SEVERITY ANALYSIS OF CRASHES USING STRUCTURAL EQUATION MODELING

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Civil, Geological, and Environmental Engineering

University of Saskatchewan

Saskatoon

By

Iman Gharraie

PERMISSION TO USE

Head of the Department of Civil, Geological and Environmental Engineering

University of Saskatchewan

57 Campus Drive, Engineering Building

Saskatoon, Saskatchewan S7N 5A9, Canada

OR

Dean

College of Graduate and Postdoctoral Studies

University of Saskatchewan

116 Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan S7N 5C9, Canada

ABSTRACT

Population growth, increased travel demand and, consequently, increased motor vehicle use has led to concerns about road safety in today's society. In transportation engineering, road safety levels are measured through frequency and severity of motor vehicle crashes. Crash data has been used in road safety modeling to analyze factors that may reduce crash frequency and severity.

Regarding crash severity analysis, modeling techniques have mainly attempted to incorporate road and traffic factors into a statistical model, building a direct relationship between independent and dependent (crash severity) variables. However, some explanatory variables can affect crash severity indirectly through one or more mediating variable. Moreover, while traditional techniques have only included measured variables, there might also be unobserved factors not included in the observed data affecting crash severity. Therefore, this thesis is aimed at investigating both observed and unobserved factors that influence the severity of crashes, directly and indirectly, using a statistical technique known as structural equation modeling (SEM). Two types of crashes that affect road safety in urban and rural areas were investigated in this thesis: red-light running related (RLR) crashes and wildlife-vehicle crashes (WVC), respectively. An SEM model was developed for each crash type.

In effect, three unobserved variables were hypothesized for RLR crashes: pre-crash travel speed (TS) of the bullet vehicle (at fault), the kinetic energy (KEs) applied from the bullet vehicle to the subject vehicle(s), and crash severity. Similarly, three latent variables were introduced for WVCs: driver's speeding attitude (SA), driver's visibility impairment (VI), and crash severity. The results show that crash data supports the main hypothesis, with measured/latent variables adequately predicting crash severity. Regarding the RLR data, results show that both TS and KE_s positively influence the overall crash severity, and that TS increase could positively affect KE_s. Regarding the WVC data, the results showed that both SA and VI positively influenced overall crash severity, and that higher VI would negatively affect SA, which would indirectly decrease crash severity. Overall, these findings could help transportation practitioners to prioritize strategies and countermeasures aimed at reducing crash severity outcomes at urban and rural road sites.

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor, Dr. Emanuele Sacchi, who has inspired my future ambitions and whose meticulous approach, expertise, and understanding, added immensely to my graduate experience. I appreciate his support throughout my Master's degree program, without which this thesis would not have been possible. The door to Dr. Sacchi's office was always open to me whenever I ran into a problem or had questions about my research, and he steered me in the right direction whenever he thought I needed it. I would also like to thank my committee members, Dr. Haithem Soliman, and Dr. James Nolan, for their support and their time to read my work and make any necessary corrections.

I would like to express my appreciation to my family for their support and encouragement during my studies. Without their continuous support, it would have been impossible for me to finish my work.

DEDICATION

This thesis is dedicated to the family I was born into, and the family I have gained along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAAM	American Association for Automotive Medicine
AASHTO	American Association of State Highway and Transportation Officials
ADF	Asymptotically Distribution-Free
AGFI	Adjusted Goodness-of-Fit Index
AIC	Akaike Information Criterion
AIS	Abbreviated Injury Scale
CAR	Crash Analysis Reporting
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
FARS	Fatality Analysis Reporting System
FHWA	Federal Highway Administration
HCM	Highway Capacity Manual
ISS	Injury Severity Score
KE_s	Kinetic Energy Transferred from the Bullet Vehicle to the Subject Vehicle
NHTSA	National Highway Traffic Safety Administration
NSC	National Safety Council
PDO	Property Damage Only
RLR	Red Light Running
RMSEA	Root Mean Square Error of Approximation
SA	Speeding Attitude
SEM	Structural Equation Modeling
SGI	Saskatchewan Government Insurance
SPF	Safety Performance Function
SRMR	Standardized Root Mean Square Residual
TAIS	Traffic Accident Information System
TLI	Tucker Lewis Index
TS	Pre-crash Travel Speed
US	United States

VI	Visibility Impairment
WLS	Weight-Least Squares
WLSMV	Weighted Least Squares Estimator Using a Diagonal Weight Matrix
WVC	Wildlife-Vehicle Crashes

CHAPTER 1

INTRODUCTION

1.1 Motor Vehicle Crashes and Their Severity

Population growth, increase of travel demand and, consequently, increase of motor vehicle use has led to concerns about road safety in today's society (AASHTO 2010). In transportation engineering, the level of road safety is generally measured in terms of the number of motor vehicle crashes that occur at a road site over a certain period of time (e.g., crashes per year). A crash is an event where one or more vehicles collide, resulting in property damage, injury, or fatality. Fatal crashes are those events that result in at least one death. Crashes that result in injuries, but no deaths, are classified as personal injury. Crashes that result in neither death nor injuries but involve damage to property are classified as property damage only (PDO) (Garber and Hoel 2014). Therefore, it is important to both investigate the frequency of motor vehicle crashes as well as their severity to evaluate road safety.

Crash frequency can be defined as the number of crashes at a road site, facility or a roadway network, under a given set of geometric design and traffic volume characteristics over a specific period of time (Lord and Mannering 2010) and crash severity as the level of injury or property damage of a crash occurrence (AASHTO 2010). Due to the enormous economic and emotional costs to society resulting from motor-vehicle crashes, researchers have been investigating methods and techniques to gain a better understanding of factors affecting them. The ultimate goal in road safety is, therefore, to improve the prediction of crash frequency and crash severity, as well as providing guidance for countermeasure and policy implementation (Lord and Mannering 2010).

While the causes of crashes are usually complex and involve several factors, they can be divided into four separate groups: factors related to driver's behavior, factors that describe the mechanical condition of the vehicle, factors related to geometric characteristics of the roadway, and factors related to the physical or climatic environment in which the crash occurred (Ogden 1996).

Therefore, it is essential to analyze different data sources to determine probable crash-related factors and develop countermeasures that will reduce the rate and severity of future crashes. These countermeasures are usually categorized into three groups: engineering countermeasures, enforcement countermeasures and driver's education countermeasures.

The focus of transportation practitioners concerned with road safety is mainly to select, design and implement engineering countermeasures, focused at reducing crash frequency and crash severity due to road and infrastructure factors. It is, therefore, essential to understand that reducing crash frequency and severity may require different approaches. For example, alternating the curvature rates, super-elevating the outer edge of a curve, providing advanced warning of an unexpected change in horizontal curves, and widening the roadways are conventional methods that reduce the frequency of crashes by lowering the tendency of vehicles to overturn and to skid laterally. Meanwhile, median barriers and roadway signs are designed to reduce crash severity if such objects are hit during a crash. Regarding intersections, modifying geometric design features (e.g. designing left turn lanes at intersections), improving sight distance (e.g. by change horizontal and/or vertical alignment of approaches to provide more sight distance), or converting the intersection to a modern roundabout can decrease both the frequency and the severity of a crash event (Garber and Hoel 2014, AASHTO 2010, Savolainen et al. 2011).

Speed limit reduction in certain areas (e.g., school zone) by using different traffic control methods (e.g., speed humps, implementing road narrowing measures, signs, flashing beacons, speed monitoring display) can reduce both severity and frequency of crashes. In addition to roadway aspects, advances in vehicle design also have the potential to reduce crash severity (e.g., electronic stability control, anti-lock brakes, safety belts or airbags). Furthermore, enforcement countermeasures (e.g., driver-training programs), which are intended to encourage drivers to adhere to the traffic laws through the threat of citation and possible fine may have potential in reducing both frequency and severity (Chen et al. 2013, Bonneson et al. 2002).

As mentioned before, it is crucial to analyze crash and traffic data in order to identify patterns and develop strategies and countermeasures that may help reducing injury and fatality rates of future crashes. For this reason, it is necessary to gain a full understanding of factors that influence the likelihood of a crash (crash frequency) or the characteristics that can mitigate crash severity faced by road users involved in a crash. Employing statistical models have always been the most reliable method in identifying and analyzing the contributions of human, environmental, roadway, and vehicle factors on crash severity (Kim et al. 2011). Unfortunately, detailed driving data (e.g., steering information, acceleration, and braking and driver response) and crash characteristics data (i.e., what might be available from vehicle black-boxes) that would help identifying cause and effect relationships concerning crashes are typically not available.

Available data on crash events are usually obtained from state and local transportation and police agencies. All relevant information is usually recorded by police officers on a crash report form. The type of form used differs from country to country; however, a typical form can include information on the location, time of the occurrence, roadway and environmental conditions, types and number of vehicles involved, a sketch showing the original paths of the maneuver or maneuvers of the vehicles involved, and the severity of the crash event (fatal, injury, or property damage only) (Garber and Hoel 2014). Using this data, researchers have developed their analytic approaches to study factors affecting the likelihood and severity of crashes over a specified period. Such approaches handle temporal and spatial elements related to crashes and ensures that explanatory variables are employed in the statistical analysis (Lord and Mannering 2010).

In order to examine the relationship between crash frequency and explanatory variables, a wide variety of statistical methods have been used. For example, predictive models known as safety performance functions (SPFs) have been employed to estimate the expected average crash frequency of a network, facility, or individual site. The predictive model is applied for a given period of time, traffic volume, and specific geometric design characteristics of the roadway. The estimated parameters rely upon regression models developed from observed crash data for several similar sites. The models developed can be applied to existing road sites, alternative designs to existing conditions (e.g., proposed upgrades or treatments), or proposed new roadways (AASHTO 2010).

Other examples of crash-based regressions are severity models, which are used to explore the relationship between crash severity outcomes and their contributing factors. Typically in these type of models, the variables related to the crash event (e.g., injury severity level, crash cost, damage level) are treated as the dependent variables and other variables related to human, environmental, roadway, and vehicle characteristics are treated as independent variables. There are several different types of statistical methods (e.g., linear regressions, multinomial logit, and ordered probit models) used according to the nature of data and the type of analysis.

Regarding crash severity models, most of the traditional statistical methodologies used in the literature are subject to limitations and shortcomings, especially because a direct relationship between dependent and independent variables is inferred and only measured variables can be employed. Therefore, in this thesis, the focus is to investigate the measured and unmeasured factors that influence the severity of crashes, directly and indirectly, using structural equation modeling (SEM). SEM is a well-known statistical analysis method, especially in the field of phycology and sociology that can overcome the limitations of more traditional methods mentioned before. By way of example, two different types of crashes were selected in this thesis that are considered to be of great importance for road safety in the urban area and in the rural area: red-light running related (RLR) crashes and wildlife-vehicle crashes (WVC), respectively. Both types of crashes have been causing considerable economic loss and emotional burden to individuals, their families, and nations as a whole (NHTSA 2012, 2015, Huijser et al. 2008, Transport Canada 2019). Studying the severity of these crash types using SEM can provide traffic engineers with important information on road safety for policymaking purposes and for designing strategies and countermeasures.

1.2 Crash Severity Models

Numerous studies have investigated the severity of crashes on road networks (Savolainen and Ghosh 2008). These researches have used a variety of methods to study the factors that influence crash severity outcomes. There are several characteristics related to the data that are crucial considerations in the development and application of an appropriate statistical method to study crash severity data. The statistical methods used by researchers have mainly relied on the nature of the dependent variable; usually, dependent variables to study crash severity are represented by discrete categories (Abdel-Aty and Keller 2005, Savolainen and Ghosh 2008, Ye and Lord 2014). Hence, discrete response models in traffic safety (often referred to as crash severity models), such as logit and probit models, are usually employed to explore the relationship between crash severity outcomes and their contributing factors such as driver characteristics, vehicle characteristics, roadway conditions, and road-environment factors (Ye and Lord 2014). Moreover, when unobserved heterogeneity is accounted for in the data set, the mixed logit model has been investigated as a good alternative. This modeling technique has, in fact, the ability to treat coefficients as random variables (Lee et al. 2018). A more detailed explanation of these models will be discussed in section CHAPTER 22.2.1.

The explanatory variables (independent variables) used in these studies can relate to human factors, road and environmental characteristics, vehicle characteristics, and specific crash information. Human factors may relate to demographics, behavior, occupant position in the vehicle, and other human characteristics. The road and environmental characteristics could relate to factors like road, weather, traffic, and trip characteristics. The vehicle characteristics could include vehicle type, safety features, size, mass, and age. The crash information relates to factors like crash type, speed, angle of the crash, and impact characteristics (Sobhani et al. 2011).

1.3 Problem Statement

While traditional modeling techniques have mainly attempted to incorporate road and traffic factors into a statistical model and build a direct relationship between independent and dependent variables (Lee et al. 2008), some explanatory variables can affect crash severity indirectly through one or more mediating variable, which may make the investigation of the relationship among explanatory variables a complex and challenging task. Moreover, while traditional techniques only analyze the observed (measured) variables, there might be unobserved factors affecting crash severity, and some latent dimensions in the data cannot be explained through observed variables.

Crashes are, in fact, multi-causal phenomena (Ogden 1996), and thus, the interaction among observed variables and the fact that some important unobserved factors can affect crash severity, should be investigated (Savolainen et al. 2011). Statistical methods that do not take into account for direct and indirect interactions among variables are likely to result in biased parameter estimates (Huang et al. 2008). In addition, if such interactions are ignored, parameters will be estimated with less precision, and there will be a loss of significance, thus making statistically defensible inferences more difficult (Anselin et al. 2013).

Therefore, it is important to explore alternative modeling techniques that can potentially unravel the complex relationship between crash severity and their contributing factors, handle indirect effects, and take into account latent (unobserved) dimensions in the modeling effort.

1.4 Research Statement

To overcome the shortcomings of traditional statistical methods, SEM is employed in this research. SEM has the advantage of representing, estimating and testing complex modeling structures, where dependent variables can be predictor variables of other dependent variables (allowing to examine indirect effects and mediation structures). With SEM, it is also possible to include both measured and latent variables (variables that are not directly observable) in a model, whereas traditional techniques can only analyze measured variables (Lee et al. 2008, Wang and Qin 2014).

Traditional statistical methods for data analysis, usually specify default models, assume that measurement occurs without error, and are somewhat inflexible. However, SEM is a statistical technique that requires specification of a model based on theory and explicitly specifies measurement error. By using SEM, it is possible to hypothesize and test the latent dimensions that can affect the severity outcome of crashes. In addition, SEM can provide the weight of variable estimates in the form of standardized results, which can be used for prioritizing the development of specific countermeasures for the type of crash under investigation. Also, by using a model diagram in SEM, it is possible to represent the complex relationships among variables in a more convenient and clear way (Suhr 2006).

For these reasons, crash severity models will be developed using SEM for two selected crash typologies (i.e., RLR-related crashes and WVCs). Latent dimensions in the data will be hypothesized and tested. Possible correlation among selected variables will be considered. Furthermore, the effect of both latent and observed variables on crash severity will be analyzed as well as direct, indirect, and total effects. Also, a comparison of the developed SEM models with traditional methods will be presented at the end of the thesis. Since SEM provides weights for each contributing factor of crash severity, the results of the study can be employed by transportation engineers and decision-makers for safety improvement and policy-making purposes to prioritize strategies and countermeasures at crash-prone locations.

1.5 Thesis Outline

Chapter 2 provides a literature review of different statistical methods used for crash severity analysis, the use of SEM in crash severity models, and existing studies on RLR-related crashes and WVCs (i.e., the two crash types analyzed in this research).

Chapter 3 provides insights into the preliminary stages of SEM analysis. Moreover, the statistical and conceptual foundations of a SEM model, interpretation of results, and goodness-of-fit criteria will be discussed.

Chapter 4 develops a crash severity model for RLR-related crashes, from data analysis to model hypothesis in order to analyze the full SEM model; a comparison with path analysis is also provided.

Chapter 5 develops a crash severity model for WVCs, from data analysis to model hypothesis, in order to analyze the full SEM model; a comparison with path analysis is also provided.

Chapter 6 compares and discusses the results from the studies found in the literature with the results of this study (SEM models for RLR-related crashes and WVCs).

Finally, Chapter 7 reports the conclusions derived from the study, limitations, and directions for future work in this field.

1.6 Publications

Research conducted in Chapter 4 generated a manuscript entitled "Severity analysis of red-light-running-related crashes using structural equation modeling" by Khaled Shaaban, Iman Gharraie, Emanuele Sacchi, and Inhi Kim, published in the Journal of Transportation Safety & Security in 2019, with DOI: 10.1080/19439962.2019.1629137.

Authors' contributions are as follows:

Study conception and design: Khaled Shaaban and Emanuele Sacchi; data collection: Khaled Shaaban, Inhi Kim and Iman Gharraie; analysis and interpretation of results: Iman Gharraie and Emanuele Sacchi; draft manuscript preparation: Iman Gharraie, Emanuele Sacchi, Khaled Shaaban, Inhi Kim.

Research conducted in Chapter 5 generated a manuscript entitled "Severity Analysis of Wildlife-Vehicle Crashes Using Generalized Structural Equation Modeling" by Iman Gharraie and Emanuele Sacchi, presented at the 2020 Annual Meeting of the Transportation Research Board.

Authors' contributions are as follows: Study conception and design: Emanuele Sacchi; data preparation: Iman Gharraie; analysis and interpretation of results: Iman Gharraie, Emanuele Sacchi; draft manuscript preparation: Iman Gharraie, Emanuele Sacchi.

CHAPTER 2

LITERATURE REVIEW

2.1 Crash Types

Crash severity modeling is usually linked to the type of crash being analyzed. It is, therefore, important to conduct a review of the particular crash type under investigation before any modeling effort is considered. There are several different crash types and they are usually grouped depending on their characteristics. For example, crashes can differ according to the location (e.g., intersection, highway), or the environmental setting (e.g., urban or rural). Moreover, crashes can be categorized according to the point of impact (e.g., head-on, rear-end, angle/side-impact), or according to the type of collision, such as collisions with fixed objects (e.g. tree, utility pole), collisions with people, or collisions with other non-fixed objects (e.g. pedestrian, bicycle, animal, motor vehicles). There are also non-collision types of crashes (e.g., run-off-the-road, jackknife). The severity of each crash type is influenced by factors related to human, environmental, roadway, and vehicle characteristics (e.g., impaired driving, speeding, restricted sight distance, slippery surface, inadequate roadway lighting) (Garber and Hoel 2014, AASHTO 2010, Ulfarsson et al. 2006).

In this thesis, two different crash types in the urban and rural setting are investigated: RLR-related crashes and WVCs, respectively. In this section a thorough review of the existing state of knowledge regarding these two crash types and their corresponding studies conducted in the road safety field will be discussed.

2.1.1 Red-Light Running-Related (RLR) Crashes

Intersections are the road sites where the majority of crashes occurs. This is because the different approach and crossing movements by motorists, bicyclists and pedestrians make at-grade intersections one of the most complex traffic situations that people encounter; moreover, their presence is much more frequent in the urban environment. In the United States (US), annually, more than 45 percent of all crashes occur at intersections (NHTSA 2015). About 2.3 million crashes occur annually at intersections in the US according to the National Highway Traffic Safety Administration, 2,850 of which were fatal and 680,000 were crashes, which caused injuries. In more details, statistics indicate that a large number of crashes occurred at signalized intersections due to traffic violations, such as running red lights (NHTSA 2012). Among many different types of crashes, running a red light is the most frequent cause of crashes in urban areas. RLR-related crashes occur when a vehicle enters an intersection any time after the signal light has turned red and conflicts with one or more vehicles with the right of way (i.e., green traffic light) (Tay and De Barros 2009). RLR-related crashes are the leading cause of injuries for road users (Retting et al. 1995).

For these reasons, the analysis of RLR-related crashes has been the focus of many highway practitioners, traffic management operators, decision-makers, and academics. Several crash models have been developed to predict the number of potential crashes and conflicts or near-misses. In early studies, the prediction models were developed based on several individual contributions such as traffic flow, the existence of signals, the number of pedestrians, and signal timing. The models have been advanced to take crash severity into account by counting the number of deaths and or injuries (Roess et al. 2004).

In the mid-1990s, Retting et al. (1999) collected 1,373 RLR crashes to classify characteristics of drivers crossing signalized intersections in Arlington, Virginia (US). The study found that most red-light runners were young drivers who wore no seat belts, had poor driving performance records, and drove over-age vehicles. Moreover, these young drivers were likely to be involved in speed violations in the past; it was also found that running traffic lights was the single most common type of crash, accounting for 22% of urban crashes and 27% of all injury crashes. This same study found that injuries were more likely in crashes involving red-light running than in other crash types; injuries were reported in 45% of RLR crashes compared with 30% for other crashes.

In 1999, Retting et al. (1999) conducted another study on fatal crashes caused by RLR drivers. RLR crashes were twice as likely to occur on urban roads compared to other fatal crashes (86% vs. 42%). RLR crashes were more likely than other fatal crashes to occur during the day (57% vs. 48%). The study also showed that 74% of RLR drivers were male. In terms of age, drivers younger than age 30 years violated the red light, about 43% of the time. It was also found that red-light runners were expected to have had traffic violation records in the past, including RLR, alcohol-impaired driving conditions, and speeding. In another research, Bonneson et al. (2002) recognized that RLR is mostly affected by signal timings, signal cycle length, or the frequency of yellow-signal presentation. Furthermore, geometric variables such as approach grade, approach width, and intersection size, may also contribute to RLR statistics (FHWA 2009).

To further investigate the traffic parameters that might influence RLR crashes, Elmitiny et al. (2014) conducted a study, which concluded that traffic volume, speed, green time phase, and traffic composition of heavy vehicles were major elements affecting RLR. It was also shown that vehicular size and the proportion of heavy vehicles in traffic were positively associated with an increase in RLR-related crashes. In a study done by Chen et al. (2017) it was concluded that increasing approaching speed, smaller headway and gap to the preceding vehicle on the adjacent lane, cycle longer length, wider exits and even snowing weather corresponds to higher RLR frequencies.

These models provide a good indication in general in order to prioritize areas that need to ensure a high level of safety. Overall several factors, such as road conditions, driver characteristics, and vehicle factors appeared to be related to RLR crashes. However, the statistical methods employed to verify these relationships had significant limitations in terms of possible inference. Despite different assumptions and model specifications, these traditional statistical methodologies incorporated only observed (available) variables into the model and built a direct relationship between independent and dependent variables related to crash severity. This limitation might be considered a drawback as some of the explanatory variables may influence crash severity indirectly through one or more mediating variables. In addition, the interactions between variables were not always taken into account. Hence, investigating the relationship between dependent, independent, and mediating variables can become a complex and challenging task (Lee et al. 2008).

It is also worth mentioning that there is a lack of studies addressing the severity pattern of RLR crashes in terms of traffic-related factors (e.g., speed level, peak/off-peak traffic conditions) that may interact in a more sophisticated way during a crash event. For this reason, it is important to explore alternative modeling techniques that could unravel the complex relationship between severity of RLR crashes and their contributing factors.

2.1.2 Wildlife-Vehicle Crashes (WVCs)

Wildlife-vehicle crashes (WVCs) are amongst the leading concerns for road safety in the rural areas of North America (Huijser et al. 2008). In addition to bringing into danger animal welfare (Sainsbury et al. 1995), WVCs pose direct risks to the life of humans (especially WVCs with large animals) and can cause significant property damage. Moose and deer, for instance, are one of the largest animals struck by vehicles in North America, causing significant damage and injuries. In the United States, 278,000 WVCs were reported by the National Highway Traffic Safety Administration (NHTSA) in 2015. Also, the NHTSA Fatality Analysis Reporting System (FARS) recorded approximately 180 fatal and 13,000 injury crashes involving animals in the same year (NHTSA 2015). The total annual cost associated with this type of crashes was estimated to be around US$8.4 billion (Huijser et al. 2008).

In Canada, over 30,000 WVCs have been reported annually resulting $200 million in costs each year (Transport Canada 2019). Also, a 9% increase in their trend was observed from 1996 to 2000. Most often, these crashes involved large animals such as moose, elk, deer, and bison (Transport Canada 2019). In Saskatchewan (Canada), presence and movement of animals (mainly moose and deer) was reported as a significant contributing factor to crashes in the province, especially on rural highways. In 2017, 64% of crashes on Saskatchewan provincial rural highways involved animals (Saskatchewan Government Insurance (SGI) 2018). According to the Saskatchewan Government Insurance, 44,464 WVCs were reported from 2010 to 2017, 5% of which resulted in either injuries or fatalities (Saskatchewan Government Insurance (SGI) 2018). These statistics necessitate further investigation on these type of crashes. A thorough understanding of the factors that influence these collisions are critical to mitigate the significant negative consequences of these crashes, which include significant socio-economic, traffic safety, and environmental costs.

Several studies suggested the absence of a unique solution to reduce WVC frequency and the importance of combining strategies that focus either on wildlife and drivers (Glista et al. 2009). Regarding wildlife, WVCs can be reduced by influencing animal behavior with measures such as animal fencing, roadside boulders and crossing structures (Huijser et al. 2008, Clevenger et al. 2002, Romin and Bissonette 1996). Alternatively, driving behavior can be modified by providing drivers with information about the presence of animals on the road or by implementing public education programs. Typical measures are simple warning signage, real-time animal reporting with mobile apps and more complex animal warning and detection systems (Romin and Bissonette 1996, Huijser et al. 2008, Mukherjee et al. 2013). Along with WVC frequency analysis and mitigation, it is also important to investigate factors that influence the severity outcomes of WVCs when the interaction with wildlife has already occurred. However, the literature is lacking concerning severity analysis studies.

Different research studies have been conducted to investigate the circumstances of WVCs. Haikonen et al. (2001) demonstrated that WVCs are more likely to happen in the early morning hours and particularly at dusk when animals are actively moving around and likely to cross the road in rural areas. It was also concluded that another critical time of the day is after sunset and drivers can effectively reduce their risk of WVCs by reducing speed and remaining alert for animal intrusions on the roadway during those times. In another study, Conn et al. (2004) showed that the highest WVCs injury rate occurred among people aged 15 to 24, where nearly 50% of them were driving the vehicle. The WVC injury rate was similar for males and females. Approximately half of these crashes involved a direct crash with the animal, and the remainder happened when the driver tried to avoid hitting the animal. The incidents in which the animal was avoided, the crash most commonly involved a vehicle leaving the roadway and hitting another object.

Langley et al. (2006) researched fatal animal-vehicle collisions over 10 years (1995-2004) and concluded that the majority of fatal crashes happened in rural areas, during the fall months, in clear weather, on straight roads, and an increasing trend for deaths was noted over this period. They also suggested that fencing appearing to be the most effective method to reduce the collisions. The use of safety equipment such as seat belts in vehicles and helmets for motorcycle riders may decrease fatalities during a collision. Rowden et al. (2008) studied road crashes in Australia and highlighted that night-time travel was found to be a significant risk factor when comparing WVCs to severe other injury crashes. There were also a significantly higher proportion of motorcyclists involved in severe crashes. It was also mentioned that there is an elevated crash count in higher speed zones. Gkritza et al. (2010a) studied the deer-animal crashes and demonstrated that crashes on dry road conditions, principal arterials as well as on roads with a posted speed limit over 55 miles per hour were more likely to be an injury. Crashes on roads with a gravel right shoulder and higher traffic volume were more likely to result in no injury. Crashes in zones with a more significant percentage of cropland was more likely to be an injury, while crashes in zones with a higher percentage of roads were more likely to be no injury.

A study by Sullivan (2011) showed that the relative occurrence risk of fatal and nonfatal WVCs in darkness is influenced by the posted speed limit, suggesting that a driver's limited forward vision at night plays a role in WVCs. In another research about animal-vehicle interactions Lao et al. (2011) found that speed limit, rural area type, and presence of species habitat have an increasing effect on animal-vehicle collision risk, whereas median width, sex of animal, high truck percentage, and high number of lanes put a decreasing effect on WVC probability.

Vanlaar et al. (2012) showed that crashes with animals are more likely to occur in the dusk and at night than during daylight hours. Inclement weather and reduced visibility are other possible factors leading to a higher frequency of WVCs. In addition, the study showed that these crashes are more recurrent in fall when animals migrate for reasons such as scarcity of food due to crop harvest and evasive action during hunting season. Regarding road infrastructure characteristics, the study concluded that posted speed limits and the number of lanes are other possible contributing factors of WVCs. Higher speeds can decrease the reaction time of drivers and provide animals with less time to avoid oncoming traffic. Wider roads would require longer crossing time to animals, thus increasing the likelihood of a crash. Visintin et al. (2018) showed that high-speed vehicles are becoming significantly problematic in WVCs regardless of the species trait. Moreover, the morning peak hour showed the highest risk of collisions, and the lowest collision risk occurred around noon. Also, it was concluded that smaller and slower-moving species are more vulnerable to these kinds of crashes.

Overall, studies show that the frequency of WVCs is higher in rural roads and higher speeds configurations. Dark periods of the day and fall months also have a higher risk of WVCs. while the past researches are mostly investigating the factors related to the frequency of WVCs, there are limited studies that investigate the severity of WVCs. For this reason, it is essential to investigate and unravel the complex relationship between severity of WVCs and their contributing factors.

2.2 Statistical Methods in Crash Severity Analysis

Traditionally, statistical methodologies have been employed to analyse and identify the contributing factors of crash severity (Savolainen et al. 2011). The explanatory variables (independent variables) used in these studies were usually related to human factors, road and environmental characteristics, vehicle characteristics, and specific crash information. Human factors may relate to demographics, behavior, occupant position in the vehicle, and anthropometric characteristics. The road and environmental characteristics could relate to factors like road, weather, traffic, and trip characteristics. The vehicle characteristics could include vehicle type, safety features, size, mass, and age. The crash information relates to factors like crash type, speed, angle of the crash, and impact characteristics (Sobhani et al. 2011).

In this section, traditional statistical methods used to investigate the severity of crashes will be discussed. Moreover, a review of SEM and the state of knowledge regarding its use in crash severity analysis will be analyzed.

2.2.1 Traditional Crash Severity Analysis Methods

A variety of techniques has been used in the literature to analyze crash severity outcomes. Identifying methodological limitations and characteristics of crash-related data are crucial considerations in the development and application of an appropriate statistical method to study crash severity. The statistical methods used by researchers have mainly relied on the nature of the dependent variable. Usually, dependent variables for crash severity are represented by discrete categories which could be either a binary response outcome or a multiple response outcome. An example of a binary response outcome could be injury versus non-injury crashes or fatal versus non-fatal crashes (Savolainen et al. 2011).

For multiple response outcome, different indicators of severity have been used. Regarding injury severity, different categories have been suggested. KABCO scale was proposed by the National Safety Council (NSC) in the USA and is typically used by law enforcement for classifying injuries where fatal injury or killed (K), incapacitating injury (A), non-incapacitating (B), possible injury (C), and property damage only (O) are the categories (Savolainen et al. 2011). Abbreviated Injury

Scale (AIS) which was developed by the American Association for Automotive Medicine (AAAM), the Organ Injury Scales (OIS) developed by the American Association for the Surgery of Trauma, and the Injury Severity Score (ISS) used by hospitals, are other types of injury severity scales (Savolainen et al. 2011).

Other studies used vehicle damage to examine crash severity. Vehicle damage can be divided into different categories such as slight or no damage, extensive damage, and total wreck (Quddus et al. 2002). A 2005 FHWA study, also provided crash cost estimates for several combinations of KABCO injury severities (Council et al. 2005). Moreover, some studies used the number of injured people in a crash as the dependent variable for crash severity.

These categories for dependent variables are typically ordinal by nature. Regarding injury severity, the levels increase from no injury to possible injury, to visible injury, to disabling injury, to fatality. Considering the ordinal nature of these data is essential in selecting an appropriate methodological approach (Savolainen et al. 2011). Furthermore, crash severity related variables might be closely correlated in some cases (for example PDO severity level may be correlated with zero number of injured people); thus, there may be shared unobserved effects among adjacent severity categories. Failing to account for such correlations can result in incorrect inferences and biased parameter estimates for certain types of model estimation methods (Savolainen et al. 2011).

To deal with the discrete and ordinal nature of crash severity outcomes, discrete response models in traffic safety (usually referred to as crash severity models), such as probit and logit models, are typically employed to explore the relationship between crash severity and its contributing factors such as roadway conditions, vehicle characteristics, driver characteristics, and environment factors (Ye and Lord 2014). Overall, based on the existing literature, the most common models are multinomial logit, nested logit, ordered logit/probit, binary logit/probit, and ordered mixed logit models (Ye and Lord 2014). Table 2.1 shows a summary listing of studies that used these models for crash severity analyses (Savolainen et al. 2011).

Table 2.1 Summary of Previous Research Investigating Crash Severity Modeling

Methodology	Previous Research	Explanation
Binary logit and binary probit	(Shibata and Fukuda 1994) (Farmer et al. 1997) (Khattak et al. 1998) (Krull et al. 2000) (Zhang et al. 2000) (Al-Ghamdi 2002) (Bedard et al. 2002) (Toy and Hammitt 2003) (Ballesteros et al. 2004) (Chang and Yeh 2006) (Sze and Wong 2007) (Lee and Abdel-Aty 2008) (Pai and Saleh 2008) (Rifaat and Tay 2009) (Liu and Dissanayake 2009) (Nevarez et al. 2009) (Haleem and Abdel-Aty 2010) (Peek-Asa et al. 2010) (Tarko et al. 2010) (Zhu et al. 2010) (Kononen et al. 2011) (Moudon et al. 2011) (Hu and Donnell 2011) (Haleem and Gan 2011) (Yuan et al. 2017)	These models are used to estimate probabilities for binary data or discrete ordinal data where there are two possible outcomes (e.g., fatal crashes and non-fatal crashes)
Mixed logit/Mixed generalized ordered logit/Mixed joint binary logit-ordered logit	(Eluru and Bhat 2007) (Eluru et al. 2008) (Moore et al. 2011) (Zhu and Srinivasan 2011b) (Shaheed et al. 2013) (Haleem and Gan 2013) (Kim et al. 2013) (Wu et al. 2014) (Cerwick et al. 2014) (Ye and Lord 2014) (Wu et al. 2016b) (Behnood and Mannering 2016) (Uddin and Huynh 2017) (Anderson and Hernandez 2017) (Li et al. 2019)	These models are used to estimate probabilities for discrete data without accounting for the ordering of the outcomes. The mixed logit model can capture heterogeneity through the use of random parameters. It also allows explanatory variables to affect the mean of the distribution of the random parameters.
Multinomial logit	(Shankar and Mannering 1996) (Carson and Mannering 2001) (Abdel-Aty and Abdelwahab 2004) (Ulfarsson and Mannering 2004) (Khorashadi et al. 2005) (Islam and Mannering 2006) (Kim et al. 2007) (Malyshkina and Mannering 2008) (Savolainen and Ghosh 2008) (Schneider et al. 2009) (Angel and Hickman 2009) (Malyshkina and Mannering 2010) (Gkritza et al. 2010b) (Rifaat et al. 2011) (Schneider and Savolainen 2011) (Ye and Lord 2011) (Tay et al. 2011) (Hu and Donnell 2011) (Hu and Donnell 2011) (Ye and Lord 2014) (Zhao and Khattak 2015) (Naik et al. 2016) (Wu et al. 2016a) (Amoh-Gyimah et al. 2017) (Chen and Fan 2019)	These models are used to estimate probabilities for discrete data with three or more outcomes. They do not account for the ordering of the severity outcomes.
Nested logit	(Shankar et al. 1996) (Chang and Mannering 1998) (Chang and Mannering 1999) (Lee and Mannering 2002) (Abdel-Aty and Abdelwahab 2004) (Holdridge et al. 2005) (Savolainen and Mannering 2007) (Haleem and Abdel-Aty 2010) (Hu and Donnell 2010) (Patil et al. 2012) (Wu et al. 2016b) (Islam et al. 2019)	These models are used to estimate probabilities for discrete data without accounting for the ordering of the outcomes. In the nested logit model, severity levels that share unobserved effects are grouped into conditional nests.

Ordered logit and ordered probit	(Khattak et al. 1998) (Klop and Khattak 1999) (Renski et al. 1999) (Khattak 2001) (Khattak et al. 2002) (Kockelman and Kweon 2002) (Quddus et al. 2002) (Abdel-Aty 2003) (Austin and Faigin 2003) (Kweon and Kockelman 2003) (Zajac and Ivan 2003) (Khattak and Rocha 2003) (Donnell and Mason Jr 2004) (Khattak and Targa 2004) (Abdel-Aty and Keller 2005) (Lee and Abdel-Aty 2005) (Shimamura et al. 2005) (Garder 2006) (Noyce et al. 2006) (Siddiqui et al. 2006) (Pai and Saleh 2007) (Das et al. 2008) (Gray et al. 2008) (Wang and Abdel-Aty 2008) (Yamamoto et al. 2008) (Pai and Saleh 2008) (Chimba and Sando 2009) (Quddus et al. 2009) (Wang et al. 2009) (Haleem and Abdel-Aty 2010) (Jung et al. 2010) (Tarko et al. 2010) (Ye and Lord 2011) (Zhu and Srinivasan 2011a) (Haleem and Gan 2011) (Hosseinpour et al. 2014) (Ye and Lord 2014) (Abegaz et al. 2014) (Naik et al. 2016) (Wang et al. 2018b) (Uddin and Huynh 2018) (Wang et al. 2018a)	These models are used to estimate probabilities for ordered discrete data where the ordering of the severity outcomes is accounted for.

Logit models, for instance, are used to estimate probabilities for binary data or discrete ordinal data. The probability of crash being more or less severe is represented as a function of highway-related variables of generalized linear type, typically a logistic function of a linear combination of these highway-related variables (Vogt and Bared 1998). The independent variables can either be continuous or categorical. The dependent (response) variable can only take the value of 0 or 1 (e.g., non-severe or severe crashes). Consider the following linear function T_{ki} to determine the severity outcome level k for crash i:

$$T_{ki} = \beta_k X_{ki} + \varepsilon_{ki} \tag{2.1}$$

where, β_k is a vector of the estimable parameters for crash severity category k; $k=0$, or 1; X_{ki} represents a vector of explanatory variables affecting the crash severity for i at severity category k (highway-related variables such as geometric variables, environmental conditions, driver characteristics, etc.); ε_{ki} is a random disturbance that account for unobserved effects; $i=1,\ldots,n$ where n is the total number of crash events included in the model. The logistic regression model to estimate the probability of crash i ending in crash severity category $k=1$, $P_i(k = 1)$, is shown in Eq. (2.2) (Hu and Donnell 2011).

$$P_i(k=1) = \frac{exp(\beta_k X_{ki})}{1+exp(\beta_k X_{ki})} \tag{2.2}$$

This functional form guarantees that $P_i(k)$ will always be a number between 0 and 1. The above form of $P_i(k)$ can be transformed into linear form which is called logit transformation (equivalent to T_{ki}) as shown below (Al-Ghamdi 2002).

$$L_{ki} = ln\frac{P_i(k)}{1-P_i(k)} = \beta_k X_{ki} + \varepsilon_{ki} \tag{2.3}$$

In case of a binary probit model, the assumption is that the observed dependent variable can be 1 if and only if its underlying continuous latent variable z takes on a positive value (Washington et al. 2003).

$$z = \beta X + \varepsilon \tag{2.4}$$

where X is a vector of variables specifying the discrete ordering for observation n, β is a vector of estimable parameters, and ε is a random disturbance. The probability of crash i, $P_i(k=1)$, ending in crash severity category $k=1$ is shown in Eq. (2.5).

$$P_i(k=1) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(\beta_1 X_{1i} - \beta_2 X_{2i})/\sigma} exp(-\frac{1}{2}\omega^2)d\omega \tag{2.5}$$

Where σ is a standard deviation used to rescale the normally distributed random variables into the standard normal distribution; and ω is a normally distributed variable (standardized).

Whenever ordering among several severity outcomes is not considered, a multinomial logit model can be used as alternative discrete outcome model (Savolainen et al. 2011). Similar to the binary logit model, in the multinomial logit model, the formulation specifies a linear function T_{ki} that determines the injury severity outcome k for observation i as shown in Eq. (2.1).

$$T_{ki} = \beta_k X_{ki} + \varepsilon_{ki} \tag{2.1}$$

In the multinomial logit, $k=1, \ldots, K$ (K is the highest crash severity ordered response) represent all the severity levels. Eq. (2.6) shows how to calculate the probability for each crash severity category. Let $P_i(k)$ be the probability of crash i ending in crash severity category k, such that

$$P_i(k) = \frac{\exp(\beta_k X_{ki})}{\sum_{\forall k} exp(\beta_k X_{ki})} \tag{2.6}$$

Multinomial logit model does not account for the ordering of severity outcomes, and is particularly susceptible to correlation of unobserved effects between injury-severity levels (Savolainen et al. 2011). To overcome the restriction of the multinomial logit model, the nested logit model might be an appropriate alternative. The nested logit model resolves this by grouping severity levels that share unobserved effects into conditional nests (Savolainen and Mannering 2007). The nested logit overcomes the limitations of multinomial logit models, allowing correlation among error terms across different severity levels. The nested logit can be written as:

$$P_i(k) = \frac{\exp(\beta_k X_{ki} + \phi_k LS_{ki})}{\sum_{\forall K} exp(\beta_K X_{Ki} + \phi_K LS_{Ki})} \tag{2.7}$$

$$P_i(j|k) = \frac{\exp(\beta_{j|k} X_i)}{\sum_{\forall J} \beta_{J|k} X_{Ji}} \tag{2.8}$$

$$LS_{ki} = Ln\left[\sum_{\forall J} exp(\beta_{j|k} X_{Ji})\right] \tag{2.9}$$

where $P_i(k)$ is the unconditional probability of crash i resulting in severity outcome k; X_{ki} are vectors of explanatory variables that determine the probability of severity category k; β are vectors of estimable parameters, $P_i(j|k)$ is the probability of crash i having injury outcome k conditional on the outcome being in outcome category k (e.g., in the nested model shown in Figure 2.1 the severity category k would be non-incapacitating injury and $P_i(j|k)$ would be the binary logit model of injury outcomes property damage only and possible injury); J is the conditional set of outcomes (conditioned on k); K is the unconditional set of outcome categories (the upper three branches of Figure 2.1); LS_{ki} is the inclusive value (logsum), and ϕ_k is an estimable parameter. Figure 2.1 shows an example of a nested model (Washington et al. 2003).

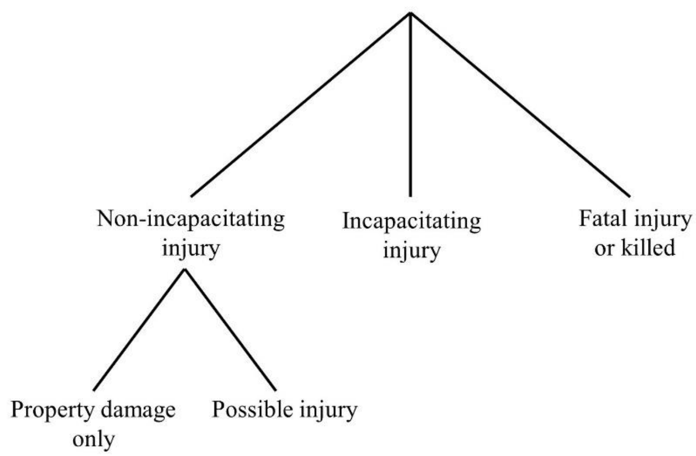


Figure 2.1 Example of Nested Model

For heterogeneous dataset, the mixed logit model might be a promising alternative as this modeling technique has the flexibility to treat coefficients as random or fixed variables (Lee et al. 2018). Mixed logit models (also commonly referred to as random parameters logit models) are a relatively recent development for the analysis of discrete data (Savolainen et al. 2011). The mixed logit model shares the same structure of linear severity function T_{ki}, utilized for the multinomial logit model, as shown in Eq. (2.1). Therefore, Eq. (2.10) shows the calculation of the probability of each crash severity category for the mixed logit model. Let $P_i(k)$ be the probability of crash *i* ending in crash severity category k, such that

$$P_i(k) = \int \frac{exp(\beta_k X_{ki})}{\sum_{\forall k} exp(\beta_k X_{ki})} f(\beta|\theta)\, d\beta \tag{2.10}$$

where $f(\beta|\theta)$ is the density function of β and θ is a vector of parameters which describe the density function (mean and variance).

In a research by Abdel-Aty (2003) three different crash severity models were compared (multinomial logit, mixed logit, and ordered probit) and the results recommended the ordered probit model over the multinomial logit models and mixed logit models. It was suggested that, although the outcome of some multinomial variables (e.g., injury severity) is discrete, these variables are inherently ordered; in these cases, the multinomial logit models would fail to account for the ordinal nature of the dependent variable. Therefore, for analyzing such responses, the ordered probit models have become the commonly used framework. Similar to the binary probit model, the ordered probit model is derived by defining an unobserved variable z, which is used to determine the ordinal ranking of crash severity data (Washington et al. 2003). This unobserved variable is typically specified as a linear function for each observation, such that

$$z = \beta X + \varepsilon \tag{2.4}$$

where X is a vector of variables specifying the discrete ordering for observation n, β is a vector of estimable parameters, and ε is a random disturbance. By applying this equation, observed ordinal data, y, for each observation are defined as

$$y = \begin{cases} 1, \text{if } z < \gamma_1 \\ k, \text{if } \gamma_{k-1} < z < \gamma_k \\ K, \text{if } z > \gamma_{K-1} \end{cases} \tag{2.11}$$

where $\gamma = \{\gamma_1, \dots, \gamma_k, \dots, \gamma_{k-1}\}$ are the threshold values for all the crash severity levels corresponding to integer ordering $k=1, \dots, K$ (K is the highest crash severity ordered response). For a given X value, the probability that the severity of an individual crash belongs to each category is

$$\begin{cases} P(y_i = 1) = \phi(-\beta X) \\ P(y_i = k) = \phi(\gamma_{k-1} - \beta X) - \phi(\gamma_{k-2} - \beta X) \\ P(y_i = K) = 1 - \phi(\gamma_{K-1} - \beta X) \end{cases} \tag{2.12}$$

where $\phi(\cdot)$ is the cumulative probability function of the standard normal distribution.

Some researchers prefer the nominal models compared to ordinal models due to the restriction placed on how variables affect the probabilities of ordered discrete outcome; that is using the same coefficient for a variable among different crash severity levels. Others still prefer ordinal models because of its overall performance and simplicity when less detailed data are available (Washington et al. 2003). While these statistical models have undoubtedly provided new insights, they have mainly attempted to incorporate road and traffic factors into a statistical model and build a direct relationship between independent and dependent variables (Lee et al. 2008).

Usually, researchers are provided with only a set of observed (measured) variables. And frequently, some of the explanatory variables may influence crash severity indirectly through one or more mediating variables; therefore, investigating the relationship among explanatory variables can become a complex and challenging task (Lee et al. 2008). Since crashes are multi-causal phenomena (Ogden 1996), the likelihood of correlation among these variables must be considered.

This correlation could be due to the fact crash-related variables are likely to share unobserved factors (factors that are not taken into consideration with available measured variables) (Savolainen et al. 2011).

Statistical methods that do not take the correlation among crash-related variables into account are likely to result in biased parameter estimates (Huang et al. 2008). Also, if these correlations are ignored, parameters will be estimated with less precision, and there will be a loss of efficiency, and thus making it more challenging to make statistically defensible inferences (Anselin et al. 2013). Therefore, there is a need to explore alternative modeling techniques, which can allow the unraveling of the complex relationship between crash severity and their contributing factors as a system, can handle the indirect effects, can take into account the latent (unobserved) dimensions in the data, and considers the correlation among variables.

2.2.2 Structural Equation Modeling (SEM)

Structural equation modeling (SEM) may be viewed as a general model of many commonly employed statistical models, such as analysis of variance and covariance, multiple regression, factor analysis, path analysis, econometric models of the simultaneous equation and non-recursive modeling, multilevel modeling, and latent growth curve modeling. Other names of SEM include covariance structural analysis, equation system analysis, and analysis of moment structures (Bowen and Guo 2012).

SEM is a statistical modeling technique that is widely used in the field of social science (Hoyle 1995, Heene et al. 2011, Maccallum and Austin 2000). For social work researchers, SEM may most often be used as an approach to data analysis that combines simultaneous regression equations and factor analysis (Cuttance and Ecob 1988). SEM has become more prevalent in investigating transportation engineering problems since1980, and the availability of improved software is rapidly accelerating its use. The number of published studies using this method has approximately doubled in the past years (Golob 2003).

SEM can account for 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). SEM allows simultaneous analysis of all dependent, independent, and mediating variables in the model in place of a separate analysis and includes any combination of three types of statistical analysis methods: path analysis (measured variables only), confirmatory factor analysis (measured and unobserved variables) and a combination of them (Kline 2005).

The general process of developing SEM is that a model structure (model hypothesis) needs to be designed based on previous theories or knowledge. The model hypothesis is usually based on latent endogenous and exogenous variables. Latent variables are the central concept in SEM. Latent constructs represent theoretical, abstract concepts or phenomena such as attitudes, behavior patterns, cognitions, social experiences, and emotions that cannot be observed or measured directly or with single items. In contrast, observed variables are variables that exist in a database or spreadsheet. They are variables whose raw scores for sample members can be seen, or observed, in a dataset. By using factor analysis, it is possible to test hypotheses about how well sets of observed variables in an existing dataset measure latent constructs (i.e., factors). Factor models are also called measurement models because they focus on how one or more latent constructs are measured, or represented, by a set of observed variables (Bowen and Guo 2012).

The distinction made between latent and observed variables represents a fundamental difference between SEM and conventional regression modeling. In the SEM framework, latent variables are of interest but cannot be directly measured. Observed variables are modeled as functions of model-specific latent constructs and latent measurement errors. SEM models are commonly presented in path diagrams. The path diagram is a summary of theoretically suggested relationships among latent variables and indicator (measured) variables, and directional (regression) and non-directional (i.e., correlational) relationships among latent variables. Importantly, correlated errors of measurement and prediction can also be modeled in SEM analyses (Bowen and Guo 2012). Usually, the measured variables are represented by rectangles, and latent variables are represented by circles.

It is critical not to omit essential variables and, afterward, a measurement model describing each latent dimension needs to be established. Ideally, SEM is conducted with large sample sizes and continuous variables with multivariate normality. The number of cases needed varies substantially based on the strength of the measurement and structural relationships being modeled and the complexity of the model being tested. The sample size should be large enough to allow statistical inference. Finally, a goodness-of-fit evaluation is performed to verify whether the model fits the data well. Typical measures are the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), Comparative Fit Index (CFI), Tucker Lewis Index (TLI), and Adjusted Goodness-of-Fit Index (AGFI) (Browne and Cudeck 1992). Once the overall goodness-of-fit is satisfactory, the significance of each model parameter is tested.

2.2.3 Severity Analysis Using SEM

As mentioned before, SEM has become a more common technique to investigate transportation and road safety problems. Lee et al. (2008) developed a SEM model to capture the complex relationship between crash severity and various factors related to road, environmental, and driver characteristics, for Korean highways. Road factors were measured through variables such as pavement category, horizontal and vertical alignment; environmental characteristics with the weather condition, day/night time, and road surface condition; and driver characteristics with the vehicle type and driver's gender and driver's age. More than 2,500 crashes were used in the analysis, and a factor analysis was done to test the model hypothesis for the latent variables. The study concluded that the effect of road factors on crash severity was the dominant one compared to the other latent dimensions. The number of involved vehicles in the crash had the most influence on the crash severity. The horizontal alignment was most influential for the road factor group. Regarding the environmental factors, road surface condition had the highest effect. Moreover, for the driver's factor, drivers in their twenties were the most influential.

In Kim et al. (2011), a SEM model was proposed to examine the severity of crashes in terms of different factors including human, vehicle, accessibility, and roadway factors. The human factor is made up of five variables, namely impaired driving, lighting condition, traffic conditions, age group of the driver, and vision obstruction. The vehicle factor is made up of variables showing car defects, the type of vehicle, and the number of people in the vehicle. The road accessibility factor is a construct made of total street length and bus route in miles, real node (intersection), and dangle (dead end) in the grid. Road construct is a combination of four variables that measure the types of defects on the road, the condition of the road surface, the horizontal and vertical alignment, and the weather conditions during the crash. Finally, the crash construct was composed of three defining variables: damage to the vehicle, injury level to the driver, and the total number of vehicles involved in the crash.

In the study, first using a confirmatory factor analysis (CFA), an acceptable measurement model is developed and tested. In the second stage, the measurement model is modified so that the predicted casual relationships between the latent variables are introduced and assessed. The results demonstrated that the human latent factor was the most influential one. Crash severity latent variable was strongly influenced by the vehicle damage variable. The human factor was most affected by light conditions. Concerning accessibility latent construct, bus route length had a strong positive influence. The road factor was heavily influenced by road surface condition.

In another study, Hassan et al. (2013) studied the factors that increase the risk of a severe roadway crashes in Riyadh using latent variables such as driver factor, road factor, and environment factor. The crash factor was measured by four observed variables, namely the number of vehicles involved in the crash, the number of injuries that resulted from the crash, damages in private properties, and damages in public properties. The results showed that the road factor was the most significant factor affecting crash severity followed by driver and environment factors, respectively. This result indicates that the crash size increases when the crash occurs on wet road surface compared to those occurring on a dry road surface. The findings also indicate that two aberrant driving behavior factors: sudden lane change and in-vehicle distraction positively affected the crash size. Also, it was found that day of the week (working day or weekend) and crash time (day or night) have a significant effect on increasing the crash severity.

Wang and Qin (2014) examined crash severity of single-vehicle crashes using SEM with injury severity and vehicle damage as main indicators of crash severity. By adopting SEM, latent variables, such as speed and force of the vehicle during the crash, were modeled jointly. These latent variables were defined using measured factors such as human, vehicle, and roadway characteristics. The study concluded that speed and force could significantly increase the injury severity and vehicle damage. Male drivers were found to drive faster than females, and older drivers to drive slower than younger drivers. Another finding was related to the fact that adverse surface and lighting characteristics were found to decrease injury severity and vehicle damage as vehicle speed was reduced. In addition, the crash severity of heavy vehicles may not only decrease because of their slower traveling speed but can also increase because of vehicle weight.

In another study, Xu et al. (2017) utilized SEM to analyze real-time crash severity in highways, where the crash risk was measured with latent variables such as congestion, speed variation, traffic volume, and occupancy variation. In this study, factor analysis was first conducted to establish initial associations between the latent traffic factors and traffic flow variables, which provided useful information for the development of the measurement equations in SEM. Each of the hypothesized latent variables was measured through different traffic variables, and the study results showed that speed variation is the most influential factor for the crash severity.

Cho et al. (2017) investigated the severity of truck-involved and non-truck-involved crashes on freeways using latent variables such as driver, road, environment, and crash characteristics. Severity was defined by observed variables such as the number of deaths, number of injured people, and the number of involved vehicles. The crash characteristics were defined by the location of the crash, crash cause, and the object that the vehicle hit. Environment factor was defined using variables such as day of the week, weather condition, time of the day. The road factor was measured through road material, road curve, and vertical alignment. Driver factor was consist of driver's age and gender. The truck-involved crashes model showed that the crash factor had the most significant impact on the severity of the crash and that crashes tend to be more severe depending on the facility, cause, and the hit object. For the non-truck-involved crashes model, environment factor showed the most significant impact on crash severity; this result suggested that non-truck-involved crashes can lead to severe crashes depending on the weather, whether it is nighttime, and whether it is a weekday.

A study done by Lee et al. (2018) investigated crash severity with rain-related factors In Seoul city. Latent variables such as rain and water depth, road, traffic environment, and human factors were modeled. Crash severity was measured by factors such as fatality rate, number of causalities, and the number of damaged vehicles. The road factor was measured by observed variables such as road type, vertical slope, super-elevation (cross slope), and curve length. Traffic environment and human factors were measured by vehicle type, time of the day, driver's age, and gender. Rain and water depth factor were modeled using variables such as rainfall intensity, drainage length, and water depth. The study concluded that traffic environment and human factors have the most effect on the crash severity levels. The number of casualties heavily influenced the severity. Road type, nighttime crashes, and water depth are the most influencing variables in each of their latent groups.

Even though the literature has addressed the analysis of crash severity with SEM, no study has investigated factors affecting the severity of RLR-related crashes or Wildlife vehicle crashes by utilizing unobserved variables related to the dynamic of a crash. Only two of the studies have focused on modeling crash severity using latent variables related to the dynamic of a crash, such as speed and force (Wang and Qin 2014, Xu et al. 2017). In fact, the majority of latent variables employed in previous SEM studies focused on categorizing road, environment, and driver variables rather than modeling dimensions (Lee et al. 2008, Kim et al. 2011, Hassan and Al-Faleh 2013, Cho et al. 2017, Lee et al. 2018). Therefore, this thesis aims to investigate the influencing factors of crash severity for RLR-related crashes as well as WVCs, through latent variables that are dynamic of a crash and not just by categorizing the observed variables.

CHAPTER 3

METHODOLOGY

SEM can include three types of statistical analysis methods: path analysis/ordinary regressions (measured variables only), confirmatory factor analysis (CFA) (measured and latent variables), and a combination of them (full SEM or, simply, SEM) (Kline 2005). The general process of developing a (full) SEM is that a model structure (model hypothesis) needs to be designed based on previous theories or knowledge. The model hypothesis is usually based on latent endogenous (dependent) and exogenous (independent) variables. The main strength of SEM is the possibility to verify the significance of hypothesized relationships between latent variables (unobserved) by means of measured variables (observed).

This chapter discusses the preparatory stages for carrying out a SEM analysis. Moreover, the statistical and conceptual foundations of a SEM model will be discussed.

3.1 Data Preparation

SEM requires data on multiple indicators (usually questionnaire items) from a large number of cases. Sample size requirements vary widely depending on characteristics of the model tested, such as model complexity and magnitude of factor loadings. The Kline book (2005) gives absolute guidelines based on the ratio of cases to estimated parameters. In absolute terms, he suggests that fewer than 100 cases is a "small" sample, 100 to 200 is "medium," and over 200 is "large." In relative terms, Kline suggests that a 20:1 case-to-parameter ratio is desirable, 10:1 "more realistic," and 5:1 "doubtful." Users with small samples (e.g., fewer than 100 cases, or only 5 cases per parameter to be estimated) may be able to proceed with a SEM analysis if factor loadings are high. In practice, even 200 cases can be inadequate for complex models or data requiring particular estimators. Analyses using methods appropriate for ordinal and non-normal data require larger sample sizes in some programs (Bowen and Guo 2012).

While using SEM, it is essential to be aware of the measurement level and distributional characteristics of data before conducting SEM analyses. SEM programs can accommodate all measurement levels and distributions; however, unique analysis properties must be selected for variables not meeting default assumptions. Default SEM procedures assume that observed variables have normal distributions. Applying the default maximum likelihood estimator in most SEM programs, for example, when data are non-normal and/or ordinal, can lead to biased estimates, misleading significance testing, and erroneous conclusions about model fit (Washington et al. 2003, Bowen and Guo 2012).

Determining the distributional qualities of one's variables is more complicated than determining the measurement level, but all general statistics programs provide the information necessary. A significant number of studies have been carried out to assess the impact of continuous and non-normal variables on SEMs. Non-normality can come from coarsely categorized continuous variables or poorly distributed continuous variables. Non-normality can be detected in several ways, including histograms, box plots, normal probability plots, and by examining the skewness and kurtosis of individual variables, which can usually be obtained as part of the descriptive procedure in statistical packages.

Conventional methods to deal with such data to obtain better univariate distributions could be to transform variables (Daniel and Wood 1980), if possible, or as suggested by Bollen (1989) and Kline (2005) to identify influential outliers before conducting SEM analyses and recode or delete as appropriate. Outliers, or influential cases, can lead to inadmissible solutions, among other undesirable consequences. Another remedy for dealing with non-normality is to use the asymptotically distribution-free estimator or weight-least squares (ADF or WLS) which is a generalized least squares estimation approach that does not rely on multivariate normality. The ADF estimator generates asymptotically unbiased estimates of the χ^2 test statistic, parameter estimates, and standard errors (Browne 1984).

The data matrix submitted for analysis to a SEM computer tool should have the property that it is positive definite, which is required for most estimation methods. A matrix that lacks this characteristic is non-positive definite, and attempts to analyze such a data matrix will probably fail (Wothke 1993). Moreover, extreme collinearity between variables should be avoided. Collinearity can occur due to what appear to be separate variables measure the same thing. Researchers may inadvertently cause extreme collinearity when they analyze composite variables and their constituent variables together (Kline 2005). In addition, researchers ideally would work with complete data sets; in case of having missing data, researchers should be aware of the extent to which their datasets have missing values and understand the mechanisms of missingness. The topic of analyzing datasets with missing observations is complicated; entire books and sections of journals are devoted to it (Kline 2005, Bowen and Guo 2012).

3.2 Path Analysis

Path analysis is the oldest member of the SEM family (Kline 2005). Path analysis was developed by geneticist Sewall Wright (1921) to specify relationships among observed variables. Wright also developed the tracing rules to calculate the model-implied correlation elements based on the proposed structural model (Mulaik 2009), which provided the foundation of SEM. One of the most popular applications of path analysis is mediation analysis (MacKinnon 2008). Path analysis can be used to test theoretical models that specify directional relationships among a number of observed variables. Path analysis determines whether the model successfully accounts for the actual relationships observed in the sample data (Norm O'Rourke and Hatcher 2013).

In path analysis, variables can be dependent in one relationship and an independent in another. These variables are known as mediating variables. In path analyses, observed dependent variables could be continuous, binary, ordered categorical (ordinal), counts, censored, or combinations of these type of variables. Also, in path analysis for non-mediating variables, observed dependent variables could be unordered categorical (nominal) (Muthén and Muthén 2017).

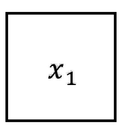
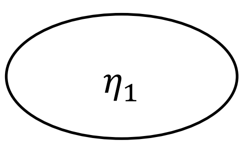
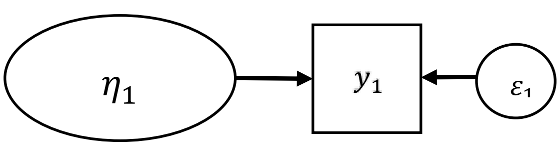
For continuous dependent variables, linear path (regression) models are used. For censored dependent variables, censored-normal path models are used, with or without inflation at the censoring point. For ordered categorical and binary dependent variables, probit or logit models are used. For ordered categorical dependent variables, logistic regression is used with the proportional odds specification. For unordered categorical dependent variables, multinomial logit models are usually used. Also, for count dependent variables, Poisson regression models are typically used, with or without inflation at the zero point (Muthén and Muthén 2017).

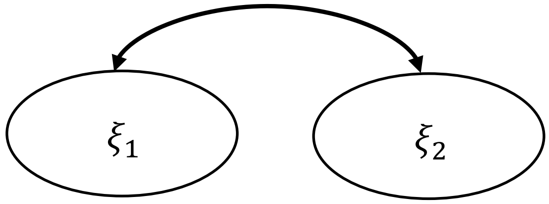
By using path analysis, it is possible to distinguish between direct, indirect, and total effects of one variable on another (Bollen 1989). Moreover, it is possible to show the model on a path diagram.

3.2.1 Path Diagram

A path diagram is the representation of a system of simultaneous equations. The primary advantage of a path diagram is that it describes the assumed relationships in a picture. For many researchers, this picture may represent the relationships more clearly than the equations. It is necessary to define the symbols involved, to understand path diagrams. Table 3.1 provides the primary symbols. The observed variables are shown by rectangles (squares). The unobserved or latent variables are shown by circles, the disturbance terms which are shown with smaller circles. Straight single-headed arrows represent causal relations between the variables connected by the arrows. A curved two-headed arrow indicates an association between two variables. The association may be due to both variables depending on some third variable(s), or the variables may have a causal relationship, but this remains unspecified (Bollen 1989).

Table 3.1 Path Analysis Symbols

Symbol	**Description**
	Rectangular or square box shows an observed variable x_1 (exogenous).
	Circle or ellipse shows a latent variable η_1 (endogenous).
	Small circle shows a disturbance term ε_1. Straight arrow shows the assumption that the latent variable (η_1) at base of the arrow "causes" the observed endogenous variable (y_1) at the head of the arrow.

	Curved two-headed arrow signifies unanalyzed association between two latent variables (ξ_1 and ξ_2).

3.2.2 Total, Direct, and Indirect Effects

Path analysis distinguishes three types of effects: direct, indirect, and total effects. The direct effect is that influence of one variable on another that is unmediated by any other variables in a path model. The indirect effects of a variable are mediated by at least one intervening variable. The sum of the direct and indirect effects is the total effects (Bollen 1989).

Total effects = Direct effect + Indirect effects

The equations for the regression model with intervening variables are shown in Eq. (3.1) and (3.2).

$$y_1 = \gamma_{11} x_1 + \varepsilon_1 \quad (3.1)$$

$$y_2 = \beta_{21} y_1 + \gamma_{21} x_1 + \varepsilon_2 \quad (3.2)$$

Figure 3.1 shows the mediation model.

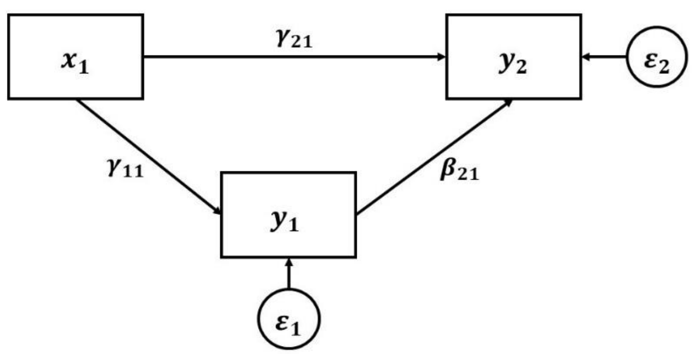


Figure 3.1 Path Analysis Model with One Mediator

As can be seen from the figure, there are two direct effects on the endogenous variable y_2 from other observed (measured) variables, one from the exogenous variable x_1 and another from the other endogenous variable, y_1. The latter specification gives y_1 a dual role as both a predictor and a criterion. This dual role is described in path analysis as an indirect effect. Indirect effects generally involve one or more intervening, or mediator variables, which are assumed to “transmit” some of the causal effects of prior variables onto subsequent variables. The product term $\gamma_{11}\beta_{21}$ represents the indirect effect via the mediator while γ_{21} is the direct effect after controlling for the mediator. The total effect between x_1 and y_2 is $\gamma_{11}\beta_{21} + \gamma_{21}$. Both ε_1 and ε_2 are disturbance terms which are independent and uncorrelated with x_1.

The critical difference between path analysis and SEM could be that the former analyzes relationships among observed (measured) variables, while the latter focuses on relationships among latent variables (constructs or factors). In order to carry out SEM, latent variables (factors) must be appropriately defined using a measurement model before they are incorporated into a SEM model. Latent variables (factors) are unobservable and must be indirectly estimated from observed variables (indicators) (Wang and Wang 2012). In the next sections, methods regarding hypothesizing latent variables/factors and testing them are discussed.

3.3 Factor Analysis

In practice, the initial steps of SEM development attempt to uncover structure in data that can then be used to formulate and specify statistical models. These situations usually arise when the study is exploratory, and there are no preliminary theories concerning the structure in the data. There are several methods to uncovering data structure. Factor analysis is a statistical approach for examining the underlying structure in multivariate data (Washington et al. 2003). Factor analysis is used to identify the number and nature of the underlying factors that are responsible for covariation in the data (Norm O’Rourke and Hatcher 2013). Factor models are also called measurement models because they focus on how one or more latent constructs are measured, or represented, by a set of observed variables (Bowen and Guo 2012).

The factor analysis aims to decrease the number of p variables to a smaller group of parsimonious $K<p$ variables. The aim is to describe the covariance among several variables in terms of a few unobservable factors (Washington et al. 2003). Similar to other statistical models, there should be a theoretical rationale for conducting a factor analysis (Pedhazur and Schmelkin 1991). It is not possible to feed all the variables into an factor analysis and hope to uncover real dimensions in the

data. First, as a model hypothesis, a theoretically motivated reason based on previous theories or knowledge should exist to suspect that some variables could be measuring a latent construct. With the factor analysis, it is possible to examine this theory about how well sets of observed variables in an existing dataset measure latent constructs (Washington et al. 2003). These factors can be latent variables in the SEM model.

The factor analysis model is developed by expressing the X_i terms as linear functions, such that

$$
\begin{aligned}
X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \varepsilon_1 \\
X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \varepsilon_2 \\
\vdots \qquad & \qquad \vdots \qquad\qquad \ddots \qquad\qquad \vdots \\
X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \varepsilon_p
\end{aligned}
\tag{3.3}
$$

Where the factor analysis model in matrix notation is given as (Washington et al. 2003).

$$(X - \mu)_{p\times 1} = L_{p\times m}\, F_{m\times 1} + \epsilon_{p\times 1} \tag{3.4}$$

Where the μ are the means for X, the F are factors and the l_{ij} are the factor loadings. The ε_i are associated solely with the X_i, and the p random errors and m factor loadings are latent (unobservable). With p equations and p+m unknowns, it is not possible to directly solve the unknowns without additional information. Restrictions are imposed to solve for the unknown factor loadings and errors. The type of restrictions defines the type of factor analysis model. Also, the type of factor rotation method determines the type of factor analysis model, oblique, or orthogonal.

The orthogonal factor analysis model hypothesizes that the factors are orthogonal or uncorrelated. Varimax rotation maximizes the sum of the variances of the factor loadings and is a standard approach for conducting an orthogonal rotation. The hypothesis behind the oblique factor analysis model is that the factors are correlated with one another with no cause-and-effect relationship specified. This model relaxes the restriction of uncorrelated factor loadings, resulting in factors that are non-orthogonal and is normally done by Promax rotation (Norm O’Rourke and Hatcher 2013, Washington et al. 2003).

Interpretation of factor analysis is simple and straightforward. Variables that have high factor loadings are considered to be highly influential in describing the factor, whereas variables with low factor loadings are considered as less influential in describing the factor. Factor loadings that are either close to 1 or close to 0 are desirable. A factor loading close to 1 indicates that F_j mainly influences a variable X_i. In contrast, a factor loading close to 0 indicates that a variable X_i is not influenced by F_j. A set of factor loadings that is as diverse as possible is desirable, and make the interpretation straightforward and simple. Investigation of the variables with high factor loadings on a particular factor is used to uncover commonality or structure among the variables. In the next step it is possible to determine the underlying constructs (dimension) that are common to variables that load highly on particular factors (Washington et al. 2003).

3.4 Confirmatory Factor Analysis

By using confirmatory factor analysis (CFA), the directional effects (relationship) between the proposed latent variables can be tested. Besides, CFA can be used to test the measurement model and see if the indicator variables effectively measure the underlying constructs of interest and that the measurement model demonstrates an acceptable fit to data (Norm O'Rourke and Hatcher 2013). In contrast to factor analysis, confirmatory factor analysis (CFA) is used in situations where the analyst has some knowledge of the number and dimensionality of the variables that are being studied either based on empirical findings or theory. The factors are defined theoretically, and how specific indicators (measurement items) are loaded onto which factors is hypothesized before testing the model. In the application of CFA, researchers are interested mainly in investigating the extent to which a set of indicators (measurement items) in a particular instrument measures the factors (latent variables) they are designed to measure (Wang and Wang 2012). In short, CFA requires a detailed and identified initial model (Bollen 1989).

Compared with factor analysis, the advantages of CFA include but are not limited to, the following: first, all factors in factor analysis are either uncorrelated (orthogonal) or correlated (oblique). In CFA, relationships among factors can be flexibly specified on a theoretical basis or based on empirical findings. Secondly, observed indicators/items in factor analysis are loaded onto all the factors; while observed indicators/items in CFA are only loaded onto factors that they are hypothesized to measure. However, an indicator may also be loaded onto one or more factors in a CFA based on a theoretical concern. As a result, a CFA model is not only theoretically more meaningful, but also is more parsimonious because the factor loadings of indicators to the irrelevant factors are all fixed, a priori, at 0 in a CFA model, thus substantially reducing the number of parameters to estimate.

Thirdly, measurement errors are not allowed to be correlated in factor analysis, yet this is not the case in CFA. However, appropriate specifications of error correlations in CFA can be used to test method effects. Fourthly, unlike the traditional factor analysis, CFA can be simultaneously conducted in multiple groups so that measurement invariance across groups can be tested. Finally, covariates can be readily included to predict the factors, thus expanding the CFA model to a SEM model (Tomas and Oliver 1999, Wang et al. 2001, Wang and Wang 2012). In CFA, the link between the observed indicators/items and the factors is represented by factor loadings that are the regression paths from the factors to the corresponding observed indicators. A slope coefficient of regressing an observed indicator on a factor is the factor loading of the indicator on that factor, and the associated residual term is the corresponding measurement error in the indicator (Wang and Wang 2012).

In doing so, the measure of an observed indicator is separated into measurement error and the measure on the underlying factor. As a result of this, the estimated relationships among latent variables would be free of the effects of measurement errors when the relationships between the factor/latent variables is modelled. CFA is fundamental to SEM. One of the prevalent uses of SEM techniques is to study construct validity or to assess the factorial structure of scales in the measuring instrument under study using the CFA model (Wang and Wang 2012). The general model for confirmatory factor analysis is similar to the SEM model, which is presented in the next section. Figure 3.2 gives an example of a CFA model with two exogenous latent factors and six exogenous observed variables.

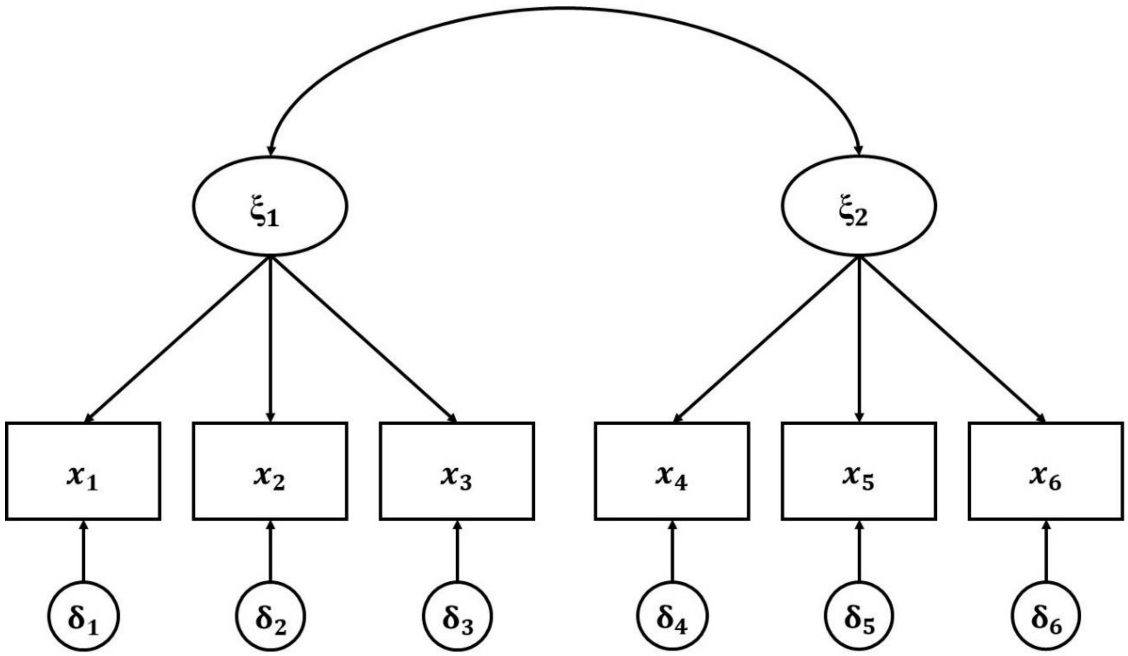


Figure 3.2 Example of CFA Model

3.5 Structural Equation Modeling (SEM)

As mentioned before, the main strength of SEM analysis is the possibility to verify the significance of hypothesized relationships between latent variables (unobserved) through measured variables (observed). SEM is able to efficiently identify the indirect, direct and total effects among variables by means of three components as can be seen in Figure 3.3: (a) a measurement model for the endogenous variables (Y measurement model), (b) a measurement model for the exogenous variable (X measurement model), and (c) a structural model (Lee et al. 2008).

In the measurement model, linear (e.g., when observed variables are continuous) or nonlinear (e.g., when observed variables are categorical) equations describe the relations between the observed variables and their underlying latent variables (factors). In the structural equations part of the model, endogenous latent variables (η) are regressed on the exogenous latent variables (ξ) and/or some other endogenous latent variables. In addition, observed variables can also be included as either independent and/or dependent variables in a SEM model (Wang and Wang 2012).

The resulting SEM model represents the regression effects of exogenous (independent) variables on endogenous (dependent) variables, as well as the effects among the endogenous variables. Figure 3.3 shows an example of a full SEM with two measurement models (each one based on two measured variables and one latent) and one structural model (based on two latent constructs). Square, oval, and circle boxes represent measured variables, latent variables, and error terms associated with variables, respectively. Arrows depict functional relationships among them that need to be validated (Lee et al. 2008).

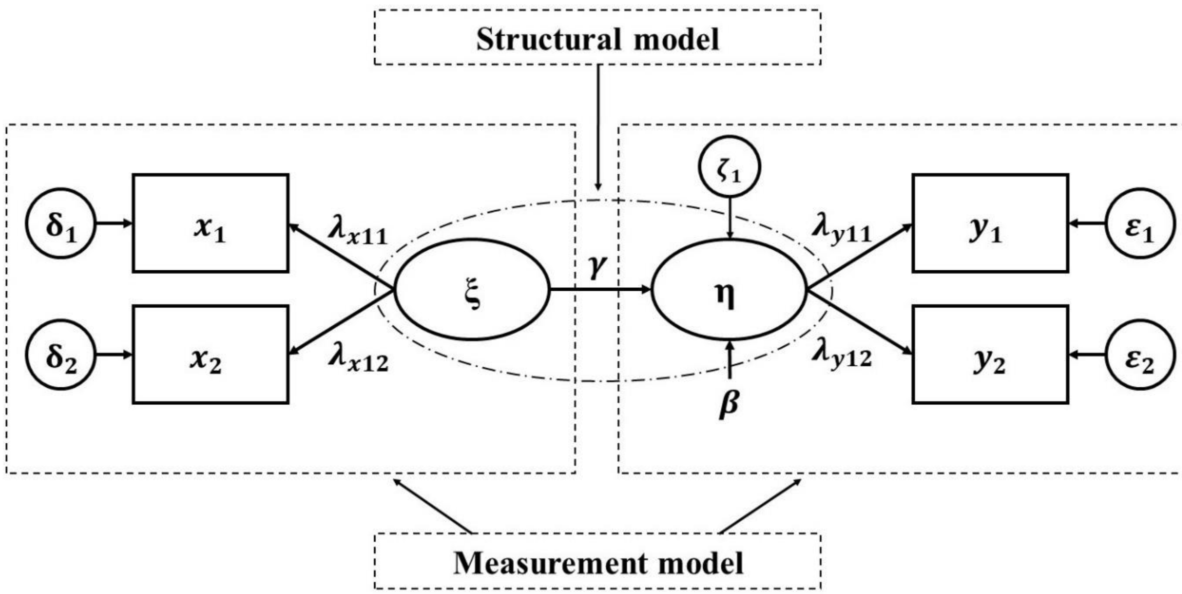


Figure 3.3 Example of SEM Diagram

In general, the basic SEM equation can be written as (Bollen 1989):

$$\eta = B \cdot \eta + \Gamma \cdot \xi + \zeta \tag{3.5}$$

Where B is an $m \times m$ matrix and Γ is an $m \times n$ matrix that contains regression parameters (i.e., β's and γ's) for the endogenous and exogenous latent variables (i.e., magnitudes of expected changes after a unit increase in η and ξ, respectively). η is an m×1 vector of endogenous latent variables, ξ is an n×1 vector of the exogenous latent variables, and $\zeta = m \times 1$ is a vector of disturbances (error terms).

The measurement models can be represented with the following equations:

$$x = \Lambda_x \xi + \delta \tag{3.6}$$

$$y = \Lambda_y \eta + \varepsilon \tag{3.7}$$

Where x is a q×1 column vector of the observed exogenous variables, y is a p×1 column vector of the observed endogenous variables, δ is a q×1 column vector of the observed exogenous errors. $\boldsymbol{\varepsilon}$ is a p×1 column vector of the observed endogenous errors. Λ_x is a $q \times n$ matrix of structural coefficient (λ_x) for the effects of the latent exogenous variables on the observed variables, and Λ_y is a $p \times m$ matrix of structural coefficient (λ_y) for the effects of the latent endogenous variables on the observed ones.

The criterion selected for parameter estimation is known as the discrepancy function, where the difference between Σ, the $p \times p$ population covariance matrix as estimated by the sample covariance matrix, and $\Sigma(\theta)$, the $p \times p$ covariance matrix from the hypnotized model, are minimized. Maximum likelihood, asymptotically distribution-free, and weighted least square are standard estimation methods for many model-fitting programs. A more in depth explanation is provided in Appendix A. Conventional SEM employs simple linear regressions as a functional relationship among variables. When measured variables are non-continuous (e.g., nominal and ordinal scale variables), the measurement models are generalized linear models, allowing to model a much more extensive range of response (generalized SEM).

3.6 SEM Results

Usually, the results of a SEM analysis is reported in unstandardized coefficients by default. For the unstandardized results, reference (measured) variables need to be selected for each latent dimension; their estimates are constrained to 1, and their effect on each measured variable is estimated (proportionally) against these reference variables. Unstandardized coefficients depend upon the units in which the variables are scaled. If two explanatory variables in an equation have the same units, then a comparison of their unstandardized coefficients provides an idea of their relative influence. However, explanatory variables often have different scales. Hence, it is difficult to compare the unstandardized coefficients among different pathways as units are various (e.g., some of them are expressed as binary and other as ordinal variables) (Bollen 1989, Grace and Bollen 2005).

To overcome this issue, analysts can adjust unstandardized coefficients to make them "dimensionless." That is, analysts modify the coefficients so that they are in similar units. The standardized coefficient shows the mean response in standard deviation units of the dependent variable for a one standard deviation change in an explanatory variable, holding constant the other variables in a model. Thus, researchers compare the shift in standard deviation units of the dependent variable that accompanies shifts of one standard deviation in the explanatory variables as a means to assess relative effects. Standardized estimates are usually employed for the comparison of parameter estimates of SEM models, which allows comparing different coefficients from different variables (Grace and Bollen 2005, Bollen 1989).

3.7 Goodness-of-Fit Criteria

Several criteria have been developed for evaluating the overall goodness-of-fit of SEM and measuring how well one model performs versus another model; most SEM scholars recommend evaluating the models by observing more than one of these indicators (Bentler and Wu 2005, Hoe 2008).

Chi-square (χ^2) is the most common method for assessing goodness-of-fit. The objective is to get a non-significant model chi-square since this indicator measures the difference between the observed variance-covariance matrix and the one reproduced by the model. The level of statistical significance shows the probability that the differences between the two matrixes are due to sampling variation (Golob 2003). As a rule of thumb, an indication of good fit is that the chi-square should be less than two times its degrees of freedom (Ullman 2013). Chi-square is formulated as:

$$\chi^2 = (N-1)F_{ML} \tag{3.8}$$

where F_{ML} is the value of the statistical criterion (fit function) minimized in ML estimation and N is the sample size. However, there might be issues associated with the use of fitting-function chi-square, which are mostly due to the influences of sample size and deviations from multi-normality. For large samples, it may be challenging to find a model that cannot be rejected, especially if the observations are more significant than 200 (Golob 2003, Hoe 2008). Other goodness-of-fit measures commonly used to measure SEM fit include RMSEA, SRMR, AGFI, CFI, and TLI.

The root mean square error of approximation (RMSEA) is a class of fit measures that is based on the population discrepancy. This measure relies on the notion of a population discrepancy function (as opposed to the sample discrepancy function) to estimate goodness-of-fit measures (Washington et al. 2003). The RMSEA value ranges from 0 to 1, and it is generally accepted that for a good model the value of RMSEA should be less than 0.08 (Browne and Cudeck 1992). RMSEA is formulated as:

$$RMSEA = \sqrt{\frac{\chi^2 - df}{df(N-1)}} \tag{3.9}$$

where N is the sample size and df is the degrees of freedom of the model.

The standardized RMR (SRMR) is an absolute measure of fit and is described as the standardized difference between the predicted correlation and observed correlation. Since the SRMR is an absolute measure of fit, it can range from 0 to 1, and a value of 0 indicates perfect fit. The SRMR has no penalty for the complexity of the model. Values less than .08 are generally considered as good fit (Hu and Bentler 1999). SRMR is formulated as:

$$SRMR = \sqrt{\frac{2\sum_{i=1}^{k}\sum_{j=1}^{i}\left[(s_{ij}-\sigma_{ij})/(s_{ii}s_{jj})\right]^2}{k(k+1)}} \tag{3.10}$$

Where k is the number of the observed variables; s_{ij} and σ_{ij} are the sample and the model-estimated covariances between the i-th and j-th variables; s_{ii} and s_{jj} are the observed standard deviations.

The Comparative Fit Index (CFI) was developed by Bentler in 1990 as a non-centrality parameter-based index to address the limitation of sample size effects. Comparative Fit Index (CFI) ranges from 0 to 1, with values closer to 1 indicating an acceptable fit (Hoe 2008). CFI is formulated as:

$$CFI = 1 - \frac{\chi^2 - df}{\chi_B^2 - df_B} \tag{3.11}$$

the numerator and the denominator of the expression in the right side of Eq. (3.11) show the chi-square non-centrality parameter for the researcher's model and the baseline model, respectively (Kline 2005).

The other fit indices are the goodness-of-fit index (GFI), adjusted GFI (AGFI), and the Tucker Lewis Index (TLI). Similar to CFI, values of GFI, AGFI, and TLI range from 0 to 1, where values closer to 1 indicate an acceptable fit (Schreiber et al. 2006). GFI is an absolute fit index that estimates the proportion of covariances in the sample data matrix explained by the model. The GFI is formulated as:

$$GFI = 1 - \frac{C_{res}}{C_{tot}} \qquad (3.12)$$

where C_{res} and C_{tot} are the residual and total variability in the sample covariance matrix, respectively (Kline 2005).

The AGFI is the goodness-of-fit index adjusted for degrees of freedom (Joreskog and Sorbom 1982), and its formulated as:

$$AGFI = 1 - \frac{k(k+1)}{2df}(1 - GFI) \qquad (3.13)$$

where k is the number of observed variables and df is the degrees of freedom of the model.

The Tucker-Lewis index (also called the non-normed fit index or NNFI), is another fit index which has a penalty for adding parameters. The TLI is computed as follows:

$$TLI = \frac{\frac{\chi_B^2}{df_B} - \frac{\chi^2}{df}}{\left(\frac{\chi_B^2}{df_B}\right) - 1} \qquad (3.14)$$

Where χ_B^2 and df_B are chi-square and degrees of freedom for the baseline model, respectively; and df is the degrees of freedom of the model (Bollen 1989).

The performance of models with a substantially different number of parameters can be compared using criteria based on Bayesian theory. The Akaike Information Criterion compares ML estimation goodness-of-fit and the dimensionality (parsimony) of each model. These criteria can be used to compare two alternative models of similar dimensionality. The model that produces the smallest value of each criterion is considered best; higher values correspond to more considerable lack of fit. In the AIC, a penalty is put on models with more significant numbers of parameters, similar to the adjusted R-squared measure in regression (Washington et al. 2003, Golob 2003). AIC is defined as:

$$AIC = 2k - 2\,ln(L) \tag{3.15}$$

where k is the number of parameters in the model and L is the maximum likelihood value for the estimated model.

CHAPTER 4

RLR CRASH SEVERITY MODEL

In this chapter, SEM was employed to investigate the severity of RLR-related crashes as introduced in the previous sections. First, a model hypothesis was established. After that, the data set used for the analysis was reviewed and converted to an acceptable format for SEM software. Finally, a SEM model was developed, and in the final step, a path analysis was performed and compared to the SEM analysis results.

4.1 Model Hypothesis

In SEM, it is essential to accurately select unobserved variables as their presence may affect the overall significance of the model. Because motor vehicle crashes are multi-causal phenomena, a broad set of factors in a crash severity model can be incorporated such as using variables from a transportation engineering and a crash dynamic viewpoint (Sobhani et al. 2011). Transportation measurements are usually available from police crash reports, but the same is not valid for those variables related to the dynamic of a crash event (e.g., kinetic energy transferred from one vehicle to another during a crash). In this part of the study for RLR-related crashes, three latent variables were proposed: the level of crash severity and two casual factors affecting crash severity, i.e., pre-crash travel speed (TS) of the bullet vehicle, and the kinetic energy (KE_s) transferred from the bullet vehicle to the subject vehicle(s) during the crash event.

These latent variables were hypothesized according to the studies in the literature. The travel speed of the vehicle is one of the major influencing factors of crash severity, and it has been investigated in many studies. (Wang and Qin 2014). Moreover the transferred kinetic energy has been investigated previously in researches and found to be a significant factor in severe crashes. Vehicles acquire kinetic energy only when in motion, and kinetic energy increases in a second-power relationship with increasing speed, but only linearly with increasing vehicle mass (Corben et al. 2004, Sobhani et al. 2011). By using SEM, it is possible to test this hypothesis and find out which variables affect the crash severity through TS and KE_s.

Both pre-crash travel speed (TS) of the bullet vehicle and the kinetic energy (KE_s) transferred from the bullet vehicle to the subject vehicle(s), contribute indirectly to property damage and injuries sustained by crash victims (crash injury severity). On the contrary, speed and kinetic energy contribute directly to crash severity, which represents the energy resulting from a crash and is distinct from crash injury severity.

TS of the bullet vehicle is a factor that can significantly affect the severity of crashes, but it is usually not reported in crash records (Sobhani et al. 2011). Nevertheless, speed can be affected by different measured variables related to infrastructure, driver, traffic, and environmental characteristics.

The kinetic energy transferred from the bullet vehicle to the subject vehicle(s) (KE_s) can be estimated using the following equation (Sobhani et al. 2011):

$$KE_s=\frac{1}{2}\times m_s\times \Delta V_s^2 \tag{4.1}$$

Where m_s is the mass of the subject vehicle, and ΔV_s is the speed change of the subject vehicle before the crash and after the crash, which is a function of crash characteristics. Similarly, to travel speed, KE_s is not reported in crash records. However, other measured variables usually present in police crash reports can be indirectly related to it. Previous studies suggested how vehicle/crash type, and the number of vehicles involved in a crash, should be accounted for in modeling crash severity outcomes. These variables are, in fact, relevant to the kinetic energy aspects of a motor vehicle crash.

Jiang et al. (2015) showed that there might be significant differences between 2- and 3-vehicle crash severities in terms of the contributing factors, the magnitude of impact, and the direction of effects. Moreover, according to Sobhani et al. (2011), the angle of the crash can substantially affect the transferred kinetic energy. These variables alongside other variables such as vehicle type, production year, and vehicle maneuver were employed as proxies for m_s and ΔV_s and afterward linked to kinetic energy (KE_s) transferred from the bullet vehicle to the subject vehicle.

4.2 Data Collection

The data for this part of the study was obtained from the Crash Analysis Reporting (CAR) system of the State of Florida (US). A sample of 2,000 crash records from 2011 to 2014 was initially used in the analysis. The data consisted of 500 crashes, randomly selected from each year, with "disregarded traffic signal" as their "first contributing cause." Initially, all 54 variables obtained from the Florida Police crash report form were considered. The majority of these variables were derived from multiple choice questions, but a few variables were also obtained from short answers entered manually in the report form. Different coding errors were identified during the data review process, possibly due to different enforcement personnel filling out forms.

Variables like crash time which require manual entry were cross-checked with variables from multiple choices like natural lighting conditions. Road surface conditions (dry/wet) were cross-checked with weather conditions. Besides, some observations with unknown entries (coded as 999) were removed from the analysis. Finally, 1,601 crash records were considered in the analysis after removing single-vehicle, pedestrian-vehicle, and bicycle-vehicle crashes (only multiple-vehicle crashes were included to investigate the model hypothesis).

All measured variables able to describe latent dimensions and support the model hypothesis were included in SEM analysis; these variables could be related to human, environmental, vehicle, and crash characteristics. Variables were also converted to binary and ordinal parameters. At first, some variables were eliminated due to collinearity issues. Afterward, a preliminary path analysis was performed to check which variables significantly affected the crash severity variables. Insignificant variables were removed with backward elimination. Finally, 17 parameters were retained in the final model. Table 4.1 shows the description of the variables used in this study.

Table 4.1 Description, Frequency, and Percentage of Selected Variables for RLR-Related Crashes

Variable	Description	Frequency	Percentage
Alcohol involvement	*Alcohol impairment of bullet vehicle driver*		
1	No alcohol involved	1540	96.19
2	Alcohol involved	61	3.81
Vertical alignment			
1	Level	1541	96.25

2	Grade	60	3.75
Gender	*Gender of the bullet vehicle driver*		
1	Male	814	50.84
2	Female	787	49.16
Weekday/weekend traffic condition			
1	Weekdays	1117	69.77
2	Weekend	484	30.23
Lighting condition			
1	Light (Day/Illuminated)	1553	97
2	Dark	48	3
Skid resistance	*Skid test number (higher is better)*		
0	0	660	41.22
1	1 - 35	285	17.8
2	35 - 40	314	19.61
3	40 <	342	21.36
Peak /off-peak hour			
1	Rest of the day	1192	74.45
2	8 AM - 9.30 AM and 4 PM - 7 PM	409	25.55
Road surface condition			
1	Dry	1431	89.38
2	Wet, Icy, Slippery	170	10.62
Age	*Age of the bullet vehicle driver*		
1	<25	441	27.55
2	25 - 35	323	20.17
3	35 - 45	250	15.62
4	45 - 55	223	13.93
5	55<	364	22.74
Number of vehicles involved			
2	2	1391	86.88
3	3	180	11.24
4	4	27	1.69
5	5	3	0.19
Subject vehicle type			
1	Automobile	1047	65.4
2	Van, Light truck	543	33.92
3	Medium, Heavy truck	11	0.69

Vehicle maneuver	*Pre-crash controlled maneuver of the bullet vehicle*		
1	Straight movement	1440	89.94
2	Turning, Backing, Lane change, U-turn, Passing	161	10.06
Point of impact			
1	Front of the vehicle	792	49.47
2	Other parts	809	50.53
Vehicle year			
1	<2000	343	21.42
2	2000 – 2005	606	37.85
3	2005 – 2010	459	28.67
4	2010 <	193	12.05
Injury severity			
1	None	561	35.04
2	Possible	480	29.98
3	Non-incapacitating	416	25.98
4	Incapacitating	130	8.12
5	Fatal	14	0.87
Total vehicle damage			
1	<5000	439	27.42
2	5000 - 10000	500	31.23
3	10000 - 15000	343	21.42
4	15000 - 23000	233	14.55
5	23000<	86	5.37
Number of injured people			
0	0	564	35.23
1	1	491	30.67
2	2	337	21.05
3	3	111	6.93
4	above 3	98	6.12

4.3 Model Development

The model was developed in SAS software version 3.71, University Edition (SAS Institute Inc 2017), which allows implementing SEM with the CALIS procedure. The weight-least squares (WLS) method with a default weight matrix was used, which yields asymptotically normal estimates regardless of the probability distribution of the population. WLS was used in order to account for possible violation of the multivariate normality assumption (Bollen 1989). Moreover, the WLS method can analyze both binary and ordinal variables (regardless of having continuous variables) which is suitable for the variables considered in this investigation. It is worth to mention that in this part of the thesis (RLR crash severity model), the regression analyses are performed using linear links, and in the next chapter the analyses are performed using probit links.

Furthermore, SAS provides a summary of the goodness-of-fit measures. The root mean square error of approximation (RMSEA) is typically used to test the model's goodness-of-fit. For RMSEA, a value of less than 0.08 indicates good fit (Lee, 1993). Other criteria used are the standardized root mean square residual (SRMR), which is less than 0.08 for a well-fitted model, and Bentler Comparative Fit Index (CFI) and Adjusted GFI (AGFI), where values lie between 0 and 1 with values closer to 1 to indicate a better model fit (Schreiber et al. 2006, Hu and Bentler 1999, Browne and Cudeck 1992). It is worth to mention that the Chi-square was not used in this study due to the influences of large sample size and deviations from multi-normality (Golob 2003).

4.4 SEM Analysis

A full SEM was developed (Figure 4.1) using unobserved variables introduced in Section 4.2 and observed variables collected in Table 4.1. Measured variables were represented with rectangles and latent variables with circles. "Speed" was set as a latent exogenous variable to represent TS of the bullet vehicle. "Kinetic energy" and "crash severity" were treated as latent endogenous variables that represent KE_s transferred to the subject vehicle(s) and the crash severity index, respectively. As Figure 4.1 shows, TS can, directly and indirectly, influence crash severity through KE_s transferred from the bullet vehicle to the subject vehicle during a collision. Hence, the model suggests that higher TS and KE_s could directly influence the overall severity of a crash, which can be observed and measured as an increased probability of crash injury severity, vehicle damage and the number of injured people.

TS was linked to measured variables representing infrastructure characteristics such as pavement skid resistance, vertical alignment, and road surface condition, which can reduce or increase speeds; driver's characteristics such as gender, age and alcohol-impaired driving conditions which may directly affect the driver's choice of speed; and other environmental and traffic characteristics such as lighting condition, peak/off-peak hour and weekday/weekend traffic conditions which are well-known factors affecting speeds (HCM 2010). These variables were chosen according to other studies in the literature and have been proven to influence the crash outcome (Lee et al. 2008, Kim et al. 2011).

Regarding kinetic energy, subject and bullet vehicles characteristics (e.g., vehicle type and production year) and crash characteristics (e.g., point of the impact, vehicle maneuver and the total number of vehicles involved) were employed as proxies for m_s and ΔV_s and afterward linked to KE_s (see Equation 4.1). Variables such as vehicle age, point of the impact, vehicle maneuver, and the total number of vehicles involved can directly affect the transferred kinetic energy and indirectly influence the severity of the crash. These variables were chosen according to studies in the literature and have been proven to influence KE_s (Sobhani et al. 2011). Moreover, measured variables that were linked to TS can indirectly affect both KE_s and crash severity latent variables. Similarly, measured variables linked to KE_s can indirectly influence the crash severity latent variable.

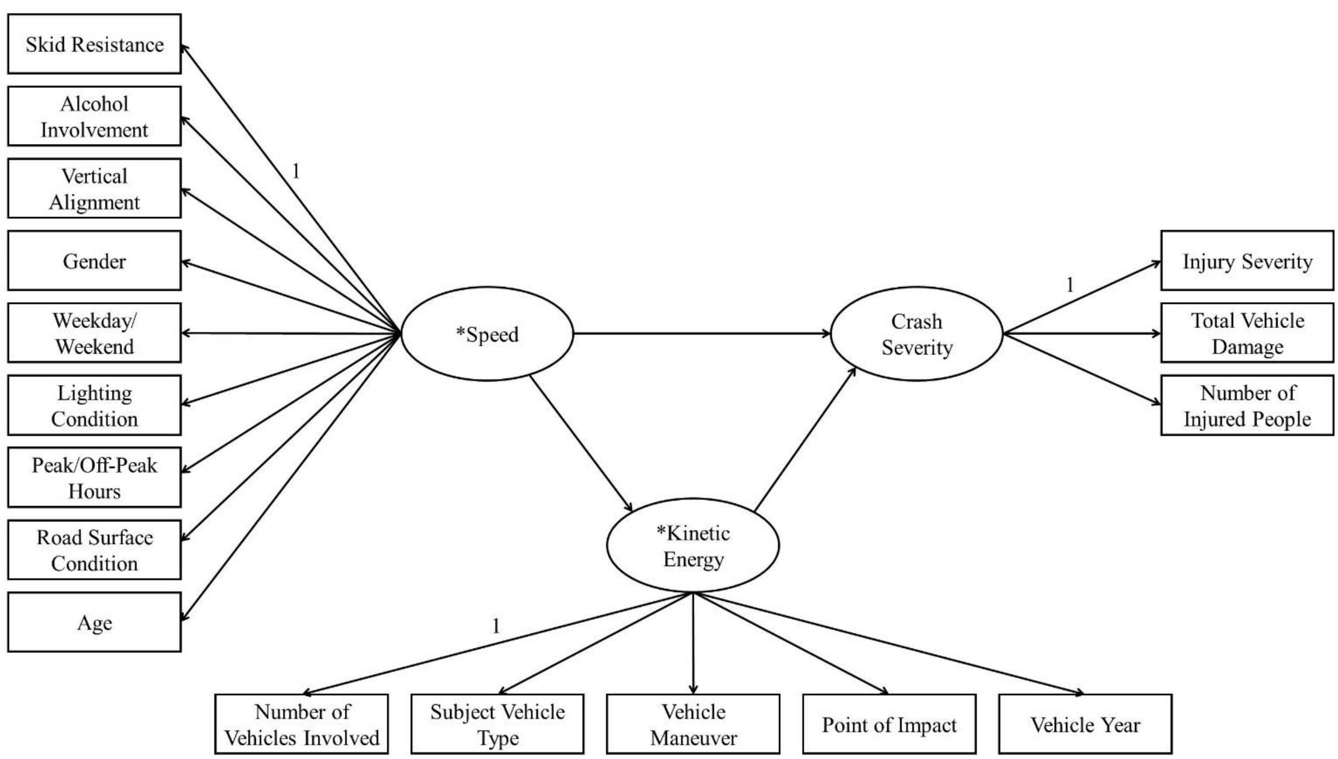


Figure 4.1 Proposed SEM with Three Latent Variables for RLR-Related Crashes

*speed= pre-crash TS of the bullet vehicle

*kinetic energy = KE_s transferred from the bullet vehicle to the subject vehicle(s) during the crash event

4.5 Estimate Results

The results of SEM analysis (unstandardized and standardized estimates) are reported in Table 4.2. For the unstandardized results, reference (measured) variables need to be selected for each latent dimension; their estimates are constrained to 1, and their effect on each measured variable is estimated (proportionally) against these reference variables. It is worth to mention that the unstandardized estimates are expressed in original units of their explanatory and dependent variables. Hence, it is difficult to compare the unstandardized coefficients among different pathways as units are various (e.g., some of them are expressed as binary and other as ordinal variables).

Standardized estimates are based on standard deviation units of the variables and are usually employed for comparison of parameter estimates of SEM models, which allows comparing different coefficients from different variables (Grace and Bollen 2005). SAS provides unstandardized and standardized outcomes for SEM models. Table 4.2 reports the standardized results as well. It is worth to note that the majority of variables apart from gender, road surface condition, and vehicle type, are significant.

Table 4.2 SEM Estimates for RLR-Related Crashes

Path			**Unstandardized Estimates**			**Standardized Estimates**		
			Estimate	**Standard Error**	**p-value**	**Estimate**	**Standard Error**	**p-value**
Crash Severity	<==	Speed	1.114	0.556	0.045	0.258	0.090	0.004*
		Kinetic energy	2.683	0.866	0.002	0.361	0.086	<.0001*
Kinetic energy	<==	Speed	0.200	0.120	0.095	0.344	0.146	0.018*
Skid resistance	<==	Speed	1.000	-	-	0.155	0.052	0.003*
Alcohol involvement			0.173	0.082	0.035	0.209	0.064	0.001*
Vertical alignment			0.114	0.066	0.083	0.119	0.050	0.017*
Gender			0.181	0.151	0.233	0.066	0.050	0.185
Weekday/weekend traffic conditions			0.340	0.172	0.048	0.137	0.051	0.008*
Lighting condition			0.115	0.060	0.055	0.161	0.056	0.004*
Peak /off-peak hour			-0.473	0.202	0.019	-0.201	0.050	<.0001*
Road surface condition			-0.114	0.094	0.227	-0.071	0.052	0.173

Age			-0.704	0.465	0.130	-0.084	0.046	0.064**
Number of vehicles involved	<==	Kinetic energy	1.000	-	-	0.342	0.061	<.0001*
Subject Vehicle type			-0.169	0.195	0.386	-0.037	0.043	0.387
Point of impact			-1.509	0.390	0.000	-0.321	0.052	<.0001*
Vehicle maneuver			-0.416	0.146	0.004	-0.156	0.045	0.001*
Vehicle year			-1.201	0.463	0.010	-0.137	0.044	0.002*
Injury severity	<==	Crash Severity	1	-	-	0.812	0.021	<.0001*
Total vehicle damage			0.640	0.039	<.0001	0.440	0.023	<.0001*
Number of injured people			1.201	0.065	<.0001	0.842	0.023	<.0001*

* Estimate significant at the 95% confidence level

** Estimate significant at the 90% confidence level

Finally, Table 4.3 shows that the developed SEM model for analyzing the severity of RLR-related crashes fitted the data well for all criteria. In particular, RMSEA lower than 0.08 is usually employed to test the overall goodness-of-fit and, in this case, the criterion was satisfied being RMSEA equal to 0.03.

Table 4.3 SEM Goodness-of-Fit Statistics for RLR-Related Crashes

Fit Summary	**Value**
Standardized RMR (SRMR)	0.07
RMSEA	0.03
Adjusted GFI (AGFI)	1.00
Bentler Comparative Fit Index (CFI)	0.86

Overall, the standardized results in Table 4.2 demonstrated that both pre-crash TS of the bullet vehicle and transferred KE_s to the subject vehicle(s) positively influence the overall crash severity, being their estimates equal to 0.258 and 0.361, respectively. These loading factors (i.e., parameter estimates) among latent variables were found statistically significant, which strongly supported the model hypothesis. Moreover, the results showed that TS increase could positively affect transferred KE_s to the subject vehicle (parameter estimate equal to 0.344), which would indirectly increase crash severity.

The indirect effect of pre-crash TS of the bullet vehicle on crash severity can be estimated by multiplying factor loadings of TS/KE_s direct relationship and KE_s/crash severity relationship, as:

$$\text{TS indirect effect} = 0.344 \times 0.361 = 0.124 \tag{4.2}$$

The total effect of TS on crash severity can be calculated by summing up direct and indirect effects of TS, as:

$$\text{TS total effect} = 0.124 + 0.258 = 0.382 \tag{4.3}$$

Therefore, while the direct effect of the transferred KE_s to the subject vehicle (0.361) is higher than the pre-crash TS of the bullet vehicle (0.258), by calculating the total effects it can be concluded that the pre-crash TS of the bullet vehicle can have more influence on crash severity (0.382), which was confirmed by Equation 4.3.

Regarding the variables that measure TS of the bullet vehicle, both peak-hour traffic conditions and older drivers showed a decrease in crash severity through reduced speeds being their factor estimates equal to -0.201 and -0.084, respectively. High pavement skid resistance, weekend traffic conditions, alcohol-impaired driving conditions, night conditions, downhill grade were found to increase TS, and subsequently cause an increase in crash severity. Alcohol-impaired driving conditions showed the most significant influence on severity being the estimate equal to 0.209; on the contrary, the age of the driver showed the least influence with an estimate of -0.084. Gender and road surface condition estimates were found not statistically significant in the SEM model.

Regarding transferred KE_s as a latent variable, the number of vehicles involved in RLR-related crashes showed a positive coefficient which is in agreement with Eq. (4.1) where m_s grows with the number of vehicles involved in a crash. The remaining variables (point of impact, vehicle maneuver, and vehicle year) showed negative standardized coefficients. Among all these variables, the number of involved vehicles showed the most considerable influence being the estimate equal to 0.342; on the contrary, vehicle year showed the least influence with a value equal to -0.137. Subject vehicle type estimate was found not statistically significant.

Finally, all variables representing crash severity, i.e., injury severity, vehicle damage, and the number of injured people showed positive relation with crash severity having values of 0.8125, 0.440, and 0.842, respectively. This implies that a higher crash severity index produces an increase in crash injury severity, property damage, and the number of injured people, as expected. Further discussion on the results is reported in Chapter 6.

4.6 Comparison with Path Analysis

To compare SEM results with traditional crash severity modeling, a path analysis was performed, where only observed variables in Table 4.1 were included. Path analysis is the application of SEM without latent variables. One of the advantages of path analysis is the inclusion of relationships among variables that serve as predictors in one single model to test the proposed theory (Cheung 2015). For path analysis, it was assumed that the parameters related to crash severity (y_i) (i.e., injury severity, number of injured people and total vehicle damage) were directly affected by measured variables related to infrastructure, driver, vehicle, traffic, environmental and crash characteristics (vector X) (Figure 4.2).

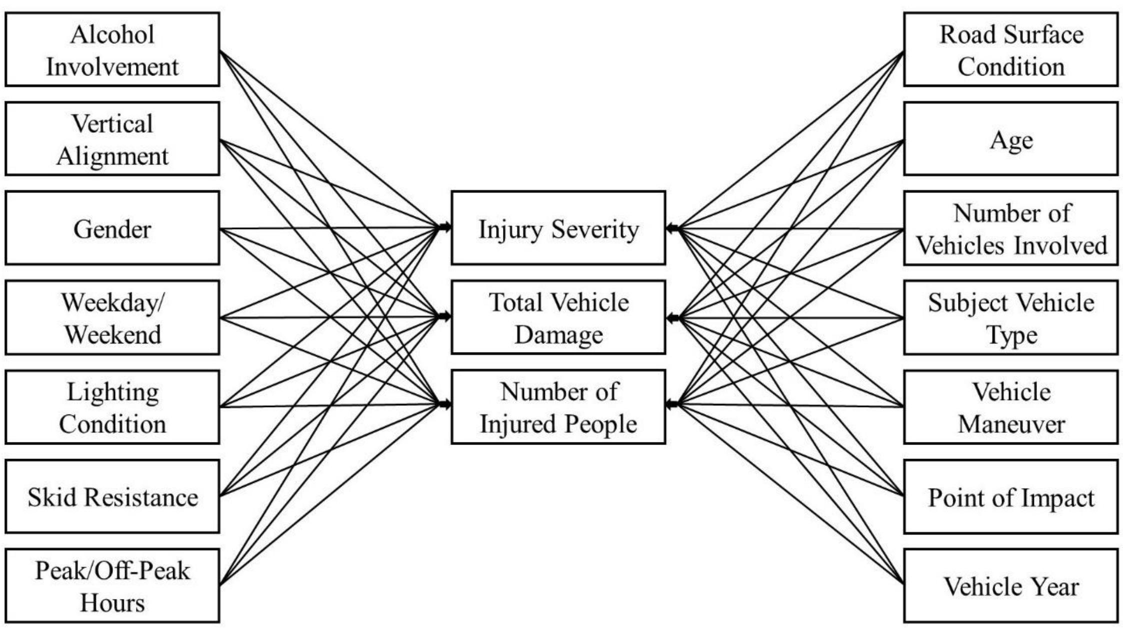


Figure 4.2 Path Analysis Model for RLR-Related Crashes

Table 4.4 shows the results for path analysis and the comparison with SEM in terms of significance, and the sign of the coefficients. Magnitudes in path analysis are expressed in original units of variables, so they are not directly comparable with SEM standardized estimates (and among different y_i). The acceptable confidence level was set at 95%.

Table 4.4 Comparison Between Path Analysis and SEM Results for RLR-Related Crashes (Unstandardized Coefficients)

Vector X	Path	y_i	Estimate (vector β)	Standard Error	p-value	Path sign	SEM sign
Alcohol involvement	===>	Injury severity	0.476	0.151	0.002	+*	+*
		Total vehicle damage	0.195	0.144	0.177	+	
		Number of injured people	0.296	0.153	0.053	+**	
Vertical alignment	===>	Injury severity	0.355	0.130	0.006	+*	+*
		Total vehicle damage	0.277	0.147	0.060	+**	
		Number of injured people	0.484	0.151	0.001	+*	
Gender	===>	Injury severity	-0.023	0.049	0.645	-	+
		Total vehicle damage	-0.040	0.057	0.481	-	
		Number of injured people	-0.022	0.056	0.690	-	
Weekday/weekend traffic condition	===>	Injury severity	0.144	0.056	0.010	+*	+*
		Total vehicle damage	0.154	0.064	0.016	+*	
		Number of injured people	0.119	0.063	0.060	+**	
Lighting condition	===>	Injury severity	0.103	0.165	0.532	+	+*
		Total vehicle damage	0.208	0.186	0.263	+	
		Number of injured people	0.212	0.201	0.291	+	
Skid resistance	===>	Injury severity	0.039	0.021	0.063	+**	+*
		Total vehicle damage	0.029	0.024	0.223	+	
		Number of injured people	0.044	0.024	0.068	+**	
Peak /off-peak hour	===>	Injury severity	-0.098	0.057	0.086	-**	-*
		Total vehicle damage	-0.220	0.066	0.001	-*	
		Number of injured people	-0.170	0.066	0.010	-*	
Road surface condition	===>	Injury severity	-0.177	0.077	0.023	-*	-
		Total vehicle damage	-0.137	0.095	0.148	-	
		Number of injured people	-0.254	0.087	0.004	-*	

Age	===>	Injury severity	0.027	0.016	0.090	-**	-*
		Total vehicle damage	-0.023	0.019	0.242	-	
		Number of injured people	0.023	0.019	0.213	+	
Number of vehicles involved	===>	Injury severity	0.461	0.062	<.0001	+*	+*
		Total vehicle damage	0.978	0.073	<.0001	+*	
		Number of injured people	0.716	0.080	<.0001	+*	
Subject Vehicle type	===>	Injury severity	0.053	0.050	0.295	+	-
		Total vehicle damage	0.011	0.060	0.851	+	
		Number of injured people	0.084	0.058	0.149	+	
Vehicle maneuver	===>	Injury severity	-0.172	0.090	0.057	-**	-*
		Total vehicle damage	-0.196	0.105	0.062	-**	
		Number of injured people	-0.237	0.109	0.029	-*	
Point of impact	===>	Injury severity	-0.180	0.050	0.000	-*	-*
		Total vehicle damage	-0.273	0.058	<.0001	-*	
		Number of injured people	-0.150	0.057	0.009	-*	
Vehicle year	===>	Injury severity	-0.068	0.026	0.010	-*	-*
		Total vehicle damage	-0.031	0.030	0.302	-	
		Number of injured people	-0.073	0.030	0.016	-*	

* Estimate significant at the 95% confidence level

** Estimate significant at the 90% confidence level

For comparison, it is possible to compare the significance, magnitude, and sign of the estimates. Table 4.5 shows the goodness-of-fit of the path analysis model. Even though the overall goodness-of-fit was found to be acceptable, out of 42 path analysis estimates only 18 showed statistical significance at 95% confidence level. For the remaining 24 path estimates, 9 were significant at 90% confidence level and 15 were found not to influence crash severity variables significantly. This might be due to the fact that a simple path analysis can account for direct relationships only and is unable to properly take into account the indirect effect of some variables on crash severity through latent and unobserved variables. Regarding the sign of the significant coefficients, both models were found to be in good agreement for variables such as alcohol involvement, vertical alignment, weekday/weekend traffic condition, peak /off-peak hour, vehicle maneuver, point of impact, and vehicle year. Further discussion on the results is reported in Chapter 6.

Table 4.5 Goodness-of-Fit Statistics for RLR Path Analysis

Fit Summary	**Value**
Standardized RMR (SRMR)	0.03
RMSEA	0.24
Adjusted GFI (AGFI)	0.99
Bentler Comparative Fit Index (CFI)	0.84

CHAPTER 5

WVC SEVERITY MODEL

In this chapter, SEM was employed to investigate the severity of WVCs as introduced in the previous sections. First, a model hypothesis was established and the data set used for the analysis was reviewed and converted to an acceptable format for SEM software. Afterwards, a factor analysis and a confirmatory factor analysis (CFA) were performed to test the model hypothesis and the directional effects (relationship) between the proposed latent variables, respectively. Finally, SEM model results were analyzed and a path analysis using probit links was performed and compared to SEM model.

5.1 Model Hypothesis

In order to better investigate the model hypothesis, three separate stages were carried out for WVCs. First, factor analysis is conducted to classify observed variables into different groups. Factor analysis is used to identify the number and nature of the underlying factors that are responsible for covariation in the data. These factors can be latent variables in the SEM model (Norm O'Rourke et al. 2013). Similar to other statistical models, there should be a theoretical rationale for conducting factor analysis (Pedhazur et al. 1991) and the analyst should suspect that some variables could measure a latent construct. It is not possible to feed all the variables into a factor analysis and hope to uncover real dimensions in the data (Washington et al. 2003). According to the literature, the two major contributing factors in more severe WVCs are speeding and reduced visibility (Haikonen and Summala 2001, Rowden et al. 2008, Gkritza et al. 2010a, Sullivan 2011, Vanlaar et al. 2012). Thus, in this study, three latent dimensions are hypothesized, namely crash severity (CS), driver's speeding attitude (SA) and driver's visibility impairment (VI). These latent variables reflect on underlying factors that are responsible for covariation in the data and are tested using factor analysis.

Second, the directional effects (relationship) between the proposed latent variables will be tested with confirmatory factor analysis (CFA). Besides, CFA can be used to test the measurement model and see if the indicator variables effectively measure the underlying constructs of interest and that the measurement model demonstrates an acceptable fit to data. In the next step, a complete SEM model with directional relationships between latent constructs is developed, and the goodness-of-fit of the model is tested. Lastly, a path analysis using probit links is performed to compare traditional severity models to the SEM results.

5.2 Data Collection

The study employed more than 40,000 WVCs from 2012 to 2018, available in the Traffic Accident Information System (TAIS) of the Saskatchewan Government Insurance (SGI). Each motor vehicle collision was recorded in TAIS with different information related to crash, environment, road, vehicle, and driver characteristics. Only single-vehicle crashes away from major urban centers in Saskatchewan and with “animal action” coded as “first major contributing factor” were considered in the analysis. It is worth to note that, apart from few exceptions, the animal type involved in the crash was not reported in TAIS dataset. However, the majority of WVCs in Saskatchewan were observed to be with ungulates (mainly deer and moose) (Saskatchewan Government Insurance (SGI) 2018).

Different coding errors were identified during the data review process, possibly due to different enforcement personnel filling out forms. Variables like “crash injury severity” were cross-checked with “number of injured people,” “road surface conditions” (dry/wet) with “weather conditions,” “accident time” with “lighting condition” and inconsistent data were removed. Also, observations with unknown entries (e.g., coded as 999) and missing data were removed from the analysis.

Finally, 10,271 crash records were considered. All categorical and nominal variables were converted into binary and ordinal parameters (where appropriate). For example, lighting conditions were converted into binary format showing dark periods of the day (corrected for sessional changes) and the rest of the day. “weather conditions” coded as clear, rain, snow, fog and others were converted into clear=0 and inclement weather=1. Hour of the crash event (peak/off-peak) and day (weekday/weekend) was included in the analysis to account for varying traffic conditions.

For rural roads, 12 PM to 6 PM is usually an appropriate timeframe with hourly volume above 6% of daily traffic (peak). Moreover, it has been shown that traffic condition varies between weekdays and weekends (Zhong et al. 2005, Cardelino 1998). Table 5.1 illustrates the sixteen measured variables representing the characteristics of crash, road, environment, and driver used in the analysis.

Table 5.1 Description, Frequency, and Percentage of Selected Variables for WVCs

Variable	**Description**	**Frequency**	**Percentage**
Road Type			
0	Undivided	8070	78.57
1	Divided	2201	21.43
Accident Site			
0	Intersection	647	6.3
1	Non-intersection	9624	93.7
Hourly Traffic Conditions	*Peak /off-peak hour traffic volumes*		
0	Rest of the day	9124	88.83
1	12 PM - 6 PM	1147	11.17
Daily Traffic Conditions	*Weekend/Weekdays traffic volumes*		
0	Weekend	2877	28.01
1	Weekdays	7394	71.99
Road Surface Conditions			
0	Dry	8255	80.37
1	Wet, icy, muddy, slippery	2016	19.63
Road Pavement Conditions			
0	Normal	10010	97.46
1	Potholes, bumps, others	261	2.54
Age	*Driver's age*		
1	<20	841	8.19
2	20 - 30	2301	22.4
3	30 - 55	4959	48.28
4	55<	2170	21.13
Vehicle Type			
0	Automobile/Motorcycle	3546	34.52
1	Other	6725	65.48
Horizontal Alignment			

0	Straight	9911	96.49
1	Horizontal curve	360	3.51
Vertical Alignment			
0	Level	9819	95.6
1	Steep incline/decline	452	4.4
Lighting Condition			
0	Daylight	4739	46.14
1	Night time	5532	53.86
Weather Condition			
0	Clear	8885	86.51
1	Rain, Snow, Fog	1386	13.49
Vehicle Damage			
1	Light or no damage	4136	40.27
2	Moderate damage	5767	56.15
3	Demolished-Write off	368	3.58
Accident Injury Severity			
1	No injury (Property damage only)	9412	91.64
2	Injury	852	8.3
3	Fatal	7	0.07
Accident Cost			
1	<$10,000	8715	84.85
2	$10,000 – 20,000	1319	12.84
3	$20,000 – 30,000	168	1.64
4	$30,000<	69	0.67
Number of Injured People			
0	0	9412	91.64
1	1	692	6.74
2	2	141	1.37
3	3	26	0.25

5.3 Model Development

The analysis was performed with MPlus 8.3 (Muthén and Muthén 2017). Variables are treated as binary or ordered categorical (ordinal) variables in the model, and probit links are used for the estimations. Model fitting was by robust weighted least squares estimator using a diagonal weight matrix (WLSMV) that is the default estimator for ordered categorical observed dependent variables. For CFA and SEM, MPlus provides a summary of the goodness-of-fit measures. The root mean square error of approximation (RMSEA) is usually employed to test model fitting. A value less than 0.08 for RMSEA indicates good fit (Browne, M. W., and Cudeck 1993). Other criteria used in Mplus are the Bentler comparative fit index (CFI) and Tucker-Lewis index (TLI): values can lie between 0 and 1, and values closer to 1 indicate a better fit. It is worth to mention that the Chi-square was not used in this study due to the influences of large sample size and deviations from multi-normality (Golob 2003).

5.4 Factor Analysis

The factor analysis performed on 16 observed variables to classify observed variables into different latent constructs. Table 5.2 reports the varimax (orthogonal) factor analysis for observed variables. The primary outcome of the analysis is factor loadings. Their interpretation has to be based on a theoretical rationale (e.g., model hypothesis). Observed variables with loading close to 1 are thought to be highly influential in describing the factor, whereas loadings close to 0 are less influential. Therefore, factor loadings lower than 0.1 are usually not reported to facilitate interpretation.

The results showed that "Factor 1" loaded high on suggested variables for crash severity (i.e., crash injury severity, accident cost, vehicle damage, and the number of injured people). Factor 4 appeared to reflect a dimension in the data related to speeding attitude. Variables related to the road (e.g., whether the location was an intersection or not, whether the road was divided or not, pavement and surface conditions); variables related to the driver (e.g. driver's age, vehicle type); environmental variables (e.g. hourly/daily traffic conditions) are related to this factor (Zhong et al. 2005, Cardelino 1998).

Horizontal/vertical alignment (which affected sight distances) showed the highest loadings on "Factors 3" which could represent a separate latent construct (driver's visibility impairment) along with lighting condition. These geometric design features (horizontal/vertical alignment), are known to affect the horizontal sightline offset and stopping sight distance (AASHTO 2011), which can eventually affect the speeding propensity through reduced visibility. Weather condition showed its highest loading on "Factor 5", but was included in the VI factor since it can be related to this dimension (Abdel-Aty et al. 2011).

Overall, the factor analysis supported the hypothesized latent variables. In the next step, the relationship between these latent variables is tested.

Table 5.2 Factor Analysis Results for WVCs

Variables	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**
Suggested Crash Severity Variables						
Accident Injury Severity	0.985					
Accident Cost	0.926					
Vehicle Damage	0.103	0.133	-0.205	-0.132		
Number of Injured People	0.113	2.021	0.162	0.224		0.15
Suggested Speeding Attitude Variables						
Road Type	0.117	0.12	0.265	0.429	0.111	
Accident Site				0.664	-0.129	0.191
Hourly Traffic Conditions			0.101			-0.171
Daily Traffic Conditions						
Road Surface Conditions				-0.174	0.875	
Road Pavement Conditions			0.245	-0.252	0.36	
Vehicle Type				0.249		-0.779
Age			-0.111	-0.153		-0.287
Suggested Visibility Impairment Variables						
Horizontal Alignment			0.536			
Vertical Alignment			0.623			
Weather Condition				0.173	0.72	
Lighting Condition			-0.147	0.115		0.204

5.5 Confirmatory Factor Analysis

Afterward, the relationships (covariance parameters) among latent variables were tested using CFA (Figure 5.1).

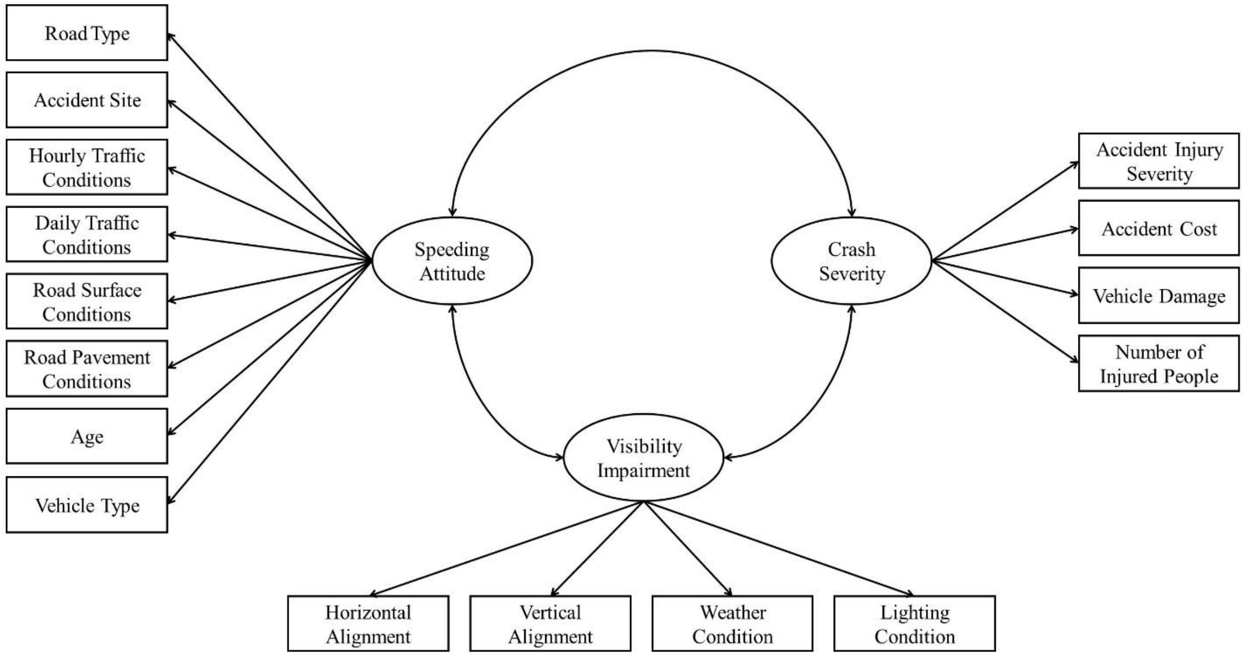


Figure 5.1 Proposed CFA Model with Three Latent Variables for WVCs

The correlations between crash severity and SA, VI and crash severity, VI, and SA were estimated and are shown in Table 5.3. Measured variables for each latent dimension were those suggested in Table 5.2.

Table 5.3 Confirmatory Factor Analysis Results for WVCs

Path			**Unstandardized Estimates**			**Standardized Estimates**		
			Estimate	**Standard Error**	**p-value**	**Estimate**	**Standard Error**	**p-value**
Speeding Attitude	<==>	Crash Severity	0.017	0.004	0	0.146	0.022	0*
Visibility Impairment	<==>	Crash Severity	0.033	0.008	0	0.398	0.053	0*
Visibility Impairment	<==>	Speeding Attitude	-0.035	0.008	0	-0.836	0.096	0*
Accident Severity	<==	Crash Severity	1	-	-	0.48	0.032	0*
Number of Injured People	<==	Crash Severity	0.996	0.011	0	0.478	0.032	0*
Accident Cost	<==	Crash Severity	0.517	0.043	0	0.248	0.022	0*
Vehicle Damage	<==	Crash Severity	1.728	0.197	0	0.829	0.046	0*

Road Type	<==	Speeding Attitude	1	-	-	0.241	0.021	0*
Accident Site			0.393	0.109	0	0.095	0.026	0*
Hourly Traffic Conditions			-0.374	0.096	0	-0.09	0.022	0*
Daily Traffic Conditions			-0.083	0.078	0.286	-0.02	0.019	0.284
Road Pavement Condition			-1.517	0.175	0	-0.366	0.033	0*
Road Surface Condition			-4.091	0.479	0	-0.988	0.048	0*
Vehicle Type			-0.454	0.082	0	-0.11	0.018	0*
Age			0.037	0.05	0.458	0.01	0.014	0.458
Horizontal Alignment	<==	Visibility Impairment	1	-	-	0.174	0.034	0*
Vertical Alignment			1.115	0.255	0	0.194	0.033	0*
Weather Condition			3.803	0.742	0	0.663	0.069	0*
Lighting Condition			0.224	0.112	0.045	0.039	0.018	0.031*

* Estimate significant at the 95% confidence level

** Estimate significant at the 90% confidence level

The goodness-of-fit values for RMSEA, CFI, and TLI are reported in Table 5.4, which shows an acceptable fit. In particular, RMSEA lower than 0.08 is usually employed to test the overall goodness-of-fit and, in this case, the criterion was satisfied being RMSEA equal to 0.04.

Table 5.4 CFA Goodness-of-Fit Statistics for WVCs

Fit Summary	**Value**
RMSEA	0.04
Bentler Comparative Fit Index (CFI)	0.85
Tucker Lewis Index (TLI)	0.82

The estimated correlations (significant at the 95% confidence level) between crash severity and SA, VI and crash severity, VI, and SA were 0.15, 0.4, and 0.84 respectively. CFA analysis also suggested the presence of a relationship between VI and SA, which was not taken into account at the modeling hypothesis stage. Now that the relationships (covariance parameters) among latent variables are tested, it is possible to move on to the SEM model.

5.6 SEM Analysis

In the next step, a full SEM was developed (Figure 5.2) using latent variables (circles) and measured variables (rectangles) introduced in the previous section. It is worth to mention that observed variables in a measurement model are obviously correlated because they are functions of the same latent dimension.

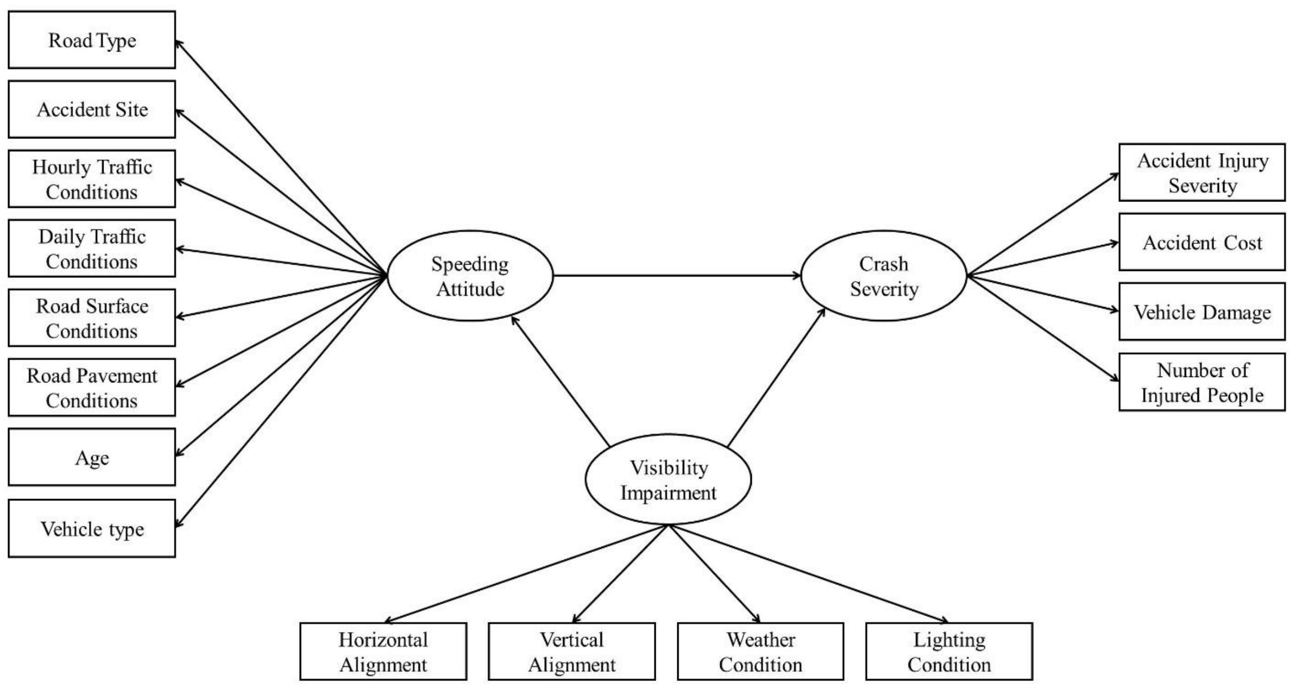


Figure 5.2 Proposed SEM Model with Three Latent Variables for WVCs

As showed in Figure 5.2, Visibility Impairment (VI) was set as a latent exogenous variable; Speeding Attitude (SA) and Crash Severity were treated as latent endogenous variables. VI can reflect the influence of lighting condition, weather condition, horizontal, and vertical alignment on the level of visibility impairment. SA was linked to measured variables representing infrastructure characteristics such as road type, accident site, road pavement condition, and road surface condition that can reduce or increase speeding attitude. Driver's characteristics such age and vehicle type which may directly affect the driver's speeding attitude as well as other environmental and traffic characteristics such as hourly and daily traffic conditions which are well-known factors affecting speeding propensity (HCM 2010).

Both VI and SA can directly influence the overall crash severity, which may be observed and measured as an increased probability of accident injury severity, vehicle damage, accident cost, and the number of injured people. Also, it was assumed from CFA that VI could, directly, influence SA and, indirectly, affect crash severity.

5.7 Estimate Results

Variables were treated as binary or ordinal in the model and ordered probit links were used for direct relationships (generalized SEM). The results of SEM analysis (unstandardized and standardized estimates) are reported in Table 5.5. For the unstandardized results, reference (measured) variables need to be selected for each latent dimension; their estimates are constrained to 1, and the effect of other measured variable is estimated (proportionally) against these reference variables.

It is worth to mention that the unstandardized estimates are expressed in original units of their variables. Therefore, it is difficult to compare the unstandardized coefficients among different pathways as units are various (e.g., some of them are expressed as binary and other as ordinal variables). Standardized estimates are based on standard deviation units of the variables and are usually used for comparison of parameter estimates of SEM models. This allows the comparison of different coefficients from different variables (Grace and Bollen 2005).

Table 5.5 SEM Estimates for WVCs

Path			**Unstandardized Estimates**			**Standardized Estimates**		
			Estimate	**Standard Error**	**p-value**	**Estimate**	**Standard Error**	**p-value**
Crash Severity	<==	Speeding Attitude	1.885	0.89	0.034	0.936	0.439	0.033*
		Visibility Impairment	4.167	2.279	0.068	1.046	0.453	0.021*
Speeding Attitude	<==	Visibility Impairment	-1.495	0.513	0.004	-0.756	0.092	<0.001*

Accident Injury Severity	<==	Crash Severity	1	-	-	0.487	0.033	<0.001*
Number of Injured People			0.998	0.011	<0.001	0.486	0.033	<0.001*
Accident Cost			0.519	0.043	<0.001	0.253	0.022	<0.001*
Vehicle Damage			1.676	0.194	<0.001	0.816	0.046	<0.001*
Road Type	<==	Speeding Attitude	1	-	-	0.242	0.022	<0.001*
Accident Site			0.395	0.109	<0.001	0.095	0.026	<0.001*
Hourly Traffic Conditions			-0.37	0.095	<0.001	-0.09	0.022	<0.001*
Daily Traffic Conditions			-0.085	0.078	0.275	-0.02	0.019	0.273
Road Pavement Conditions			-1.506	0.174	<0.001	-0.364	0.033	<0.001*
Road Surface Conditions			-4.114	0.485	<0.001	-0.994	0.049	<0.001*
Vehicle Type			-0.455	0.082	<0.001	-0.11	0.018	<0.001*
Age			0.035	0.05	0.482	0.01	0.014	0.482
Horizontal Alignment	<==	Visibility Impairment	1	-	-	0.122	0.036	0.001*
Vertical Alignment			1.535	0.492	0.002	0.188	0.034	<0.001*
Weather Condition			6.04	1.88	0.001	0.738	0.082	<0.001*
Lighting Condition			0.318	0.183	0.082	0.039	0.019	0.04*

* Estimate significant at the 95% confidence level

** Estimate significant at the 90% confidence level

Table 5.6 shows that the developed Generalized SEM model for analyzing the severity of WVCs fitted the data well for all the goodness-of-fit criteria. In particular, RMSEA lower than 0.08 is usually employed to test the overall goodness-of-fit and, in this case, the criterion was satisfied being RMSEA equal to 0.04.

Table 5.6 SEM Goodness-of-Fit Statistics for WVCs

Fit Summary	**Value**
RMSEA	0.04
Bentler Comparative Fit Index (CFI)	0.85
Tucker Lewis Index (TLI)	0.82

Standardized results demonstrated that both SA and VI positively influenced the overall crash severity, with estimates equal to 0.936 and 1.046, respectively. These coefficients were found to be statistically significant at the 95% confidence level, which supported the model hypothesis. Moreover, the results showed that higher VI would negatively affect SA (parameter estimate equal to -0.756), which would indirectly decrease crash severity. These results are in agreement with Visintin et al. (2018) and Vanlaar et al. (2012) which demonstrated that pre-crash speed and visibility are major influencing factors of crash severity.

The indirect effect of VI on crash severity can be estimated by multiplying factor loadings of VI/SA direct relationship and SA/crash severity relationship, as:

$$\text{VI indirect effect} = -0.756 \times 0.936 = -0.708 \tag{5.1}$$

The total effect of VI on crash severity can be calculated by summing up direct and indirect effects of VI, as:

$$\text{VI total effect} = -0.708 + 1.046 = 0.338 \tag{5.2}$$

Therefore, while visibility impairment can cause more severe crashes, it can also decrease the severity through reduced speeding attitude. However, the adverse effect of visibility impairment was found to be greater than its benefits, which was confirmed by the positive total effect on crash severity Eq. (5.2).

Regarding standardized regression coefficients of SA measurement model, divided roadways (road type variable) and non-intersection sites (accident site variable) predicted higher crash severity through increased speeding attitude than undivided roads and intersection sites, being their estimates equal to 0.242 and 0.095, respectively. On the contrary, a decrease of crash severity through less speeding attitude was recorded for peak (hourly) traffic volumes (-0.09), poor pavement conditions (-0.364), wet road surface (-0.994), and heavy vehicles (-0.11). For this latent variable, road surface conditions showed the most considerable influence on speeding attitude being the estimate equal to -0.994; on the contrary, the peak hour traffic showed the least influence with an estimate of -0.09. Driver's age and daily traffic condition estimates were not statistically significant in the SEM model.

Regarding visibility impairment as a latent variable, all variables showed positive and significant regression parameters. Roadways curves, rolling terrain could cause more severe crashes having estimated equal to 0.122 and 0.188 respectively. In addition, inclement weather and dark periods of the day also showed to results in more severe crashes having estimates equal to 0.738 and 0.039 respectively. Among all these variables, inclement weather showed the most considerable influence being the estimate equal to 0.738; on the contrary, lighting condition showed the least influence with a value equal to 0.039.

Finally, all variables measuring crash severity, i.e., accident injury severity, vehicle damage, accident cost, and the number of injured people showed positive and significant relation with crash severity having values of 0.487, 0.816, 0.253, and 0.486 respectively. This implies that a higher WVC severity is resulted from an increase in crash injury severity, vehicle damage, accident cost, and the number of injured people, as expected. Further discussion on the results is reported in Chapter 6.

5.8 Comparison with Traditional Crash Severity Modeling

To compare SEM results with traditional crash severity modeling, a path analysis was performed using probit links, which would mimic an ordered probit crash severity model. As mentioned before, path analysis is the application of SEM without latent variables (ordinary regression analysis). Therefore, only observed variables in Table 5.1 were employed, and each parameter measuring crash severity (y_i) (i.e., crash injury severity, number of injured people, accident cost, and vehicle damage) was directly affected by all measured variables related to infrastructure, driver, vehicle, traffic, and environmental factors (vector X) (Figure 5.3).

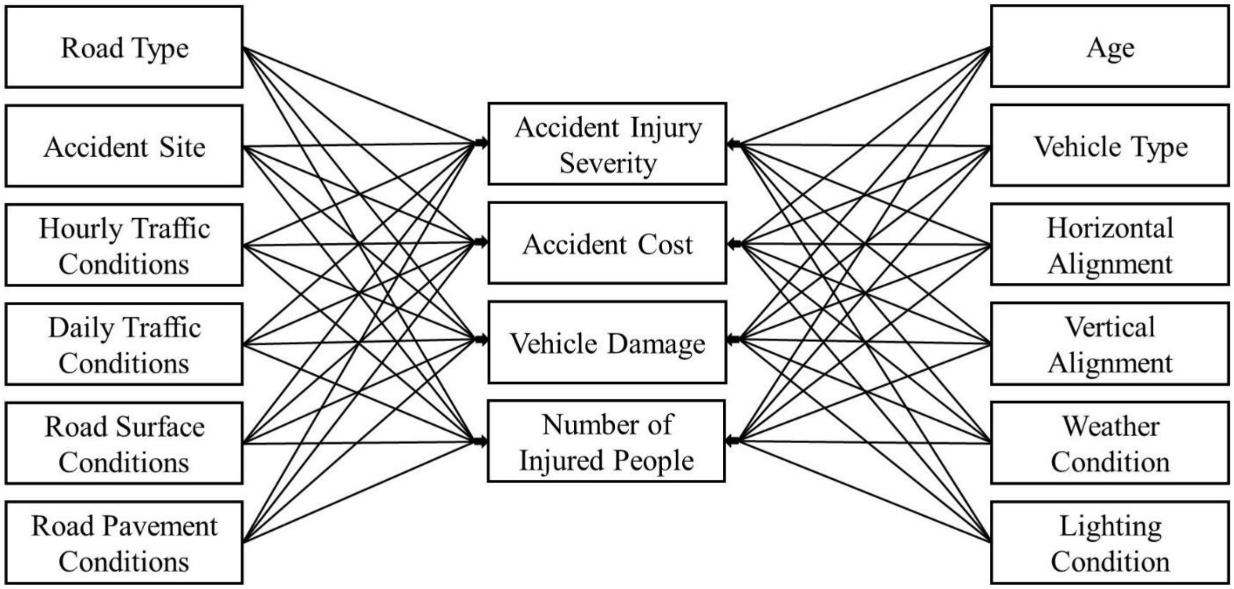


Figure 5.3 Path Analysis Model for WVCs

The use of ordered probit models has been a common practice in traditional crash severity analysis (Abdel-Aty 2003, Kockelman and Kweon 2002) and, therefore, path relationships with ordered probit links were employed to build the model. As mentioned before, for vector X, the probability that the parameter measuring crash severity of an individual crash (y_i) belongs to each category k=1, ..., K (e.g., K=4 for accident cost in this study) is:

$$\begin{cases} P(y_i = 1) = \phi(-\beta X) \\ P(y_i = k) = \phi(\gamma_{k-1} - \beta X) - \phi(\gamma_{k-2} - \beta X) \\ P(y_i = K) = 1 - \phi(\gamma_{k-1} - \beta X) \end{cases} \tag{5.3}$$

where β is a vector of the estimable parameters for each y_i, $\gamma=\{\gamma_1,...,\gamma_k,\ldots,\gamma_{K-1}\}$ are the threshold values for all crash severity categories, and ϕ $(\cdot)$ stands for the cumulative probability function of the standard normal distribution.

Table 5.7 shows the results of path analysis and the comparison with SEM in terms of significance, and the sign of the coefficients. Magnitudes in path analysis are expressed in original units of variables, so they are not directly comparable with SEM standardized estimates (and among different y_i).

Table 5.7 Comparison Between Path Analysis and SEM Results for WVCs (Unstandardized Coefficients)

Vector X	Path	y_i	Estimate (vector β)	Standard Error	p-value	Path sign	SEM sign
Road Type	==>	Accident Injury Severity	0.072	0.043	0.098	+**	+*
		Number of Injured People	0.061	0.043	0.155	+	
		Accident Cost	-0.085	0.037	0.023	-*	
		Vehicle Damage	0.278	0.028	<0.001	+*	
Accident Site	==>	Accident Injury Severity	0.508	0.108	<0.001	+*	+*
		Number of Injured People	0.507	0.107	<0.001	+*	
		Accident Cost	-0.106	0.062	0.087	-**	
		Vehicle Damage	0.873	0.052	<0.001	+*	
Hourly Traffic Conditions	==>	Accident Injury Severity	-0.208	0.069	0.003	-*	-*
		Number of Injured People	-0.202	0.069	0.003	-*	
		Accident Cost	-0.027	0.053	0.608	-	
		Vehicle Damage	-0.068	0.039	0.08	-**	
Daily Traffic Conditions	==>	Accident Injury Severity	-0.007	0.04	0.86	-	-
		Number of Injured People	-0.007	0.04	0.859	-	
		Accident Cost	0.03	0.033	0.371	+	
		Vehicle Damage	-0.007	0.025	0.795	-	
Road Pavement Conditions	==>	Accident Injury Severity	0.264	0.103	0.011	+*	-*
		Number of Injured People	0.229	0.107	0.031	+*	
		Accident Cost	0.044	0.091	0.628	+	
		Vehicle Damage	0.44	0.075	<0.001	+*	
Road Surface Conditions	==>	Accident Injury Severity	-0.15	0.053	0.005	-*	-*
		Number of Injured People	-0.142	0.052	0.006	-*	
		Accident Cost	0.076	0.04	0.059	+**	
		Vehicle Damage	-0.216	0.031	<0.001	-*	
Vehicle Type	==>	Accident Injury Severity	-0.275	0.037	<0.001	-*	-*
		Number of Injured People	-0.25	0.037	<0.001	-*	
		Accident Cost	0.24	0.033	<0.001	+*	
		Vehicle Damage	-0.233	0.024	<0.001	-*	

Age	==>	Accident Injury Severity	0.047	0.022	0.033	+*	+
		Number of Injured People	0.048	0.022	0.027	+*	
		Accident Cost	0.012	0.018	0.522	+	
		Vehicle Damage	-0.05	0.013	<0.001	-*	
Horizontal Alignment	==>	Accident Injury Severity	0.149	0.09	0.097	+**	+*
		Number of Injured People	0.147	0.088	0.092	+**	
		Accident Cost	-0.111	0.082	0.178	-	
		Vehicle Damage	0.397	0.066	<0.001	+*	
Vertical Alignment	==>	Accident Injury Severity	0.183	0.08	0.023	+*	+*
		Number of Injured People	0.208	0.076	0.006	+*	
		Accident Cost	-0.107	0.077	0.163	-	
		Vehicle Damage	0.284	0.055	<0.001	+*	
Weather Condition	==>	Accident Injury Severity	0.232	0.054	<0.001	+*	+*
		Number of Injured People	0.227	0.053	<0.001	+*	
		Accident Cost	-0.097	0.047	0.037	-*	
		Vehicle Damage	0.422	0.036	<0.001	+*	
Lighting Condition	==>	Accident Injury Severity	0.06	0.039	0.126	+	+*
		Number of Injured People	0.068	0.039	0.079	+**	
		Accident Cost	0.104	0.033	0.001	+*	
		Vehicle Damage	0.063	0.025	0.012	+*	

* Estimate significant at the 95% confidence level

** Estimate significant at the 90% confidence level

It is noteworthy to mention that since the path model is just identified with zero degrees of freedom, the software does not provide the goodness of the fit for the path analysis. Thus, it is not possible to compare the fit of the models. However, it is possible to compare the significance, magnitude, and the signage of the estimates. The results showed that out of 48 path estimates, 30 were significant at 95% confidence level. For the remaining 18 path estimates, 7 were significant at 90% confidence level and 11 were found not to influence crash severity variables significantly.

Accident cost as severity crash outcome, was not statistically significant for 6 out of 12 explanatory variables, while it was found to be significant in SEM. This might be due to the fact that a simple path analysis can account for direct relationships only and is unable to properly account for the indirect effect of variables on crash severity through latent and unobserved variables. Regarding the sign of significant estimates, both models were found to be in good agreement for variables such as accident site, hourly traffic condition, road surface condition, horizontal and vertical alignment, and lighting condition. Further discussion on the results is presented in Chapter 6.

CHAPTER 6

DISCUSSION

In this chapter, a discussion on the results of SEM analysis and the findings in the literature for RLR-related crashes and WVCs will be carried out, respectively. Moreover, a comparison between path analysis and SEM analysis for both crash types will be conducted. Finally, a comparison between the two SEM analyses for RLR and WVC models will be presented, respectively.

6.1 RLR-Related Crashes Results

Three latent variables were hypothesized for RLR-related crashes: the level of crash severity and two casual factors affecting crash severity, i.e., pre-crash travel speed (TS) of the bullet vehicle, and the kinetic energy (KE_s) transferred from the bullet vehicle to the subject vehicle(s) during the crash event. The results demonstrated that both pre-crash TS of the bullet vehicle and transferred KE_s to the subject vehicle(s) positively influence the overall crash severity, being their estimates equal to 0.26 and 0.36, respectively. These results are in agreement with Wang et al. (Wang and Qin 2014) which demonstrated that pre-crash speed (as a latent variable) was a major influencing factor of crash severity. Regarding the transferred KE_s to the subject vehicle(s), the results are in agreement with findings of Sobhani et al. (2011) and Corben et al. (2004) where it was shown that the transferred KE_s is a significant factor for increased crash severity.

The results showed that TS increase could positively affect transferred KEs to the subject vehicle (parameter estimate equal to 0.344), which would indirectly increase crash severity. These results are in agreement with the fundamental of physics where the kinetic energy of each vehicle is related to the mass and the speed of the vehicle (Corben et al. 2004, Halliday et al. 2004).

Regarding the variables that measure TS of the bullet vehicle, both peak-hour traffic conditions and older drivers showed a decrease in crash severity through reduced speeds being their factor estimates equal to -0.201 and -0.084, respectively. Regarding driver’s age, results are consistent with other studies (Kim et al. 2011, Wang and Qin 2014, Lee et al. 2018) where it was found that lower crash severity was associated to older drivers. As for peak/off-peak hour of traffic, the sign of the parameters is supported by traffic operations principles where lower travel speeds are associated with peak-hour traffic (HCM 2010). High pavement skid resistance, weekend traffic conditions, alcohol-impaired driving conditions, night conditions, downhill grade were found to increase TS, and subsequently cause an increase in crash severity.

Regarding skid resistance, higher scores represent the better condition of the pavement surface, which may result in an increase of speeds by drivers. Additionally, the probability of higher speeds in the weekend traffic conditions can be explained by the fact that less traffic is usually recorded in the weekend than during weekdays which is supported by the studies in the literature (Zhong et al. 2005, Cardelino 1998). Concerning lighting condition, low luminance and limited visibility can affect driver's perception and expectations; this can result in different actions and behavior between daytime and nighttime driving, especially in terms of speeds perceived, desired, and adopted (Planek et al. 2014). Besides, according to Retting et al. (1999), red-light runners at night time were more likely to be young drivers with poor driving records and alcohol impairment (58% of night-time crashes), all of which can result in the inability to reduce speed and higher crash severity.

Alcohol-impaired driving conditions showed the most considerable influence on severity being the estimate equal to 0.209; on the contrary, the age of the driver showed the least influence with an estimate of -0.084. This latter result is in agreement with Retting et al. (1999), which found that the majority of red-light runners (two-thirds of the sample analyzed) were under the influence of alcohol, which led to an increase of crashes. Gender and road surface condition estimates were found to not be statistically significant in the SEM model.

Regarding transferred KE_s as a latent variable, the number of vehicles involved in RLR-related crashes showed a positive coefficient which is in agreement with Equation 4.1 where m_s grows with the number of vehicles involved in a crash. These results are in agreement with fundamental of physics where the kinetic energy of each vehicle is related to the mass and the speed of the vehicle (Halliday et al. 2004), in this case, the number of vehicles involved in the crash can affect the mass and eventually increase the transferred KE_s. The remaining variables (point of impact, vehicle maneuver and vehicle year) showed negative standardized coefficients.

With respect to the point of impact, the negative coefficient means that head-on impacts could transfer more kinetic energy to the subject vehicle(s) compared to impacts with other parts of the vehicle. Similarly, straight movements demonstrated more influence on transferred kinetic energy compared to other types of maneuvers. These two variables were included in the model as a proxy of ΔV_s (Equation 4.1) since speed variation is a function of crash characteristics (Sobhani et al. 2011). Sobhani et al. (2011) demonstrated that KE_s varies with crash angle and the impacted part of the subject vehicle, where head-on crashes were found to cause the most severe consequences; this confirmed the results obtained in this study. Moreover, in research carried out by Evans (1994), it was shown that higher speed changes occurred during frontal impacts which cause higher injury severity.

Vehicle year showed to negatively affect the transferred $KE_{s,}$ implying that the newer the vehicle, the less the crash severity. This finding is supported by the findings of studies done by National Highway Traffic Safety Administration (NHTSA 2004) and Huelke et al. (1968) where it was concluded that less energy would be transferred and absorbed by newer vehicles during frontal impacts.

Among all these variables, the number of involved vehicles showed the most considerable influence being the estimate equal to 0.342; on the contrary, vehicle year showed the least influence with a value equal to -0.137. The estimated coefficient for the subject vehicle type was found to not be statistically significant.

Finally, all variables representing crash severity, i.e., injury severity, vehicle damage, and the number of injured people showed positive relation with crash severity having values of 0.812, 0.440, and 0.842, respectively. These findings are in agreement with the papers in the literature which had similar conclusions (Manner and Wünsch-Ziegler 2013, Lee et al. 2008, Kim et al. 2011, Wang and Qin 2014). This implies that a higher crash severity index produces an increase of crash injury severity, property damage and the number of injured people, as expected.

6.1.1 Comparison of SEM with Path Analysis Results (RLR Model)

For path analysis, it was assumed that the parameters related to crash severity (i.e., injury severity, number of injured people, and total vehicle damage) were directly affected by measured variables related to infrastructure, driver, vehicle, traffic, environmental, and crash characteristics. For SEM analysis, it was hypothesized that the observed parameters describe latent dimensions (TS or the transferred KE_s) in the data and could indirectly affect the crash severity. Both of these analyses were performed using the weight-least squares (WLS) method with the default weight matrix, which yields asymptotically normal estimates regardless of the probability distribution of the population. WLS was used to account for possible violation of the multivariate normality assumption (Bollen 1989). Moreover, the WLS method can analyze binary and ordinal variables (regardless of having continuous variables) which is suitable for the variables considered in this investigation.

To compare SEM analysis results with path analysis, it is possible to evaluate the goodness-of-fit, significance, and the signage of the estimates for both models. Regarding the goodness-of-fit for path analysis the values for four different criteria were as follows: SRMR = 0.03, RMSEA = 0.24, AGFI = 0.99, and CFI = 0.84. For the SEM analysis the estimated goodness-of-fit values were as follows: SRMR = 0.07, RMSEA = 0.03, AGFI = 1.00 and CFI = 0.86. For SEM analysis, all the criteria showed an acceptable fit. For path analysis, the majority of the criteria showed an acceptable fit except for the RMSEA value, which was not in the acceptable range. This result showed that the SEM analysis outperforms the path analysis by having a better fit for the data.

The results for the path analysis showed that only 18 path estimates out of 42, were statistically significant at 95% confidence level. For the remaining 24 path estimates, 9 were significant at 90% confidence level, and 15 were found not to influence crash severity variables significantly. Regarding SEM analysis, only 3 path estimates were insignificant, i.e., gender, road surface condition, and subject vehicle type. Gender and subject vehicle type were insignificant in both models, which showed that these variables do not affect the crash severity regardless of the analysis method. Regarding road surface condition, while SEM model showed that this variable is not significant, path analysis showed significant negative values for injury severity and number of injured people and an insignificant negative value for total vehicle damage.

Both models were found to be in good agreement for variables such as vertical alignment, weekday/weekend traffic condition, peak /off-peak hour, vehicle maneuver, the number of involved vehicles and the point of impact, which had a significant effect on all the crash severity related dependent variables. Apart from these 6 variables, no other variable showed significant coefficients for all three of the crash severity variables. However, regarding the signage of the significant coefficients, both models were found to be in good agreement for variables such as alcohol involvement, vertical alignment, weekday/weekend traffic condition, peak /off-peak hour, vehicle maneuver, point of impact, and vehicle year. Moreover, for the path analysis, variables such as lighting condition, skid resistance, and age, showed insignificant estimates for all the dependent variables.

Overall, the SEM model was able to explain observed data better, having less insignificant variables. Besides, the causality in path analysis was less straightforward than SEM. For example, the sign of variables such as alcohol involvement, vertical alignment, weekday/weekend traffic condition, and the number of involved vehicles, indicated that these variables contributed directly to crash severity increase; on the contrary, the sign of variables such as peak /off-peak hour, road surface condition, vehicle maneuver, point of impact, and vehicle year indicated that these variables are decreasing the crash severity. However, these direct paths did not provide clear evidence that the occurrence of one variable caused the other. On the contrary, causal relationships in SEM seemed more straightforward as latent variables as mediating factors indicated that the observed variables could influence the crash severity through either pre-crash travel speed (TS) or the transferred kinetic energy (KE_s).

6.2 WVCs Results

In this part of the study, three latent variables were hypothesized: the level of crash severity (CS), driver's speeding attitude (SA) and driver's visibility impairment (VI). The standardized results demonstrated that both SA and VI positively influenced the overall crash severity, with estimates equal to 0.936 and 1.046, respectively. These coefficients were found to be statistically significant at the 95% confidence level, which supported the model hypothesis. Moreover, the results showed that higher VI would negatively affect SA (parameter estimate equal to -0.756), which would indirectly decrease crash severity. These results are in agreement with Visintin et al. (2018) and Vanlaar et al. (2012) which demonstrated that pre-crash speed and visibility are a major influencing factor of crash severity.

Regarding standardized regression coefficients of SA measurement model, divided roadways (road type variable) and non-intersection sites (accident site variable) predicted higher crash severity through increased speeding attitude than undivided roads and intersection sites, being their estimates equal to 0.242 and 0.095, respectively. Regarding divided roads, results are consistent with other studies where it was found that the severity of crashes on divided roads are higher due to higher speeds (Rifaat et al. 2011). As for accident site, for intersection traffic controls could usually be seen well ahead of the intersections and drivers might be more cautious when approaching the intersection. Consequently, drivers might be able to respond quickly and reduce their vehicle speed significantly if needed, thereby resulting in lower crash severity (Rifaat et al. 2011).

On the contrary, a decrease of crash severity through less speeding attitude was recorded for peak (hourly) traffic volumes (-0.09), poor pavement conditions (-0.364), wet road surface (-0.994), and heavy vehicles (-0.11). Peak hour traffic (-0.09), roads with potholes (-0.364), roads with a wet surface (-0.994), and trucks (-0.11) showed a decrease the crash severity through having a less speeding propensity. Regarding hourly traffic conditions, the sign of the parameters is supported by traffic operations principles where free-flow conditions were associated with off-peak traffic hours when drivers can assume higher speed levels (HCM 2010). As for lower speeding attitude for roads with wet surface and potholes, the results are consistent with the research in the literature (Edwards 1999, Setyawan et al. 2015).

For vehicle type, heavier vehicles showed a decrease in speeding attitude, which is in agreement with the findings of the study done by Wang et al. (2014). For this latent variable, road surface conditions showed the most considerable influence on speeding attitude being the estimate equal to -0.994; on the contrary, the of the peak hour traffic showed the least influence with an estimate of -0.9. Driver's age and daily traffic condition estimates were found to not be statistically significant in the SEM model.

Regarding visibility impairment as a latent variable, all variables namely horizontal and vertical alignment, inclement weather, and dark periods of the day showed positive and significant regression parameters; having estimates equal to 0.122, 0.188, 0.738, and 0.039 respectively. Concerning geometric features (horizontal and vertical alignment), the positive coefficient means that the roadways curves, rolling terrain could cause more severe crashes by having adverse effects on the horizontal sightline offset and stopping sight distance (AASHTO 2011). Besides, inclement weather and dark periods of the day also showed to result in more severe crashes by impairing the visibility.

Regarding the inclement weather, the results are in agreement with previous studies (Andrey and Yagar 1993, Abdel-Aty et al. 2011), where it was concluded that inclement weather could cause higher injury and fatality rates by impairing the visibility. Regarding the lighting condition, the results are in agreement with studies in the literature (Lachenmayr et al. 1998, Owens and Sivak 1993), where it was established that nighttime driving results in impaired visibility, less reaction time, and more severe crashes. Among all these variables, inclement weather showed the most significant influence having estimate equal to 0.738; on the contrary, lighting condition showed the least influence with a value equal to 0.039.

Finally, all variables measuring crash severity, i.e., accident injury severity, vehicle damage, accident cost, and the number of injured people showed positive and significant relation with crash severity having values of 0.487, 0.816, 0.253, and 0.486 respectively. These findings are in agreement with the papers in the literature which had similar conclusions (Kim et al. 2011, Manner and Wünsch-Ziegler 2013, Wang and Qin 2014). This implies that a higher WVC severity is produced with an increase in crash injury severity, vehicle damage, accident cost, and the number of injured people, as expected.

6.2.1 Comparison of SEM with Path Analysis Results (WVC Model)

For this part, a path analysis using ordered probit links were performed. The use of ordered probit models has been a common practice in traditional crash severity analysis (Abdel-Aty 2003, Kockelman and Kweon 2002). It was assumed that the parameters related to crash severity (i.e., accident injury severity, accident cost, number of injured people, and vehicle damage) are ordered categorical and are directly affected by measured variables related to infrastructure, driver, vehicle, traffic, environmental, and crash characteristics.

For the SEM analysis, it was hypothesized that the observed parameters are describing latent dimensions (SA and VI) in the data and are indirectly affecting the crash severity. The crash severity latent variable was measured by accident injury severity, accident cost, number of injured people, and vehicle damage which were assumed as ordered categorical. Both of these analyses were performed using a robust weighted least squares estimator using a diagonal weight matrix (WLSMV) which is the default estimator for ordered categorical observed dependent variables (Muthén and Muthén 2017).

To compare the SEM analysis results with the path analysis, it is possible to evaluate the goodness-of-fit, significance, and the signage of the estimates for both models. For the SEM analysis the estimated goodness-of-fit values were as follows: RMSEA = 0.04, CFI = 0.85 and TLI = 0.82. For the SEM analysis, all the criteria show an acceptable fit. However, the path model was just identified with zero degrees of freedom, and the software did not provide the goodness of the fit for the path analysis in this case. Thus, it was not possible to compare the fit of the models.

The results of the path analysis showed 11 insignificant path estimates, while the SEM model only showed two insignificant variables. Daily traffic conditions were insignificant in both models, which showed that this variable does not affect the crash severity regardless of the analysis method. Regarding driver's age, while the SEM model showed that driver's age is not significantly influential on the speeding propensity, the path analysis showed significant positive values for injury severity and the number of injured people and a significant negative value for vehicle damage. Accident cost was not significantly affected by the driver's age.

Concerning the sign of significant estimates, both models were found to be in good agreement for variables such as accident site, hourly traffic condition, road surface condition, horizontal and vertical alignment, and lighting condition. However, the path analysis showed conflicting signs for some variables. For example, path analysis estimate for roads with potholes showed to increase all the crash severity indicators. In the SEM model, however, this variable showed a decrease of crash severity for poor pavement conditions through lower speeding attitude.

Divided roads in the path analysis showed a positive effect on vehicle damage and a negative effect on accident cost, while the effect on injury severity and the number of injured people was insignificant. This variable showed a positive effect on crash severity through higher speeding propensity in the SEM model. Heavier vehicles showed a decrease in all the severity indicators except for accident cost. In the SEM model, this variable decreased the crash severity through lower speeding propensity. Similarly, inclement weather had the same positive signage for both SEM and path analysis except for the accident cost, which showed a negative sign.

Overall, the SEM model is able to explain the data better, with less insignificant variables and more consistent signs among severity-related variables. Besides, the causality in path analysis was less straightforward than SEM. The direct paths did not provide clear evidence that the occurrence of one variable caused the other. On the contrary, causal relationships in SEM seemed more straightforward as latent variables as mediating factors indicated that the observed variables could influence the crash severity through either speeding attitude (SA) or visibility impairment (VI).

6.3 Comparison of SEM Results between RLR and WVC Model

To investigate and compare the two different SEM analyses between RLR-related crashes and WVCs, it is essential to highlight the differences between these two models. The modeling and analysis for the RLR-related crashes were performed in SAS software version 3.71, University Edition (SAS Institute Inc 2017), which allows implementing SEM analysis through the CALIS procedure. A dataset of 1,601 RLR-related crash records in Florida (US) from 2011 to 2014 was used in the analysis. Only multiple-vehicle crashes were included to investigate the model hypothesis. Variables were treated as binary and ordinal parameters in the model, and linear links were employed for the estimations. The weight-least squares (WLS) method with the default weight matrix was used, which yields asymptotically normal estimates regardless of the probability distribution of the population. Furthermore, SAS provides a summary of the goodness-of-fit measures; namely, the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), Bentler comparative fit index (CFI), and adjusted GFI (AGFI).

Regarding WVCs, the modeling and analysis were performed with MPlus 8.3 (Muthén and Muthén 2017). A dataset of 10,271 wildlife-vehicle crash records in Saskatchewan (Canada) from 2012 to 2018 were used in the analysis. Only single-vehicle crashes away from major urban centers with “animal action” coded as “first major contributing factor” were considered in the analysis. Variables were treated as binary or ordered categorical (ordinal) variables in the model, and probit links were employed for the estimations. Model fitting was by robust weighted least squares estimator using a diagonal weight matrix (WLSMV) which is the default estimator for ordered categorical observed dependent variables. Furthermore, MPlus provides a summary of the goodness-of-fit measures; namely, the root mean square error of approximation (RMSEA), Bentler comparative fit index (CFI), and Tucker-Lewis index (TLI).

Considering the fundamental differences between the two models, it is possible to evaluate the significance, and sign of the estimates for both models. It is noteworthy that comparing the goodness-of-fit between the two models was not possible since the models were developed in different software using different types of links and estimators.

For the RLR model, three latent variables were hypothesized (crash severity, pre-crash travel speed (TS), and transferred kinetic energy (KE_s)). The results demonstrated that both TS and KE_s positively influence the overall crash severity, having significant estimates equal to 0.258 and 0.361, respectively. Moreover, the results showed that TS increase could positively influence transferred KE_s (parameter estimate equal to 0.344), which would indirectly increase crash severity. By calculating the total effect of TS (0.382), it was shown that the pre-crash TS of the bullet vehicle could have more influence on crash severity than the transferred KE_s. For the RLR model, the measured crash severity related variable that had the most influence on the overall crash severity latent variable was the number of the injured people (0.842).

For the WVC model, three latent variables were hypothesized (crash severity, speeding attitude (SA), and visibility impairment (VI)). The results demonstrated that both SA and VI positively influence the overall crash severity, with significant estimates equal to 0.936 and 1.046, respectively. Moreover, the results showed that higher VI would negatively affect SA (parameter estimate equal to -0.756, which would indirectly decrease crash severity. By calculating the total effect of VI (0.338), it was shown that the speeding attitude (SA) could have more influence on crash severity than the visibility impairment. For the WVC model, the measured crash severity related variable that had the most influence on the overall crash severity latent variable was the vehicle damage (0.816).

Regarding the variables that are similar in both models, the hourly traffic conditions (peak/off-peak hour) showed to affect the crash severity through reduced TS for the RLR-related crashes model; this variable showed to affect the crash severity through reduced SA for WVC model. It is worth to mention that the peak hour for RLR model was chosen according to the urban setting, whereas in the WVC model, this variable was defined according to the rural setting. Vertical alignment (rolling terrain) showed to positively affect the crash severity through increased TS and increased VI for RLR and WVC model, respectively. Lighting condition was shown to influence the crash severity through VI for WVC model. For the RLR model lighting condition (night time driving) was showed to affect the crash severity through increased TS.

Daily traffic condition (weekend/weekday) was not significant in the WVC model; however, in the RLR model, this variable was found to positively affect crash severity through higher TS in weekends. Road surface condition (dry/wet) was not found significant in affecting crash severity in the RLR model; this variable was found to negatively affect crash severity through reduced SA in the WVC model. Driver's age was not found to significantly affect crash severity in the WVC model; this variable was affecting the crash severity through lower TS for the older drivers in the RLR model at 90% confidence level. Accident injury severity showed to positively affect the overall crash severity latent variable in both models at 95% confidence level. Similarly, vehicle damage showed to have a positive significant effect on the overall crash severity latent variable in both models. Vehicle damage was measured in monetary value in the RLR model, this variable was measured by qualitative assessment of the functional damage severity sustained by a vehicle.

The rest of the variables were not comparable between the two models. Concerning the RLR model, only three path estimates were insignificant, namely gender, road surface condition, and subject vehicle type. For the WVC model, only two path estimates were insignificant namely daily traffic conditions and driver's age.

CHAPTER 7

CONCLUSIONS

7.1 Summary of Findings

Overall, both SEM specifications (RLR and WVC) showed an acceptable fit, while the hypothesized latent dimensions in both models seemed to fit the data well. Regarding the RLR model, standardized results demonstrated that TS and KE_s positively influenced overall crash severity. Their respective were equal to 0.26 and 0.36, respectively. The loading factors (i.e., parameter estimates) among latent variables were found to be statistically significant, which again supported the model hypothesis.

Moreover, the results showed that TS increase could positively affect KE_s (parameter estimate equal to 0.344), which would in turn indirectly increase crash severity. Regarding variables measuring TS, alcohol-impaired driving conditions showed the greatest influence on severity with an estimate equal to 0.209; while on the contrary, driver age showed the least influence with an estimate of 0.084. Regarding the variables measuring KE_s, the number of involved vehicles showed most considerable influence, with an estimate equal to 0.342; while vehicle model year showed the least influence with a value equal to 0.137. The number of injured people was found to be the most influential variable for crash severity, and total vehicle damage was the least influential, with estimates of 0.842 and 0.440, respectively.

Regarding the WVC specification, standardized results demonstrated that both SA and VI positively influenced overall crash severity, with estimates equal to 0.936 and 1.046, respectively. These coefficients were found to be statistically significant at the 95% confidence level, supporting the model hypothesis. Moreover, the results showed that higher VI would negatively affect SA (parameter estimate of -0.756), which would indirectly decrease crash severity. Regarding the variables measuring SA, road surface condition showed the most influence on severity, with an estimate of -0.994; conversely, peak hour traffic showed the least influence with an estimate of

-0.09. Regarding the variables measuring VI, inclement weather showed the greatest influence with an estimate equal to 0.738. Lighting condition was found to have the least influence with an estimate of 0.039. Vehicle damage was found to be the most influential variable on crash severity, while accident cost was the least influential. These variables had estimates equal to 0.816 and 0.253, respectively.

In comparison to path analyses (i.e. a traditional severity model), both SEM analyses generated superior results with respect to goodness-of-fit, significance, and signs of the estimates. Also, the causal relationships estimated in SEM seemed more straightforward since the observed variables could influence crash severity through latent variables as mediating factors. Therefore, I conclude that SEM analysis outperformed traditionally used severity models.

7.2 Research Implications

As suggested in Savolainen et al. (2011), a thorough understanding of factors that affect the frequency and severity of crashes is critical for the development of effective countermeasures and strategies that may mitigate or exacerbate the likelihood or degree of injury sustained by crash-involved road users. Since SEM provides weights for each contributing factor of crash severity, traffic engineers and traffic-related decision makers can readily employ the proposed methodology and generated results for policymaking purposes and to prioritize infrastructure strategies at crash-prone locations.

For instance, regarding RLR-related crashes this analysis indicated that alcohol consumption had the most substantial direct effect on the speed of RLR vehicles, a factor that also indirectly affected the severity of the crash. Even though in many countries regulations against impaired driving already exist, a study by Elvik et al. (2009) reported the limited effect of current legislation on this controversial issue, leading to an insignificant deterrence effect on young drivers and even in some cases an increase in fatal crashes. The findings of this research further support the importance of preventing such behavior, possibly by raising driver’s awareness of the continued risks associated with “drinking and driving” (possibly through educational publicity campaigns targeted at specific risk groups like novice/young drivers).

Another contributing factor affecting RLR crash severity was non-peak traffic conditions and associated higher travel speeds. This finding would suggest a need to increase traffic enforcement levels during particular times of the day. Other variables, such as involvement of more than two vehicles and frontal impacts in a RLR-related crashes, were found to contribute indirectly to crash severity through the increase of transferred kinetic energy. This latter finding would support the implementation of strategies and countermeasures to improve driver compliance at signalized intersections, especially with devices that can prevent violations and reduce conflict occurrence (e.g., red-light cameras) (Shaaban and Pande 2018).

Regarding KE_s, the findings of this research further support the importance of improving car-front design features by optimizing the style, geometry, and stiffness, as well as developing improved energy-absorbing features and devices (Corben et al. 2004). To this end, in-vehicle technologies, such as intelligent speed adaptation or crash warning/avoidance systems, skid-resistant pavements, improved sight distances, traffic-calming measures (e.g., roundabouts, medians, road narrowing), narrower roads and lower speed limits are among the more promising options to lower crash severities by lowering speeds and transferred kinetic energy .

Regarding WVCs, the results of this study offer improved knowledge about the relationship between WVCs and other factors which contribute to the increase or reduction of severity. As suggested in Rowden et al. (2008), analyzing travel speeds in rural areas is crucial for mitigating the severity of WVCs. The results of this study showed that road surface condition and pavement conditions possessed the two highest direct effects on speeding attitude in rural areas, indirectly affecting the severity of the crash. While better surface and pavement conditions is desirable, one study done by Brubacher et al. (2018) found that the number of fatal crashes has doubled on rural highways in British Columbia (Canada), in a situation where the speed limit was raised to 120 kilometers per hour in 2014. Considering this, results from this research further support the importance of preventing excess driving speeds while raising driver's awareness of the risks involved with speeding. I conclude that technologies such as dynamic speed limit signs which adjust posted limits based on factors such as weather and traffic, or stricter law enforcement, and/or educational publicity campaigns would likely be useful to help mitigate the severity of WVCs by reducing impact velocity (Sullivan 2011).

Knowing the weight of variables that influence visibility can have implications for prioritizing specific countermeasures designed to extend the driver's forward view of the roadway. A study done by Brooks et al. (2011) showed that although drivers reduce speed to some degree in minimal visibility, this reduction is typically not enough to react to and avoid roadway hazards like pedestrians, stopped vehicles, or animals. Therefore, additional countermeasures such as reducing posted speed limits in places with low visibility, road lighting improvement, use of variable message signs, dynamic control of the forward beam pattern to extend the driver's view of the road, night vision systems, or other advanced detection systems that assist drivers in identifying the position of animals in the roadway would be helpful in such situations (Mahlke et al. 2007, Birch 2001). Driving simulators are ideal environments to train new drivers using virtual low visibility/high-risk situations (such as fog or snow) without the risk of physical harm to them or others on the road (Mueller and Trick 2012, Sullivan 2011).

Overall, the findings of this study can be employed by transportation practitioners to support safety improvement and policymaking purposes and also to prioritize strategies and countermeasures aimed at reducing crash severity outcomes at road sites. Moreover, these findings can motivate the use of programs to help train, promote, or encourage drivers to obey traffic rules as well as to educate them about conditions that can aggravate the severity of crashes.

7.3 Limitations and Future Work

In this research, two different types of crashes (RLR-related crashes and WVCs) were investigated, and the datasets used in this research were limited to the data from the Crash Analysis Reporting (CAR) system of the State of Florida (US), and the Traffic Accident Information System (TAIS) of the Saskatchewan Government Insurance (SGI) (Canada). Since each county, government, police, or insurance company has its own way of collecting crash records, other key or standardized variables could be gainfully included in the proposed SEM model. Also, I expect that somewhat different results would likely have been obtained if other regions were included in the analysis. Obtaining similar data from other jurisdictions would be beneficial to validate the basic results of this analysis. For example, WVCs in Saskatchewan mainly occur with ungulates like deer and moose. It would be insightful to test the model hypotheses and validate the results for other regions of North America where different wildlife spices exist.

Moreover, the SEM analysis for WVCs relied on data covering a 7-year period (2012 to 2018), whereas for RLR crashes, my dataset covered just over 4 years (2011 to 2014). Additional crash records and the analysis of more extended periods would further support the significance of this research and its findings. Another major issue is the under-reporting of crashes, as clearly not all significant crashes are reported and collected. Some jurisdictions only report crashes which are above a specific threshold amount of property damage, while others require the degree of vehicle damage to be above a certain level (Hauer 2006). And according to Vanlaar et al. (2012), WVCs generally are only reported when there is a claim or a police report, while WVC data on small animals are limited or non-existent. In some jurisdictions across Canada, there is speculation that under-reporting may be as high as 40 - 50% (Transport Canada 2019).

Sampling is also an issue. Most crash databases are obtained from police reports, and it is well-known that individuals part of no injury or minor injury crashes are less likely to report them to police or insurance companies. In 2009, the National Highway Traffic Safety Administration conducted a study that estimated that about 25 percent of minor injury crashes and half of the no-injury crashes are unreported. This is in sharp contrast to fatal crashes for which the reporting rate is nearly 100 percent (National Highway Traffic Safety Administration 2009, Blincoe et al. 2002). Thus crash records are generally referred to as outcome-based samples. In this type of sample, the injury severities reflected by the police-reports is not the same as the actual sample of crashes due to underreporting of less severe injuries. Outcome-based samples may result in biased parameter estimates (Savolainen et al. 2011).

Overall, this study focused mainly on transportation engineering aspects contributing to crash severity outcomes. As future work, other important human factors that could play an important role in severity outcomes (e.g., driving while on the phone, fatigue driving, or not wearing a seatbelt), should also be considered. Ultimately, the investigation of crash severities using SEM analysis on crash data from different regions will help to validate the results obtained in this thesis. Moreover, the severity of other types of crashes can be investigated with the same methodology, and it would be informative to compare such results. And further research on the magnitude of the crash underreporting issue and its impact on statistical estimation of contributing factors is recommended.

7.4 Closing Remarks

Annually, motor vehicle crashes impose an enormous economic and emotional burden on society. Researchers have been investigating ways to gain a better understanding of the factors that affect the likelihood and severity of crashes as well as provide direction for policies and countermeasures aimed at reducing both the number and severity of crashes (Lord and Mannering 2010). Employing statistical modeling has always been the most popular way to identify and analyze the contributions of human, environmental, roadway, and vehicle factors on crash severity (Kim et al. 2011). However, many of the traditional statistical methodologies applied in the literature are subject to limitations and shortcomings.

Modeling techniques in this area have mainly attempted to incorporate road and traffic factors into a statistical model and then identify relationships between independent and dependent variables (Lee et al. 2008). But there may be certain explanatory variables that can affect crash severity indirectly through one or more mediating variables (measured or latent). Investigating the relationship among explanatory variables can become a complex and challenging task. Moreover, while each method provides meaningful information, there are still some unexplained aspects of crashes, and some latent dimensions in the data cannot be explained through observed variables.

In this thesis, the objective was to investigate factors that influence the severity of crashes using the technique known as SEM. Compared to more traditional statistical analysis, SEM provides the added advantage of representing, estimating and testing complex modeling structures, where dependent variables can also be identified as predictor variables of other dependent variables (allowing examination of indirect effects and mediation structures). With SEM, it is also possible to include in a model both measured and latent variables.

By way of illustration, two different types of crashes were selected. There were RLR-related crashes as well as WVCs, both of which represent essential and frequent crash types in different environmental settings. RLR-related crashes are most likely to happen in an urban environment. WVCs, on the other hand are most likely to occur in a rural environment. To help identify the influencing factors on crash severity, an SEM model for each type of crash was developed. Here, I hypothesized that two latent dimensions, namely TS and KE_s strongly affect the severity of RLR-related crashes. For WVCs, I hypothesized that crash severity is most influenced by two latent dimensions, namely SA and VI.

Both SEM model specifications (RLR, and WVC) had an acceptable overall fit, while the hypothesized latent dimensions in both models fit the data as well. Moreover, SEM results demonstrated positive or negative marginal effects of these variables on crash severity. Regarding the RLR model, I found that both TS and KEs positively influence overall crash severity. In addition, a TS increase of the bullet vehicle can positively affect transferred KE to the subject vehicle(s), which indirectly affects crash severity.

Moreover, it was demonstrated that measured variables such as road, driver and environmental factors could indirectly affect crash severity through speed, and that vehicle factors and crash characteristics can also affect transferred kinetic energy to the subject vehicle(s). Regarding the WVC model, the results showed that SA and VI positively influenced overall crash severity. Also, VI was found to negatively affect speeding attitude, and this would indirectly decrease crash severity. Regarding measured variables, it was demonstrated that road, driver, and environmental factors could indirectly affect crash severity through SA and VI. Road surface and weather conditions showed the most influence on SA and VI, respectively.

In comparison with path analyses (i.e. a traditional severity model), both these SEM analyses generated superior results over goodness-of-fit, significance, and the signage of the estimates. Also, the causal relationships in SEM seemed more intuitive since the observed variables can influence the crash severity through latent variables as mediating factors. Therefore, it is possible to conclude that SEM analysis outperformed traditionally used severity models. In this regard, it can also help researchers better understand the mechanism of crashes using both measured and latent variables.

It appeared from this research that a reduction of travel speeds both in urban/rural areas is crucial for mitigating the severity of crashes. Moreover, the development of strategies for improving visibility and for having less transferred kinetic energy during a crash event would also be beneficial. Since SEM provides weights for each contributing factor of the latent variables, the results of the study can be employed by transportation engineers and decision-makers for safety improvement and policy-making purposes to prioritize strategies and countermeasures at crash-prone locations. Moreover, I offer that these findings can be used to help train, promote, or encourage drivers to obey traffic rules as well as to educate them about the conditions that can aggravate the severity of being involved in crashes.

REFERENCES

AASHTO. 2010. Highway Safety Manual. 1st ed.. American Association of State Highway and Transportation Officials, Washington, D.C.

Abdel-Aty, M. 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of safety research 34:597–603.

Abdel-Aty, M., and H. Abdelwahab. 2004. Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. Accident Analysis & Prevention 36:447–456.

Abdel-Aty, M., A.-A. Ekram, H. Huang, and K. Choi. 2011. A study on crashes related to visibility obstruction due to fog and smoke. Accident Analysis & Prevention 43:1730–1737.

Abdel-Aty, M., and J. Keller. 2005. Exploring the overall and specific crash severity levels at signalized intersections. Accident Analysis & Prevention 37:417–425.

Abegaz, T., Y. Berhane, A. Worku, A. Assrat, and A. Assefa. 2014. Effects of excessive speeding and falling asleep while driving on crash injury severity in Ethiopia: a generalized ordered logit model analysis. Accident; analysis and prevention 71:15–21.

Al-Ghamdi, A. S. 2002. Using logistic regression to estimate the influence of accident factors on accident severity. Accident Analysis & Prevention 34:729–741.

American Association of State Highway and Transportation Officials. 2011. A Policy on Geometric Design of Highways and Streets, 6th Edition.

Amoh-Gyimah, R., E. N. Aidoo, M. A. Akaateba, and S. K. Appiah. 2017. The effect of natural and built environmental characteristics on pedestrian-vehicle crash severity in Ghana. International Journal of Injury Control and Safety Promotion 24:459–468.

Anderson, J., and S. Hernandez. 2017. Roadway classifications and the accident injury severities of heavy-vehicle drivers. Analytic Methods in Accident Research.

Andrey, J., and S. Yagar. 1993. A temporal analysis of rain-related crash risk. Accident Analysis & Prevention 25:465–472.

Angel, A., and M. D. Hickman. 2009. Estimating occupant injury severity in two-vehicle crashes.

Anselin, L., R. Florax, and S. J. Rey. 2013. Advances in spatial econometrics: methodology, tools and applications. Springer Science & Business Media.

Austin, R. A., and B. M. Faigin. 2003. Effect of vehicle and crash factors on older occupants. Journal of Safety Research 34:441–452.

Ballesteros, M. F., P. C. Dischinger, and P. Langenberg. 2004. Pedestrian injuries and vehicle type in Maryland, 1995--1999. Accident Analysis & Prevention 36:73–81.

Bedard, M., G. H. Guyatt, M. J. Stones, and J. P. Hirdes. 2002. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accident Analysis & Prevention 34:717–727.

Behnood, A., and F. L. Mannering. 2016. An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. Analytic Methods in Accident Research.

Bentler, P. M., and E. J. C. Wu. 2005. EQS 6 for Windows guide. Multivariate Software, Encino, CA.

Birch, S. 2001. Adaptive Front Lighting. Automotive Engineering International.

Blincoe, L. J., A. G. Seay, E. Zaloshnja, T. R. Miller, E. O. Romano, S. Luchter, R. S. Spicer, and others. 2002. The economic impact of motor vehicle crashes.

Bollen, K. A. 1989. Structural Equations with Latent Variables. Hoboken, NJ, USA: John Wiley & Sons, Inc., Hoboken, NJ, USA.

Bonneson, J. A., K. Zimmerman, and M. A. Brewer. 2002. Engineering countermeasures to reduce red-light-running.

Bowen, N. K., and S. Guo. 2012. Structural equation modeling.

Brooks, J. O., M. C. Crisler, N. Klein, R. Goodenough, R. W. Beeco, C. Guirl, P. J. Tyler, A. Hilpert, Y. Miller, J. Grygier, B. Burroughs, A. Martin, R. Ray, C. Palmer, and C. Beck. 2011. Speed choice and driving performance in simulated foggy conditions. Accident Analysis and Prevention.

Browne, M. W., and Cudeck, R. 1993. Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), Testing structural equation models. Newbury Park, CA: Sage.

Browne, M. W. 1984. Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology 37:62–83.

Browne, M. W., and R. Cudeck. 1992. Alternative Ways of Assessing Model Fit. Sociological Methods & Research 21:230–258.

Brubacher, J., H. Chan, S. Erdelyi, G. Lovegrove, and F. Faghihi. 2018. Road Safety Impact of Increased Rural Highway Speed Limits in British Columbia, Canada. Sustainability 10:3555.

Cardelino, C. 1998. Daily Variability of Motor Vehicle Emissions Derived from Traffic Counter Data. Journal of the Air & Waste Management Association 48:637–645.

Carson, J., and F. Mannering. 2001. The effect of ice warning signs on ice-accident frequencies and severities. Accident Analysis & Prevention 33:99–109.

Cerwick, D. M., K. Gkritza, M. S. Shaheed, and Z. Hans. 2014. A comparison of the mixed logit and latent class methods for crash severity analysis. Analytic Methods in Accident Research 3–4:11–27.

Chang, H.-L., and T.-H. Yeh. 2006. Risk factors to driver fatalities in single-vehicle crashes: comparisons between non-motorcycle drivers and motorcyclists. Journal of transportation engineering 132:227–236.

Chang, L.-Y., and F. Mannering. 1999. Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents. Accident Analysis & Prevention 31:579–592.

Chang, L.-Y., and F. L. Mannering. 1998. Predicting vehicle occupancies from accident data: An accident severity approach. Transportation Research Record 1635:93–104.

Chen, L., C. Chen, R. Ewing, C. E. McKnight, R. Srinivasan, and M. Roe. 2013. Safety countermeasures and crash reduction in New York City - Experience and lessons learned. Accident Analysis and Prevention.

Chen, P., G. Yu, X. Wu, Y. Ren, and Y. Li. 2017. Estimation of red-light running frequency using high-resolution traffic and signal data. Accident Analysis and Prevention 102:235–247.

Chen, Z., and W. (David) Fan. 2019. A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. International Journal of Transportation Science and Technology 8:43–52.

Cheung, M. W. L. 2015. Meta-analysis : a structural equation modeling approach.

Chimba, D., and T. Sando. 2009. Neuromorphic prediction of highway injury severity. Advances in Transportation Studies 19:17–26.

Cho, S., D. Kim, and S. Kho. 2017. Latent Factors of Severity in Truck-Involved and Non-Truck-Involved Crashes on Freeways 11:920–927.

Clevenger, A. P., B. Chruszcz, K. Gunson, and M. Brumfit. 2002. Highway mitigation monitoring-Three Sisters Parkway interchange. Final report (Aug 1999-Jul 2002). Report prepared for Alberta Sustainable Resource Development, Canmore, Alberta, Canada.

Conn, J. M., J. L. Annest, and A. Dellinger. 2004. Nonfatal motor-vehicle animal crash-related injuries—United States, 2001–2002. Journal of Safety Research 35:571–574.

Corben, B., M. Cameron, T. Senserrick, and G. Rechnitzer. 2004. Development of the Visionary Research Model- Application to the car/pedestrian conflict. Monash University Accident Research Centre Reports:73.

Council, F. M., E. Zaloshnja, T. Miller, and B. N. Persaud 1947-. 2005. Crash cost estimates by maximum police-reported injury severity within selected crash geometrics. Page (U. S. F. H. A. O. of S. R. and Development, Ed.).

Cuttance, P., and R. Ecob. 1988. Structural modeling by example: applications in educational, sociological, and behavioral research.

Daniel, C., and F. S. Wood. 1980. Fitting equations to data: computer analysis of multifactor data. John Wiley & Sons, Inc.

Das, A., A. Pande, M. Abdel-Aty, and J. Santos. 2008. Urban arterial crash characteristics related with proximity to intersections and injury severity. Transportation Research Record 2083:137–144.

Donnell, E. T., and J. M. Mason Jr. 2004. Predicting the severity of median-related crashes in Pennsylvania by using logistic regression. Transportation Research Record 1897:55–63.

Edwards, J. B. 1999. Speed adjustment of motorway commuter traffic to inclement weather. Transportation Research Part F: Traffic Psychology and Behaviour 2:1–14.

Elmitiny, N., R. C. Harb, E. Radwan, and M. Ahmed. 2014. Traffic Operation Factors Related to Red-Light Running: An Empirical Analysis.

Eluru, N., and C. R. Bhat. 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis & Prevention 39:1037–1049.

Eluru, N., C. R. Bhat, and D. A. Hensher. 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis & Prevention 40:1033–1054.

Elvik, R., A. Hoye, T. Vaa, and M. Sorensen. 2009. The Handbook of Road Safety Measures : Second Edition. Emerald Publishing Limited.

Evans, L. 1994. Driver injury and fatality risk in two-car crashes versus mass ratio inferred using Newtonian mechanics. Accident Analysis and Prevention.

Farmer, C. M., E. R. Braver, and E. L. Mitter. 1997. Two-vehicle side impact crashes: the relationship of vehicle and crash characteristics to injury severity. Accident Analysis & Prevention 29:399–406.

FHWA. 2009. Engineering Countermeasures to Reduce Red-Light Running. FHWA-SA-10-005 Washington, D.C.

Garber, N. J., and L. A. Hoel. 2014. Traffic and highway engineering. Cengage Learning.

Garder, P. 2006. Segment characteristics and severity of head-on crashes on two-lane rural highways in Maine. Accident Analysis & Prevention 38:652–661.

Gkritza, K., M. Baird, and Z. N. Hans. 2010a. Deer-vehicle collisions, deer density, and land use in Iowa's urban deer herd management zones. Accident Analysis and Prevention 42:1916–1925.

Gkritza, K., C. R. Kinzenbaw, S. Hallmark, and N. Hawkins. 2010b. An empirical analysis of farm vehicle crash injury severities on Iowa's public road system. Accident Analysis and Prevention 42:1392–1397.

Glista, D. J., T. L. DeVault, and J. A. DeWoody. 2009. A review of mitigation measures for reducing wildlife mortality on roadways. Landscape and Urban Planning 91:1–7.

Golob, T. F. 2003. Structural equation modeling for travel behavior research. Transportation Research Part B: Methodological 37:1–25.

Grace, J. B., and K. A. Bollen. 2005. Interpreting the Results from Multiple Regression and Structural Equation Models. The Bulletin of the Ecological Society of America 86:283–295.

Gray, R. C., M. A. Quddus, and A. Evans. 2008. Injury severity analysis of accidents involving young male drivers in Great Britain. Journal of Safety Research 39:483–495.

Haikonen, H., and H. Summala. 2001. Deer-vehicle crashes: extensive peak at 1 hour after sunset. American journal of preventive medicine 21:209–13.

Haleem, K., and M. Abdel-Aty. 2010. Examining traffic crash injury severity at unsignalized intersections. Journal of safety research 41:347–357.

Haleem, K., and A. Gan. 2011. Identifying Traditional and Nontraditional Predictors of Crash Injury Severity on Major Urban Roadways. Traffic Injury Prevention 12:223–234.

Haleem, K., and A. Gan. 2013. Effect of driver's age and side of impact on crash severity along urban freeways: A mixed logit approach. Journal of Safety Research 46:67–76.

Halliday, D., R. Resnick, and J. Walker. 2004. Fundamentals of Physics. 1st editio. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Hassan, H. M., and H. Al-Faleh. 2013. Exploring the risk factors associated with the size and severity of roadway crashes in Riyadh. Journal of Safety Research 47:67–74.

Hauer, E. 2006. The frequency-severity indeterminacy. Accident Analysis and Prevention.

HCM. 2010. highway capacity manual. Fifth edition. Washington, D.C. : Transportation Research Board, c2010-.

Heene, M., S. Hilbert, C. Draxler, M. Ziegler, and M. Bühner. 2011. Masking Misfit in Confirmatory Factor Analysis by Increasing Unique Variances: A Cautionary Note on the Usefulness of Cutoff Values of Fit Indices. Psychological Methods 16:319–336.

Hoe, S. L. 2008. Issues and procedures in adopting structural equation modeling technique. Journal of applied quantitative methods 3:76–83.

Holdridge, J. M., V. N. Shankar, and G. F. Ulfarsson. 2005. The crash severity impacts of fixed roadside objects. Journal of Safety Research 36:139–147.

Hosseinpour, M., A. S. Yahaya, and A. F. Sadullah. 2014. Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian Federal Roads. Accident Analysis & Prevention 62:209–222.

Hoyle, R. H. 1995. The Structural Equation Modeling Approach. Basic Concepts and Fundamental Issues.

Hu, L. T., and P. M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling.

Hu, W., and E. T. Donnell. 2010. Median barrier crash severity: Some new insights. Accident Analysis & Prevention 42:1697–1704.

Hu, W., and E. T. Donnell. 2011. Severity models of cross-median and rollover crashes on rural divided highways in Pennsylvania. Journal of Safety Research 42:375–382.

Huang, H., H. C. Chin, and M. M. Haque. 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis. Accident Analysis & Prevention 40:45–54.

Huelke, D. F., and W. A. Chewning. 1968. The energy-absorbing steering column: a study of collision performance in fatal and nonfatal accidents.

Huijser, M. P., P. T. McGowen, J. Fuller, A. Hardy, and A. Kociolek. 2008. Wildlife-Vehicle Collision Reduction Study: Report to Congress. Page Western Transportation Institute Montana State University.

Islam, M., A. Ermagun, and S. Banerjee. 2019. Gender Differences in Injury Severity Risk of Single-Vehicle Crashes in Virginia: A Nested Logit Analysis of Heterogeneity. arXiv preprint arXiv:1901.03289.

Islam, S., and F. Mannering. 2006. Driver aging and its effect on male and female single-vehicle accident injuries: Some additional evidence. Journal of safety Research 37:267–276.

Jiang, X., H. Zheng, Y. Qiu, and W. Fan. 2015. Differences in injury severities between 2-vehicle and 3-vehicle crashes. Traffic injury prevention 16:289–297.

Joreskog, K. G., and D. Sorbom. 1982. Recent Developments in Structural Equation Modeling. Journal of Marketing Research.

Jung, S., X. Qin, and D. A. Noyce. 2010. Rainfall effect on single-vehicle crash severities using polychotomous response models. Accident Analysis & Prevention 42:213–224.

Khattak, A. J. 2001. Injury severity in multivehicle rear-end crashes. Transportation Research Record 1746:59–68.

Khattak, A. J., P. Kantor, and F. M. Council. 1998. Role of adverse weather in key crash types on limited-access: roadways implications for advanced weather systems. Transportation research record 1621:10–19.

Khattak, A. J., M. D. Pawlovich, R. R. Souleyrette, and S. L. Hallmark. 2002. Factors related to more severe older driver traffic crash injuries. Journal of Transportation Engineering 128:243–249.

Khattak, A. J., and M. Rocha. 2003. Are SUVs "supremely unsafe vehicles"?: analysis of rollovers and injuries with sport utility vehicles. Transportation Research Record 1840:167–177.

Khattak, A. J., and F. Targa. 2004. Injury severity and total harm in truck-involved work zone crashes. Transportation research record 1877:106–116.

Khorashadi, A., D. Niemeier, V. Shankar, and F. Mannering. 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. Accident Analysis & Prevention 37:910–921.

Kim, J.-K., S. Kim, G. F. Ulfarsson, and L. A. Porrello. 2007. Bicyclist injury severities in bicycle--motor vehicle accidents. Accident Analysis & Prevention 39:238–251.

Kim, J.-K., G. F. Ulfarsson, S. Kim, and V. N. Shankar. 2013. Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. Accident Analysis and Prevention 50:1073–1081.

Kim, K., P. Pant, and E. Yamashita. 2011. Measuring Influence of Accessibility on Accident Severity with Structural Equation Modeling. Transportation Research Record: Journal of the Transportation Research Board 2236:1–10.

Kline, R. B. 2005. Principles and practice of structural equation modeling, 2nd ed. Page Principles and practice of structural equation modeling, 2nd ed. Guilford Press, New York, NY, US.

Klop, J. R., and A. J. Khattak. 1999. Factors influencing bicycle crash severity on two-lane, undivided roadways in North Carolina. Transportation Research Record 1674:78–85.

Kockelman, K. M., and Y.-J. Kweon. 2002. Driver injury severity: an application of ordered probit models. Accident Analysis & Prevention 34:313–321.

Kononen, D. W., C. A. C. Flannagan, and S. C. Wang. 2011. Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. Accident Analysis & Prevention 43:112–122.

Krull, K. A., A. J. Khattak, and F. M. Council. 2000. Injury effects of rollovers and events sequence in single-vehicle crashes. Transportation Research Record 1717:46–54.

Kweon, Y.-J., and K. M. Kockelman. 2003. Overall injury risk to different drivers: combining exposure, frequency, and severity models. Accident Analysis & Prevention 35:441–450.

Lachenmayr, B., J. Berger, A. Buser, and O. Keller. 1998. Reduced visual function causes higher risks of traffic accidents. Der Ophthalmologe 95:44–50.

Langley, R. L., S. A. Higgins, and K. B. Herrin. 2006. Risk factors associated with fatal animal-vehicle collisions in the United States, 1995-2004. Wilderness and Environmental Medicine 17:229–239.

Lao, Y., G. Zhang, Y. J. Wu, and Y. Wang. 2011. Modeling animal-vehicle collisions considering animal-vehicle interactions. Accident Analysis and Prevention 43:1991–1998.

Lee, C., and M. Abdel-Aty. 2005. Comprehensive analysis of vehicle--pedestrian crashes at intersections in Florida. Accident Analysis & Prevention 37:775–786.

Lee, C., and M. Abdel-Aty. 2008. Presence of passengers: does it increase or reduce driver's crash potential? Accident Analysis & Prevention 40:1703–1712.

Lee, J., J. Chae, T. Yoon, and H. Yang. 2018. Traffic accident severity analysis with rain-related factors using structural equation modeling – A case study of Seoul City. Accident Analysis and Prevention 112:1–10.

Lee, J., and F. Mannering. 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis & Prevention 34:149–161.

Lee, J. Y., J. H. Chung, and B. Son. 2008. Analysis of traffic accident size for Korean highway using structural equation models. Accident Analysis and Prevention 40:1955–1963.

Li, Z., Y. Ci, C. Chen, G. Zhang, Q. Wu, Z. (Sean) Qian, P. D. Prevedouros, and D. T. Ma. 2019. Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. Accident Analysis & Prevention 124:219–229.

Liu, L., and S. Dissanayake. 2009. Factors affecting crash severity on gravel roads. Journal of Transportation Safety & Security 1:254–267.

Lord, D., and F. Mannering. 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives.

Maccallum, R. C., and J. T. Austin. 2000. Applications of structural equation modeling in psychological research. Page Annu. Rev. Psychol.

MacKinnon, D. P. 2008. Introduction to statistical mediation analysis. APA handbook of research methods in psychology Vol 2 Research designs Quantitative qualitative neuropsychological and biological.

Mahlke, S., D. Rösler, K. Seifert, J. F. Krems, and M. Thüring. 2007. Evaluation of six night vision enhancement systems: Qualitative and quantitative support for intelligent image processing. Human Factors 49:518–531.

Malyshkina, N. V, and F. Mannering. 2008. Effect of increases in speed limits on severities of injuries in accidents. Transportation research record 2083:122–127.

Malyshkina, N. V, and F. L. Mannering. 2010. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents. Accident Analysis & Prevention 42:131–139.

Manner, H., and L. Wünsch-Ziegler. 2013. Analyzing the severity of accidents on the German Autobahn. Accident Analysis and Prevention 57:40–48.

Moore, D. N., W. H. Schneider IV, P. T. Savolainen, and M. Farzaneh. 2011. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. Accident Analysis & Prevention 43:621–630.

Moudon, A. V., L. Lin, J. Jiao, P. Hurvitz, and P. Reeves. 2011. The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington. Accident Analysis & Prevention 43:11–24.

Mueller, A. S., and L. M. Trick. 2012. Driving in fog: The effects of driving experience and visibility on speed compensation and hazard avoidance. Accident Analysis and Prevention.

Mukherjee, A., S. Stolpner, X. Liu, U. Vrenozaj, C. Fei, and A. Sinha. 2013. Large animal detection and continuous traffic monitoring on highways. Pages 1–3 SENSORS, 2013 IEEE.

Mulaik, S. A. 2009. Linear causal modeling with structural equations. Chapman and Hall/CRC.

Muthén, L. K., and B. O. Muthén. 2017. Mplus User's Guide. Eighth Edition. Los Angeles, CA.

Naik, B., L.-W. Tung, S. Zhao, and A. J. Khattak. 2016. Weather impacts on single-vehicle truck crash injury severity. Journal of Safety Research 58:57–65.

National Highway Traffic Safety Administration. 2009. Traffic safety facts: Motorcycles. DOT HS 811:159.

Nevarez, A., M. Abdel-Aty, X. Wang, and J. Santos. 2009. Multi-level analysis of severe crashes along arterials including the effect of road entity and collision type. Journal of Transportation Safety and Security.

NHTSA. 2004. Lives saved by the Federal Motor Vehicle Safety Standards and other vehicle safety technologies, 1960-2002-Passenger cars and light trucks-with a review of 19 FMVSS and their effectiveness in reducing fatalities, injuries and crashes.

NHTSA. 2012. Traffic safety facts 2012, National Highway Traffic Safety Administration, US.

NHTSA. 2015. Traffic safety facts 2015, National Highway Traffic Safety Administration, US. Nhtsa:2015.

Norm O'Rourke, P. D. R. P., and L. Hatcher. 2013. A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling, Second Edition. SAS Institute.

Noyce, D. A., G. X. Lu, and R. J. McKendry. 2006. Analysis of the magnitude and predictability of median crossover crashes utilizing logistic regression.

Ogden, K. W. 1996. Safer roads : a guide to road safety engineering. Avebury Technical.

Owens, D., and M. Sivak. 1993. The role of visibility in nighttime traffic accidents. Pages 133–147 2nd International Symposium on visibility and luminance in Roadway lighting.

Pai, C.-W., and W. Saleh. 2007. An analysis of motorcyclist injury severity under various traffic control measures at three-legged junctions in the UK. Safety science 45:832–847.

Pai, C.-W., and W. Saleh. 2008. Exploring motorcyclist injury severity in approach-turn collisions at T-junctions: Focusing on the effects of driver's failure to yield and junction control measures. Accident Analysis & Prevention 40:479–486.

Patil, S., S. R. Geedipally, and D. Lord. 2012. Analysis of crash severities using nested logit model—Accounting for the underreporting of crashes. Accident Analysis & Prevention 45:646–653.

Pedhazur, E. J., and L. P. Schmelkin. 1991. Measurement, design, and analysis : an integrated approach. Lawrence Erlbaum Associates.

Peek-Asa, C., C. Britton, T. Young, M. Pawlovich, and S. Falb. 2010. Teenage driver crash incidence and factors influencing crash injury by rurality. Journal of safety research 41:487–492.

Planek, T. W., S. Sinelnikov Associate Editor Jonathan Thomas Associate Editor Kenneth Kolosh Associate Editor Kathleen Porretta Managing Editor, F. Bella, and A. Calvi. 2014. Analysis of driver speeds under night driving conditions using a driving simulator. Journal of Safety Research 49:45.e1-52.

Quddus, M. A., R. B. Noland, and H. C. Chin. 2002. An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of safety research 33:445–462.

Quddus, M. A., C. Wang, and S. G. Ison. 2009. Road traffic congestion and crash severity: econometric analysis using ordered response models. Journal of Transportation Engineering 136:424–435.

Renski, H., A. J. Khattak, and F. M. Council. 1999. Effect of speed limit increases on crash injury severity: analysis of single-vehicle crashes on North Carolina interstate highways. Transportation Research Record 1665:100–108.

Retting, R. A., R. G. Ulmer, and A. F. Williams. 1999. Prevalence and characteristics of red light running crashes in the United States. Page Accident Analysis and Prevention.

Retting, R. A., A. F. Williams, D. F. Preusser, and H. B. Weinstein~. 1995. Classifying urban crashes for countermeasure development*. Page Accid. Anal. and Prev.

Rifaat, S. M., and R. Tay. 2009. Effects of street patterns on injury risks in two-vehicle crashes. Transportation Research Record 2102:61–67.

Rifaat, S. M., R. Tay, and A. de Barros. 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. Accident Analysis and Prevention 43:276–283.

Roess, R. P., E. S. Prassas, and W. R. McShane. 2004. Traffic engineering. Pearson/Prentice Hall.

Romin, L. A., and J. A. Bissonette. 1996. Deer: vehicle collisions: status of state monitoring activities and mitigation efforts. Wildlife Society Bulletin:276–283.

Rowden, P., D. Steinhardt, and M. Sheehan. 2008. Road crashes involving animals in Australia. Accident Analysis and Prevention 40:1865–1871.

Sainsbury, A. W., P. M. Bennett, and J. K. Kirkwood. 1995. the Welfare of Free-Living Wild Animals in Europe - Harm Caused By Human Activities. Animal Welfare 4:183–206.

SAS Institute Inc. 2017. SAS Institute Inc. Page SAS Institute Inc. MarketLine Company Profile. Cary, NC.

Saskatchewan Government Insurance (SGI). 2018. 2017 Saskatchewan Traffic Accident Facts.

Savolainen, P., and I. Ghosh. 2008. Examination of factors affecting driver injury severity in Michigan’s single-vehicle--deer crashes. Transportation Research Record 2078:17–25.

Savolainen, P., and F. Mannering. 2007. Probabilistic models of motorcyclists’ injury severities in single-and multi-vehicle crashes. Accident Analysis & Prevention 39:955–963.

Savolainen, P. T., F. L. Mannering, D. Lord, and M. A. Quddus. 2011. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accident Analysis & Prevention 43:1666–1676.

Schneider, W. H., P. T. Savolainen, and K. Zimmerman. 2009. Driver injury severity resulting from single-vehicle crashes along horizontal curves on rural two-lane highways. Transportation Research Record 2102:85–92.

Schneider, W., and P. Savolainen. 2011. Comparison of Severity of Motorcyclist Injury by Crash Types. Transportation Research Record 2265:70–80.

Schreiber, J. B., F. K. Stage, J. King, A. Nora, and E. A. Barlow. 2006. Reporting structural equation modeling and confirmatory factor analysis results: A review.

Setyawan, A., I. Kusdiantoro, and Syafi’i. 2015. The Effect of Pavement Condition on Vehicle Speeds and Motor Vehicles Emissions. Procedia Engineering 125:424–430.

Shaaban, K., and A. Pande. 2018. Evaluation of red-light camera enforcement using traffic violations. Journal of Traffic and Transportation Engineering (English Edition).

Shaheed, M. S. B., K. Gkritza, W. Zhang, and Z. Hans. 2013. A mixed logit analysis of two-vehicle crash severities involving a motorcycle. Accident Analysis and Prevention 61:119–128.

Shankar, V., and F. Mannering. 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. Journal of safety research 27:183–194.

Shankar, V., F. Mannering, and W. Barfield. 1996. Statistical analysis of accident severity on rural freeways. Accident Analysis & Prevention 28:391–401.

Shibata, A., and K. Fukuda. 1994. Risk factors of fatality in motor vehicle traffic accidents. Accident Analysis & Prevention 26:391–397.

Shimamura, M., M. Yamazaki, and G. Fujita. 2005. Method to evaluate the effect of safety belt use by rear seat passengers on the injury severity of front seat occupants. Accident Analysis & Prevention 37:5–17.

Siddiqui, N. A., X. Chu, and M. Guttenplan. 2006. Crossing locations, light conditions, and pedestrian injury severity. Transportation research record 1982:141–149.

Sobhani, A., W. Young, D. Logan, and S. Bahrololoom. 2011. A kinetic energy model of two-vehicle crash injury severity. Accident Analysis & Prevention 43:741–754.

Suhr, D. 2006. The basics of structural equation modeling. Presented: Irvine, CA, SAS User Group of the Western Region of the United States (WUSS).

Sullivan, J. M. 2011. Trends and characteristics of animal-vehicle collisions in the United States. Journal of Safety Research 42:9–16.

Sze, N.-N., and S. C. Wong. 2007. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. Accident Analysis & Prevention 39:1267–1278.

Tarko, A. P., H. Bar-Gera, J. Thomaz, and A. Issariyanukula. 2010. Model-Based application of abbreviated injury scale to police-reported crash injuries. Transportation Research Record 2148:59–68.

Tay, R., and A. De Barros. 2009. Minimizing red light violations: How many cameras do we need for a given number of locations? Journal of Transportation Safety and Security.

Tay, R., J. Choi, L. Kattan, and A. Khan. 2011. A Multinomial Logit Model of Pedestrian-Vehicle Crash Severity. International Journal of Sustainable Transportation 5:233–249.

Tomas, J. M., and A. Oliver. 1999. Rosenberg's self-esteem scale: Two factors or method effects. Structural Equation Modeling: A Multidisciplinary Journal 6:84–98.

Toy, E. L., and J. K. Hammitt. 2003. Safety impacts of SUVs, vans, and pickup trucks in two-vehicle crashes. Risk Analysis: An International Journal 23:641–650.

Transport Canada. 2019. National Collision Database Online (NCDB). http://wwwapps2.tc.gc.ca/Saf-Sec-Sur/7/NCDB-BNDC/p.aspx?l=en.

Uddin, M., and N. Huynh. 2017. Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways. Accident Analysis and Prevention.

Uddin, M., and N. Huynh. 2018. Factors influencing injury severity of crashes involving HAZMAT trucks. International Journal of Transportation Science and Technology 7:1–9.

Ulfarsson, G. F., S. Kim, and E. T. Lentz. 2006. Factors affecting common vehicle-to-vehicle collision types: Road safety priorities in an aging society. Transportation research record 1980:70–78.

Ulfarsson, G. F., and F. L. Mannering. 2004. Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. Accident Analysis & Prevention 36:135–147.

Ullman, J. B. 2013. Structural Equation Modeling, Using multivariate statistics.

Vanlaar, W., K. E. Gunson, S. W. Brown, and R. D. Robertson. 2012. Wildlife-vehicle collisions in canada: a review of the literature and a compendium of existing data sources. Page Traffic Injury Research Foundation.

Visintin, C., N. Golding, R. van der Ree, and M. A. McCarthy. 2018. Managing the timing and speed of vehicles reduces wildlife-transport collision risk. Transportation Research Part D: Transport and Environment 59:86–95.

Vogt, A., and J. Bared. 1998. Accident Models for Two-Lane Rural Segments and Intersections. Transportation Research Record: Journal of the Transportation Research Board 1635:18–29.

Wang, J., H. A. Siegal, R. S. Falck, and R. G. Carlson. 2001. Factorial structure of Rosenberg's Self-Esteem Scale among crack-cocaine drug users. Structural Equation Modeling 8:275–286.

Wang, J., and X. Wang. 2012. Structural Equation Modeling. John Wiley & Sons, Ltd, Chichester, UK.

Wang, K., and X. Qin. 2014. Use of Structural Equation Modeling to Measure Severity of Single-Vehicle Crashes. Transportation Research Record: Journal of the Transportation Research Board 2432:17–25.

Wang, P., X.-Z. Lu, Z.-C. Yan, and M.-J. Zhang. 2018a. Analysis on Influencing Factors of Rear-end Crash Severity Based on Ordered Probit Model. Gonglu Jiaotong Keji = Journal of Highway and Transportation Research and Development:102.

Wang, T., J. Chen, C. Wang, and X. Ye. 2018b. Understand e-bicyclist safety in China: Crash severity modeling using a generalized ordered logit model. Advances in Mechanical Engineering 10.

Wang, X., and M. Abdel-Aty. 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accident Analysis & Prevention 40:1674–1682.

Wang, Z., H. Chen, and J. J. Lu. 2009. Exploring impacts of factors contributing to injury severity at freeway diverge areas. Transportation Research Record 2102:43–52.

Washington, S. P., M. G. Karlaftis, and F. L. Mannering. 2003. Statistical and econometric methods for transportation data analysis.

Wothke, W. 1993. Nonpositive definite matrices in structural modeling. Sage Focus Editions 154:256.

Wu, Q., F. Chen, G. Zhang, X. C. Liu, H. Wang, and S. M. Bogus. 2014. Mixed logit model-based driver injury severity investigations in single- and multi-vehicle crashes on rural two-lane highways. Accident Analysis & Prevention 72:105–115.

Wu, Q., G. Zhang, Y. Ci, L. Wu, R. A. Tarefder, and A. "Dely" Alcántara. 2016a. Exploratory multinomial logit model-based driver injury severity analyses for teenage and adult drivers in intersection-related crashes. Traffic Injury Prevention 17:413–422.

Wu, Q., G. Zhang, X. Zhu, X. C. Liu, and R. Tarefder. 2016b. Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways. Accident Analysis & Prevention 94:35–45.

Xu, C., D. Li, Z. Li, W. Wang, and P. Liu. 2017. Utilizing structural equation modeling and segmentation analysis in real-time crash risk assessment on freeways. KSCE Journal of Civil Engineering 22:1–9.

Yamamoto, T., J. Hashiji, and V. N. Shankar. 2008. Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. Accident Analysis & Prevention 40:1320–1329.

Ye, F., and D. Lord. 2011. Investigation of effects of underreporting crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit, and mixed logit. Transportation Research Record 2241:51–58.

Ye, F., and D. Lord. 2014. Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models. Analytic methods in accident research 1:72–85.

Yuan, Q., M. Lu, A. Theofilatos, and Y.-B. Li. 2017. Investigation on occupant injury severity in rear-end crashes involving trucks as the front vehicle in Beijing area, China. Chinese Journal of Traumatology 20:20–26.

Zajac, S. S., and J. N. Ivan. 2003. Factors influencing injury severity of motor vehicle--crossing pedestrian crashes in rural Connecticut. Accident Analysis & Prevention 35:369–379.

Zhang, J., J. Lindsay, K. Clarke, G. Robbins, and Y. Mao. 2000. Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario. Accident Analysis & Prevention 32:117–125.

Zhao, S., and A. Khattak. 2015. Motor vehicle drivers' injuries in train-motor vehicle crashes. Accident Analysis and Prevention.

Zhong, M., S. Sharma, and P. Lingras. 2005. Refining Genetically Designed Models for Improved Traffic Prediction on Rural Roads. Transportation Planning and Technology 28:213–236.

Zhu, H., K. K. Dixon, S. Washington, and D. M. Jared. 2010. Single-Vehicle Fatal Crash Prediction for Two-Lane Rural Highways in the Southeastern United States.

Zhu, X., and S. Srinivasan. 2011a. A comprehensive analysis of factors influencing the injury severity of large-truck crashes. Accident Analysis & Prevention 43:49–57.

Zhu, X., and S. Srinivasan. 2011b. Modeling occupant-level injury severity: An application to large-truck crashes. Accident Analysis & Prevention 43:1427–1437.

APPENDIX A – SEM MODEL FORMULATION

An example of a structural equation model is shown in Figure A. 1. As mentioned before, the latent variables are presented with circles (or ellipse), and the observed variables are shown in rectangular or square boxes. The measurement of a latent variable is done through one or more observable variables (Wang and Wang 2012).

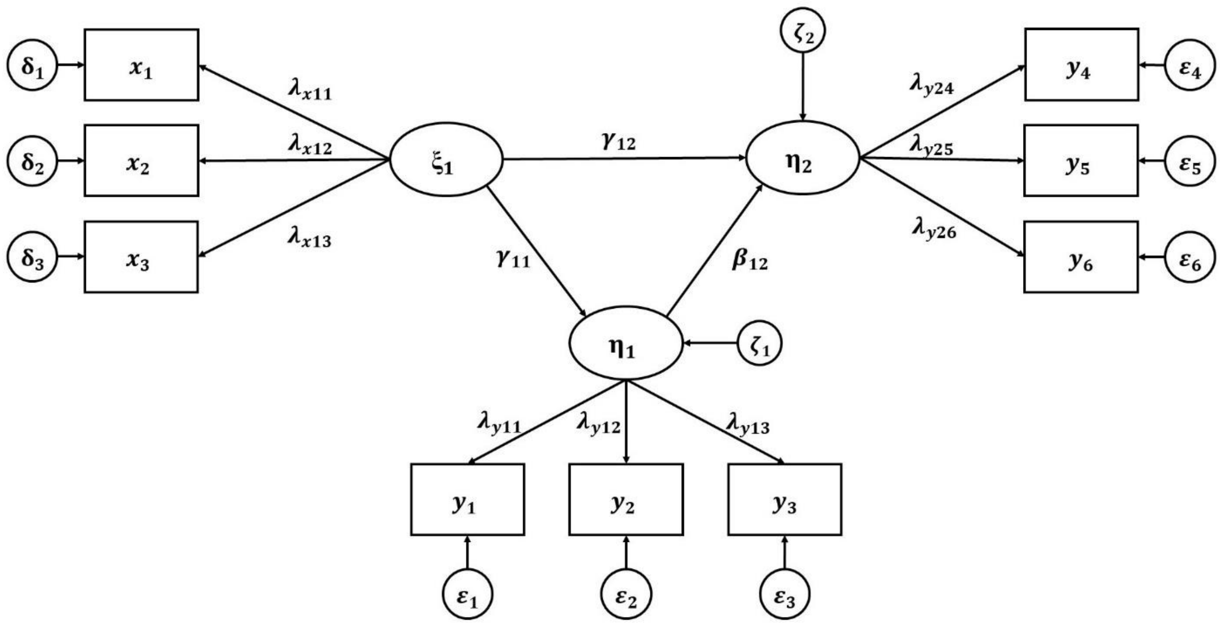


Figure A. 1 General SEM.

In this example three observed variables ($x_1 - x_3$) are used as indicators of the latent variable ξ_1; three indicators ($y_1 - y_3$) are used for latent variable η_1; three indicators ($y_4 - y_6$) are used for latent variable η_2. The latent variables that are measured by variables within the model are called endogenous latent variables (η); the latent variables, whose causes lie outside the model, are exogenous latent variables (ξ). Indicator variables that measure the exogenous latent variables are called exogenous indicators ($x_1 - x_3$) and indicator variables that measure the endogenous latent variables are endogenous indicators ($y_1 - y_6$). δ shows the measurement error term for the exogenous indicators and ε shows the measurement error term for the endogenous indicators.

The coefficients β and γ in the diagram are path coefficients. β indexes the path coefficient of the dependent endogenous variable, and γ indexes the causal variable (either endogenous or exogenous). If the causal variable is exogenous (ξ), γ would be used for the path coefficient; if the causal variable is another endogenous variable (η), β would be used for the path coefficient. For example, β_{12} shows the effect of endogenous variable η_1 on the endogenous variable η_2. Since the predictions are not perfect, there are always residuals or errors. The ζ's in the model, show structural equation residual terms (Wang and Wang 2012).

As mentioned in section 3.5, a SEM model consists of three components: (a) a measurement model for the endogenous variables (Y measurement model), (b) a measurement model for the exogenous variable (X measurement model), and (c) a structural model (Lee et al. 2008). In the measurement model, linear (e.g., when observed variables are continuous) or nonlinear (e.g., when observed variables are categorical) equations describe the relations between the observed variables and their underlying latent variables (factors). In the structural equations part of the model, endogenous latent variables (η) are regressed on the exogenous latent variables (ξ) and/or some other endogenous latent variables. In addition, observed variables can also be included as either independent and/or dependent variables in a SEM model (Wang and Wang 2012). The general SEM can be expressed by three equations (Bollen 1989):

$$\eta = B \cdot \eta + \Gamma \cdot \xi + \zeta \tag{3.5}$$

$$x = \Lambda_x \xi + \delta \tag{3.6}$$

$$y = \Lambda_{\mathrm{y}} \eta + \varepsilon \tag{3.7}$$

Where in Eq. (3.5), η is an m×1 vector of endogenous latent variables, ξ is an n×1 vector of the exogenous latent variables; B is an $m \times m$ matrix and Γ is an $m \times n$ matrix that contains regression parameters (i.e., β's and γ's) for the endogenous and exogenous latent variables (i.e., magnitudes of expected changes after a unit increase in η and ξ, respectively). $\zeta = m \times 1$ is a vector of disturbances (error terms) that is assumed to have an expected value of zero [$E(\zeta)=0$].

The measurement models is represented with Eq. (3.6) and Eq. (3.7). where x is a q×1 column vector of the observed exogenous variables, y is a p×1 column vector of the observed endogenous variables, δ is a q×1 column vector of the observed exogenous errors. ε is a p×1 column vector of the observed endogenous errors. The expected value of ε and δ is zero. Λ_x is a $q \times n$ matrix of structural coefficient (λ_x) for the effects of the latent exogenous variables on the observed variables, and Λ_y is a $p\times m$ matrix of structural coefficient (λ_y) for the effects of the latent endogenous variables on the observed ones.

A full SEM model is defined by the specification of eight parameter matrixes ($\Lambda_x, \Lambda_y, \Gamma$, $B, \Phi, \Psi, \Theta_\varepsilon$, and Θ_δ). Φ (phi), Ψ (psi), Θ_ε (theta-epsilon), and Θ_δ (theta-delta) are the four variance/covariance matrixes in SEM. Φ is the $n\times n$ variance/covariance matrix for the latent exogenous variables (ξ). Ψ is the $m\times m$ variance/covariance matrix for the residual terms of the structural equations (ζ). Θ_ε and Θ_δ are the $p\times p$ and $q\times q$ variance/covariance matrixes for the measurement errors of the observed variables y and x, respectively. All of these matrixes are symmetric square matrixes, where the number of rows and columns in each of the matrixes are equal. The main diagonal elements of each matrix show the variances that should always be positive and the off-diagonal elements, show the covariances related to the pairs of variables in the matrix (Wang and Wang 2012).

The hypothesized model showed in Figure A. 1 can be specified in matrix notation based on the three basic equations (3.5, 3.6, 3.7). Eq. (3.5) can be expressed as:

$$\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix} = \begin{bmatrix}0 & 0\\ \beta_{12} & 0\end{bmatrix}\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix} + \begin{bmatrix}\gamma_{11}\\ \gamma_{12}\end{bmatrix}[\xi_1] + \begin{bmatrix}\zeta_1\\ \zeta_2\end{bmatrix} \tag{A.1}$$

From Eq. (A.1) it is possible to derive the following two structural equations:

$$\eta_1 = \gamma_{11}\xi_1 + \zeta_1 \tag{A.2}$$

$$\eta_2 = \beta_{12}\eta_1 + \gamma_{12}\xi_1 + \zeta_2$$

The measurement equation (3.7) can be expressed as:

$$\begin{bmatrix}y_1\\ y_2\\ y_3\\ y_4\\ y_5\\ y_6\end{bmatrix} = \begin{bmatrix}\lambda_{y11} & 0\\ \lambda_{y12} & 0\\ \lambda_{y13} & 0\\ 0 & \lambda_{y24}\\ 0 & \lambda_{y25}\\ 0 & \lambda_{y26}\end{bmatrix}\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix} + \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \varepsilon_4\\ \varepsilon_5\\ \varepsilon_6\end{bmatrix} \tag{A.3}$$

From Eq. (A.3) it is possible to derive the following six measurement structural equations:

$y_1 = \lambda_{y11}\eta_1 + \varepsilon_1$

$y_2 = \lambda_{y12}\eta_1 + \varepsilon_2$

$y_3 = \lambda_{y13}\eta_1 + \varepsilon_3$

$y_4 = \lambda_{y24}\eta_2 + \varepsilon_4$ (A.4)

$y_5 = \lambda_{y25}\eta_2 + \varepsilon_5$

$y_6 = \lambda_{y26}\eta_2 + \varepsilon_6$

The measurement equation (3.6) can be expressed as:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \lambda_{x11} \\ \lambda_{x12} \\ \lambda_{x13} \end{bmatrix} [\xi_1] + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} \quad \text{(A.5)}$$

From Eq. (A.5) it is possible to derive the following three measurement structural equations:

$x_1 = \lambda_{x11}\xi_1 + \delta_1$

$x_2 = \lambda_{x12}\xi_1 + \delta_2$ (A.6)

$x_3 = \lambda_{x13}\xi_1 + \delta_3$

It is assumed that E (ζ) = 0, E (ε) = 0, and E(δ) = 0, Cov (ζ, ξ) = 0, Cov (ε, η) = 0, and Cov (δ, ξ) = 0. In addition, multivariate normality is assumed for the observed and latent variables.

The criterion selected for parameter estimation is known as the discrepancy function, where the difference between Σ, the $p \times p$ population covariance matrix as estimated by the sample covariance matrix, and $\Sigma(\theta)$, the $p \times p$ covariance matrix from the hypnotized model, are minimized.

The aim of model estimation is to find a set of model parameters θ to produce $\Sigma(\theta)$ so that [Σ-$\Sigma(\theta)$] can be minimized. The discrepancy between Σ and $\Sigma(\theta)$ indicates if the model fits the data well (Wang and Wang 2012).

Since both Σ and $\Sigma(\theta)$ are unknown, [S-$\Sigma(\hat{\theta})$] or (S-$\hat{\Sigma}$) is minimized in SEM; where S is the sample variance/covariance matrix, $\hat{\theta}$ are the model parameter estimates; $\Sigma(\hat{\theta})$ or $\hat{\Sigma}$ is the model estimated/implied variance=covariance matrix. To minimize the difference between S and $\hat{\Sigma}$

during the estimation process, a particular fitting function should be utilized. Maximum likelihood, asymptotically distribution-free, and weighted least square are standard estimation methods for many model-fitting programs. Eq. (A.7) shows the maximum likelihood (ML) fitting function:

$$F_{ML}(\hat{\theta}) = \ln|\hat{\Sigma}| + tr(S\hat{\Sigma}^{-1}) - \ln|S| - (p+q) \tag{A.7}$$

where S and $\hat{\Sigma}$ are the sample and model estimated variance/covariance matrixes, respectively; and $(p + q)$ is the number of observed variables involved in the model (yielding (p + q)(p + q + 1)/2 unique variances and covariances).

APPENDIX B – SOFTWARE CODES

For the first part of the thesis, SAS software version 3.71, University Edition (SAS Institute Inc 2017) was used, which allows implementing SEM with the CALIS procedure. The weight-least squares (WLS) method was used in order to account for possible violation of the multivariate normality assumption. The code for this model is as follows:

```
PROC IMPORT DATAFILE="/FOLDERS/MYFOLDERS/RLR_CRASH.XLSX"

OUT=WORK.MYEXCEL

DBMS=XLSX

REPLACE;

RUN;

PROC CALIS DATA=WORK.MYEXCEL MAXITER=50000 PLOT=PATHDIAGRAM MODIFICATION METHOD=WLS;

PATH

TRAVEL_SPEED ---> SKID_RESISTANCE ALCOHOLE_INVOLVED VERTI-CAL_ALIGNMENT GENDR DAY_OF_WEEK LIGHTING_CONDITION PEAK_HOUR ROAD_SURFACE_CONDITION AGE = 1.,

KINETIC_ENERGY ---> NUMBER_OF_VEHICLES VEHICLE_TYPE POINT_OF_IMPACT VEHICLE_MANUVER VEHICLE_YEAR = 1.,

CRASH_SEVERITY ---> SEVERITY TOTAL_VEHICLE_DAMAGE NUMBER_OFI_INJURED = 1.,

TRAVEL_SPEED ---> KINETIC_ENERGY,

TRAVEL_SPEED KINETIC_ENERGY ---> CRASH_SEVERITY;

FITINDEX ON(ONLY) = [CHISQ DF PROBCHI RMSEA SRMSR BENTLERCFI AGFI AIC] NOINDEXTYPE;

RUN;
```

For the second part of the thesis, MPlus 8.3 (Muthén and Muthén 2017) was used and probit links were used for the estimations. Model fitting was by robust weighted least squares estimator using a diagonal weight matrix (WLSMV) that is the default estimator for ordered categorical observed dependent variables. The code for this model is as follows:

```
DATA: FILE IS C:\ANIMAL CRASHES\WVC_CRASH.CSV;

VARIABLE: NAMES ARE

SEVERITY NUMBEROFINJURED ACCIDENTCOST DAMAGE PEAKHOUR LIGHTINGCONDITION WEATHERCONDITION SURFACECONDITION PAVEMENTCONDITION ROADTYPE ACCIDENTSITE HORIZONTALALIGNMENT VERTICALALIGNMENT;

USEVARIABLES ARE

SEVERITY NUMBEROFINJURED ACCIDENTCOST DAMAGE PEAKHOUR LIGHTINGCONDITION WEATHERCONDITION SURFACECONDITION PAVEMENTCONDITION ROADTYPE ACCIDENTSITE HORIZONTALALIGNMENT VERTICALALIGNMENT AGE VEHICLETYPE DAYOFWEEK;

CATEGORICAL ARE

SEVERITY NUMBER_OF_INJURED ACCIDENT_COST DAMAGE PEAK_HOUR LIGHTING_CONDITION WEATHER_CONDITION SURFACE_CONDITION PAVEMENT_CONDITION ROAD_TYPE ACCIDENT_SITE HORIZONTAL_ALIGNMENT VERTICAL_ALIGNMENT AGE VEHICLE_TYPE DAY_OF_WEEK;

MODEL:

CRASH_SEVERITY BY SEVERITY NUMBEROFINJURED ACCIDENTCOST DAMAGE;

SPEEDING_ATTITUDE BY ROAD_TYPE ACCIDENT_SITE PEAK_HOUR DAY_OF_WEEK SURFACE_CONDITION PAVEMENT_CONDITION AGE VEHICLE_TYPE;

BY HORIZONTAL_ALIGNMENT VERTICAL-ALIGNMENT LIGHTINGCONDITION WEATHERCONDITION;

CRASH_SEVERITY ON SPEEDING_ATTITUDE VISIBILITY_IMPAIRMENT;

SPEEDING_ATTITUDE ON VISIBILITY_IMPAIRMENT;

MODEL INDIRECT:

CRASH_SEVERITY IND VISIBILITY_IMPAIRMENT;
```

```
CRASH_SEVERITY IND SPEEDING_ATTITUDE;

OUTPUT: STANDARDIZED;
```