Use of Structural Equation Modeling to Measure Severity of Single-Vehicle Crashes

Kai Wang and Xiao Qin

Injury severity and vehicle damage are two of the main indicators of the level of crash severity. Other factors, such as driver characteristics, roadway conditions, highway geometry, environmental factors, vehicle type, and roadside objects, may also be directly or indirectly related to crash severity. All these factors interact in such complicated ways that it is often difficult to identify their interrelationships. The aim of this study was to examine the relationships between these contributors and the severity of single-vehicle crashes. Structural equation modeling (SEM) offers the opportunity to explore the complex relationships between variables by handling endogenous variables and exogenous variables simultaneously. Furthermore, SEM allows latent variables to be included in the model and bridges the gap between dependent and explanatory variables. In this study, the number of latent variables was defined by the understanding of collision force, kinetic energy, and mechanical process of a collision, as well as statistical goodness of fit that was based on available data. Three SEM models (one with one latent variable, one with two, and one with three) representing the hypothesized relationships between collision force, speed of a vehicle, and severity of a crash were developed and evaluated in an attempt to unravel the relationships between exogenous factors and severity of single-vehicle crashes. On the basis of goodness of fit and model predictive power, the model with two latent variables outperformed the other two. Additional insights about model selection were provided through the development and comparison of the three models.

Because, in large part, of the significant impacts of injuries, deaths, and economic losses from motor vehicle accidents, traffic safety continues to be the top priority on the national transportation agenda. AASHTO's Strategy Highway Safety Plan, entitled Toward Zero Deaths: A National Strategy on Highway Safety, outlined a new road safety campaign focused on reducing fatal and severe-injury crashes. The campaign demands high levels of interagency cooperation between state departments of transportation and public safety, health, and other safety stakeholders (*1*). Under the collective improvements in highway engineering, vehicle technologies, driver education, and traffic enforcements, traffic fatalities in the United States have declined sharply. The number of single-vehicle fatal crashes in 2009 was 18,748 compared with 23,445 in 1990, a 20% reduction in 20 years (*2*). To accelerate implementation of the Toward Zero Deaths plan, the critical factors contributing to fatal and severe crashes must be identified and effective safety countermeasures must move forward.

The use of statistical methodologies has become common in identifying and analyzing the contributions of human, environmental, roadway, and vehicle factors on crash severity (*3*). Among all methods, the discrete choice model has been employed extensively to show the relationship between injury severity and vehicle damage. Representative logistic and probit models have been developed to examine the relationship between both injury severity and vehicle damage and their contributing factors. The mixed logic model is a good alternative when data sets are heterogeneous, as this modeling technique has the flexibility to treat coefficients as random or fixed variables.

Despite different assumptions and model specifications, all these modeling methodologies attempt to incorporate all available factors into the model and build a direct relationship between independent and dependent variables. However, this procedure may be problematic because the interrelationships among these factors can be intricate and therefore difficult to observe or verify (*4*). New modeling techniques are needed to unravel the complex relationships between crashes and their contributing factors.

The complex causes of a crash may be resolved by structural equation modeling (SEM), which can handle complex relationships among endogenous variables (i.e., variables that can be regressed on other variables) and exogenous variables (i.e., variables that are simultaneously independent). Moreover, SEM can include latent variables that are expected to bridge the knowledge gaps between dependent and explanatory variables in the model (*4*).

The primary objective of this study is to identify the causal relationships between exogenous factors and the severity of single-vehicle crashes by using structural equation models with different model structures. The secondary objective is to offer additional insights about model selection.

LITERATURE REVIEW

The discrete choice model is a popular choice for modeling the severity of a crash injury. This type of model can be treated as a nonordinal model by using the multinomial logistic or probit model (*5–8*). It can also be treated as an ordinal by using the ordered logistic or probit model (*9–14*).

However, both ordered and unordered logistic or probit models are fixed-parameter models in which each parameter is consistent across observations. The fixed-parameter models may not be reasonable when the data sets are heterogeneous or unobserved factors exist. Researchers have attempted to use the heterogeneity models to account for the issue of heterogeneity of crash data. Malyshkina and Mannering used a Markov switching multinomial logistic model to

K. Wang, Room 315, and X. Qin, Room 148, Crothers Engineering Hall, Department of Civil and Environmental Engineering, South Dakota State University, Box 2219, Brookings, SD 57007. Current affiliation for K. Wang: Department of Civil and Environmental Engineering, University of Connecticut, Unit 3037, 261 Glenbrook Road, Storrs, CT 06269-3037. Corresponding author: X. Qin, xiao.qin@sdstate.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2432, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 17–25.
DOI: 10.3141/2432-03

account for the possibility of unobserved factors that affect crash severity (*15*). Another appropriate alternative when the data sets are heterogeneous is the random-parameter (mixed) logistic model, as this approach has the flexibility to treat the parameters as fixed or random variables (*5, 7, 16–19*). Savolainen et al. summarized numerous discrete choice models currently used in modeling crash severity and provided additional insights about model assessment and selection (*20*).

Irrespective of the disparity between methodologies, the severity of personal injuries might not be the most honest variable when the severity of a crash is being examined because (*a*) the safety design of a vehicle can provide superior protection to the occupants by keeping intrusion into the passenger space to a minimum and (*b*) the resultant reported injury severity can be biased by the accident victim's descriptions, complaints, and responses (*21*). Therefore, many researchers have attempted to use vehicle damage as an indicator of the magnitude of the collision or to combine vehicle damage with injury severity (*21–23*).

Although considering or incorporating vehicle damage into injury severity models is a step toward accurately and impartially identifying the impacts of factors on crash severity and heterogeneity models can overcome the issue of data heterogeneity, discrete choice models may not explicitly capture the interrelationships between variables when the factors interact in such complicated ways. Both logistic or probit models and heterogeneity models tend to impose direct relationships on the independent and dependent variables; in contrast, SEM can effectively establish multiple relationships between or within endogenous and exogenous variables simultaneously and incorporate latent variables into the model to bridge the gaps between them. In past research, SEM has been used to describe the index, quality, or level of service that is difficult to measure directly.

Kim et al. used SEM to test the number of crashes influenced by accessibility, which is not directly measured but implied via the latent variable defined as a factor influenced by road length, bus route length, number of intersections, and number of dead ends in a given area (*3*). They found that accessibility had the opposite effect on crash severity, in that increased accessibility reduces crash severity (*3*). Khattak and Targa applied SEM to examine the risk factors affecting severity of injuries caused by large trucks by introducing the latent variable truck rollover in single-vehicle crashes (*24*). In that case, truck rollover was not directly observed, but they suspected strong interrelationships between the occurrence of a rollover, injury severity, and other factors. The results showed that some factors, such as dangerous truck-driving behaviors, speeding, and reckless driving, can definitely cause severe injuries to occupants by increasing the probability of truck rollover (*24*).

Lee et al. used SEM to investigate "traffic accident size," defined as the number of involved vehicles, the number of damaged vehicles, and the number of deaths, injuries, or both (*4*). The model suggested that road, driver, and environmental factors are strongly related to accident size and that roadway factors are significantly higher than driver and environmental factors. Schorr et al. used SEM to develop a collision propensity index for unsignalized intersections in California (*25*). They found that the three most populated counties in California (Los Angeles, Orange, and San Diego) yielded low roadway safety; this finding is of great concern because nearly half the state's population resides in these three counties.

SEM has also been used to predict people's demeanors, attitudes, behaviors, and opinions. Hassan and Abdel-Aty used SEM to quantify the effect of young drivers' behaviors, attitudes, and perceptions on crash involvement (*26*). The study identified that aggressive driving violations, in-vehicle distractions, and exceeding the speed limit significantly increase the probability of involvement in crashes. Hamdar et al. applied SEM to develop a quantitative aggressiveness propensity index for driver behavior at intersections (*27*). The index was treated as a latent variable and was intended to capture the propensity for aggressive driving at an intersection. The study demonstrated that drivers' tendencies for driving aggressively in an intersection can be influenced by the number of heavy vehicles, the number of pedestrians, traffic volume, and the like. Ambak et al. used SEM in Malaysia to predict a motorcyclist's intention to use a helmet properly (*28*). The study illustrated that an increase in a motorcyclist's positive attitude, subjective norms, and perceived behavioral control can increase that person's intention of properly using a helmet (*28*). Farag et al. examined the relationship between e-shopping and in-store shopping by means of an SEM model (*29*). The latent variables were positive in-store shopping attitude, positive e-shopping attitude, Internet experience, and the personal characteristic of being adventurous. The results showed that searching online positively affects the number of shopping trips and the frequency of shopping online. The researchers also found that time constraints had an indirect positive effect on online shopping and that searching online had an indirect negative effect on shopping duration.

Although all these SEM studies included latent variables and therefore succeeded in explaining the complexity of relationships between variables, none of those studies discussed the effect of the number of latent variables in relation to their meaning. The underlying relationships may involve more than one latent variable because the latent variables cannot be directly observed or measured. The meaning and purpose of introducing a latent variable hinges on the effectiveness of measurement of the dependent variable.

STATISTICAL METHODOLOGY

An SEM model tests and estimates the complicated relationships between variables through a combination of statistical methods and qualitative causal assumptions (*30*). Two types of variables are used in such a model: observed ones, which are directly collected or measured, and latent ones, which are not directly observed or measured (*31*). A structural equation model includes any combination of three types of statistical analysis methods: path analysis, confirmatory factor analysis, and hybrid. A path analysis models the directed dependencies among variables, equivalently to any form of multiple regression analysis, factor analysis, discriminant analysis, and the like. Confirmatory factor analysis is often used to test whether a hypothesized relationship structured between observed and latent variables is consistent. The hybrid model combines the path analysis model and the confirmatory factor analysis model (*31–33*).

Elements of SEM

In an SEM model, a variable can be both a dependent variable and an independent variable simultaneously. SEM can effectively distinguish direct, indirect, and total effects between variables through three major components, as shown in Figure 1 (*4*):

- A measurement model for the independent variable or exogenous variable (x-measurement model),
- A measurement model for the dependent variable or endogenous variables (y-measurement model), and
- A structural model between latent endogenous and exogenous variables.

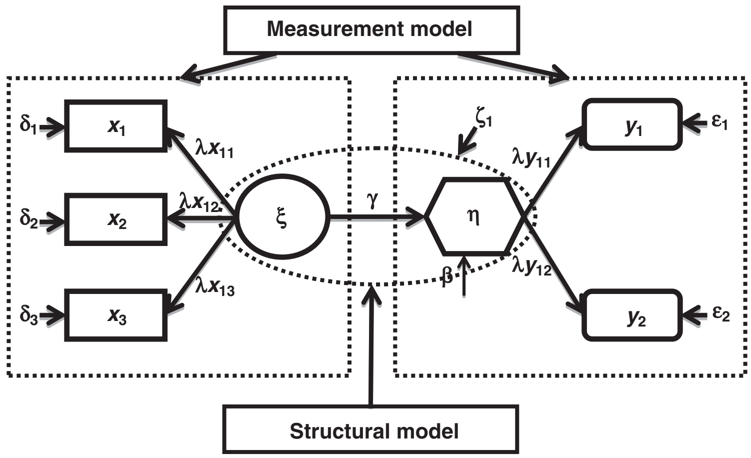

FIGURE 1 Example of SEM (see Table 1 for variable definitions).

In both measurement models, the contribution of latent variables to observed variables can be measured; in the structural model, the hypothesized relationship between latent variables can be built on the basis of theoretical or empirical knowledge.

Combining the measurement and the structural components, SEM articulates the regression effects of exogenous or independent variables on endogenous or dependent variables, as well as the effects between endogenous variables, also called "autocorrelation."

The measurement model can be expressed as a matrix format (*4*):

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_y & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_x \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{bmatrix} \tag{1}$$

The structural model can be expressed as

$$\boldsymbol{\eta} = \boldsymbol{B\eta} + \boldsymbol{\Gamma\xi} + \boldsymbol{\zeta} \tag{2}$$

where the variables of SEM from Equations 1 and 2 are defined in Table 1 (*4*).

Model Development and Goodness of Fit

Coefficients of an SEM model are estimated by covariance analysis. The differences between observations and predictions are minimized through iterations. In this study, STATA, Version 12, was used to develop the model, and the asymptotically distribution-free method was used to estimate the coefficients to account for the violation of the multivariate normality assumption (*34, 35*).

SEM relies heavily on empirical assumptions that should be satisfied to ensure model accuracy. The causal relationships between the multiple variables should be specified a priori. The goal of doing so is to determine, on the basis of model performance, whether a hypothesized model is consistent with the data collected (*35*). The root mean square error of approximation (RMSEA) is usually used to test a model's goodness of fit. A value of RMSEA less than .08 indicates goodness of fit in a model (*36*), and smaller is better. RMSEA is described in Equation 3 as a function of a chi-square value and the degrees of freedom. Specifically, RMSEA measures the difference between observed and predicted values per degree of freedom (*36*).

$$\text{RMSEA} = \sqrt{\frac{\chi_M^2 - M}{M(N-1)}} \tag{3}$$

where

M = degrees of freedom,
χ_M = chi-square test of model, and
N = sample size.

TABLE 1 Elements of SEM

Model	Variable	Description
Measurement	x	$q \times 1$ column vector of observed exogenous variables
	y	$p \times 1$ column vector of observed endogenous variables
	ξ	$n \times 1$ column vector of latent exogenous variables
	η	$m \times 1$ column vector of latent endogenous variables
	δ	$q \times 1$ column vector of measurement error terms for observed variables x
	ε	$p \times 1$ column vector of measurement error terms for observed variables y
	Λ_x	The matrix ($q \times n$) of structural coefficients for latent exogenous variables to their observed indicator variables
	Λ_y	The matrix ($p \times m$) of structural coefficients for latent endogenous variables to their observed indicator variables
Structural	Γ	The matrix ($m \times n$) of regression effects for exogenous latent variables to endogenous latent variables
	B	The coefficient matrix ($m \times m$) of direct effects between endogenous latent variables
	ζ	$m \times 1$ column vector of error terms

NOTE: The βs (components of B matrix) and the γs (components of Γ matrix) are magnitudes of expected changes after a unit increases in η or ξ. Similarly, λs (components of Λ matrix) are expected changes of observed variables with respect to a unit change in the latent variable.

Another criterion provided by STATA to evaluate the model's goodness of fit is the Akaike information criterion (AIC) formulated in Equation 4.

$$\mathrm{AIC} = 2k - 2\ln(L) \tag{4}$$

where k is the number of parameters in the model and L is the maximum likelihood value for the estimated model. A model with a smaller AIC value performs better.

Model Hypothesis

The hypothesized relationships between latent variables can be materialized in the structural component of an SEM. The latent variable is unobserved and is used to connect two observed variables if an indirect relationship between them is suspected. The nature and number of latent variables may affect the goodness of fit as well as the goodness of logic of the models. Three latent variables are proposed here: two are causal factors to the level of injury severity, and one is the indexing variable for injuries. The two causal factors are vehicular travel speed (speed) and force of collision (force). Generally, speed and force are believed to contribute directly to injuries sustained by accident victims. The indexing latent variable is crash severity (crashsvr), which represents the overall severity of a crash, as it accounts for both the severity of personal injury and vehicle damage.

An SEM model was developed for each latent variable. The first model contains speed and force, both of which contribute to the third latent variable, crashsvr. The second model includes both speed and force without crashsvr. The third model includes only speed as a latent variable. All three latent variables are associated with respective observed variables. For example, a driver's choice of speed may be affected by weather, light, pavement conditions, or a combination of those factors. Speed is also restricted by the roadway geometries and affected by a driver's characteristics, such as age and gender. According to Newton's second law, the force acting upon an object equals the product of the acceleration rate and mass. Therefore, the latent variable force can be associated with vehicle type (a proxy for vehicle weight), the roadside object being struck, and speed. All the speculated relationships between variables are presented later in a directed acyclic graph in the section on results.

DATA COLLECTION AND ANALYSIS

Single-vehicle crashes were used in this study because of their simplicity compared with multivehicle crashes. A single-vehicle crash is defined as a collision between a vehicle and one or more animals, pedestrians, bikes, or fixed obstacles (*37*). In 2008 and 2009, 2,286 single-vehicle crashes occurred in Wisconsin, accounting for 16.2% of all crashes. When the KABCO scale from the Model Minimum Uniform Crash Criteria was applied to all those single-vehicle crashes (*38*), 1,216 (53.2%) crashes were O (no apparent injury); 919 (40.2%) crashes were either Type B (suspected minor injury) or C (injury possible); and 151 (6.6%) were either Type K (fatal injury) or A (suspected serious injury). For vehicle damage, based on the Wisconsin Motor Vehicle Accident Report Form (*39*), 706 (30.9%) were none (no damage) or minor (cosmetic damage); 683 (29.9%) were moderate (broken or missing parts); 755 (33.0%) were severe (salvageable or total loss); the rest (6.2%) were missing values.

Crash data elements were classified into five categories: driver characteristics, highway characteristics, environmental factors, vehicle types, and types of struck objects. Human factors include driver's age and gender. Highway characteristics include highway geometry, which is based on an officer's opinion. Environmental factors include weather, light, and roadway surface condition. Vehicle types are ranked from low to high on the basis of vehicle weight. Object types are the kind of obstacle struck by a vehicle, including a pedestrian, a bicyclist, an animal, or a fixed object. The descriptions of the variables are listed in Table 2.

RESULTS

Directed acyclic graphs (Figures 1 through 5 here) illustrate the direct and indirect relationships between observed and latent variables, with the coefficients next to the arrow. To interpret a parameter, a positive value means a positive influence on the resulting outcome, and a negative value means a negative influence on the resulting outcome. The magnitude of impact is measured by the value.

Path Analysis Model

To serve as a benchmark, path analysis was performed only where observed variables were included. The authors assumed that both injury severity and vehicle damage were directly influenced by human factors, environmental factors, highway characteristics, the object that the vehicle struck, and vehicle types. The results with an acceptable goodness of fit measured by RMSEA (= .065) are illustrated in Figure 2. Only 11 variables were statistically significant. The others were not sufficiently effective to quantify either the injury severity or vehicle damage. The reason for this situation may be that some variables affect crash outcomes in an indirect fashion. Unobserved factors cannot be identified by using the path analysis model but can be effective models in SEM. The rest of this section presents the SEM models with one, two, and three latent variables in an attempt to unravel the complex relationships between the observed factors.

SEM with Three Latent Variables

In the SEM model with three latent variables (Figure 3), speed was set as a latent exogenous variable that explains the travel speed of the involved vehicle. Force and crashsvr were treated as latent endogenous variables that, respectively, represent the kinetic energy and the severity index, which is categorized by injury severity and vehicle damage.

As Figure 3 shows, high speed positively influences the force of a collision when a crash occurs. Both high speed and strong crash collision force could positively influence the overall severity of a crash, which can be observed as an increased probability of injury severity and vehicle damage.

The indirect impact between variables can be explained by the multiplicative relationship. For example, both horizontal and vertical alignment may decrease injury severity and vehicle damage through reduced vehicle speed. Similarly, compared with normal circumstances, inclement weather, poor pavement surface conditions, and poor lighting conditions decrease both injury severity and vehicle damage through reduced vehicle speed. Apparently, this finding would not be available if speed had not been introduced as the latent

TABLE 2 Description of Selected Variables

Variable	Description, Variable Type	Frequency	Percentage
Driver Characteristics			
Age	Driver age, ordinal	na	na
1	Young (<25)	1,010	44.2
2	Middle (25–55)	1,037	45.4
3	Old (>55)	238	10.4
Gender	Driver gender, binary	na	na
1	Female	747	32.7
2	Male	1,539	67.3
Highway Characteristics			
Roadhor	Horizontal curve, dummy	531	23.2
Roadvert	Vertical curve, dummy	502	22.0
Environmental Factors			
Wthrcond	Weather condition, categorical	na	na
Clear	Clear	1,753	76.7
Windy	Windy	6	0.2
Rain	Rain	203	8.9
Snow	Snow	292	12.8
Sleet	Sleet or hail	19	0.8
Fog	Fog	13	0.6
Lgtcond	Lighting condition, categorical	na	na
Day	Day	1,144	50.0
Without	Night without street light	825	36.1
Light	Night with street light	317	13.9
Roadcond	Road surface condition, categorical	na	na
Dry	Dry	1,414	61.9
Wet	Wet	344	15.0
Snow	Snow or slush	380	16.6
Ice	Ice	148	6.5
Vehicle Types			
Vehtype	Vehicle type, ordinal	na	na
1	Passenger car	1,802	85.1
2	Light truck	240	11.3
3	Heavy truck	75	3.6
Objects			
Objhit	Object type vehicle hit on, categorical	na	na
Other	Other objects (bike or pedestrian or animal)	1,621	71.0
Pole	Pole (traffic sign or utility pole	122	5.3
Guardrail	Guardrail	102	4.5
Medbar	Median barrier	115	5.0
Tree	Tree	272	11.9
Ditch	Ditch	37	1.6
Bridge	Bridge (parapet or pier or rail)	17	0.7
Injury Severity			
Injsvr	Level of injury severity, ordinal	na	na
1	O	1,216	53.2
2	E + C	919	40.2
3	K + A	151	6.6
Vehicle Damage			
Vehdmg	Level of vehicle damage, ordinal	na	na
1	None or minor	706	30.9
2	Moderate	683	29.9
3	Severe	755	33.0

variable. On the contrary, to conclude that inclement weather, poor pavement conditions, and poor lighting conditions reduce injury severities may be fallacious when only the direct effects are present. Similarly, for vehicle type, heavy vehicles can decrease crash severity through decreasing travel speed but increase crash severity through increasing the force of a collision.

SEM with Two Latent Variables

In the SEM model with two latent variables, speed and force were treated as latent endogenous and exogenous variables, respectively. As Figure 4 shows, vehicle speed and collision force have a positive impact on injury severity and vehicle damage. Except for object type,

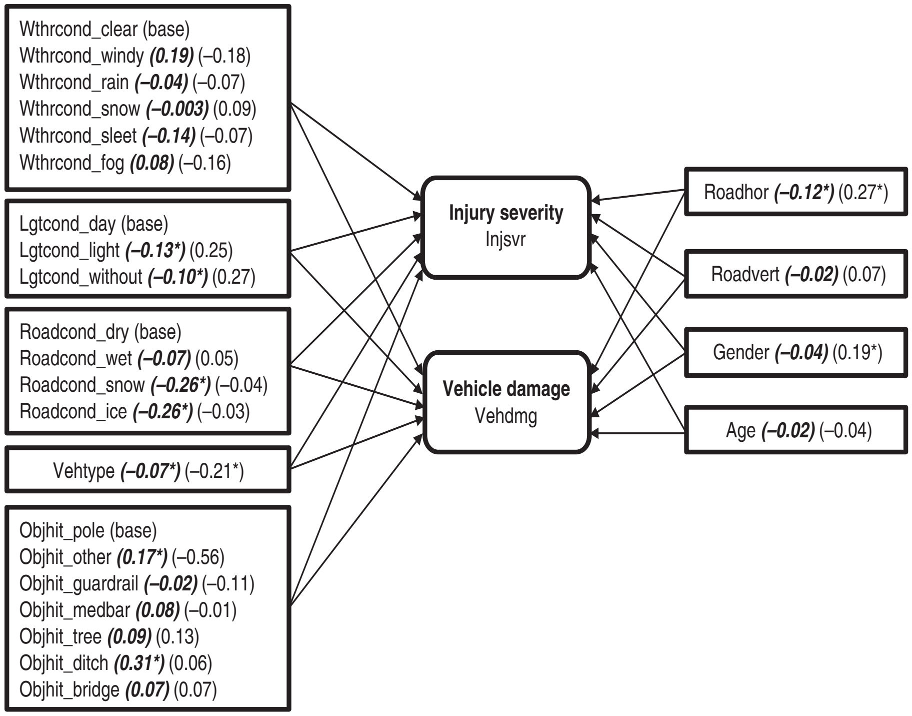

FIGURE 2 Path analysis model [*significant at 5% level; log likelihood (LL) = −10,055.3; RMSEA = 0.065; AIC = 20,178.6; boldface italic = coefficient related to injury severity; other coefficients = related to vehicle damage].

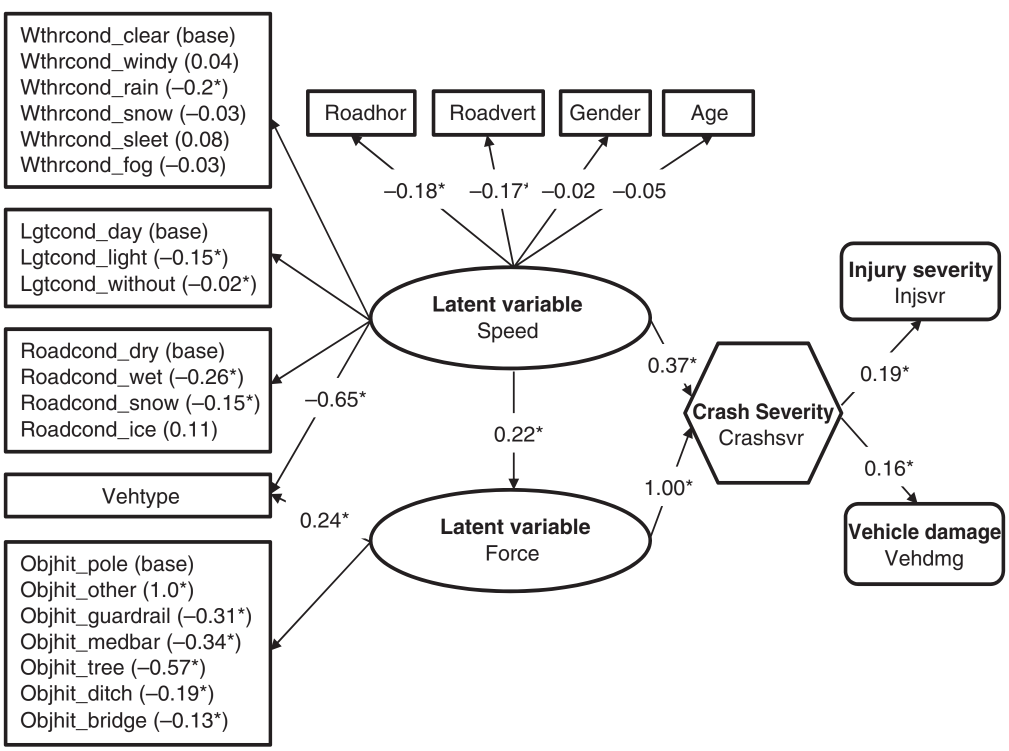

FIGURE 3 SEM model with three latent variables (*significant at 5% level; LL = −6,180.0; RMSEA = 0.1; AIC = 12,475.9).

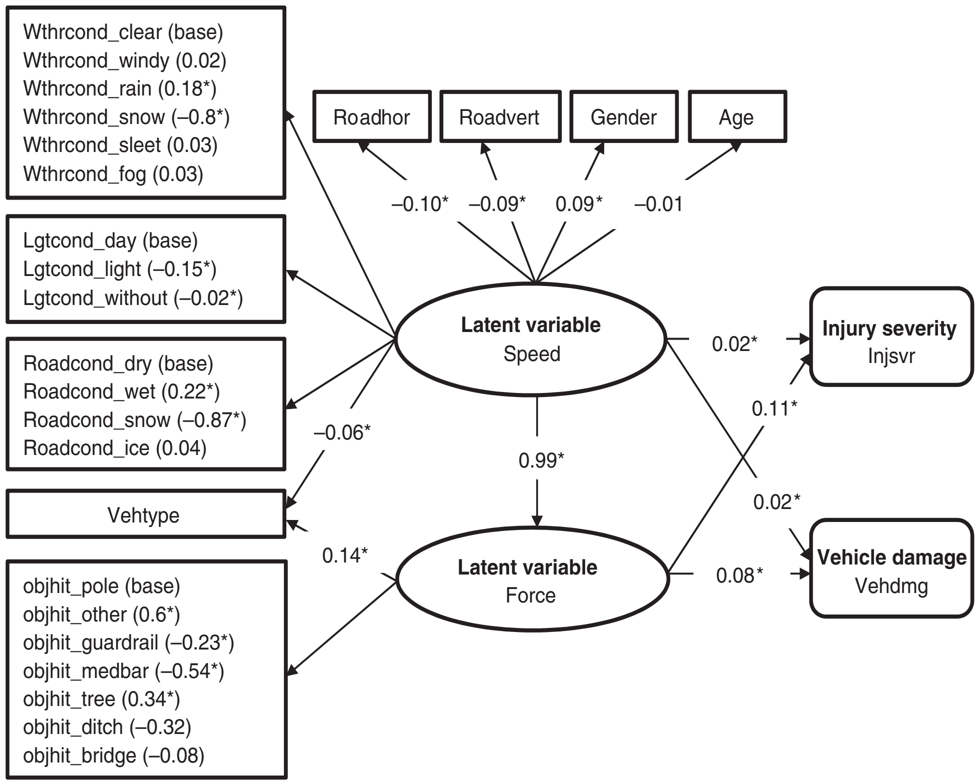

FIGURE 4 SEM model with two latent variables (*significant at 5% level; LL = −4,567.7; RMSEA = 0.08; AIC = 9,269.5).

the interrelationships between the observed independent variables, latent variables, and dependent variables are consistent with the SEM model that has three latent variables. The results of the SEM model with three latent variables show that hitting a tree decreases collision force compared with hitting a utility pole. But in the SEM model with two latent variables, hitting a tree increases the collision force compared with hitting a utility pole. The finding of the SEM model with two latent variables is logically sound and consistent with the findings in a previous study conducted by Qin et al., in which the findings were related to the use of a breakaway design for roadside utility poles (*21*). Comparing the results of the first two SEM models, the one with two latent variables offers more reasonable explanations than the one with three latent variables.

SEM with One Latent Variable

In the SEM model with one latent variable, speed was treated as a latent exogenous variable because the magnitude of a collision force may be subject to variations such as the weight of the subject, the deformation of the vehicle compartment, the vehicle's safety design, or speed. Moreover, the collision force is proportional to the speed. Hence, speed as the sole latent variable is used in an attempt to explain the intriguing relationships between observed variables. Figure 5 presents some results similar to those of other SEM models in that speed increases the possibility of more severe injury and vehicle damage; horizontal and vertical curves reduce speed because of limited slight distance; snowy weather, snowy pavement, and poor lighting conditions reduce speed; and heavy vehicles travel at a lower speed. But some other statistically significant variables are puzzling: rainy weather and wet and icy pavements increase speed. These counterintuitive coefficients are cautions that the selection of the number of latent variables as well as definition of the relationship between observed and latent variables may affect the regression results. In this model, the object type was not included because no direct relationship exists between vehicle speed and the object that the vehicle is striking.

MODEL COMPARISON

Despite many similarities between the results from the three SEM models, each model presents a unique perspective and understanding of the way that the factors interact and the ways that they affect injury severity and vehicle damage. Some factors have consistent impacts across the three SEM models. The SEM models with both one and three latent variables have some counterintuitive coefficients that are quite difficult to explain. The SEM model with the two latent variables speed and force seems to have the most meaningful results.

According to the statistical goodness of fit, the path analysis model has the lowest RMSEA but the highest AIC value. The SEM with two latent variables outperforms others, as the model has the lowest RMSEA and AIC values. Results of model comparison are shown in Table 3.

From the information available, the SEM with two latent variables is recommended as the best model for predicting crash severity because it achieves acceptable RMSEA and the lowest AIC value, with most variables being statistically significant at the 5% level.

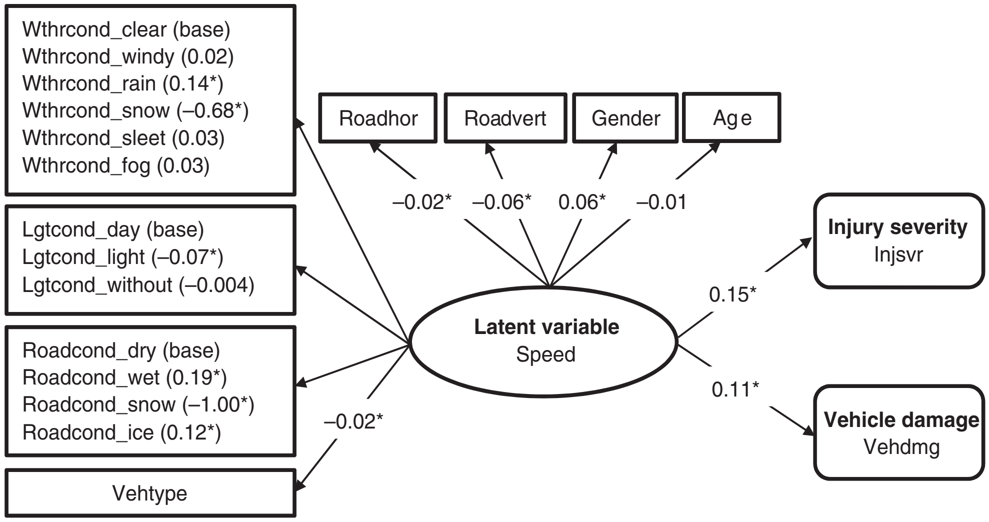

FIGURE 5 SEM model with one latent variable (*significant at 5% level; LL = −8,184.9; RMSEA = 0.1; AIC = 16,465.7).

CONCLUSIONS AND RECOMMENDATIONS

Traffic safety is an important issue affecting the service capability and efficiency of the nation's highways, and reducing injuries and deaths has always been the first priority among researchers and agencies. Among all highway crash types, the single-vehicle crash is one that occurs most frequently, even for states and areas with low traffic volume. To reduce the severity of single-vehicle crashes, the causes of crash severity must be explored and relevant countermeasures implemented to mitigate crash risk.

Crash severity issues have been extensively studied over the past several decades, and a variety of statistical methodologies (e.g., logistic and probit models) have been used to identify the relationships between independent factors and crash severities. Although both logistic and probit models can directly explain the impacts of indicator factors on crash severities, they might fail to unravel the complex interrelationships between the variables by overlooking unobserved factors. SEM is appealing in that the hypothesized causal relationships can be constructed through latent variables in a structural model. The relationship between latent and observed variables can also be conveniently established by using the measurement model. SEM combines both the structural and measurement models in one modeling process. Moreover, in an SEM model, a variable can be both dependent and independent simultaneously. Therefore, SEM can effectively distinguish direct, indirect, and synergic effects between variables and thus more accurately capture the underlying relationships between factors.

TABLE 3 Model Performances

Model	RMSEA	AIC	Number of Variables
SEM			
Three latent variables	0.1	12,475.9	20
Two latent variables	0.08	9,269.5	20
One latent variable	0.1	16,465.7	12
Path analysis	0.065	20,178.6	11

Determining the number of latent variables and defining their relationships with the observable variables are critical steps in an SEM model. In this study, SEM models with one, two, and three latent variables were developed and compared. The comparison shows that the SEM model with two latent variables (speed and force) had the best statistical goodness of fit and the most statistically significant variables at a significance level of 5%.

The SEM results revealed that vehicle speed can positively influence collision force, and both vehicle speed and collision force can significantly increase injury severity and vehicle damage. Males are more likely to drive faster than females, and older drivers tend to drive slower than younger drivers, although this variable is not significant in any of the three at the 5% level. When compared with normal roadway conditions, adverse surface and lighting characteristics decrease both injury severity and vehicle damage because vehicle speed is reduced. In addition, the crash severity of heavy vehicles may be decreased because of their slower traveling speed, but it can also be increased because of vehicle weight. The authors anticipate that the results of this study can unravel complex relationships between injury severity, vehicle damage, and contributing factors via different SEMs and offer additional insights about the model choices for safety analysis.

Irrespective of the consistent impacts of the factors among the three models, selection of the latent variables was subjectively hypothesized by theory and empirical research. To validate the hypothesis of the model, future work on the collection of data related to latent variables is recommended. Furthermore, the underreporting issue could significantly affect the model because the assumption about a random sample of crash data is biased without consideration of the underreporting issue related to crash severity level (*7*). Future research is recommended to estimate the magnitude of the underreporting issue and to address the impact of underreporting on model accuracy.

ACKNOWLEDGMENT

The authors thank the University of Wisconsin–Madison Traffic Operations and Safety Laboratory (TOPS) for providing the data.

REFERENCES

1. *Toward Zero Deaths: A National Strategy on Highway Safety.* U.S. Department of Transportation and AASHTO, 2009. http://safety.fhwa.dot.gov/tzd/docs/tzd_summary_v3.pdf.
2. *Fatality Analysis Reporting System Annual.* National Center for Statistic and Analysis, NHTSA, 2012. http://www.census.gov/compendia/statab/2012/tables/12s1105.pdf.
3. Kim, K, P. Pant, and E. Yamashita. Measuring Influence of Accessibility on Accident Severity with Structural Equation Modeling. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2236,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 1–10.
4. Lee, J., J. Chung, and B. Son. Analysis of Traffic Accident Size for Korean Highway Using Structural Equation Models. *Accident Analysis and Prevention,* Vol. 40, 2008, pp. 1955–1963.
5. Qin, X., K. Wang, and C.E. Cutler. Logistic Regression Models of the Safety of Large Trucks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2392,* Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 1–10.
6. Dissanayake, S. Comparison of Severity Affecting Factors Between Young and Older Drivers Involved in Single Vehicle Crashes. *International Association of Traffic and Safety Sciences,* Vol. 28, 2004, pp. 48–54.
7. Ye, F., and D. Lord. Investigation of Effects of Underreporting Crash Data on Three Commonly Used Traffic Crash Severity Models: Multinomial Logit, Ordered Probit, and Mixed Logit. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2241,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 51–58.
8. Ghulam, H., J. Bhanu, and M. Uday. Multinomial Logistic Regression Model for Single-Vehicle and Multivehicle Collisions on Urban U.S. Highways in Arkansas. *Journal of Transportation Engineering,* Vol. 138. No. 6, 2012, pp. 786–797.
9. Zajac, S.S., and J.N. Ivan. Factors Influencing Injury Severity of Motor Vehicle–Crossing Pedestrian Crashes in Rural Connecticut. *Accident Analysis and Prevention,* Vol. 35, No. 3, 2003, pp. 369–379.
10. Khattak, A.J., P. Kantor, and F.M. Council. Role of Adverse Weather in Key Crash Types on Limited-Access Roadways: Implications for Advanced Weather Systems. In *Transportation Research Record 1621,* TRB, National Research Council, Washington, D.C., 1998, pp. 10–19.
11. Kockelman, K.M., and Y.J. Kweon. Driver Injury Severity: An Application of Ordered Probit Models. *Accident Analysis and Prevention,* Vol. 34, No. 3, 2002, pp. 313–321.
12. Abdel-Aty, M., and J. Keller. Exploring the Overall and Specific Crash Severity Levels at Signalized Intersections. *Accident Analysis and Prevention,* Vol. 37, 2005, pp. 417–425.
13. Christoforou, Z., S. Cohen, and G. Karlaftis. Vehicle Occupant Injury Severity on Highways: An Empirical Investigation. *Accident Analysis and Prevention,* Vol. 42, No. 6, 2010, pp. 1606–1620.
14. O'Donnell, C.J., and D.H. Connor. Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice. *Accident Analysis and Prevention,* Vol. 28, No. 6, 1996, pp. 739–753.
15. Malyshkina, N., and F. Mannering. Markov Switching Multinomial Logit Model: An Application to Accident-Injury Severities. *Accident Analysis and Prevention,* Vol. 41, No. 4, 2009, pp. 829–838.
16. Chen, F., and S. Chen. Injury Severities of Truck Drivers in Single- and Multi-Vehicle Accidents on Rural Highways. *Accident Analysis and Prevention,* Vol. 43, No. 5, 2011, pp. 1677–1688.
17. Moore, D.N., W. Schneider, P.T. Savolainen, and M. Farzaneh. Mixed Logit Analysis of Bicycle Injury Severity Resulting from Motor Vehicle Crashes at Intersection and Non-Intersection Locations. *Accident Analysis and Prevention,* Vol. 43. No. 3, 2011, pp. 621–630.
18. Milton, J.C., V.N. Shankar, and F.L. Mannering. Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. *Accident Analysis and Prevention,* Vol. 40, No. 1, 2008, pp. 260–266.
19. Kim, J.K., G. Ulfarsson, V. Shankar, and F. Mannering. A Note on Modeling Pedestrian Injury Severity in Motor Vehicle Crashes with the Mixed Logit Model. *Accident Analysis and Prevention,* Vol. 40, No. 5, 2010, pp. 1695–1702.
20. Savolainen, P.T., F.L. Mannering, D. Lord, and M.A. Quddus. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis and Prevention,* Vol. 43, No. 5, 2011, pp. 1666–1676.
21. Qin, X., K. Wang, and C.E. Cutler. Analysis of Crash Severity Based on Vehicle Damage and Occupant Injuries. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2386,* Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 95–102.
22. Huang, H., H.C. Chin, and M.M. Haque. Severity of Driver Injury and Vehicle Damage in Traffic Crashes at Intersections: A Bayesian Hierarchical Analysis. *Accident Analysis and Prevention,* Vol. 40, 2008, pp. 45–54.
23. Quddus, M., R.B. Noland, and H.C. Chin. An Analysis of Motorcycle Injury and Vehicle Damage Severity Using Ordered Probit Models. *Journal of Safety Research,* Vol. 33, 2002, pp. 445–462.
24. Khattak, A.J., and F. Targa. Injury Severity and Total Harm in Truck-Involved Work Zone Crashes. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1877,* Transportation Research Board of the National Academics, Washington, D.C., 2004, pp. 106–116.
25. Schorr, J., S. Hamdar, and T. Vassallo. Collision Propensity Index for Unsignalized Intersections: Structural Equation Modeling Approach. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.
26. Hassan, H., and M. Abdel-Aty. Exploring the Safety Implications of Young Drivers' Behavior, Attitudes, and Perceptions. *Accident Analysis and Prevention,* Vol. 50, 2012, pp. 361–370.
27. Hamdar, S., H. Mahmassani, and R. Chen. Aggressiveness Propensity Index for Driving Behavior at Signalized Intersections. *Accident Analysis and Prevention,* Vol. 40, 2008, pp 315–326.
28. Ambak, K., R. Ismail, R. Abdullah, and M. Borhan. Prediction of Helmet Use Among Malaysian Motorcyclists Using Structural Equation Modeling. *Australian Journal of Basic and Applied Sciences,* Vol. 4, No. 10, 2010, pp. 5263–5270.
29. Farag, S., T. Schwanen, M. Dijst, and J. Faber. Shopping Online and–or In-Store? A Structural Equation Model of the Relationship Between e-Shopping and In-Store Shopping, *Transportation Research Part A,* Vol. 41, No. 2, 2007, No. 2, pp. 125–141.
30. Wright, S. Correlation and Causation. *Journal of Agricultural Research,* Vol. 20, 1921, pp. 557–585.
31. Sato, T., and M. Akamatsu. Modeling and Prediction of Driver Preparations for Making a Right Turn Based on Vehicle Velocity and Traffic Conditions While Approaching an Intersection. *Transportation Research Part F,* Vol. 11, No. 4, 2008, pp. 242–258.
32. Bollen, K.A., and J.S. Long (eds.). *Testing Structural Equation Models,* Vol. 154. Sage Publications, Newbury Park, Calif., 1993.
33. Kline, R.B. Principles and Practice of Structural Equation Modeling, 2nd ed. Guilford Press, New York, 2005.
34. *Structural Equation Modeling,* Rel. 12. Stata Press, College Station, Tex., 2011.
35. Lei, W., and Q. Wu. An NCME Instructional Module on Introduction to Structural Equation Modeling: Issues and Practical Considerations. *Educational Measurement: Issues and Practice,* Fall 2007, pp. 33–43.
36. Browne, W., and R. Cudeck. Alternative Ways of Assessing Model Fit. In *Testing Structural Equation Models* (K.A. Bollen and J.S. Long, eds.), Sage Publications, Newbury Park, Calif., 1993, pp. 445–455.
37. *Safe Car Guide.* http://www.safecarguide.com/exp/statistics/statistics.htm. Accessed April 4, 2009.
38. *Model Minimum Uniform Crash Criteria,* 4th ed. U.S. Department of Transportation, June 2012. www.mmucc.us.
39. *Law Enforcement Officer's Instruction Manual for Completing the Wisconsin Motor Vehicle Accident Report Form (MV 4000).* Division of Motor Vehicles, Wisconsin Department of Transportation, Madison, 1998.

The Safety Data, Analysis, and Evaluation Committee peer-reviewed this paper.