2020

Investigating the Leading Causes of Fatalities of Aging Pedestrians Using Bayesian Network Model

Luciano Lalika
University of North Florida, lucianollalika@gmail.com

INVESTIGATING THE LEADING CAUSES OF AGING PEDESTRIAN FATALITIES USING THE BAYESIAN NETWORK MODEL

By

Luciano Lalika

A thesis submitted to the School of Engineering

In partial fulfillment of the requirements for the degree of

Master of Science in Civil Engineering

UNIVERSITY OF NORTH FLORIDA

COLLEGE OF COMPUTING, ENGINEERING, AND CONSTRUCTION

December 2020

THESIS CERTIFICATE OF APPROVAL

The thesis of Luciano Lalika is approved:

Dr. Thobias Sando, Thesis Advisor and Committee Chairperson	**Date**
Dr. Ramin Shabanpour, Committee Member	**Date**
Dr. Cigdem Akan, Committee Member	**Date**

DEDICATION

... to the Almighty God and my family.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to everyone whose contribution is immensely appreciated.

Earnest thanks goes to my supervisor, Dr. Thobias Sando, for his immense support. He spent much of his time to instruct, assist, encourage, and advise me. I really appreciate his merits and devotional services.

I also extend my sincere thanks to my committee members, Dr. Ramin Shabanpour and Dr. Cigdem Akan, for their support as well.

Blessed are my family members for their outstanding devotion, care, encouragement, and prayers as well. They awesomely contributed a lot to the completion of this handwork.

Lastly, I would like to express my appreciation to all of the UNF Transportation Lab members for their support during of the 2019/2020 academic year.

Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AADT	Annual Average Daily Traffic
BCI	Bayesian Credible Interval
BDeu	Bayesian Dirichlet equivalent uniform
BLR	Bayesian Logistic Regression
BN	Bayesian Network
DAG	Directed Acyclic Graph
EBGM	Empirical Bayes Geometric Mean
FDOT	Florida Department of Transportation
GHC	Greedy Hill-Climbing
GOLM	Generalized Ordered Logit Model
HMC	Hamiltonian Monte Carlo
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation
NUTS	No U-Turn Sampling
PO	Proportional Odds
PPO	Partial Proportional Odds
U.S.	United States

ABSTRACT

Identifying factors associated with older pedestrian fatalities is key to implementing strategies aimed at improving pedestrian safety. This study focused on investigating the leading risk factors for older pedestrian fatalities at Florida roadway intersections. Analyses consisted of a Bayesian logistic regression (BLR) model to identify significant factors influencing pedestrian fatality, followed by a Bayesian Network (BN) model to identify the leading cause of pedestrian fatality among the statistically significant risk factors. Furthermore, the probabilistic inference of the leading causes of older pedestrian fatalities obtained from the BN was conducted through individual evidence predictive inference, diagnostic inference, and combined evidence predictive inference to understand the association with fatality. The models were developed with data from 913 pedestrian-vehicle crashes involving older pedestrians (65 years and older) that occurred at Florida roadway intersections from 2016 through 2018. Among the statistically significant factors retrieved by the BLR, vehicle maneuver, lighting condition, road type, posted speed, and driver age were found to have a direct probabilistic association with older pedestrian fatality. The diagnostic inference revealed that when a fatal pedestrian crash occurs, it is most likely associated with a vehicle moving straight, with a probability of 85.01%. Findings from this study can be used to inform the next step before developing the effective countermeasures for reducing the number of fatalities of older pedestrians in pedestrian-vehicle crashes.

CHAPTER 1 INTRODUCTION

Background

Each year, 1.35 million fatalities are recorded on roadways around the world, and more than half of those fatalities are pedestrians, motorcyclists, and cyclists (World Health Organization [WHO], 2018). In the United States (U.S.), traffic fatalities involving motor vehicle occupants decreased by 1% between 2009 and 2018. However, pedestrian fatalities increased by 53% over the same duration, despite an increase of less than 1% of the proportion of trips made by walking (Smart-Growth-America and National-Complete-Street-Coalition, 2020). Due to the rising trend of pedestrian fatalities and the increasing number of urban residents that choose to walk, improving pedestrian safety is a primary focus area for transportation agencies. However, pedestrian safety programs typically focus on specific population groups based on crash statistic demographics (New York City Department of Transportation [NYCDOT], 2010; Pedestrian and Bicycle Information Center [PBIC], 2007).

Older adults are overrepresented in deaths involving pedestrians. While the older population in the U.S. increased by 32% between 2009 and 2018, fatalities for 65-and-older pedestrians increased by 65% (National Highway Traffic Safety Admistration, 2020). This over-representation of older pedestrians in serious injury and fatal crashes compared to younger adults may be due, in part, to an age-related diminished ability to select gaps in oncoming traffic (Oxley et al., 2005). Moreover, increased fragility with age may exacerbate injury severity. Older pedestrians are slower to react to unexpected vehicle movements, thus, collisions with vehicles are relatively more severe (J. K. Kim et al., 2008). Older adult vulnerabilities, and the associated fatality statistics, require special consideration in developing pedestrian crash countermeasures, especially with ongoing changes in the population.

Today, the average life expectancy in the U.S. is higher than in any other period in history (Mather et al., 2019). In 2018, the older population comprised about 16% of the total U.S. population, or approximately one in seven Americans (Mather et al., 2019). By 2030, more than 20% of the total U.S. population will be over age 65 (Colby & Ortman, 2014). Florida, being a popular retirement destination, leads the nation with a population age 65 and older of 20%, higher than the national average of 16%. Moreover, this population group is expected to increase. In 2018, Florida also had the highest number of fatalities of older pedestrians, accounting for 11% of all older pedestrian fatalities in the U.S. (National Highway Traffic Safety Administration, 2020). Nevertheless, older adults are now more active and mobile and would prefer to maintain that lifestyle for as long as possible. Therefore, fatalities of older pedestrians in Florida warrant an in-depth examination that utilizes newer data resources and accurate analytical tools to better understand older pedestrian fatality risk and contributing factors.

Identification of factors contributing to the risk of older pedestrian fatality, along with their possible interdependency, is the first step towards improving safety. While the contribution of individual factors has been well explored, the majority of variables are interdependent. Although geometric characteristics may determine traffic capacity, traffic volume also varies with time of day and day of the week, and the inter-relationship among these variables may influence the risk of older pedestrian fatalities. Hence, using a regression model that ignores the relationship of contributing factors may limit potential inferences (de Oña et al., 2011; Kidando et al., 2017). Therefore, applying the Bayesian network model will help to reveal the probabilistic (interdependency) relationship among explanatory variables themselves (i.e., roadway type and traffic volume, time of day and traffic volume, and crash season and traffic volume) and explanatory variables and the response variable (i.e., lighting condition and crash severity), and thus, will enhance the selection and prioritization of the appropriate countermeasures to improve older pedestrian safety by the transportation agencies.

To identify the direct association among the contributing risk factors themselves (crash and roadway characteristics), together with the crash severity of the older pedestrians, this study evaluated crashes involving pedestrians age 65 and older that occurred at intersections on Florida roadways maintained by the Florida Department of Transportation (FDOT)) from 2016 to 2018.

Study Objective

The objective of this study was to identify the leading risk variables influencing older pedestrian fatalities at roadway intersections in Florida using Bayesian Logistic Regression (BLR) and Bayesian Network (BN) models. The BLR model was selected to estimate significant variables that influence the severity of crashes for aging pedestrians. Meanwhile, the BNs model was used to address three research elements not addressed by the BLR model: 1) which variables are associated with each other, among the significant risk factors, 2) which set of variables have the highest combined likelihood of association with an older pedestrian fatality, given combined evidence of two or more factors, and 3) diagnostics analysis capability to evaluate the most likely cause of the fatality, given previous occurrences of the corresponding crash type existed on the highway. The diagnostics analysis is also referred to as backward reasoning, from outputs to inputs, which is one of the benefits of using the BN model. This analysis cannot be estimated in the basic regression models, including the BLR, as they only perform forward predictions.

Moreover, the BN model provides a visual presentation of the network structure, which shows a coherent interpretation of the cause and effect connection (Kidando et al., 2019; Kutela & Teng, 2019; Xie & Waller, 2010). The network structure also offers a convenient way of reasoning and predicting events, given some evidence. Findings from this study may assist transportation officials in the design and prioritization of measures intended to improve older pedestrian safety.

Thesis Organization

This thesis consists of six chapters. Chapter 1 discusses general statistics of the aging pedestrian fatalities, their population growth and vulnerability, and the research objectives. Chapter 2 provides an extensive literature review on the contributing factors of the crash severity of aging pedestrians involved in vehicle-pedestrian crashes, different approaches used by previous studies to evaluate the contributing factors of older pedestrian crash severity, and the proposed approach in this study. Chapter 3 presents a detailed study methodology and a proposed algorithm. Chapter 4 discusses the study site, data collection, and data description, and Chapter 5 presents the analysis results. Finally, Chapter 6 provides the conclusion and highlights important findings.

CHAPTER 2 LITERATURE REVIEW

This chapter provides an extensive literature review on the factors influencing the crash severity of aging pedestrians involved in vehicle-pedestrian crashes, including various methodologies adopted by previous studies to evaluate the contributing factors of older pedestrian crash severity. The approach proposed in this study to evaluate contributing factors is also presented.

Factors Contributing to the Aging Pedestrian Crash Severity

Because of pedestrian vulnerability, vehicle-pedestrian crashes often result in severe injury to the pedestrian, or fatality (Pour-Rouholamin & Zhou, 2016). Several studies investigated factors that contribute to the severity of pedestrian crashes. As shown in Table 1, these studies explored the effects of pedestrian and driver attributes, demographic and socio-economic characteristics of crash locations, roadway features, vehicle attributes, and the nature of collisions on the severity of pedestrian crashes. Several previous studies explored the relationship between pedestrian age and fatal accidents, and found that the risk, of fatal and severe injuries resulting from a crash, increases with pedestrian age (Haleem et al., 2015; J. K. Kim et al., 2008; Pour-Rouholamin & Zhou, 2016; Salon & McIntyre, 2018; Sarkar et al., 2012; Sasidharan & Menéndez, 2014). However, among these studies, different age thresholds have been used to classify the population group perceived to be most susceptible to fatality and severe injuries, e.g., 55 years and older (Sarkar et al., 2012), 65 years and older (Pour-Rouholamin & Zhou, 2016; Zhai et al., 2019), 70 years and older (Salon & McIntyre, 2018), and 75 years and older (Sasidharan & Menéndez, 2014).

Several studies focused on understanding the severity of pedestrian crashes involving older adults (see Table 1). Using the Bayesian complementary log-log, one study found that the likelihood of fatal crashes involving older adult pedestrians increases with alcohol use, first harmful event, vehicle movement, shoulder type, and posted speed (Kitali et al., 2017).

Table 1: Summary of related studies in older pedestrian crash injury severity analysis

Reference	Subject	Method	Notable significant factors increasing the severity of pedestrian crashes
Oh et al. (2005)	All pedestrians	Logistic regression	Heavy vehicles
Gorrie et al. (2008)	Older adults	Logistic regression	Older pedestrians with moderate to high neurofibrillary tangles were more likely to be: at least partially responsible for the incident, injured while in low complexity situations, involved in impacts with reversing vehicles, impacted in near lanes of traffic, and struck by a vehicle off-road.
Kim et al (2008)	All pedestrians	Heteroskedastic Logit model	Increasing pedestrian age, intoxicated driver, traffic sign, two-way divided roadways, speeding, male driver, vehicle type, both driver and pedestrian at fault, roadway lighting, vehicle movement, and pedestrian-only at fault.
Sarkar et al. (2012)	All pedestrians	Logistic regression	Older adult pedestrians (>55 years), children (<15 years of age), jaywalking, no traffic control, stop control, pedestrian crossings, rainy season, and pedestrian collisions with trucks, buses, baby taxis or tempos (auto-rickshaws), and tractors.
Sasidharan and Menendez (2014)	All pedestrians	PPO model	Pedestrian age and gender, roadway lighting, and jaywalking.
Haleem et al. (2015)	All pedestrians	Mixed logit model	At signalized intersections: Higher truck percentage, higher AADT, speed limit, pedestrian age, rainy weather, roadway lighting, and speed limit. At unsignalized intersections: Pedestrian walking along the roadway, pedestrian age, at-fault pedestrians, vehicle type, roadway lighting, and higher speed limit.
Pour-Rouholamin and Zhou (2016)	All pedestrians	PO, PPO and GOLM model	Older adult pedestrians (>65 years), pedestrians not wearing contrasting clothing, adult drivers (16–24), drunk drivers, time of day, divided highways, multilane highways, darkness, and heavy vehicles.
Kitali et al. (2017)	Older adults	Bayesian complementary log-log	Pedestrian age, alcohol involvement, first harmful event, vehicle movement, shoulder type, and posted speed.
Salon and McIntyre (2018)	All pedestrians	Logistic regression	Pedestrians age (>70 years), driver sobriety, daylight, rain or fog, speed limit (>25 mph) and larger vehicles.
Das et al. (2019)	Older adults	EBGM data mining	Backing vehicles, at night segment-related crashes, failure to yield while crossing at intersections, roadway lighting.
Li and Fan (2019)	All pedestrians	Latent class clustering and PPO model	Driver sobriety, heavy vehicles, and pedestrian or driver gender.
Zhai et al. (2019)	All pedestrians	Mixed logit model	High temperature than 30 °C, rainy conditions, pedestrian age (66–80 years), negligent driving behaviors, vehicle type, and crashes at crosswalks.

Note: The literature review in the table is not exhaustive of previous works on the severity of pedestrian crashes; EBGM = Empirical Bayes Geometric Mean; PPO = Partial Proportional Odds; PO = Proportional Odds; GOLM = Generalized Ordered Logit Model.

Another study adapted the Empirical Bayes Geometric Mean (EBGM) data mining method to determine the key associations between contributing factors of pedestrian crashes involving older adults (Das et al., 2019). Study findings revealed several patterns, such as female pedestrians age 79 and over and male pedestrians age 65 to 69 being significantly involved in fatal crashes that are vehicle-backing related and segment related, respectively. A study by Kim et al. (2008) explored the effects of age-specific heteroscedasticity in the severity of pedestrian crashes, using a heteroskedastic generalized extreme value model, and concluded that pedestrian age, especially over age 65, introduces heteroscedasticity in the probability of a fatal crash.

Another study by Gorrie et al. (2008) investigated the role of cognitive decline associated with dementia neuropathology in fatalities of older adult pedestrians. Results indicated that the cognitive decline increased the likelihood of older pedestrians being involved in fatal crashes in specific situations, including low complexity situations, impacts with reversing vehicles, impacts near traffic lanes, or being struck by a vehicle off-road. In addition to identifying contributing factors, the previous studies reviewed introduced analysis approaches that are more suitable for evaluating the severity of pedestrian crashes.

Approaches Used to Evaluate Factors Contributing to Pedestrian Crash Severity

Various methodological approaches have been used to evaluate factors that contribute to the severity of pedestrian crashes (see Table 1). Generalized linear models for binary response data, such as logistic regression model (Oh et al., 2005; Salon & McIntyre, 2018; Sarkar et al., 2012) and complementary log-log model (Kitali et al., 2017), were used to associate the severity of crashes and contributing factors by categorizing severity into two groups (minor injuries and severe injuries). One study analyzed more than two levels of crash severity (minor injuries, severe injuries, and fatal) using three ordered-response models, including the partial proportional odds

(PPO) model, the conventional ordered (proportional odds – PO) model, and the generalized ordered logit model (GOLM) (Pour-Rouholamin & Zhou, 2016). Results revealed that, among the three models, the PPO model better explained the severity of pedestrian crashes.

Another study applied a latent class clustering to classify crashes with different distribution characteristics of contributing factors and subsequently used the PPO model to identify factors that influence the severity (Li & Fan, 2019). Results showed that the latent class segmented sub-models were more effective in identifying the contributing factors than a full model without classes. A few studies also used a mixed logit model to identify factors that influence the severity of pedestrian-vehicle crashes (Haleem et al., 2015; Zhai et al., 2019). Despite the benefits, previously applied methods failed to consider the possible interdependence between explanatory variables influencing the risk of pedestrian injury severity. Furthermore, using a model that ignores interdependence may lead to wrong inferences being made (de Oña et al., 2011; Kidando et al., 2017).

Proposed Approach

Bayesian networks have been used to analyze the influence of factors on the response variables and the interrelationship of factors graphically (Stylianou & Dimitriou, 2018). The BN model presents the probabilistic relationship among variables as a graphical model and is easy to understand models presented as a function. The graph of the BN model includes nodes that represent random variables and edges that show the conditional dependence between the variables (Kidando et al., 2019; Kutela & Teng, 2019). This graphical representation is also known as the directed acyclic graph (DAG).

Contrary to binary regression models, the BN model can incorporate the use of prior knowledge of current situations and use the evidence to update model scenarios (Y. Kim & Park, 2019). Literature also suggests that BNs perform the same or better than binary regression models

(Lee et al., 2005; Witteveen et al., 2018). The BN model is superior in the interpretation of the results, compared to the parametric regression model, as it visually describes the probabilistic relationship among variables in the model (Cong et al., 2018; Kutela & Teng, 2019). It relaxes the independence assumption of variables, seen in regression models, which in turn yields better performance (Kidando et al., 2019; Kutela & Teng, 2019; Xie & Waller, 2010).

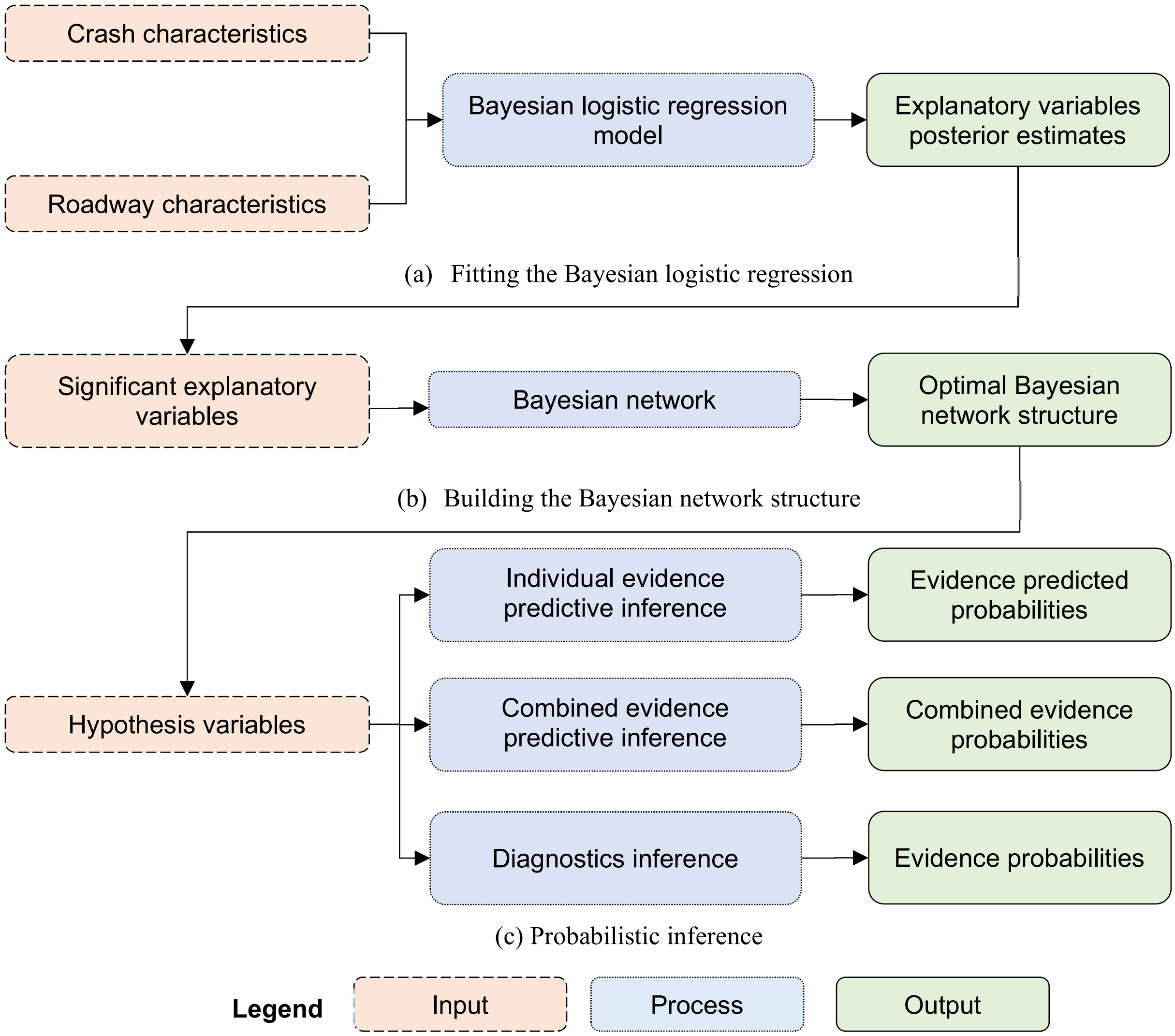
In summary, over the past few decades, several studies have focused on investigating factors contributing to the crash severity of pedestrians in vehicle-pedestrian crashes using ordinary models (Z. Chen & Fan, 2019; Moudon et al., 2011). The conventional models are categorized as ordered response models, such as the logit model, and non-ordered response models, such as the multinomial logistic model. The ordered models assume that the predictors can only have the same effect on different levels of the dependent variable, which is often not the case with crash injury severities. On the other hand, non-ordered analyses completely ignore the inherently hierarchical nature of crash injury severities. Therefore, treating the crash severity as either ordered or non-ordered ignores some of the essential principles (Sasidharan & Menéndez, 2014). To address these concerns, in recent years more sophisticated approaches, such as the partial proportional odds model, have been recently applied to bridge the gap between ordered and non-ordered crash severity modeling frameworks by considering the inherent ordered nature of the traffic crash severity and allowing at least some of the predictors to have different impacts on different levels of crash injury severity (Sasidharan & Menéndez, 2014). However, these advanced methods also ignore the interdependency relationship between explanatory variables themselves and between explanatory variables and the response variable. Alternatively, the Bayesian network model allows probabilistic inference, which utilizes the benefits of the independence relationships among the variables (explanatory variables themselves and explanatory variables and response variable) (Kidando et al., 2019).

Although previous studies have used various methods to analyze the severity of pedestrian crashes involving older adults, to the best of authors' knowledge, none have applied the BN model. The present study developed a BN model to evaluate the contributing factors of fatal pedestrian crashes involving older adults and understand the interrelationship between the contributing factors. To reduce the complexity of the developed BN model, the model was built using only significant variables identified from the BLR.

CHAPTER 3 METHODOLOGY

This thesis proposed BLR and BNs to investigate the leading causes of fatalities of aging pedestrians involved in vehicle-pedestrian crashes. The BLR model was used to determine the significant variables influencing the severity of crashes for aging pedestrians, and the BN model was adopted to establish the probabilistic relationship between the factors influencing fatal crashes of aging pedestrians revealed by the BLR model.

Figure 1 presents the approach used to investigate the probabilistic relationship between the factors contributing to fatalities of older pedestrians at intersections. Specifically, the workflow indicated in Figure 1 can be divided into three main steps: (a) fitting the BLR model; (b) building the BN structure; and (c) implementing probabilistic inferences in the trained BN, which are prediction and diagnostics reasoning. With prediction inference, the model can estimate the impact of certain evidence, or evidence set, on the change in response variable outcomes. Meanwhile, diagnostic reasoning, given a response variable outcome, determines the probability that it is certain evidence.



Note: Evidence = A condition that has been observed, e.g., road type; Hypothesis variable = A variable that has a direct probabilistic relationship with the occurrence of fatal older pedestrian-vehicle.

Figure 1. Methodology workflow.

Bayesian Logistic Regression (BLR) Model

To identify significant variables contributing to the crash severity of aging pedestrians in the vehicle-pedestrian crashes, the Bayesian logistic regression model was applied. The binary BLR for determining an aging pedestrian crash severity, Y_i, can be defined with a vector of explanatory variables, $\boldsymbol{X}_i$, as expressed in Equation 1.

$$Y_i = \begin{cases} 1 \; if \; the \; crash \; severity \; is \; fatality \\ 0 \qquad else \; injury \end{cases}$$

$$Y_i \sim Bernoulli(\lambda_i) \tag{1}$$

$$logit\,(\lambda_i) = \beta_0 + \boldsymbol{\beta_j X_i} + \varepsilon_i$$

$$\beta_0,\ \beta_j \sim\ N(0, 100)$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$\sigma \sim HalfNormal(0, 100)$$

where, λ_i is the severity function of observation i, $\boldsymbol{\beta_j}$ is a vector of the parameters, and ε_i is the stochastic error term.

For the calibration of the BLR model parameters, a full Bayes approach through Markov Chain Monte Carlo (MCMC) simulation was adopted. No U-Turn sampling (NUTS) steps were adopted in the analysis as well. The NUTS is based on the Hamiltonian Monte Carlo (HMC) that avoids the random walk behavior, which has a greater advantage over convergence during sampling compared to other sampling techniques, such as Metropolis. More information regarding the comparison of NUTS and other techniques for sampling the posterior distribution can be found in the study by Hoffman and Gelman (2014).

With the full Bayes approach, the prior distribution of each model parameter in the model must be defined. Non-informative priors were specified for this purpose. In the absence of informative priors, which are generally obtained from previous studies that performed similar analyses, non-informative priors were assigned to the model parameters, a common practice in the Bayesian paradigm (Kruschke, 2013). The non-informative priors impose a minimal influence over the estimates and allow the data characteristics to dominate instead (Ntzoufras, 2009).

As indicated in Equation 1, the prior distribution, specified for the regression coefficients β_0 and β_j , is a normal distribution with a mean of zero, and a variance of 100. The convergence of the MCMC simulations was assessed using the Gelman-Rubin Diagnostic statistic. Also, a visual diagnostics approach was used to assess the convergence of the chains, including the use of the autocorrelation plot and trace plot of each parameter.

Bayesian Network (BN)

To integrate subjectivity, as well as reveal hidden probabilistic relationships among variables, the structure learning of the BN was conducted using an algorithm and expert knowledge. More specifically, the BN structure was trained, using the Bayesian Dirichlet equivalent uniform (BDeu) as the search algorithm. After the BN structure was developed, the expert knowledge and findings from previous studies were applied to refine the trained BN structure by only changing some of the arrow directions, such as the cause-effect direction. A similar approach was adopted in several previous studies (Cong et al., 2018; Stylianou & Dimitriou, 2018; Xie & Waller, 2010).

The greedy hill-climbing (GHC) algorithm was adopted as the search strategy to retrieve the optimal network structure from the data. The GHC algorithm iteratively adds, removes, and reverses edges to find a network with the highest score (Kidando et al., 2019). The best network structure is obtained once the score cannot be improved further in the search process. Assume dataset *T* is used to train the network structure *B*, the BN structure then obtains the best network structure *B* by maximizing the scoring value, $\mathrm{BDeu}(B,T)$. The BDeu metric can be expressed using Equation 2.

$$\mathrm{BDeu}(B,T) = \log\big(P(B)\big) + \sum_{i=1}^{n}\sum_{j=1}^{q_i}\left(log\left(\frac{\Gamma\left(\frac{N'}{q_i}\right)}{\Gamma\left(N_{ij}+\frac{N'}{q_i}\right)}\right) + \sum_{k=1}^{r_i} log\left(\frac{\Gamma\left(N_{ijk}+\frac{N'}{r_i q_i}\right)}{\Gamma\left(\frac{N'}{r_i q_i}\right)}\right)\right) \tag{2}$$

where,

N' is the equivalent sample size,

N_{ij} is the number of instances in the data *T*, where variable $\prod_{X_i}$ takes their *j-th* configuration,

such that $\sum_{k=1}^{r_i} N_{ijk} = N_{ij}$,

N_{ijk} is the number of instances in the data T,

r_i is the number of states of the finite random variable $\boldsymbol{X}_i$,

$q_i = \prod_{\boldsymbol{X}_i \in \boldsymbol{X}_i} r_i$ is the number of possible configurations of the parent set $\prod_{\boldsymbol{X}_i}$ of $\boldsymbol{X}_i$, and

n is the number of observations.

Given the estimated optimal BN structure, and the evidence associated with the hypothesis variables, the model parameters, which are the discrete probability values in the conditional probability tables, were estimated using the maximum likelihood estimation (MLE) method.

Probabilistic Inference

Using the optimal network retrieved in the analysis, the probabilistic inference through prediction reasoning and diagnostics reasoning was conducted. The predictive inference involves valuing the probability of the occurrence of an event, e.g., fatal older pedestrian-vehicle crash given one evidence (individual evidence predictive inference) or more (combined evidence predictive inference). In other words, prediction inference is the process of reasoning forward from inputs, i.e., hypothesis variables, to outputs, such as injury or fatal crash. Unlike prediction, diagnostics reasoning focuses on determining the likely value of inputs given what we know about outputs, i.e., injury or fatal crash. Thus, diagnostics can be thought of as the process of reasoning backward from outputs to inputs. Predictive reasoning answers questions, such as what is the probability of an older pedestrian-vehicle fatal crash to occur when the vehicle maneuver at the time of the crash was straight? Meanwhile, diagnostics reasoning answers questions, such as what is the probability that the vehicle was moving straight if a fatal older pedestrian-vehicle crash occurred?

Individual Evidence Predictive Inference

The predictive inference can be performed using one evidence at a time, and in this case, referred to as the individual evidence predictive inference. In other words, the BN model can be used to estimate the impact of certain evidence on the change in the response variable outcomes. As depicted in Equation 3, prediction inference is conducted by assigning evidence of a hypothesis variable.

$$P(Fatality = i | e_x = 1) \tag{3}$$

where, i is the probability of the older-pedestrian-vehicle crash resulting in fatality, and e_x represents the evidence of a hypothesis variable x, e.g., lighting condition.

To provide a better interpretation of the hypothesis variables on how they influence fatality prediction, a variable impact analysis was performed using the direct pseudoelasticity. Using Equation 4, the direct pseudoelasticity quantifies how the probability of an event (e.g., fatal crash occurrence) varies with the evidence without requiring direct comparison to the probability of another event.

$$E_x^{P_i} = \frac{P_i[x=1] - P_i[x=0]}{P_i[x=0]} \tag{4}$$

where, $E_x^{P_i}$ is the direct pseudoelasticity of variable x for crash i from the indicator variable x, with $x = 0$ set as the base condition.

For a variable with multiple categorical values, a variable influence analysis was conducted for each of its values, where one of its values was set as the base condition. The relative variable influence for each other value on all categories was calculated accordingly (Cong et al., 2018).

Combined Evidence Predictive Inference

In addition to estimating the influence of each hypothesis variable on the likelihood of a fatal crash, the impact of concurrent evidence was also studied. For instance, one might be interested in determining the likelihood of fatal crash occurrence given that the vehicle is moving straight, and the lighting condition is nighttime at the time of the crash. The predicted probability of a fatal crash to occur, based on the combined evidence, was estimated using Equation 5.

$$P\left(Fatality = i | e_{x_1} = x_1, e_{x_2} = x_2 , \ldots , e_{x_h} = X_h\right) \quad (5)$$

where, e_x is the evidence of a hypothesis variable x, and x_h is the observed evidence of hypothesis variables X. Similar to individual hypothesis variable analysis, for the combined evidence, each observed evidence was assigned a certainty value of 1, i.e., $P\left(Fatality = i | e_{x_1} = x_1, e_{x_2} = x_2, \ldots , e_{x_h} = X_h = 1\right)$.

Diagnostics Inference

Given a fatal or injury crash has occurred, the diagnostics inference determines the likelihood of hypothesis variables. As indicated in Equation 6, diagnostics inference is conducted by assigning certainty of the crash severity and observing the result of the hypothesis variables.

$$P(e_x = i | Fatality = 1) \quad (6)$$

where, e_x represents the evidence of a hypothesis variable x, e.g., lighting condition, and i is the probability of the evidence of a hypothesis variable.

The conditional probability distributions of the trained BNs structure were estimated using the maximum likelihood approach. Moreover, both BN structure training and inferences were implemented using the pyAgrum 0.15.2 program, a Python open-source package (Wuillemin, 2019).

CHAPTER 4 STUDY SITE AND DATA COLLECTION

This chapter presents the location of the analyzed area and describes various sources where data was obtained. Also discussed are the definitions of the crash-related variables used in this study and the descriptive statistics of the data used for analysis.

Study Site and Data Collection

Crash data were retrieved from the Signal Four Analytics database, a statewide interactive web-based geospatial crash analytical tool, developed and hosted by the University of Florida. Data acquired consisted of pedestrian-vehicle crashes that occurred at Florida intersections and involved older pedestrians (age 65+) during the years 2016, 2017, and 2018. Dataset attributes consisted of crash type, driver characteristics, and some of the roadway features. Also, a police crash report of each recorded incident was extracted from the database to manually collect missing information, such as vehicle maneuver, traffic control device, road type, pedestrian gender, and driver gender.

Furthermore, other roadway information on the crash locations was collected from the 2017 Florida Department of Transportation (FDOT) Roadway Characteristics Inventory (RCI). Roadway characteristics retrieved from the RCI database included shoulder type, annual average daily traffic (AADT), posted speed limit, and roadway surface width. Data for a total of 913 crashes involving older pedestrians at intersections in Florida were collected and used in the analysis.

Data Description

This section presents the definitions of the variables and their respective attributes used in this research, together with the descriptive statistics of the variables. The variables were classified into two major groups: crash characteristics and roadway characteristics. Crash characteristics included vehicle maneuver, crash season, time of day, alcohol involvement, lighting condition, road surface condition, pedestrian age, pedestrian gender, driver age, and driver gender. On the other hand, roadway characteristics comprised road surface width, shoulder type, traffic control device, road type, posted speed limit, and AADT. Table 2 shows the descriptive statistics of all variables used for analysis in this study. It shows the variables and their corresponding attributes, count, and percentage for both crash severity levels (injury and fatality) and the total count of each category.

Crash characteristics

As presented in Table 2, the vehicle maneuver was categorized into two movements: straight-ahead or turning. Approximately 87% of older pedestrian fatal crashes occurred when the vehicles were moving straight, while only 13% of fatalities occurred when vehicles were making turning movements. The crash season was grouped into two seasons: spring or summer-fall. About 68% of older pedestrian fatal crashes occurred during the summer-fall season, while only 32% occurred during the spring season. Time of day was categorized as off-peak hours or peak hours. Half (50%) of older pedestrian fatal crashes occurred during off-peak hour periods, while the remaining half (50%) occurred during peak hour periods. Alcohol involvement was categorized as to whether alcohol was involved in either the pedestrian or the driver during the crash occurrence. In approximately 92% of older pedestrian fatal crashes, neither the driver nor the pedestrian had used alcohol, while only 8% involved alcohol use by either the driver or the pedestrian. Lighting condition was categorized as daytime or nighttime lighting conditions. Data showed that 63% of crashes that resulted in older pedestrian fatalities occurred at night, and only 37% occurred during the daytime.

Road surface condition was classified as to whether the road surface was dry or wet during the crash occurrence. About 89% of older pedestrian fatal crashes occurred on a dry road surface, and only 11% occurred on a wet road surface.

Table 2: Descriptive statistics of analysis variables

	Variable	Category	Crash Severity Type				Total
			Injury		Fatality		
			Count	Percent (%)	Count	Percent (%)	
Crash characteristics	Vehicle maneuver	Straight ahead	387	51	135	87	522
		Turn	370	49	21	13	391
	Crash season	Spring	294	39	50	32	344
		Summer-Fall	463	61	106	68	569
	Time of day	Off-peak hour	408	54	78	50	486
		Peak hour	349	46	78	50	427
	Alcohol involvement	No	739	98	144	92	883
		Yes	18	2	12	8	30
	Lighting condition	Daytime	523	69	57	37	580
		Nighttime	234	31	99	63	333
	Road surface condition	Dry	701	93	139	89	840
		Wet	56	7	17	11	73
	Pedestrian age	65-74	474	63	88	56	562
		>74	283	37	68	44	351
	Pedestrian gender	Male	385	51	108	69	493
		Female	372	49	48	31	420
	Driver age	Junior (≤59)	490	65	124	79	614
		Senior (≥60)	267	35	32	21	299
	Driver gender	Male	432	57	104	67	536
		Female	325	43	52	33	377
Roadway characteristics	Road surface width (feet)	<22	163	22	23	15	186
		≥22	594	78	133	85	727
	Shoulder type	Curbed	447	59	103	66	550
		Non-curbed	310	41	53	34	363
	Traffic control device	Controlled	337	45	42	27	379
		No Control	420	55	114	73	534
	Road type	Undivided	474	63	58	37	532
		Divided	293	37	98	63	381
	Posted speed (mph)	Low (<45)	488	64	73	47	561
		High (≥45)	269	36	83	53	352
	AADT (vpd)	≤10,000	205	27	25	16	230
		10,000-20,000	167	22	29	19	196
		>20,000	385	51	102	65	487
	Response variable	Crash Severity	757	83	156	17	913

The data showed that 69% of older pedestrian fatal crashes involved male pedestrians, while only 31% involved female pedestrians. Also, more than half (56%) of older pedestrian fatal crashes involved pedestrians aged 65 to 74, and 44% involved pedestrians aged 75 and older. Over two thirds (67%) of fatal crashes of older pedestrians involved male drivers, while the remaining one third (33%) involved female drivers. Also, 79% of older pedestrian fatal crashes involved drivers age 59 or below and only 21% involved drivers age 60 or older.

Road characteristics

As shown in Table 2, road surface width was categorized into two groups: a width of less than 22 feet or greater than or equal to 22 feet. About 85% of older pedestrian fatal crashes occurred on the roadways with a surface width greater or equal to 22 feet, while only 15% of older pedestrian fatal crashes occurred on a roadway with a surface width of less than 22 feet. Shoulder type was categorized into two groups depending on the presence of curbs (curbed or uncurbed shoulders). Data showed that nearly two-thirds of older pedestrian fatal crashes (66%) occurred on curbed shoulder roadways, while only one third (34%) occurred on non-curbed shoulder roadways. The traffic control device variable indicates whether the intersection was controlled or not. This variable was categorized into two groups: controlled or uncontrolled intersections. Approximately, 73% of all older pedestrian fatal crashes occurred at uncontrolled intersections, while only 27% occurred at controlled intersections. The road type variable defines whether the highway where a crash occurred was a divided or undivided roadway. Descriptive statistics revealed that 63% of fatal crashes occurred on divided highways, and only 37% occurred on undivided sections. The posted speed limit was categorized into two groups: a posted speed limit of less than 45 mph or greater than or equal to 45 mph. About 47% of all older pedestrian fatal crashes occurred on roadways with a posted speed limit of less than 45 mph, and 53% of all older pedestrian fatal crashes occurred on roadways with a posted speed limit greater than or equal to 45 mph. Intersection approaches with an AADT greater than 20,000 vehicles/day was associated with 65% of older pedestrian fatal crashes, while 19% and 16% of fatal crashes involving older pedestrians occurred on approaches with an AADT of 10,000 to 20,000 vehicles/day and less than or equal to 10,000 vehicles/day, respectively.

CHAPTER 5 RESULTS AND DISCUSSION

This chapter discusses the analysis results from the BLR and BN models. Also discussed are the individual evidence predictive inferences, combined evidence predictive inferences, and diagnostic inferences included in the BN model.

Bayesian Logistic Regression Model Results

Table 3 shows the BLR model results. Of the 16 explanatory variables included in the model, the following eight variables were found to be statistically significant at a 90% Bayesian Credible Interval (BCI): vehicle maneuver, road type, pedestrian gender, lighting condition, shoulder type, AADT, posted speed, and driver age. The following sections discuss these variables further.

Vehicle Maneuver

Results from Table 3 reveal that older pedestrian crashes involving turning vehicles have a 78.4% lower risk of older pedestrian fatality compared to crashes involving vehicles moving straight [Odds ratio 0.216, 90% BCI (-2.008, -1.067)]. Vehicles traveling straight generally have higher impact energy at the time of the crash, compared to turning vehicles. Thus, older pedestrians involved in crashes with straight-moving vehicles have an increased risk of sustaining a fatality, compared to crashes involving turning vehicles. This finding is consistent with the previous research (Kitali et al., 2017; Moudon et al., 2011b; Zahabi et al., 2011).

Road Type

Compared to pedestrian-vehicle crashes that occurred on undivided roadways, older pedestrians involved in crashes on divided roadways are about 2 times more likely to result in fatality [Odds ratio 1.998, 90% BCI (0.368, 1.054)]. This finding is supported by a several previous studies (J. K. Kim et al., 2008; Pour-Rouholamin & Zhou, 2016; Chen et al., 2019). These studies suggested that the higher likelihood of older pedestrian fatalities may be attributed to vehicles traveling at relatively higher speeds on divided roadways, as well as longer crossing distances, which expose older pedestrians to a greater risk of a severe crash.

Pedestrian Gender

As indicated in Table 3, crashes involving female older pedestrians have a 33.1% lower risk of resulting in a pedestrian fatality, compared to crashes involving male older pedestrians [Odds ratio 0.669, 90% BCI (-0.741, -0.048)]. This finding is similar to several previous studies (Hu and Cicchino 2018; Onieva-Gracia et al., 2016; Zhu et al., 2013), and may be attributed to the walking exposure of male pedestrians, compared to female pedestrians.

Lighting Condition

Nighttime lighting conditions also pose a greater risk of pedestrian-vehicle crashes. As suggested in Table 3, crashes involving older pedestrians at night are 2.5 times more likely to result in a pedestrian fatality, compared to the crashes that occur during the daytime [Odds ratio 2.447, 90% BCI (0.563, 1.263)]. Consistent with previous studies, this finding was expected since nighttime is associated with reduced visibility, compared to daytime lighting conditions (Hu & Cicchino, 2018; J. K. Kim et al., 2008; Montella et al., 2011; Zahabi et al., 2011).

Table 3: Results of the Bayesian logistic regression model

Variables	Factors	Estimate	Odds Ratio	90% BCI	
				L-90% CI	U-90% CI
Vehicle maneuver	Straight ahead				
	Turn	**-1.532**	**0.216**	**-2.008**	**-1.067**
Road type	Undivided				
	Divided	**0.692**	**1.998**	**0.364**	**1.054**
Traffic control device	Controlled				
	Uncontrolled	-0.146	0.864	-0.551	0.253
Driver gender	Male				
	Female	-0.274	0.76	-0.618	0.058
Pedestrian gender	Male				
	Female	**-0.402**	**0.669**	**-0.741**	**-0.048**
Season	Spring				
	Summer-Fall	0.268	1.307	-0.08	0.61
Time of the day	Off-peak hours				
	Peak hours	0.115	1.122	-0.226	0.458
Alcohol related	No				
	Yes	0.514	1.672	-0.233	1.243
Light condition	Daytime				
	Nighttime	**0.907**	**2.477**	**0.563**	**1.263**
Type of shoulder	Curbed				
	Noncurbed	**-0.418**	**0.658**	**-0.774**	**-0.076**
Road surface condition	Dry				
	Wet	0.364	1.439	-0.195	0.906
AADT (vpd)	<10,000				
	10,000-20,000	0.453	1.573	-0.073	0.952
	>20,000	**0.586**	**1.797**	**0.126**	**1.05**
Pedestrian age	65-74				
	>74	0.333	1.395	-0.022	0.68
Driver age	Junior (<60)				
	Senior (≥60)	**-0.621**	**0.537**	**-0.989**	**-0.254**
Road surface width (feet)	<22				
	≥22	0.017	1.017	-0.442	0.486
Posted speed (mph)	Low (<45)				
	High (≥45)	**0.400**	**1.492**	**0.053**	**0.745**

Note: Values in bold are significant at the 90% Bayesian credible interval (BCI). vpd stands for vehicles per day and mph represents miles per hour.

Shoulder Type

The likelihood of older pedestrian fatality is 34.2% lower with crashes that occur on roadways with non-curbed shoulders, compared to crashes on roadways with curbed shoulders [Odds ratio 0.658, 90% BCI (-0.774, -0.076)] (Table 3). This finding was contrary to expectations, as well as observations by Kitali el al. (2017), which suggested that curbed shoulders act as a barrier to pedestrians by discouraging vehicles from departing the roadway, and therefore, may minimize the injury severity for pedestrians. Moreover, previous research on the association of shoulder type and pedestrian crashes is scarce; therefore, more efforts are required to determine the evidence of the relationship observed in this study.

AADT

Results presented in Table 3 suggest that roadways with higher traffic volume (AADT greater than 20,000) are associated with a higher number of older pedestrian fatalities (79.7%), compared to roadways with lower AADT (less than or equal to 10,000) [Odds ratio 1.797, 90% BCI (0.126, 1.050)]. This observation is consistent with previous studies (Obeng & Rokonuzzaman, 2013; Haleem et al., 2015), revealing that an increase of traffic volume increases the likelihood of pedestrian fatality, especially at signalized intersections.

Posted Speed

As indicated in Table 3, crashes occurred on roadways with high speed limit (≥45 mph) are more likely to result in an older pedestrian fatality, compared to roadways with lower speed limit (<45 mph), by 49.2% [Odds ratio 1.492, 90% BCI (0.053, 0.745)]. This finding was expected since crashes that occur at higher speeds result in a higher impact to aging pedestrians, therefore the chance of fatality increases (Kitali et al., 2017).

Driver Age

Table 3 indicates that older pedestrian-vehicle crashes involving drivers age 60 and over are 46.3% less likely to result in a pedestrian fatality, compared to crashes involving drivers age below 60 [Odds ratio 0.537, 90% BCI (-0.989, -0.254)]. This finding is consistent with previous research that older drivers (65+ years) tend to be more cautious and drive at relatively lower speeds roads, compared to the younger age groups (J. K. Kim et al., 2008; Pour-Rouholamin & Zhou, 2016).

Discrete Bayesian Networks Results

Optimal Bayesian Network Structure

Figure 2 illustrates the optimal BN structure that was developed. The hybrid approach revealed that four nodes were directly associated (dependence relationship) with older pedestrian crash severity. These factors are also referred to as hypothesis variables. The hypothesis variables determined by the model include lighting condition, vehicle maneuver, road type, posted speed and driver age. Also, the optimal structure showed that road type, AADT and posted speed have direct relationship with each other. On the other hand, based on the BN results, pedestrian gender and shoulder type had neither a direct nor indirect association with older pedestrian crash severity.

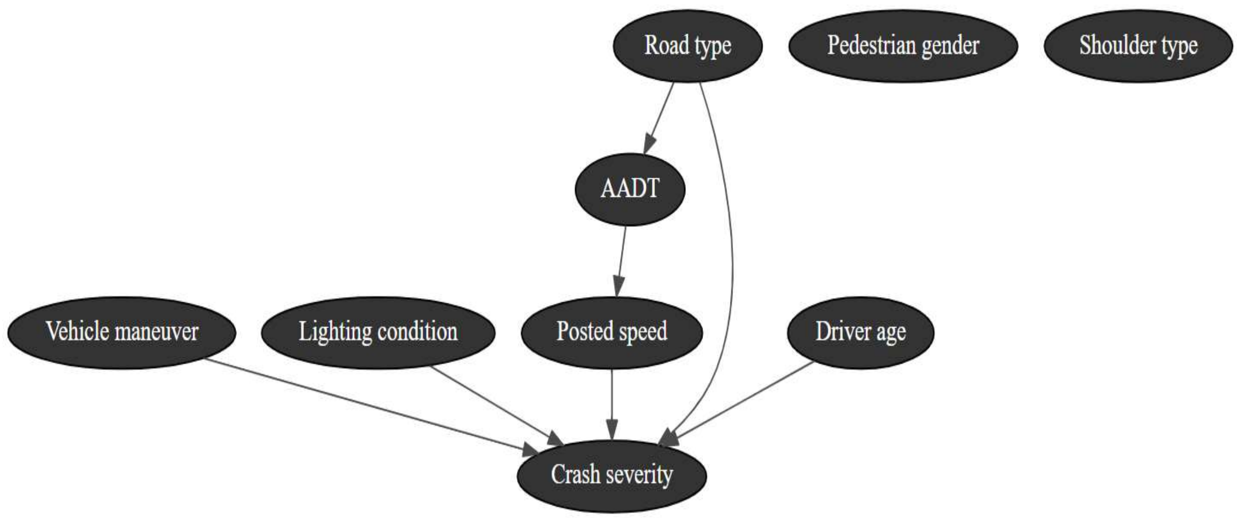


Figure 2. Optimal Bayesian network structure.

Individual Evidence Predictive Inference Results

Based on the optimal BN structure shown in Figure 2, a sensitivity analysis was conducted to query the network structure given a set of evidence. The analysis focused on variables that were revealed to have a direct association with older pedestrian crash severity. Table 4 provides the results of the sensitivity analysis of the hypothesis variables.

Table 4: Predicted probability of older pedestrian injury and fatality from crashes

Variable	Observed evidence	Predicted probabilities			
		Injury	Pseudoelasticity	Fatality	Pseudoelasticity
Vehicle maneuver	Straight ahead	0.7866	-17.2%	0.2134	325.1%
	Turning	0.9498	20.7%	0.0502	-76.5%
Lighting condition	Daytime	0.8918	12.2%	0.1082	-47.2%
	Nighttime	0.7951	-10.8%	0.2049	89.4%
Road type	Undivided	0.8953	11.6%	0.1047	-47.0%
	Divided	0.8023	-10.4%	0.1977	88.8%
Posted speed	Low	0.8700	4.2%	0.1300	-21.3%
	High	0.8349	-4.0%	0.1651	27.0%
Driver age	Junior	0.8361	-6.9%	0.1639	61.3%
	Senior	0.8984	7.5%	0.1016	-38.0%

An evaluation of the individual evidence was conducted to identify how likely individual evidence affects the risk of older pedestrian fatalities. After assigning the hypothesis variables, i.e., vehicle maneuver, lighting condition, road type, posted speed, and driver age, as evidence in the analysis, results shown in Table 4 revealed that all hypothesis variables exhibit a similar trend to the BLR model results. In other words, crashes are more likely to be fatal for older pedestrians when the vehicle is moving straight, at night, on a divided highway, on high posted speed limit road, and involving a junior driver (age 59 or younger).

Diagnostic Inference Results

Unlike forward prediction, a task performed by regression models, backward reasoning involves prediction from the outcome to the cause. This reasoning analysis in the BNs is conducted by assigning certainty in one of the outcome categories. As indicated in Figure 3, the analysis of the diagnostic of the fatality category was performed to identify its leading cause. The 'Fatality' crash severity was assigned a 100% probability, which imposes a belief about the fatality. For the vehicle maneuver category in Figure 3, when a fatal crash is observed on a highway, the highest likelihood is estimated on a vehicle moving straight (85.01%). This shows that when a crash has been observed and is fatal, it is most likely associated with a vehicle that was moving straight. Crashes involving junior drivers were estimated to be the second likely category. This finding suggests that given a fatal crash has occurred, the second most likely associated age group is junior drivers, with a 76.82% probability, followed by the probability that it was divided highway (57.50%), during nighttime (52.09%), and on a high posted speed roadway (44.35%).

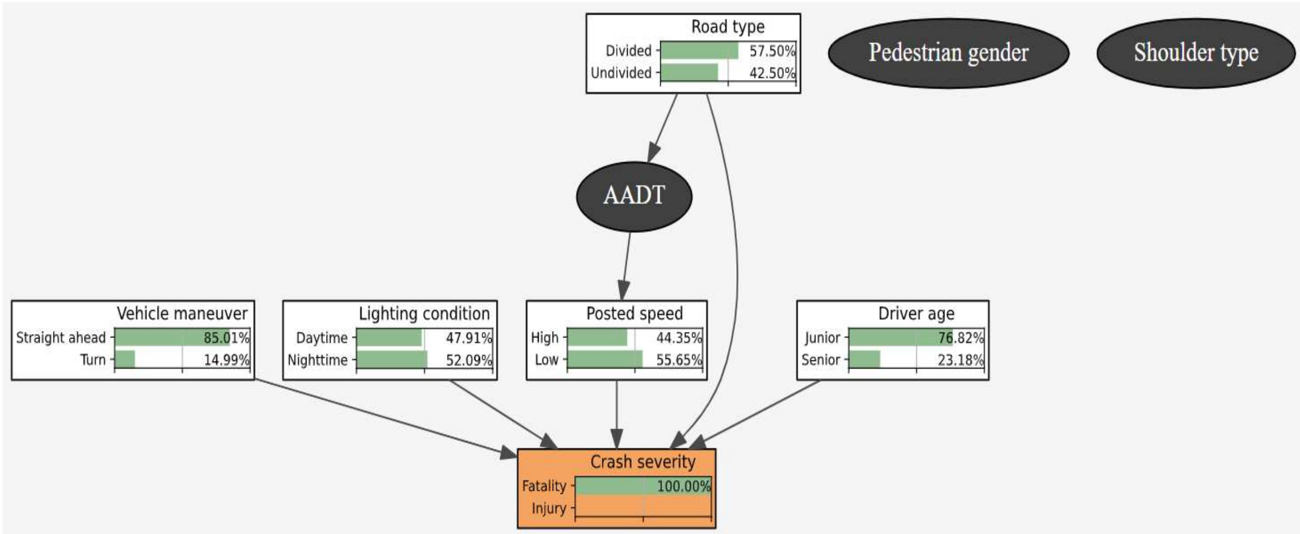


Figure 3. Diagnostic inference analysis results.

Combined Evidence Predictive Inference Results

In addition to assessing the influence of individual evidence on predictive and diagnostics inferences, the impact of concurrent evidence was also assessed. Table 5 shows the top five combined scenarios with the highest likelihood of aging pedestrian fatal crashes. All scenarios involved straight moving vehicles. Out of five scenarios, only one includes daytime lighting conditions, while the remaining four involve nighttime lighting conditions. Four scenarios include divided highways, and only one consists of an undivided roadway. Also, two scenarios consist of a junior driver and two includes a senior driver. Lastly, four scenarios consist of a high posted speed, while only one include a low posted speed.

Figure 4 shows the Bayesian network structure with the combination of evidence that resulted in the highest likelihood of fatal crashes. As indicated in Figure 4, fatal crashes are more likely to occur given that the vehicle maneuver was straight at the time of the crash, the crash occurred during nighttime, involved a junior driver, and occurred on a divided roadway with a posted speed of at least 45 mph. This type of analysis is important to transportation officials for prioritizing countermeasures. Multiple risk factors can be assessed to identify the most likely causes of pedestrian fatalities.

Table 5: Combined evidence predictive inference results

Evidence					Predicted probabilities
Vehicle maneuver	Lighting conditions	Road type	Driver age	Posted speed	Fatality
Straight ahead	Daytime	Divided	Junior	High speed	35.9%
	Nighttime	Undivided	Senior	High speed	41.7%
		Divided	Junior	Low speed	31.0%
				High speed	61.3%
			Senior	High speed	53.3%

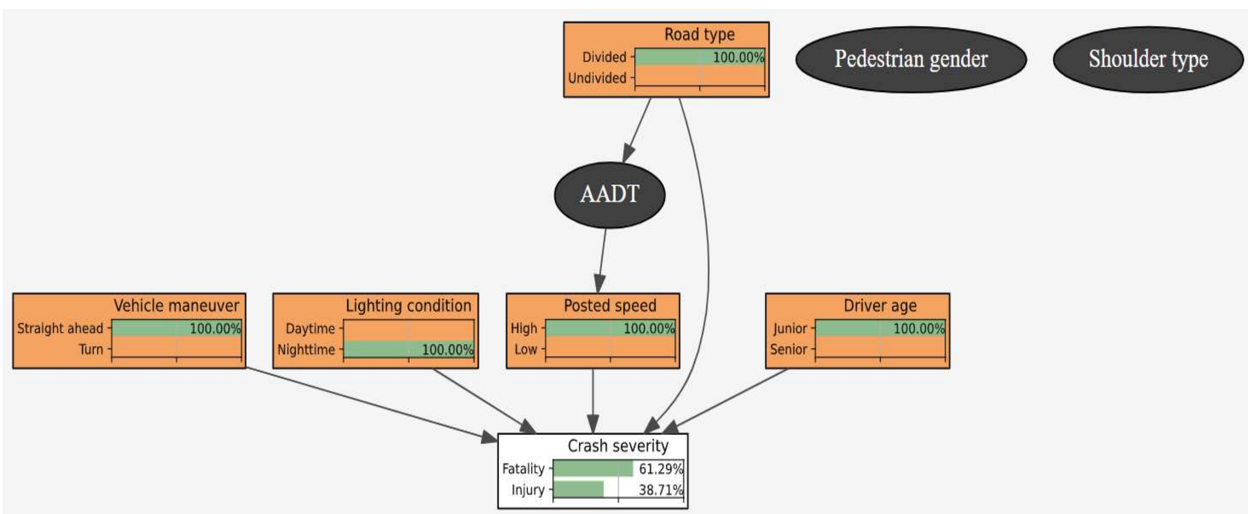


Figure 4. Combined evidence predictive inference analysis results.

Overall, the results of the BLR model, individual evidence analysis, diagnostic analysis, and combined evidence analysis concur with each other. A straight moving vehicle, nighttime lighting condition, junior driver group, on a divided roadway with high posted speed increase the likelihood of a pedestrian-vehicle crash resulting in an older pedestrian fatality.

CHAPTER 6 CONCLUSION AND RECOMMENDATION

Unlike vehicle occupants, pedestrians are more likely to sustain severe injuries in vehicle-pedestrian crashes because they are unshielded. This likelihood is greater for older pedestrians. Older pedestrians are associated with age-frailty, reduced physical strength, and greater risk for bone fracture, and thus, have a higher risk of injury/fatality if involved in a crash. Also, sensorial, cognitive, and self-perception abilities change due to age and age-related problems and are major issues that contribute to older pedestrian-vehicle crash occurrences.

This study presented an improved analysis approach for identifying risk variables that influence older pedestrian fatalities. The proposed approach used the Bayesian Logistic Regression (BLR) model and the Bayesian Network (BN) model. The BLR model was selected to estimate the significant variables influencing fatal crashes, while the BN model was used to establish a probabilistic dependence structure of older pedestrian fatality causes.

Initially, the study identified the significant factors influencing the risk of fatalities of older pedestrians in pedestrian-vehicle crashes at Florida intersections. Results reveal that turning vehicles decrease the risk of fatal crashes of older pedestrians, compared to straight moving vehicles. Crashes occurring during the night have higher risk of fatality to aging pedestrians, compared to those occurring during daylight conditions. Crashes occurring on the divided roadways also pose a higher risk of fatality for older pedestrians than the other roadway configurations.

Female older pedestrians are less likely to sustain fatal crashes, compared to male older pedestrians. Non-curbed shoulder roadways decrease the risk of older pedestrian fatal crashes, than roadways with curbed shoulders. Crashes occurring on roadways with higher traffic volumes (AADT > 20,000) are more likely to cause older pedestrian fatality, compared to crashes on lower traffic volume roadways. With respect to age, older pedestrian fatalities are less likely when a senior driver (60 years and older) is involved, compared to another driver age group. And, roads with high posted speed (45 mph or above) are associated with a higher likelihood of older pedestrian fatalities, compared to low posted speed.

The leading causes of older pedestrian fatalities were investigated from the significant factors influencing the risk of older pedestrian fatalities at intersections. This was performed using the optimal network structure and conducting the probabilistic inference through individual evidence predictive inference, predictive evidence predictive inference, and diagnostic inference. The optimal network structure revealed that vehicle maneuver, lighting condition, road type, posted speed, and driver age are directly associated with the crash severity of older pedestrians. Also, it was revealed that there is a direct relationship between the road type and traffic volume (AADT), and traffic volume and posted speed. From the individual evidence predictive inference results, it was found that crashes are more likely to be fatal for older pedestrians when the vehicle is moving straight, compared to when the vehicle is turning, at night, compared to daylight, on a divided roadway, compared to an undivided roadway, on high posted speed, compared to low posted speed, and involving a junior driver (age 59 or younger), compared to senior drivers (60 years or older).

For the combined evidence predictive inference, the results indicated that fatal older pedestrian crashes are more likely to occur given that the vehicle maneuver was straight at the time of the crash, the crash occurred during nighttime, involved a junior driver, and occurred on a divided roadway with high posted speed. Lastly, the diagnostic inference suggested that when a fatal crash is observed on a highway, the highest likelihood is estimated on a vehicle moving straight. Crashes involving junior drivers were estimated to be the second likely category, followed by the probability that it was on a divided roadway, at nighttime, and on a high posted speed roadway.

Based on analysis, the findings of this study will help to inform the next step before developing the countermeasures to improve the safety of older pedestrians.

In future work, it is recommended to investigate the leading causes of older pedestrian fatalities at intersections using an algorithm that can evaluate both continuous and categorical data simultaneously, since the BN experiences difficulty in handling continuous and discrete data simultaneously. For this study, the continuous data, especially AADT, pedestrian age, and driver age had to be discretized. To accurately estimate the risk of older pedestrian fatalities at intersections, it is recommended that other pedestrian age groups also be considered for analysis. Also, it is recommended that additional variables, such as population density (i.e., rural and urban), type and size of vehicles, and physical restrictions of older pedestrians. Lastly, exploring models that work well with fat-tail distributions is also recommended because of the nature of pedestrian crash severity.

REFERENCES

Chen, Y., Ma, J., & Chen, N. (2019). Analyzing Pedestrian Fatality Risk in Accidents at Mid-Blocks. *Journal of Transportation Technologies*, *09*(02), 171–192. https://doi.org/10.4236/jtts.2019.92011

Chen, Z., & Fan, W. (David). (2019). A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. In *International Journal of Transportation Science and Technology* (Vol. 8, Issue 1, pp. 43–52). https://doi.org/10.1016/j.ijtst.2018.10.001

Colby, S. L., & Ortman, J. M. (2014). The Baby Boom Cohort in the United States : 2012 to 2060. *US Census Bureau*, 16. https://doi.org/P25-1141

Cong, H., Chen, C., Lin, P.-S., Zhang, G., Milton, J., & Zhi, Y. (2018). Traffic Incident Duration Estimation Based on a Dual-Learning Bayesian Network Model. *Transportation Research Record*, *2672*(45), 196–209. https://doi.org/10.1177/0361198118796938

Das, S., Bibeka, A., Sun, X., & Zhou, H. T. (2019). *Elderly Pedestrian Fatal Crash-Related Contributing Factors : Applying Empirical Bayes Geometric Mean Method*. *2673*(8), 254–263. https://doi.org/10.1177/0361198119841570

de Oña, J., Mujalli, R. O., & Calvo, F. J. (2011). Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. *Accident Analysis & Prevention*, *43*(1), 402–411. https://doi.org/https://doi.org/10.1016/j.aap.2010.09.010

Haleem, K., Alluri, P., & Gan, A. (2015). Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accident Analysis & Prevention*, *81*, 14–23. https://doi.org/10.1016/J.AAP.2015.04.025

Hoffman, M. D., & Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(2008), 1593–1623.

Hu, W., & Cicchino, J. B. (2018). An examination of the increases in pedestrian motor-vehicle crash fatalities during 2009–2016. *Journal of Safety Research*, *67*, 37–44. https://doi.org/10.1016/j.jsr.2018.09.009

Kidando, E., Moses, R., Ozguven, E. E., & Sando, T. (2017). Evaluating Traffic Congestion Using the Traffic Occupancy and Speed Distribution Relationship: An Application of Bayesian Dirichlet Process Mixtures of Generalized Linear Model. *Journal of Transportation Technologies*, *07*(03), 318–335. https://doi.org/10.4236/jtts.2017.73021

Kidando, E., Moses, R., Sando, T., & Ozguven, E. E. (2019). Assessment of factors associated with travel time reliability and prediction: an empirical analysis using probabilistic reasoning approach. *Transportation Planning and Technology*, *42*(4), 309–323. https://doi.org/10.1080/03081060.2019.1600239

Kim, J. K., Ulfarsson, G. F., Shankar, V. N., & Kim, S. (2008). Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accident Analysis and Prevention*, *40*(5), 1695–1702. https://doi.org/10.1016/j.aap.2008.06.005

Kim, Y., & Park, J. (2019). Incorporating prior knowledge with simulation data to estimate PSF multipliers using Bayesian logistic regression. *Reliability Engineering and System Safety*, *189*(April), 210–217. https://doi.org/10.1016/j.ress.2019.04.022

Kitali, A. E., Kidando, E., Sando, T., Moses, R., & Ozguven, E. E. (2017). Evaluating Aging Pedestrian Crash Severity with Bayesian Complementary Log–Log Model for Improved Prediction Accuracy. *Transportation Research Record: Journal of the Transportation Research Board*, *2659*(1), 155–163. https://doi.org/10.3141/2659-17

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/a0029146

Kutela, B., & Teng, H. (2019). Prediction of drivers and pedestrians' behaviors at signalized mid-block Danish offset crosswalks using Bayesian networks. *Journal of Safety Research*, *69*, 75–83. https://doi.org/10.1016/J.JSR.2019.02.008

Lee, S.-M., Abbott, P., & Johantgen, M. (2005). Logistic Regression and Bayesian Networks to Study Outcomes Using Large Data Sets. *Nursing Research*, *54*(2), 133???138. https://doi.org/10.1097/00006199-200503000-00009

Li, Y., & Fan, W. (David). (2019). Modelling severity of pedestrian-injury in pedestrian-vehicle crashes with latent class clustering and partial proportional odds model: A case study of North Carolina. *Accident Analysis and Prevention*, *131*(April), 284–296. https://doi.org/10.1016/j.aap.2019.07.008

Mather, M., Jacobsen, L. A., & Pollard, K. M. (2019). *Aging in the United States*.

Montella, A., Aria, M., D'Ambrosio, A., & Mauriello, F. (2011). Data-mining techniques for exploratory analysis of pedestrian crashes. *Transportation Research Record*, *2237*, 107–116. https://doi.org/10.3141/2237-12

Moudon, A. V., Lin, L., Jiao, J., Hurvitz, P., & Reeves, P. (2011a). 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 and Prevention*, *43*(1), 11–24. https://doi.org/10.1016/j.aap.2009.12.008

Moudon, A. V., Lin, L., Jiao, J., Hurvitz, P., & Reeves, P. (2011b). 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 and Prevention*, *43*(1), 11–24. https://doi.org/10.1016/j.aap.2009.12.008

New York City Department of Transportation (NYCDOT). (2010). *Safe Streets for Seniors: Chinatown, Manhattan*.

NHTSA (National Highway Traffic Safety Administration). (2020). *2018 Traffic Safety Facts: Older population (DOT HS 812 928)*.

Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. John Wiley & Sons, Inc.

Obeng, K., & Rokonuzzaman, M. (2013). Pedestrian Injury Severity in Automobile Crashes. *Open Journal of Safety Science and Technology*, *03*(02), 9–17. https://doi.org/10.4236/ojsst.2013.32002

Oh, C., Kang, Y., Kim, B., & Kim, W. (2005). Analysis of Pedestrian-Vehicle Crashes in Korea: Focused on Developing Probabilistic Pedestrian Fatality Model. *Proceedings: International Technical Conference on the Enhanced Safety of Vehicles*, *2005*, 8p-8p.

Onieva-Gracia, M. A., Martinez-Ruiz, V., Lardelli-Claret, P., Jimenez-Moleon, J. J., Amezcua-Prieto, C., Luna-del-Castillo, J. de D., & Jemenes-Mejias, E. (2016). Gender and Age Differences in Components of Traffic-related Pedestrian Death Rates: Exposure, Risk of Crash and Fatality Rate. *Inury Epidemiology*, *3*(14). https://doi.org/doi 10.1186/s40621-016-0079-2

Oxley, J. A., Ihsen, E., Fildes, B. N., Charlton, J. L., & Day, R. H. (2005). Crossing roads safely: An experimental study of age differences in gap selection by pedestrians. *Accident Analysis & Prevention*, *37*(5), 962–971. https://doi.org/https://doi.org/10.1016/j.aap.2005.04.017

Pedestrian and Bicycle Information Center (PBIC). (2007). *Safe Routes to School Guide*.

Pour-Rouholamin, M., & Zhou, H. (2016). Investigating the risk factors associated with pedestrian injury severity in Illinois. *Journal of Safety Research*, *57*, 9–17. https://doi.org/10.1016/j.jsr.2016.03.004

Salon, D., & McIntyre, A. (2018). Determinants of pedestrian and bicyclist crash severity by party at fault in San Francisco, CA. *Accident Analysis & Prevention*, *110*, 149–160. https://doi.org/https://doi.org/10.1016/j.aap.2017.11.007

Sarkar, S., Tay, R., & Hunt, J. D. (2012). Logistic Regression Model of Risk of Fatality in Vehicle–Pedestrian Crashes on National Highways in Bangladesh. *Transportation Research Record: Journal of the Transportation Research Board*, *2264*(1), 128–137. https://doi.org/10.3141/2264-15

Sasidharan, L., & Menéndez, M. (2014). Partial proportional odds model—An alternate choice for analyzing pedestrian crash injury severities. *Accident Analysis & Prevention*, *72*, 330–340. https://doi.org/https://doi.org/10.1016/j.aap.2014.07.025

Smart-Growth-America and National-Complete-Street-Coalition. (2020). *2020 Dangerous by design*.

Stylianou, K., & Dimitriou, L. (2018). Analysis of Rear-End Conflicts in Urban Networks using Bayesian Networks. *Transportation Research Record*, *2672*(38), 302–312. https://doi.org/10.1177/0361198118790843

Witteveen, A., Nane, G. F., Vliegen, I. M. H., Siesling, S., & IJzerman, M. J. (2018). Comparison of Logistic Regression and Bayesian Networks for Risk Prediction of Breast Cancer Recurrence. *Medical Decision Making*, *38*(7), 822–833. https://doi.org/10.1177/0272989X18790963

World Health Organization (WHO). (2018). *Global status report on road safety 2018 (CC BYNC-SA 3.0 IGO)*.

Wuillemin, P.-H. (2019). *pyAgrum Documentation: Release 0.15.2*.

Xie, C., & Waller, S. (2010). Estimation and application of a Bayesian network model for discrete travel choice analysis. *Transportation Letters*, *2*(2), 125–144. https://doi.org/10.3328/TL.2010.02.02.125-144

Zahabi, S. A. H., Strauss, J., Manaugh, K., & Miranda-Moreno, L. F. (2011). Estimating Potential Effect of Speed Limits, Built Environment, and Other Factors on Severity of Pedestrian and Cyclist Injuries in Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2247*(1), 81–90. https://doi.org/10.3141/2247-10

Zhai, X., Huang, H., Sze, N. N., Song, Z., & Hon, K. K. (2019). Diagnostic analysis of the effects of weather condition on pedestrian crash severity. *Accident Analysis and Prevention*, *122*(October 2018), 318–324. https://doi.org/10.1016/j.aap.2018.10.017

Zheng, Y., Author, C., Manjunatha, P., Elefteriadou, L., & Ponnaluri, R. (2017). Empirical assessment of sdaptive Signal Control Technologies in Florida. In *Transportation Research Board, 96th Annual Meeting* (pp. 1–15).

Zhu, M., Zhao, S., Coben, J. H., & Smith, G. (2013). Why More Male Pedestrians Die in Vehicle-pedestrian Collisions than Female Pedestrians: A Decompositional Analysis. *Inury Prevention*, *19*(4), 227–231. https://doi.org/doi: 10.1136/injuryprev-2012-040594