2013); this is not the case when comparing different types of signalization. This consistency is a possible explanation for why the number of collisions at signalized and unsignalized intersections is nearly identical both on a national level (NHTSA, 2012) and within the data set.

In terms of the individual variables, the influence within the main street characteristics dimension is almost the same in both models – increasing the value of each variable influences safety in the same fashion and in approximately the same proportion in both models. Observation of the intersection characteristics dimension demonstrates that the only commonality is that increasing the number of intersection legs will decrease safety in both models. Increasing the design speed or level of actuation/control for unsignalized intersection will negatively influence safety, while an increase in those variables will have a positive effect on safety in the signalized model. Comparing the results for actuation and control is a difficult task given the non-encompassing nature of the variables – specifically actuation. As pointed out by Nittymaki (2001), there are multiple aspects to signal control (maximizing safety while minimizing delay and environmental impacts) making it difficult to consider them all together and leading to a situation where cause–consequence relationships are impossible to explain. Exacerbating this issue is the fact that data from the HSIS does not speak to cycle length, permissive left turns or signal coordination for pre-timed signals (for actuated signals these characteristics are inherently highly variable based on the signals response to traffic conditions, Zheng and Recker, 2013). To a lesser extent there are similar difficulties associated with the interpretation of stop control as land use issues can play a role in the sense that there may be an increase in the number of 4-way stops when comparing urban areas to their rural counterparts. Observation of covariance terms for actuation and control with that of intersection legs indicates a negative valued relationship between actuation and intersection legs and a positive valued relationship between stop control and intersection legs. Similarly, for the unsignalized model, design speed and intersection legs have a positive valued relationship while the opposite is true for the signalized model. Coupled with the fact that an increase in the number of intersections legs is one of the main indicators of a decrease in safety for both models, these covariance terms provide additional insights into the differing impacts of increasing the level of actuation and stop control. The limitations of these variables along with these associated correlations make it difficult to draw comparative conclusions about the effects of flow conditions at both types of intersections. This was also a finding of Greibe (2003) who points out that internal correlations of variables associated with traffic flow make it difficult to identify the safety effects of one explanatory variable since it may be influenced by others in the model. For the cross street characteristics dimension, increasing values for all variables has an adverse effect on safety for signalized intersections and a positive effect for unsignalized intersections. Referring to Tables 7 and 8, this discrepancy between the models is likely a result of the substantial increase in the average amount of cross street AADT (as well as an increased average number of cross street lanes and channelization) in the signalized data set as compared to the unsignalized one.

6. Concluding remarks

This study features the development of a safety propensity index for both signalized and unsignalized intersections through the use of a structural equation modelling approach. Data sets provided by the HSIS for the California transportation network were vetted and 22,422 collisions at unsignalized intersections and 20,215 collisions occurring at signalized intersections occurring between 2006 and 2010 were considered for analysis. Through the use of a factor analysis (performed using the SAS software), pertinent variables were identified and possible dimensions were postulated. Several structures were then tested before statistically significant models were achieved using the LISREL software for both intersection types. A number of checks were carried out to validate the models significance and the results were then analyzed. The manner in which each variable and dimension effects safety at both types of intersections was outlined and the models were then compared to provide additional insight.

The statistically significant results produced in this study are specific to the analyzed network. Future research should examine the application of this method to a different transportation network (especially signalized versus unsignalized intersections) as it is unclear whether or not the model structure and coefficient values would remain consistent. Furthermore, additional considerations should be given to the impacts of land use as identified throughout this paper. Though a universal model would be difficult to achieve, the continued application of the conceptual and quantitative framework utilized in this study provides an opportunity for researchers to gain new perspectives on safety.

Acknowledgements

The authors would like to thank the Highway Safety Information System for their efforts in proving the data utilized in this study. This material is based upon work supported by the National Science Foundation under Grant No. 0927138. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Abdel-Aty, M., Abdelwahab, H., 2004. Predicting injury severity levels in traffic crashes: a modeling comparison. J. Transport. Eng., http://dx.doi.org/10.1061/(ASCE)0733-947X(2004)130:2(204).

Abdel-Aty, M., Haleem, K., 2011. Analyzing Angle Crashes at Unsignalized Intersections Using Machine Learning Techniques. Accid. Anal. Prev. 43 (1), 461–470.

Abdel-Aty, M., Keller, J., 2005. Exploring the Overall and Specific Crash Severity Levels at Signalized Intersections. Accid. Anal. Prev. 37 (3), 417–425.

Bham, G., Javvadi, B., Manepalli, U., 2012. Multinomial Logistic Regression Model for Single-Vehicle and Multivehicle Collisions on Urban U.S. Highways in Arkansas. J. Transport. Eng., http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000370.

Caliendo, C., Guida, M., 2012. Microsimulation approach for predicting crashes at unsignalized intersections using traffic conflicts. J. Transport. Eng., http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000473.

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to traffic accident occurrence at signalized intersections. Accid. Anal. Prev. 35 (2), 253–259.

Cortina, J., 1993. What is coefficient alpha? An examination of theory and application. J. Appl. Psychol. 78 (1), 98–104.

Fitzpatrick, K., Carlson, P., Brewer, M., Wooldridge, M., 2003. Design speed, operating speed, and posted speed limit practices. In: 82nd Annual Meeting of the Transportation Research Board, January.

Forbes, G., Gardner, T., McGee, H., Srinivasan, R., 2012. Methods and Practices for Setting Speed Limits: An Informational Report. Federal Highway Administration, Report No. FHWA-SA-12-004.

Golob, T.F., Meurs, H., 1986. Development of Structural Equation Models of the Dynamics of Passenger Travel Demand, Irvine: Institute of Transportation Studies. University of California Irvine, Irvine, California.

Golob, T.F., 2001. Structural Equation Modelling for Travel Behavior Research, Irvine: Institute of Transportation Studies. University of California Irvine, Irvine, California.

Greibe, P., 2003. Accident prediction models for urban roads. Accid. Anal. Prev. 35 (2), 273–285.

Haleem, K., Abdel-Aty, M., 2010. Examining traffic crash injury severity at unsignalized intersections. J. Saf. Res. 41 (4), 347–357.

Hamdar, S.H., Schorr, J., 2013. Interrupted versus uninterrupted flow: a safety propensity index for driver behavior. Accid. Anal. Prev. 55, 22–33.

Hamdar, S.H., Mahmassani, H.S., Chen, R.B., 2008. Aggressiveness propensity index for driving behavoir at signalized intersections. Accid. Anal. Prev. 40, 315–326.

Helland, I., 1987. On the interpretation and used of R^2 in regression analysis. Biometrics 43, 61–69.

Hu, L., Bentler, P.M., 1998. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. Psychol. Methods, 424–453.

Isebrands, H., Hallmark, S., Li, W., McDonald, T., Storm, R., Preston, H., 2010. Roadway lighting shows safety benefits at rural intersections. J. Transport. Eng., http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000161.

Jung, S., Xiao, Q., Noyce, D., 2012. Injury severity of multivehicle crash in rainy weather. J. Transport. Eng., http://dx.doi.org/10.1061/(ASCE).TE.1943-5436.0000300.

Karlaftis, M., Golias, I., 2001. Effects of road geomertry and traffic volumes on rural roadway accident rates. Accid. Anal. Prev. 34 (3), 357–365.

Kim, D., Washington, S., Oh, J., 2006. Modeling crash types: new insights into the effects of covariates on crashes at rural intersections. J. Transport. Eng., http://dx.doi.org/10.1061/(ASCE)0733-947X(2006)132:4(282).

Kvalseth, T., 1985. Cautionary note about R^2. Am. Stat. 39 (4 Part 1), 279–285.

Lee, J., Chung, J., Son, B., 2008. Analysis of traffic accident size for Korean highway using structural equation models. Accid. Anal. Prev. 40, 1955–1963.

Li, J., Abdelwahab, W., Brown, G., 1994. Joint effects of access and geometry on two-lane rural highway safety in British Columbia. Can. J. Civil Eng. 21 (6), 1012–1024.

National Cooperative Highway Research Program, 2009. Report 500: Guidance for Implementation of the AASHTP Strategic Highway Safety Plan: A Guide for Reducing Speeding-Related Crashes. Transportation Research Board, Washington, DC.

National Highway Traffic Safety Administration, 2012. Traffic Safety Facts 2010. U.S. Department of Transportation, Washington DC.

Nittymaki, J., 2001. General fuzzy rule base for isolated traffic signal control – rule formulation. Transport. Plan. Technol. 24, 227–247.

Paleti, R., Eluru, N., Bhat, C., 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. Accid. Anal. Prev. 42 (6), 1839–1854.

Pirdavani, A., Brijs, T., Bellemans, T., Wetz, G., 2010. Evaluation of traffic safety at unsignalized intersections using microsimulation: a utilization of proximal safety indicators. Adv. Transport. Stud. Int. J. 22 (A), 43–50.

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. J. Transport. Eng. 122 (2), 105–113.

Quddus, M., Wang, C., Ison, S., 2010. Road traffic congestion and crash severity: econometric analysis using ordered response models. J. Transport. Eng., http://dx.doi.org/10.1061/(ASCE).TE.1943-5436.0000044.

Reynaldo, J., Santos, A., 1999. Cronbach's Alpha: a tool for assessing the reliability of scales. J. Exten. 37 (2), 2TOT3.

Sijtsma, K., 2009. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika 74 (1), 107–120, http://dx.doi.org/10.1007/s11336-008-9101-0.

Vogt, A., Bared, J., 1998. Accident models for two-lane rural segments and intersections. Transport. Res. Rec. 1635, 18–29, http://dx.doi.org/10.3141/1635-03.

Wang, Y., Ieda, H., Mannering, F., 2003. Estimating rear-end accident probabilities at signalized intersections: occurance-mchanism approach. J. Transp. Eng. 129 (4), 377–384, http://dx.doi.org/10.1061/(ASCE)0733-947X(2003)129:4(377).

Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analysis of rear-end crashes at signalized intersections. Accid. Anal. Prev. 38 (6), 1137–1150.

Wu, K., Donnell, E., Himes, S., Sasidharan, L., 2013. Exploring the association between traffic safety and geometric design consistency based on vehicle speed metrics. J. Transport. Eng., http://dx.doi.org/10.1061/(ASCE).TE.1943-5436.0000553.

Xie, Y., Zhang, Y., Liang, F., 2009. Crash injury severity analysis using bayesian ordered probit models. J. Transport. Eng., http://dx.doi.org/10.1061/(ASCE)0733-947X(2009)135:1(18).

Zheng, X., Recker, W., 2013. An adaptive control algorithm for traffic-actuated signals. Transport. Res. C 30, 93–115, http://dx.doi.org/10.1016/j.trc.2013.02.007.