An In-Depth Examination of Pedestrian Crashes Via a Data-Driven Framework: A Case Study of Signalized Intersections in Austin, Texas

November 2019

Center for Transportation Safety
Texas A&M Transportation Institute

An In-Depth Examination of Pedestrian Crashes Via a Data-Driven Framework: A Case Study of Signalized Intersections in Austin, Texas

by

Sirajum Munira
Graduate Research Assistant

Ipek N. Sener, Ph.D.
Associate Research Scientist

Boya Dai
Associate Transportation Researcher

November 2019

Center for Transportation Safety
Texas A&M Transportation Institute
College Station, Texas

Disclaimer

Acknowledgments

This research project was supported by TTI CTS. The authors would like to acknowledge the invaluable assistance of the City of Austin in furnishing data required for this research. The research team also thanks TTI researchers Ben Ettelman and Joan Hudson for their valuable discussions during the project, as well as TTI editors Dawn Herring and Michelle Benoit for their editorial review.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

Walking and bicycling make up a relatively small portion of transportation in the United States and yet account for a disproportionate share of the total fatal and serious injury crashes. In 2018, 6,227 pedestrians were killed in traffic-related crashes, which was around 15% of all traffic fatalities in the United States (Governors Highway Safety Association [GHSA], 2019). Despite the efforts of many U.S. cities to promote pedestrian safety, national crash statistics for pedestrians show an upward trend: 2018 and 2017 had a 4% and 1.7% increase in pedestrian fatalities, respectively, compared to previous years. The number of pedestrian fatalities in 2018 was the highest since 1990 (GHSA, 2019). Austin, Texas, is no exception to this overwhelming nonmotorized crash trend. In 2018, pedestrians made up 42% of all traffic fatalities in Austin, the highest number of pedestrian deaths in almost 10 years (Bradshaw, 2019).

Many big cities in the United States, including Austin, are endeavoring to adopt a holistic approach to increase safety and mobility for pedestrians of all ages. To develop and implement effective strategies to reduce pedestrian crashes, preferably to zero, a better understanding of the causes and consequences of pedestrian crashes is essential. Although pedestrian-related crashes occur on various road facilities, such as intersections, driveways, and midblock locations, safety planners often focus on intersection-related crashes because a significantly large proportion of crashes are observed in or near intersections (Choi, 2010). The Texas Strategic Highway Safety Plan reported that more than one-third of fatal and incapacitating injury crashes in Texas in 2013 were identified as intersection related (Texas Department of Transportation [TxDOT], 2016a). A report analyzing crashes in the Capital Area Metropolitan Planning Organization (CAMPO) region estimated that the total cost of intersection crashes was around $3.3 billion from 2010 to 2014 (TxDOT, 2016a). The same report also revealed that more than one of every seven severe crashes (fatal and suspected serious injury) at intersections in the CAMPO region involved a pedestrian or bicyclist.

In addition to discerning the location of pedestrian crashes, effective countermeasures warrant a profound understanding of the role of multiple exogenous factors (such as exposure or traffic condition) affecting crash occurrence. While considering policies to reduce the frequency of crashes, especially involving vulnerable road users such as pedestrians and bicyclists, planners must also contemplate countermeasures to minimize the severity of those crashes. Bicyclists and pedestrians are 2.3 and 1.5 times more likely, respectively, to be fatally injured in a trip than passenger-vehicle occupants are, according to a study by Beck et al. (2007).

Safety advocates in multiple areas are persistent in their efforts to develop evidence-based data-driven strategies to reduce pedestrian fatalities. The literature is replete with studies of various aspects of pedestrian crash risk, type, and severity at different geographic scales including intersections, census tracts, and block groups. However, most of these studies suffer from two limitations.

The first limitation is the absence of pedestrian demand or exposure data, which, despite being some of the most important inputs when analyzing pedestrian safety, are often not available. Although models (such as the direct demand model) are available to estimate nonmotorized demand at a specific spatial scale that can be used as an exposure measure for safety analysis, only a handful of studies have used these models (Hasani et al., 2019; Lee et al., 2019).

The second limitation is the inadequacy of separate models to capture the potential correlations across various crash aspects. For example, when analyzing crash frequency and severity, the traditional univariate modeling approach leaves room for error by ignoring the presence of common unobserved factors that simultaneously influence the occurrence of crashes by severity type at a spatial scale. Studies have argued that crash frequency across different attributes (mode involved, severity, crash type, and damage) tend to be correlated and are thus multivariate in nature (Yasmin & Eluru, 2018). Therefore, the univariate models, which analyze crash attributes separately, increase the risk of potential biases, leading to inaccurate estimation (Ma et al., 2008). For this reason, for analyzing multiple crash attributes (such as severity), developing and deploying a multivariate model is recommended to obtain a reliable, robust estimate of the impacts of various factors on crash frequencies for different severities (Liu & Sharma, 2018; Park & Lord, 2007).

Acknowledging the superiority of multivariate models, a number of studies have developed crash models by type, mode, and severity at various geographic scales (Lee et al., 2015; Ma et al., 2008; Park & Lord, 2007; Wang et al., 2014; Xie et al., 2019; Ye, Pendyala, et al., 2009; Zeng et al., 2017). Although a number of researchers have attempted to develop multivariate models to investigate intersection-related crashes for motorized vehicles (Alarifi et al., 2018; Cheng et al., 2018; Huang et al., 2017; Park & Lord, 2007), multivariate analyses for pedestrian-involved crashes at intersections are rare (Heydari et al., 2017).

In light of this discussion, this study focused on the development of a data-driven framework for analyzing multiple pedestrian crash severities at signalized intersections, in a joint context, incorporating pedestrian exposure. Austin was selected as the study area given its strong commitment to its Vision Zero goals yet lack of data and tools to facilitate strategic data-informed decisions. The study area covered the entire city area with a total of 409 intersections identified for the analysis. To the authors' knowledge, no studies have estimated pedestrian demand or exposure at Austin intersections, despite recent studies having shown that disregarding pedestrian exposure could significantly affect the crash analysis model (Fitzpatrick et al., 2018).

This study was performed in two parts:

1. Using available pedestrian count data from the City of Austin, a direct demand model was developed for estimating pedestrian volume/exposure at the intersection (signalized) level.

2. The exposure information was integrated into the development of a multivariate model for analyzing pedestrian crash frequency at signalized intersections for three severity levels: fatal crash, suspected serious injury or incapacitating injury crash, and non-incapacitating injury crash.

Chapter 2. Literature Review

This chapter provides an overview of the different approaches of crash analysis using a multivariate perspective. In addition, the chapter discusses the importance of and approaches to considering the spatial dependency of the crash data, multivariate crash models focused on nonmotorized modes, and use of exposure variables in nonmotorized crash analysis.

2.1 Multivariate Crash Analysis

To inform and design safety-related policies guided by models with superior predictive power and accuracy, researchers have given significant attention to advanced statistical modeling techniques such as multivariate models, random parameter models, finite mixture/Markov switching models, hierarchical models, neural and Bayesian neural network models, and so forth (Lord & Mannering, 2010). Among these advanced models, research on joint or multivariate modeling of correlated outcomes has been particularly popular in the last few years. The key strength of the multivariate modeling approach is its ability to handle correlations across different levels of crash attributes (such as crash occurrence and severity), which are likely to be affected by common unobserved factors simultaneously (Mannering et al., 2016). Several studies (Huang & Abdel-Aty, 2010; Xie et al., 2013) have suggested that these models provide more reliable and accurate estimation than traditional univariate models.

The multivariate modeling approaches for crash analysis, attempted by several studies, generally vary in terms of crash attributes investigated, modeling structures, and aggregation level. In terms of crash attributes, studies have used multivariate models to examine crash frequency by severity level (El-Basyouny & Sayed, 2009; Ma et al., 2008; Park & Lord, 2007; Wang et al., 2014; Xie et al., 2019; Zeng et al., 2017; Zhan et al., 2015), crash frequency by collision type (Bhowmik et al., 2018; El-Basyouny et al., 2014; Song et al., 2006; Ye et al., 2009), crash frequency by transportation mode (Lee et al., 2015; Huang et al., 2017), injury severity and driver error (Wali et al., 2018), and injury severity and vehicle damage (Wang et al., 2015).

In terms of modeling structure, researchers have examined a number of approaches based on their data collection and analysis requirements. Studies have used Poisson gamma models (Abdel-Aty & Radwan, 2000; Poch & Mannering, 1996; Xie et al., 2019), Poisson lognormal models (Alarifi et al., 2017; El-Basyouny & Sayed, 2009; Huang et al., 2017; Lee et al., 2015; Park & Lord, 2007; Wang & Kockelman, 2013), copula-based approaches (Nashad et al., 2016; Rana et al., 2010; Wang et al., 2019a; Yasmin et al., 2014; Yasmin et al., 2018), and multivariate random-parameter zero-inflated negative binomial models (Anastasopoulos, 2016). A few studies have also used the fractional split approach for modeling crash frequency by different attributes (Bhowmik et al., 2018; Yasmin et al., 2016). Studies have employed both Bayesian (Cheng et al., 2018; Ma et al., 2017) and frequentist (Narayanamoorthy et al., 2013) estimation techniques to make statistical inferences under the multivariate setting.

Studies analyzing crashes in a multivariate context also vary in aggregation level. Similar to the univariate approach, multivariate models can be categorized into two types: macro-level (such as a regional or traffic analysis zone level) and micro-level (such as intersections or the road segment level). Macro-level models can examine the influence of sociodemographic, land use, or road network characteristics on crash attributes, which can evaluate safety conditions from a planning perspective (Wang & Huang, 2016). Micro-level models focus on intersection- or road-segment-related characteristics and are often used to identify black spots. Examples of aggregation for macroscopic models (in a multivariate context) include the traffic analysis zone level (Bhowmik et al., 2018), county level (Song et al., 2006), and census tract level (Wang & Kockelman, 2013; Xie et al., 2019). Acknowledging the need for models to design safety-related countermeasures at the microscopic level, numerous studies have applied the multivariate modeling approach to examine crashes at the level of the intersection (Huang et al., 2017; Park & Lord, 2007; Strauss et al., 2014; Wang et al., 2019a; Ye et al., 2009), roadway segment (El-Basyouny & Sayed, 2009; Wang et al., 2014), highway corridor (Ma et al., 2017), and intersection and road segment simultaneously (Zeng & Huang, 2014).

Another important aspect of modeling crash frequency is considering the spatial dependence of the observations, which is often ignored (Mannering et al., 2016). Research has shown that crash models need to account for spatial dependency because spatial correlation exists extensively among adjacent locations in road networks and neighborhood zones (Quddus, 2008; Wang et al., 2019b). For example, the frequency of crashes at a location may sharply change with the distance from the central business district, or crashes on road segments in close proximity can be clustered together because they have similar traffic flow characteristics. Consideration of spatial correlation while examining crash models for intersections in the urban road network is particularly crucial because the intersections at close proximity are more likely to share similar land use and traffic characteristics (Abdel-Aty & Wang, 2006; Xie et al., 2013; Xie et al., 2014). The spatial model can handle the spatial interaction and spatial structure in crash data, which leads to improved model parameter estimation and can reflect unmeasured confounding variables (Huang et al., 2017; Wang et al., 2019b).

Wang et al. (2017) discussed the four approaches that are generally used for developing spatial models for multivariate count data:

- The conditional autoregressive model (CAR).
- The multivariate finite mixture model.
- The generalized ordered response model.
- The spatio-temporal model.

Studies have indicated that the most popular approach is the CAR model, probably because it takes advantage of the flexibility of the Bayesian hierarchical framework to account for the spatial correlation (Ma et al., 2017; Zeng & Huang, 2014). Examples of studies involving spatial

components in the multivariate crash analysis context are by Xie et al. (2019); Huang et al. (2017); Wang and Huang (2016); Zeng and Huang (2014); Aguero-Valverde (2013); and Barua, et al. (2014).

2.2 Multivariate Model Nonmotorized Crash Analysis

Crash analysis for nonmotorized modes has received significantly less attention than crash analysis for motorized vehicles, under the multivariate setting. A number of studies aimed to formulate joint models for nonmotorized crashes at the zonal level and recognized the dependency of various crash attributes. Wang and Kockelman (2013) aggregated crash data at the census tract and developed a multivariate Poisson lognormal CAR model for pedestrian crashes across different severity levels using 3 years of crash data (2007 to 2009) for the Austin area. Nashad et al. (2016) developed a copula-based bivariate negative binomial model for pedestrian and bicycle crash frequency analysis at the macro level (the statewide traffic analysis zone). The variables used in the model included exposure measures (vehicle miles traveled), socioeconomic characteristics, road network characteristics, and land use attributes. The study concluded that macro-level nonmotorized crash analysis needs to accommodate the dependence between pedestrian and bicycle crash count events. Cai et al. (2017) developed a joint model for crash frequency and the proportion of nonmotorists at the traffic analysis district level in Florida. Narayanamoorthy et al. (2013) developed a spatial multivariate count model to examine the number of pedestrian and bicyclist injuries by injury severity at the census tract level for New York City. The study used various risk factors such as sociodemographic characteristics (population density and distribution of population based on age, income, and race), land use variables (the proportion of commercial and industrial land use), activity intensity characteristics (the number of schools and universities), road network characteristics (the proportion of highways and bicycle route length), commute mode shares, and transit supply characteristics (the number of bus stops).

Although the research was not specific to pedestrian crash severity, Huang et al. (2017) simultaneously analyzed the occurrence of motor vehicle, bicycle, and pedestrian crashes at urban intersections in Florida using multiple explanatory variables such as annual daily traffic for major/minor roads, population density, leg number, speed limit, and presence of a traffic signal, pedestrian signal, crosswalk, bus stop, and median. Heydari et al. (2017) investigated the crash correlates of walking and cycling at signalized intersections in Montreal, Canada, using a flexible multivariate latent class approach; the explanatory variables were motorized volume by turning direction, nonmotorized volume, maximum speed limit, leg number, and presence of a pedestrian signal, subway station, bus stop, and median, along with land use characteristics such as employment, commercial area, land use mix, and number of schools.

2.3 Exposure Measure for Models

Acknowledging that the exposure measure is an essential element for modeling nonmotorized crash frequency, studies have used a number of methods to quantify the exposure to crash risk. Wang and Kockelman (2013) used walk miles traveled as an exposure measure, which was estimated using household travel survey data and least squares regression. Vehicle miles traveled was another exposure measure used by studies for nonmotorized crash analysis through a joint model (Nashad et al., 2016). Cai et al. (2017) found that the product of the log of population and the log of vehicle miles traveled was the best exposure variable to examine pedestrian crashes for a zip code–level analysis. Given it is often difficult to quantify the number of pedestrian/bicyclist miles of travel and motorized vehicle miles of travel at a zonal level, studies, in both a multivariate and non-multivariate context, have used surrogate measures such as population density (LaScala et al., 2000; Narayanamoorthy et al., 2013), income (Loukaitou-Sideris et al., 2007), activity intensity characteristics (Mitra & Washington, 2012), and so forth. In other studies, bicycle and pedestrian count data obtained from both signalized and unsignalized intersections were incorporated as exposure measures (Heydari et al., 2017; Strauss et al., 2014).

2.4 Current Study in Context

The current study builds on earlier research and proposes a joint modeling approach for analyzing intersection pedestrian crashes for three severity levels. To achieve the objective, a multivariate Poisson lognormal model was used to accommodate the overdispersion issue in crash data and account for potential correlation among the crash severities. The potential presence of spatial correlations among the intersections was also accounted for to develop a robust model for pedestrian crash severity at the intersections because previous studies have suggested that ignoring spatial correlation may lead to biased model parameters and inferior model performance (Aguero-Valverde & Jovanis, 2010; LeSage & Pace, 2009). The Bayesian framework using the Markov chain Monte Carlo simulation method was used for model estimation.

Chapter 3. Model Description and Formulation

3.1 Mixed Poisson Model for Crash Analysis

Researchers have extensively used mixed Poisson models to accommodate the overdispersion in crash counts. These mixed Poisson models, hierarchical in nature, accommodate the observed crashes (conditional on the mean) that are mutually independent and Poisson distributed at the first level. The mixed Poisson models allow the unobservable mean of crashes to vary across locations with an assumed probability distribution at the second level.

Most highway safety researchers have used two types of mixed Poisson models: Poisson gamma and Poisson lognormal. Studies have suggested that the Poisson lognormal model is more flexible than the Poisson gamma in accommodating the multivariate structure and spatial correlation (Aguero-Valverde & Jovanis, 2009; Ma et al., 2008). In the Poisson lognormal model, the Poisson parameter is assumed to follow a lognormal distribution. The marginal distribution of this model does not have a closed form, and the maximum likelihood estimates approach cannot be directly used to estimate model parameters, unlike the Poisson gamma model. For this reason, Markov chain Monte Carlo simulation methods from the Bayesian perspective have been used for model estimation.

In this study, a multivariate spatial Poisson lognormal model was developed to observe the pedestrian crash frequency across different severities at the intersections in the Austin area. The model estimation was conducted through a full Bayesian approach, which considers the uncertainty related to model parameters and provides exact measures of uncertainty (Miaou & Lord, 2003). In order to compare model performance, a univariate spatial Poisson lognormal model was also developed for the same region. The specifications of the multivariate spatial Poisson lognormal model and the specifications and priors of the univariate Poisson lognormal model are provided in the next sections.

3.2 Multivariate Spatial Poisson Lognormal Model

Let Y_{ik} denote the number of pedestrian crashes observed at the ith intersection (=1, 2, . . ., 409) of kth severity (=1, 2, 3) during the study period. In the Poisson hierarchical models, Y_{ik}, when conditional on the mean crash rate λ_{ik}, is assumed to be Poisson distributed, which can be expressed as:

$$Y_{ik} \sim Poisson(\lambda_{ik}) \tag{1}$$

The mean crash rate (λ_{ik}) can be specified at the second level of hierarchy:

$$log(\lambda_{ik}) = \alpha_k + X_i\,\beta_k + S_{ik} + U_{ik} \tag{2}$$

where:

- α_k is the intercept term of severity *k*.
- X_i indicates a column vector of covariates (pedestrian volume, intersection features, traffic condition, etc.).
- $\beta_k = (\beta_{k1}, \beta_{k2}, \ldots, \beta_{km})$ denotes an *m* dimensional regression coefficient vector specific to each observation type *k*. For example, *m* = 7 (because seven explanatory variables have been used in the final model) for this study.
- U_{ik} represents the error term that captures site-specific heterogeneity not explained by spatial effects. It is assumed to be multivariate normally distributed with a mean vector of 0 and a variance-covariance matrix of Σ . This is equivalent to $e^{U_{ik}} \sim Lognormal\ (0, \Sigma)$, where Σ is the variance-covariance matrix for heterogeneous effects.

$$U_{ik} = \begin{pmatrix} U_{i1} \\ U_{i2} \\ U_{i3} \end{pmatrix} \text{ and } \Sigma = \begin{cases} \sigma^2{}_{11} & \sigma^2{}_{12} & \sigma^2{}_{13} \\ \sigma^2{}_{21} & \sigma^2{}_{22} & \sigma^2{}_{23} \\ \sigma^2{}_{31} & \sigma^2{}_{32} & \sigma^2{}_{33} \end{cases} \tag{3}$$

Here, the diagonal elements, $\sigma^2{}_{kk}$, represent the heterogeneous variance of U_{i1}, U_{i2}, and U_{i3}. The off-diagonal elements denote the heterogeneous covariance among U_{i1}, U_{i2}, and U_{i3}. For the precision matrix Σ^{-1}, the most commonly used noninformative Wishart distribution is specified as the prior, written as Wishart (*I*, *r*). Here, *I* denotes the identity matrix, and *r* (≥ *K*) denotes the degrees of freedom, set at 3 to make the prior minimally informative (Gelman, 2006).

S_{ik} is a spatially structured random effects term that accounts for spatial autocorrelation, which cannot be incorporated by the Poisson lognormal model alone (Huang et al., 2017). Numerous previous studies have indicated that geographic area, road segments, or intersections that are closer to one another tend to have common features affecting their collision severity (Liu & Sharma, 2017; Ma et al., 2017; Wang & Kockelman, 2013). To explore the spatial correlation between adjacent intersections, S_{ik} is assigned an intrinsic conditional autoregressive (ICAR) (Besag et al., 1991) prior for each severity level *k*. The multivariate ICAR model is the intrinsic version of the multivariate conditional autoregressive model (Lawson, 2013) and has been used by several studies for multivariate spatial analysis (Liu & Zhu, 2017; Ma et al., 2017). For the spatially structured random effects $S_i = (S_{i1,} S_{i2,} S_{i3})^T$, the multivariate ICAR specification can be expressed as:

$$S_i | (S_{-i1}, S_{-i2}, S_{-i3}) \sim MN(\overline{S_i}, \Sigma_s / n_i) \tag{4}$$

where:

$$\overline{S_i} = \left(\sum_{j \neq i} \frac{\omega_{i,j} S_{j1}}{n_i}, \sum_{j \neq i} \frac{\omega_{i,j} S_{j2}}{n_i}, \sum_{j \neq i} \frac{\omega_{i,j} S_{j3}}{n_i} \right)^T \tag{5}$$

where:

- $\omega_{i,j}$ denotes the weight that intersection *j* has on intersection *i*: $\omega_{i,j}$ = 1 if *i* and *j* are adjacent and 0 otherwise.
- n_i is the number of intersections adjacent to intersection *i*.
- Σ_s is the covariance matrix where the diagonal elements represent the conditional variance of $S_{i1,}S_{i2,}S_{i3}$. The off-diagonal elements represent the conditional within-intersection covariance. Σ_s is also assumed to follow a Wishart distribution.

3.3 Prior Specification for Univariate and Multivariate Models

Prior specification is a crucial component of Bayesian modeling approaches. Owing to the lack of sufficient prior knowledge of the distributions for individual parameters, uninformative (vague) prior distributions are usually specified (Ma et al., 2017). The intercept term α_k was assigned a uniform prior *dflat*( ). The regression coefficient β_{km} (for *m* number of predictors) was specified to follow a noninformative normal distribution with a mean of 0 and a variance of 10,000.

3.4 Prior Specification Specific to Univariate Model

The key differences between the multivariate and univariate models lie in the prior specifications of the random effects. For the univariate model, which cannot accommodate the dependence between severity types, the random effects for different severities of crashes are independent. Therefore, for the univariate model, U_{ik} was assumed to follow an independent normal distribution as follows:

$$U_{ik} \sim N(0, \sigma_{uk}^2) \tag{6}$$

S_{ik} was determined through an ICAR distribution, as expressed by:

$$S_{ik} | S_{-ik} \sim N(\overline{S_{ik}}, \sigma_{sk}^2/n_i) \text{ for } k=1,2,3 \tag{7}$$

$$\overline{S_{ik}} = \sum_{j \neq i} \frac{\omega_{i,j} S_{jk}}{n_i} \tag{8}$$

where:

- $\omega_{i,j}$ denotes the weight that intersection *j* has on intersection *i*. $\omega_{i,j}$ = 1 if *i* and *j* are adjacent and 0 otherwise.

According to the ICAR model, the conditional distribution of S_{ik}, given the remaining components (S_{-ik}), is normal with mean $\overline{S_{ik}}$ and variance σ_{sk}^2/n_i. Here, n_i is the number of

intersections adjacent to intersection *i*. The variation of S_{ik} is controlled by the overall variance parameter σ^2_{sk}. The hyper-parameters for $1/\sigma^2_{sk}$ and $1/\sigma^2_{uk}$ are *Gamma* (0.5, 0.0005).

3.5 Model Evaluation

The deviance information criteria (DIC) were used as the goodness-of-fit measures for model comparisons. The DIC are a generalization of Akaike information criteria proposed by Speigelhalter et al. (2002) and provide a Bayesian measure of model complexity and fitting.

The DIC can be defined as:

$$DIC = D(\bar{\theta}) + 2\rho D = \bar{D} + \rho D \tag{9}$$

where:

- $D(\bar{\theta})$ is the deviance using the posterior mean values of the parameters of interest ($\bar{\theta}$).
- $\bar{D}$ is the posterior mean of deviances.
- ρD is the effective number of parameters in the model.

Lower DIC values for the model are preferred. Generally, differences in DIC of more than 10 point to keeping the model with the lower DIC; differences between 5 and 10 are considered substantial; differences less than 5 suggest that the models are not statistically different (MRC Biostatistics Unit, 2004).

Chapter 4. Data Description

Austin, Texas, was selected as the study area for this study to facilitate the city's newly adopted holistic approach (Austin Transportation Department, 2018) to improve citywide pedestrian safety and promote walking for transport and physical activity. The city also adopted the Vision Zero initiative to reduce traffic-related deaths and injuries to zero by the year 2025.

This chapter explains the data compiled and processed for the development of the crash model described in Chapter 3. The data used include crash data as the main variable of interest as well as exposure data and other explanatory variables used to develop the pedestrian crash model.

4.1 Crash Data

The traffic crash data for this study were taken from TxDOT's Crash Records Information System (CRIS) (TxDOT, 2016b). The data obtained included the disaggregated crash data for all locations within the study area, collected over 8 years (2011 to 2018). Along with various crash features, CRIS reports the severity of crashes (not injured, possible injury, non-incapacitating injury, incapacitating injury, and killed) and units involved (such as motor vehicle or pedestrian) and identifies the coordinates of each crash location. CRIS also separates crash data based on the location where the traffic crash occurred:

- Intersection crash: a traffic crash that occurs within the limits of an intersection.
- Intersection-related crash: a traffic crash that occurs on an approach to or exit from an intersection and results from an activity, behavior, or control related to the movement of traffic units through the intersection.
- Driveway access crash: a traffic crash that occurs on a driveway access or involves a road vehicle entering or leaving another roadway by way of a driveway access.
- Non-intersection crash: a traffic crash that is not an intersection crash, intersection-related crash, or driveway access crash.

To meet the objectives of this study, pedestrian-involved intersection or intersection-related crashes were extracted from the dataset. After obtaining the crashes based on the specified criteria, the location of each crash was matched or spatially joined with the nearest intersections of the network. The intersection map was generated using the city's comprehensive transport network data. Only crashes that were within a 300-foot buffer (Fitzpatrick et al., 2018) of the nearest intersection were considered. A total of 655 crashes at 409 signalized intersections were identified. The crashes were categorized into three types based on severity: fatal crashes, incapacitating injury or suspected serious injury crashes, and non-incapacitating injury crashes. Table 1 presents the crash frequency of each severity level.

Table 1. Description and Frequency of Crash Severity Types.

Severity	Description	Frequency
Fatal	Fatal (killed) crash	30
Incapacitating	Suspected serious injury/incapacitating injury crash	119
Non-incapacitating	Non-incapacitating injury crash	506
	Total crashes	655

4.2 Pedestrian Exposure Data

Given the lack of pedestrian exposure or volume data across the region, a direct demand model was developed to estimate pedestrian volume in all crash locations based on the available count data. This particular type of model is one of the most frequently used modeling approaches in the area of pedestrian/bicyclist safety (Turner et al., 2017). The modeling framework uses count observations from limited locations and estimates demand at a specific location (midblock or intersection) by directly relating the counts to mode, trip, and traveler attributes using a form of regression analysis (Ortuzar & Willumsen, 2011). A comprehensive literature review of the direct demand model is beyond the scope of this study but is provided by Munira and Sener (2017).

The following subsections provide information on data gathering and processing for estimating the exposure or pedestrian volume, which was used as an input in the crash model.

4.2.1 Data Used for Estimating Pedestrian Volume

This study collected actual volume data from two sources:

- Short-duration count data from the City of Austin Transportation Department.
- Continuous count data from Eco-Counter.

The City of Austin Transportation Department collected 24-hour short count data for pedestrians from 44 intersections in the study area. Following the standard data method, the pedestrian volume data were collected on typical weekdays distributed over 5 months (April, May, June, August, and October) in 2017. The continuous count data were obtained from Eco-Counter, which has been collecting pedestrian and bicycle data in 11 locations in the Austin area since 2012. The count data from the permanent counter were taken to estimate the daily and monthly factors (Nordback et al., 2013), which were used to calculate the annual average daily pedestrian volume for the 44 locations. The final annual average daily pedestrian volume was used as the dependent variable of the pedestrian direct demand model.

To use as an input in the pedestrian direct demand model, a rich set of explanatory variables was created with data from multiple sources such as the Data and Technology Services of the City of

Austin Transportation Department, the City of Austin public data portal, the 2017 American Community Survey, the City of Austin Planning and Development Review Department, the Texas Education Agency, the Austin Transportation Department Arterial Management Division, and the Capital Metropolitan Transportation Authority (Capital Metro) data portal. All of these datasets were then analyzed, cleaned, and processed to bring them to homogenous spatial scales (buffer levels around the intersection). Over 300 variables for three buffer zones—0.25 miles, 0.5 miles, and 1 mile—were created for this study. The variables were categorized into seven groups: demographics, socioeconomics, network/interaction with vehicle traffic, pedestrian- or bicycle-specific infrastructure, transit facilities, major generators, and land use. For a detailed description of each variable category, see Munira and Sener (2017).

4.2.2 Estimation of Pedestrian Volume

Based on the estimated pedestrian volume for the 44 intersections and the explanatory variables created at three buffer zones, a negative binomial model was developed. As discussed by Munira and Sener (2017), the negative binomial model has been used frequently in estimating nonmotorized volume due to the discrete nature of the volume data and since the data are overdispersed in nature. Different combinations of explanatory variables were examined, and the best model was selected based on goodness of fit and predictive accuracy as well as intuitive considerations and parsimony in specification. Table 2 presents the results of the final pedestrian direct demand model.

Table 2. Pedestrian Direct Demand Model Results.

Variable (Buffer Radius)	Estimate	T-Stat	Pr(>|z|)
Intercept	4.088	8.21	0.00
Paved and unpaved trail length (0.5 miles)	6.37e-05	4.86	0.00
Number of commercial establishments (0.1 miles)	2.39e-02	2.44	0.01
Total population under 5 years (0.5 miles)	−3.72e-03	−3.17	0.00
Population work at home (0.1 miles)	6.10e-02	4.14	0.00
Number of transit stops (1.0 miles)	8.96e-03	2.61	0.01
Model Statistics			
Sample Size (N)	44		
Root Mean Square Error (RMSE)	598.17		
R-squared (R^2)	0.77		
Mean Absolute Error (MAE)	379.35		

As illustrated in Table 2, the best model was obtained with variables of different buffer levels. This finding is consistent with previous studies (Liu & Griswold, 2009; Miranda-Moreno & Fernandes, 2011) and confirms the need for developing model variables at different buffer scales. Further investigation into the model variables revealed that while some variables

conformed to previous studies, some variables provided unique insights into the pedestrian behavior for the Austin region.

The model suggested that with the increasing length of paved and unpaved trail around 0.5 miles from the intersection, pedestrian volume increases. Previous studies have also indicated a significant positive relationship between pedestrian activity and trail length (Hankey & Lindsey, 2016). Similarly, a significant positive influence of commercial space (Miranda-Moreno & Fernandes, 2011; Tabeshian & Kattan, 2014) and transit stops (Hankey et al., 2017; Pulugurtha & Repaka, 2008) on pedestrian volume was also observed. This finding is intuitive because people are expected to walk to and from transit stops to access their final destinations. Furthermore, commercial spaces, such as shopping areas, are likely to attract pedestrians. The negative relationship between the population of small children and pedestrian volume may be attributed to people's unwillingness to walk when they have to travel with small children (Jones et al., 2010). Moreover, an interesting relationship was observed between the home-based worker population and walking activity. The *work-at-home population* refers to a worker's lack of travel from home to a separate workplace (U.S. Census Bureau, 2017). The positive relationship between this population group and pedestrian volume implies that home-based employees, who save a significant amount of commute time and energy, are probably more likely to walk for physical activity, daily chores, or recreation, contributing to the increasing pedestrian volume in their nearby areas.

Figure 1 illustrates pedestrian volume estimated by the model at the intersections of the study area. While the figure provides the pedestrian volume for the 409 signalized intersections of the study area (i.e. the entire city area), the inset map provides a close-up examination of the central region of the city.

Overall, the findings from the model provide valuable insights into the factors affecting pedestrian activity at intersections in the Austin region. In addition to safety analysis, this model can be used to estimate pedestrian demand at other intersections of the same region, which is often needed for planning purposes. The model result can also be used as an exposure measure in future pedestrian crash models if needed.

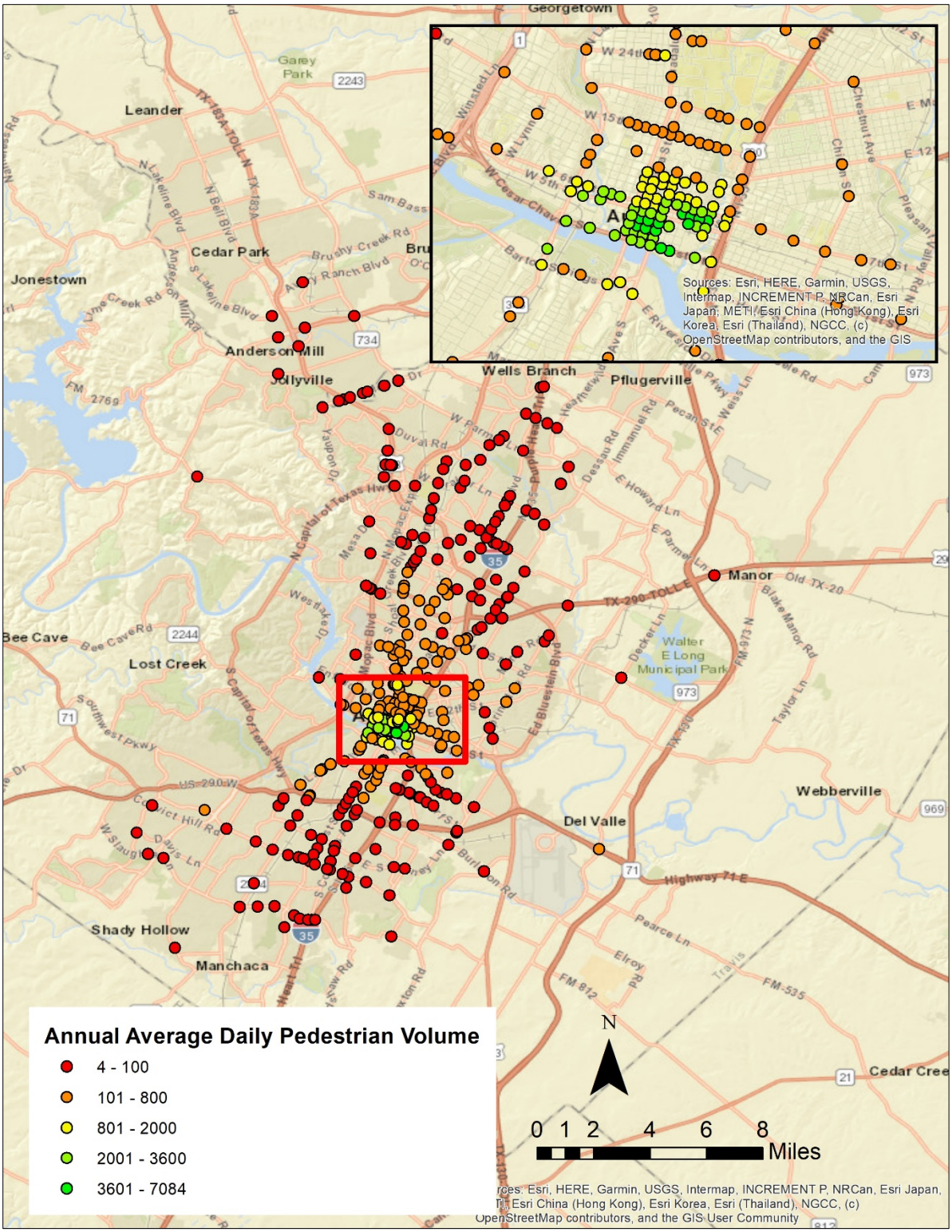

Figure 1. Pedestrian Volume at the 409 Signalized Intersections of the Study Area. (magnified central region in the inset)

4.3 Explanatory Variables

The pedestrian crash data were integrated with traffic characteristics, road geometry, and built environment features on the intersection approaches. Data for explanatory variables for each intersection were obtained from the TxDOT Roadway-Highway Inventory (RHiNO), the City of Austin's signal data and sidewalk data, and Capital Metro's transit data. Data preparation was mainly performed in ArcGIS.

The 2017 TxDOT RHiNO was used to obtain roadway features for the target intersections. Because the RHiNO network often did not perfectly align with the City of Austin's comprehensive transport network data, extensive manual data processing was needed to match the intersections and minimize error. Data were gathered for both major and minor approaches of the intersections; the major approach to the intersecting street was the street with the greater traffic volume, larger cross section, and/or higher functional class. Issues in the topology of RHiNO added to the difficulties in identifying the major and minor streets for each intersection. The key issues faced during this process were:

- Intersecting streets did not always intersect in the RHiNO data. Often, tiny gaps existed between the roadway segments of the intersections.
- RHiNO roadway segments did not always break at intersections.
- The length of roadway segments varied significantly from less than 1 foot to more than 100 feet.

To correctly categorize roadway segments as part of the major or minor street at each target intersection, the following solutions were applied:

1. The missing intersections in RHiNO, caused by the gaps between roadway segments around intersections, were manually added to include all the target intersections.
2. The complete intersection network was used to break RHiNO segments at target intersections. Only RHiNO segments within 50 feet of the target intersections were kept for the following analysis.
3. Major and minor streets for each intersection were identified based on traffic volume and geometric characteristics.

Each of the processes went through extensive quality control.

To obtain traffic and pedestrian signal data in the intersection, the City of Austin's signal data were processed for each intersection. Similarly, Capital Metro's bus stop data were gathered to determine the presence of a bus stop (within 300 feet) for each location. To obtain the sidewalk data, first the presence of a sidewalk for both major and minor streets was confirmed. Using the

data, a new variable was created that identified whether a sidewalk was present in at least one approach of the intersection.

Table 3 presents the descriptive statistics of the variables created for the model. Although variables were created for identifying truck percentage, functional classification, and presence of a median for the minor approach, they were not considered for the model due to the presence of excessive missing values in the created dataset.

Table 3. Descriptive Statistics of the Crash Model Variables.

Variables	Description	Mean	Minimum	Maximum
Ped_Vol	Annual average daily pedestrian volume	605	4	7,083
Leg_Num	Number of legs at the intersection	3.73	3	5
Spd_Max	Maximum speed limit at the intersection	44	25	65
Num_Lanes_Maj	Number of through lanes on the major road	4	2	6
Num_Lanes _Min	Number of through lanes on the minor road	2	1	6
ADT_Major	Adjusted average daily traffic volume on the major road	22,742	405	52,976
Truck_per_Maj	Percent of trucks in the average daily traffic at the major approach	3.33	1	8.9
Categorical Variable				
Bus_Stop	Presence of a bus stop (within 300 feet of the intersection)	No bus stop (42%)		
		Bus stop present (58%)		
Side_walk	Presence of a sidewalk on one approach	No sidewalk on any approaches (11%)		
		Sidewalk on one approach (89%)		
Ped_Sig	Presence of a pedestrian signal	No signal (7%)		
		Signal present (93%)		
F_ System _Maj	Functional system of the major road	Other principal arterial (60.4%)		
		Minor arterial (23.7%)		
		Major collector (14.2%)		
		Minor collector (0.2%)		
		Local (1.5%)		
F_System_Min	Functional system of the minor road	Other principal arterial (10.0%)		
		Minor arterial (9.8%)		
		Major collector (36.4%)		
		Minor collector (2.9%)		
		Local (40.8%)		
Med_ Maj	Presence of median on the major approach	0 = no median (94.0%)		
		1 = median exits (5.9%)		

The correlations between independent variables were checked. In cases where the absolute value of the correlation coefficient between pairs of variables was greater than or equal to 0.6, the variables were considered highly correlated (Evans, 1996), and inclusion of these variables might have led to an unreliable estimate of model parameters. Therefore, several checks were conducted to examine potential multicollinearity issues due to these variables. Accordingly, the functional system variable for both major and minor approaches and the variable for the number of lanes for both major and minor approaches were excluded from the final model. The median and pedestrian signal variables were also excluded because they exhibited low variation.

Chapter 5. Crash Model Results

5.1 Model Performance

The final model developed is a Bayesian multivariate Poisson lognormal CAR model. For comparison purposes, a Bayesian univariate Poisson lognormal CAR model was also estimated with the same variable specification.

The models were estimated using WinBUGS software and statistical software R (R Core Team, 2016). Package R2WinBUGS (Sturtz et al., 2005) was also used to run WinBUGS from R software and to estimate the parameters. The posterior summaries were obtained via 100,000 iterations with 50,000 burn-in samples. The convergence of the model was assessed by inspecting the trace plots and ensuring that the Monte Carlo error for each parameter of interest was less than about 5% of the sample standard deviation.

Table 4 presents the goodness of fit of both the multivariate and univariate models. The table illustrates that the multivariate model outperformed the univariate model, with the multivariate model having lower $\bar{D}$ (1803 versus 1899 for the univariate model) and DIC values (1970 versus 1981 for the univariate model) than the univariate model. The multivariate model exhibited a significant drop in $\bar{D}$ and DIC values compared to the univariate model.

Table 4. Summary of Model Performance.

Model Type	$\bar{D}$	ρ_D	DIC
Univariate model	1899.40	82.069	1981.47
Multivariate model	1803.11	167.771	1970.88

The results suggest that the multivariate model accounting for correlation among different crash types provides better model fitting, and the use of the multivariate spatial model instead of the univariate spatial model is more appropriate.

5.2 Model Results

5.2.1 Explanatory Variables (Observed)

Since the multivariate model outperformed the univariate model, the discussion of the explanatory variables is based on the multivariate model results. Table 5 presents the coefficient estimates for the multivariate spatial model. To observe the significance level, the 95% credible intervals of the posterior sampled parameters were checked. The 95% credible interval contains the sampled data values from the 2.5th percentile to the 97.5th percentile of the posterior distributions. Similarly, the 90% credible interval from the 5th percentile and 95th percentile

values was assessed. A variable is statistically insignificant if the confidence interval contains zero (Gelman, 2004).

Table 5. Estimated Coefficients of the Multivariate Spatial Model.

Variable	Fatal Crash	Incapacitating Injury Crash	Non-incapacitating Injury Crash
	Mean (Sd)	Mean (Sd)	Mean (Sd)
Intercept	−8.19 (3.26)**	−3.404 (1.3)**	−0.63 (0.72)
Ped_Vol (in 100)	−0.04 (0.03)	0.01 (0.01)	0.01 (0.005)**
Truck_per_Maj	−0.13 (0.20)	−0.10 (0.1)	0.01 (0.05)
Bus_Stop	0.09 (0.47)	−0.38 (0.22)*	0.28 (0.12)**
Side_walk	0.74 (0.94)	−0.19 (0.3)	−0.01 (0.17)
Leg_Num	−0.36 (0.44)	−0.02 (0.25)	0.03 (0.12)
Spd_Max	0.10 (0.05)**	0.002 (0.02)	0.003 (0.01)
ADT_Maj (in 1000 vehicles per day)	−0.001(0.03)	0.03 (0.01)**	0.02 (0.01)**

*Significant coefficients at the 90% confidence level.

**Significant coefficients at the 95% confidence level.

The variation in significance and the magnitude of explanatory variables across different crash severity types emphasize the need for multivariate models by severity type in order to provide accurate guidelines for designing countermeasures.

The results showed that for the fatal crash severity level, the only significant variable (at 95% confidence level) was the maximum speed limit. This significant positive influence of speed limit on fatal pedestrian crashes suggests that pedestrians are more at risk of being killed in a crash when vehicles are driving at a higher speed. Previous studies have also revealed that higher speed limits are associated with a greater risk of pedestrian crashes, including severe pedestrian injuries (Davis, 2001; Jensen, 1998; Zegeer et al., 2006). According to model results, holding all other variables constant, the relative risk of a fatal crash for pedestrians at signalized intersections increased by around 10% with an increase of one standard deviation in speed limit.

When the variables for incapacitating injury crashes and non-incapacitating injury crashes were investigated, both crash severities had two common significant variables: average daily traffic volume on the major approach and the presence of a bus stop.

As might be expected, the daily traffic volume on the major approach had a positive influence on both incapacitating injury crashes and non-incapacitating injury crashes, which conforms to previous studies (Harwood et al., 2008; Huang et al., 2017; Zegeer et al., 2001; Zhao et al., 2018). The relative risk of incapacitating injury crashes and non-incapacitating injury crashes

increased by 3.4% and 2.4%, respectively, with an increase of one standard deviation of traffic volume (in 1000 vehicles per day).

In contrast to some earlier studies (Huang et al., 2017; Strauss et al., 2014), the effect of the presence of a bus stop around the intersection showed an interesting difference across the two crash types. The model results showed that the presence of a bus stop decreased (by 31%) the risk of incapacitating injury crashes but increased (by 32%) the risk of non-incapacitating injury crashes for pedestrians. This finding may be attributed to the fact that an intersection with bus stops, accommodating a lot of pedestrian and bicycle traffic, may decrease the risk of incapacitating injury crashes because the speed is generally low and motorists are careful and vigilant, but may increase the risk of non-incapacitating injury crashes due to the decreased visibility.

In terms of pedestrian exposure measure, the results indicated that increasing pedestrian volume contributes to increasing non-incapacitating injury crashes. Previous studies have also suggested a similar relationship (Amoh-Gyimah et al., 2016; Cai et al., 2016; Osama & Sayed, 2017). Although at a lower confidence level (t-stat > 1), the pedestrian volume appeared to have a negative effect on fatal crashes, which indicates that intersections with higher pedestrian activity exhibit lower risk of fatal pedestrian crash. Previous studies have suggested that the relationship between pedestrian volume and crashes is complex. While the total number of pedestrian crashes at a particular location tends to increase with increasing pedestrian volume, the increase is nonlinear in nature (Leden, 2002; Jacobsen, 2003).

5.2.3 Heterogeneous and Spatial Effects (Unobserved)

Table 6 presents the variance-covariance and correlation of heterogeneous effects across crash severities within intersections. Table 7 presents the variance-covariance and correlation matrix of spatial effect for crash counts for three severity types. The diagonal cells of the table indicate the variance for each crash severity. The covariance matrix is presented in the upper part of the matrix in each table. The correlation matrix is presented in the lower part of the matrix in each table. The effect is significant when the standard deviation is lower than the mean and not significant when the standard deviation is higher than the mean (Huang et al., 2017).

Table 6. Variance-Covariance and Correlation Matrix for Heterogeneous Effects.

	Fatal	Incapacitating	Non-incapacitating
		Mean (Standard Deviation)	
Fatal	0.52 (0.4)[a]	0.01 (0.19)[b]	−0.05 (0.1)[b]
Incapacitating	0.01 (0.39)[c]	0.38 (0.19)[a]	0.06 (0.07)[b]
Non-incapacitating	−0.15 (0.3)[c]	0.22 (0.24)[c]	0.17 (0.05)[a]

[a] Variance.
[b] Covariance.
[c] Correlation.

Table 7. Variance-Covariance and Correlation Matrix for Spatial Effects.

	Fatal	Incapacitating	Non-incapacitating
		Mean (Standard Deviation)	
Fatal	0.44 (0.36)[a]	0.06 (0.22)[b]	0.02 (0.16)[b]
Incapacitating	0.09 (0.39)[c]	0.51 (0.33)[a]	0.17 (0.18)[b]
Non-incapacitating	0.04 (0.38)[c]	0.37 (0.32)[c]	0.32 (0.16)[a]

[a] Variance.
[b] Covariance.
[c] Correlation.

As Table 6 shows, the variance of heterogeneous effects for the crash count of each severity is significant and indicates the need to incorporate a heterogeneous error term in the model. Moreover, the value of heterogeneous variance is the highest for fatal crashes, which suggests that fatal crashes exhibit more randomness than incapacitating and non-incapacitating injury crashes. However, the covariance for heterogenous effects is not significant. In addition, the correlation between crash counts of all severity types is not significant, indicating no significant unobserved common factor contributing to fatal, incapacitating, and non-incapacitating crash counts for pedestrians.

Similar to heterogenous effects, the results of Table 7 also indicate the significant variance of spatial effects for crash counts of each crash severity and suggest that the crash observations of different severities exhibit a significant correlation between adjacent intersections. The results also indicate the insignificance of the covariance for spatial effects. Furthermore, the correlation between fatal and incapacitating crashes and fatal and non-incapacitating crashes for the spatial residual is also not significant. This suggests that a higher number of fatal crashes occurring at a particular intersection is significantly correlated with a higher number of fatal crashes at its adjacent intersection, but is not significantly correlated with crash frequency of incapacitating and non-incapacitating crashes at the adjacent intersection. However, a significant correlation exists between the crash frequency of incapacitating and non-incapacitating crashes at the adjacent intersections.

Chapter 6. Discussion

The findings of the study offer valuable insights into both pedestrian demand and crashes of different severity levels in the study area. One of the noteworthy contributions of this study is that it incorporated pedestrian volume at all intersections of the study area in the crash analysis—an aspect most often missing or represented via surrogate measures. Even the 2018 Pedestrian Safety Action Plan of the City of Austin (Austin Transportation Department, 2018) highlighted the need to incorporate reliable pedestrian volume for safety analysis in the area. The direct demand model for pedestrians not only proved to be a crucial component in the crash model but also can be used as a standalone pedestrian demand estimation tool needed for nonmotorized policy formulation, project prioritization, pollution analysis, and so forth.

Additionally, the direct demand model yielded valuable insights into the factors affecting pedestrian demand in the Austin area. The positive influence of trails (both paved and unpaved) shows how the urban trail facilities contribute to pedestrian traffic. The city's Urban Trails Master Plan (City of Austin, 2014) outlines the plan for future expansion of the urban trail network throughout the city, and the related model result underscores the promising potential of the expansion projects to encourage pedestrian activity. In addition, given that commercial establishments and transit stops attract pedestrian activity, city officials should consider promoting mixed-use developments and provide facilities such as well-buffered sidewalks near transit stops and commercial areas. Another significant determinant of pedestrian activity at intersections in the Austin area is the population of employees working remotely. As shown by the model, the population working at home contributes to increasing pedestrian activity at nearby locations. This finding indicates that home-based employees, who save a significant amount of commute time and energy, are probably more likely to walk for physical activity, daily chores, or recreation. Employers seeking policies to boost the physical and mental well-being of employees may consider flexible work arrangements so that employees can engage in physical activity.

The pedestrian crash model developed for this study, along with the subsequent results, can be beneficial in helping policy makers create both short- and long-term strategies to reduce pedestrian crashes of all severity types. While some variables conform to previous studies, other variables offer new insights into the crash patterns of the study area. The positive relationship between speed limit and fatal crash risk is well supported by observations reported by the Austin Transportation Department (2018), which indicated that although crashes are more frequent in locations with lower speed limits (30–45 mph), the risk of fatal crashes is the highest (64% from 2010 to 2015) when the speed limits increase beyond 45 mph. The findings of the model in this study highlight the need to perform road safety audits of high-speed roadways and to develop criteria to promote safe design speeds of city streets. Educational campaigns to promote safe driving and walking behavior may also prove beneficial.

The significant positive influence of traffic volume on both incapacitating and non-incapacitating injury crashes at intersections also warrants specific policy attention. A road diet strategy that involves narrowing or eliminating travel lanes has proven to be an effective strategy for improving pedestrian safety conditions by lowering vehicle speed and reducing crossing distance (Zegeer et al., 2001). In addition, installation of bike lanes, which provides a buffer between the street and the sidewalk, increases driver awareness and expectation as well as reduces potential conflict between pedestrians and bicyclists on the sidewalk—both encouraging nonmotorized activity and improving road safety conditions.

The influence of the presence of bus stops at the intersections across incapacitating and non-incapacitating injury crashes offers interesting insights as well. The negative and positive influence of bus stops on incapacitating injury and non-incapacitating injury crashes, respectively, can be explained by findings from previous studies. For instance, Clifton et al. (2009) found a significant negative influence of bus stops on injury crashes but not on fatal crashes. The authors suggested that better transit access might be representative of an urban center area where motorists travel at slow speeds and there are fewer crashes with severe injuries. On the other hand, the higher risk of incapacitating injury crashes near the transit stops may be attributed to more pedestrians walking to board or exit buses. The positive influence of bus stops on pedestrian crashes at intersections was also reported by Pulugurtha and Sambhara (2011), but the severity of crashes was not differentiated.

The policies regarding bus stops should aim to reduce pedestrian crashes of all severity types and need to consider a combination of design elements. For example, increasing pedestrian crossing time, installing high-visibility crosswalks and refuge islands, and ensuring adequate light have been proven to be effective in reducing pedestrian crashes (Chen et al., 2013). At the same time, extra caution should be exercised before implementing engineering measures such as relocation of bus stops (such as to midblock locations), only considering the measure's positive influence on incapacitating injury crashes. A report by the Austin Transportation Department (2018) also indicated that despite having more crashes at intersection locations, crashes at midblock locations are more often severe. This finding might be attributed to the higher speed and lower expectation of pedestrians crossing.

Finally, the influence of pedestrian volume on different crash severity types provides interesting insight. Although at a lower confidence level (t-stat > 1), pedestrian volume appears to have a negative effect on fatal crashes, which indicates that intersections with higher pedestrian activity exhibit lower risk of fatal pedestrian crashes. This phenomenon might be because severe crashes (in the Austin area) are more likely to occur on non-local roads, which generally accommodate lower pedestrian volumes than do local roads. On the other hand, intersections with higher pedestrian demand tend to experience higher non-incapacitating injury crashes (significant at the 95% confidence level). The relationship between pedestrian severity type and pedestrian demand at intersections merits further investigation.

Chapter 7. Conclusion

Intersection-level crash analyses are essential to obtaining deeper insights into the factors affecting safety conditions in order to facilitate policy decisions. The traffic operations at signalized intersections are complex, and the pedestrian crash risk of different severities at signalized intersections can be influenced by many operational and geometric factors, which require profound understanding to design countermeasures.

This study contributed to the field of research by developing an integrated analysis framework to examine the impacts of various factors on crash frequencies across fatal, incapacitating injury, and non-incapacitating injury crashes involving pedestrians in the Austin area. The ultimate objective was to demonstrate the usability of direct demand models to develop exposure measures—a key feature of crash analysis—and to illustrate the potential of complex multivariate models to accurately estimate crash parameters to help develop policy-based countermeasures aimed at reducing pedestrian crash risk at intersections.

A multivariate Poisson lognormal spatial model was developed. The results showed that the multivariate model accounting for correlation among different crash types provided better model fitting, and the use of the multivariate spatial model instead of the univariate spatial model was more appropriate in the study context. Moreover, the model could distinguish the difference in influence of multiple explanatory variables across the crash types at the intersections in the Austin area.

The analysis is not without limitations. Because the crash data do not identify the name of the intersection related to a specific crash, the geo-referenced crash data were used to join and identify the nearest intersection. Therefore, the research depended on the accuracy of the coordinates input. Issues in the topology of RHiNO data added to the challenge of processing data.

Future studies need to gather more intersection-related features to observe their influence on pedestrian crashes of different severity types. In addition to intersection location, crash risk on midblock locations should be investigated to design policy measures.

References

Abdel-Aty, M., & Wang, X. (2006). Crash estimation at signalized intersections along corridors: Analyzing spatial effect and identifying significant factors. *Transportation Research Record, 1953*, 98–111.

Abdel-Aty, M. A., & Radwan, A. E. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention, 32*, 633–642.

Aguero-Valverde, J. (2013). Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. *Accident Analysis & Prevention, 59*, 365–373.

Aguero-Valverde, J., & Jovanis, P. (2009). Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. *Transportation Research Record, 2136*, 82–91.

Aguero-Valverde, J., & Jovanis, P. P. (2010). Spatial correlation in multilevel crash frequency models: Effects of different neighboring structures. *Transportation Research Record, 2165*, 21–32.

Alarifi, S. A., Abdel-Aty, M., & Lee, J. (2018). A Bayesian multivariate hierarchical spatial joint model for predicting crash counts by crash type at intersections and segments along corridors. *Accident Analysis & Prevention, 119*, 263–273.

Alarifi, S. A., Abdel-Aty, M. A., Lee, J., & Park, J. (2017). Crash modeling for intersections and segments along corridors: A Bayesian multilevel joint model with random parameters. *Analytic Methods in Accident Research, 16*, 48–59.

Amoh-Gyimah, R., Saberi, M., & Sarvi, M. (2016). *Modelling crash spatial heterogeneity using semi-parametric geographically weighted Poisson regression*. Paper presented at the Australasian Road Safety Conference, Canberra, ACT, Australia.

Anastasopoulos, P. C. (2016). Random parameters multivariate Tobit and zero-inflated count data models: Addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. *Analytic Methods in Accident Research, 11*, 17–32.

Austin Transportation Department. (2018). *City of Austin Pedestrian Safety Action Plan 2018: Vision Zero*. Retrieved from https://austintexas.gov/sites/default/files/files/Transportation/Pedestrian_Safety_Action_Plan_1-11-18.pdf

Barua, S., El-Basyouny, K., & Islam, M. T. (2014). A full Bayesian multivariate count data model of collision severity with spatial correlation. *Analytical Methods in Accident Research, 3*, 28–43.

Beck, L. F., Dellinger, A. M., & O'Neil, M. E. (2007). Motor vehicle crash injury rates by mode of travel, United States: Using exposure-based methods to quantify differences. *American Journal of Epidemiology, 166*(2), 212–218.

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics, 43*(1), 1–20.

Bhowmik, T., Yasmin, S., & Eluru, N. (2018). A joint econometric approach for modeling crash counts by collision type. *Analytic Methods in Accident Research, 19*, 16–32.

Bradshaw, K. (2019). *Pedestrian deaths reached record levels in 2018, police data show*. Retrieved from https://www.statesman.com/news/20190215/pedestrian-deaths-reached-record-levels-in-2018-police-data-show

Cai, Q., Abdel-Aty, M., & Lee, J. (2017). Macro-level vulnerable road users crash analysis: A Bayesian joint modeling approach of frequency and proportion. *Accident Analysis & Prevention*, *107*, 11–19.

Cai, Q., Lee, J., Eluru, N., & Abdel-Aty, M. (2016). Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. *Accident Analysis and Prevention*, *93*, 14–22.

Chen, L., Chen, C., Ewing, R., McKnight, C. E., Srinivasan, R., & Roe, M. (2013). Safety countermeasures and crash reduction in New York City—Experience and lessons learned. *Accident Analysis & Prevention*, *50*, 312–322.

Cheng, Z., Zu, Z., & Lu, J. (2018). Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification of urban road intersections. *Sustainability*, *11*(1), 1–17.

Choi, E. H. (2010). *Crash factors in intersection-related crashes: An on-scene perspective* (Report No. HS-811 366). Washington, DC: National Highway Traffic Safety Administration.

City of Austin. (2014). *Urban trails master plan*. Retrieved from http://www.austintexas.gov/edims/document.cfm?id=210984

Clifton, K. J., Burnier, C. V., & Akar, G. (2009). Severity of injury resulting from pedestrian-vehicle crashes: What can we learn from examining the built environment? *Transportation Research Part D: Transport and Environment*, *14*(6), 425–436.

Davis, G. A. (2001). Relating severity of pedestrian injury to impact speed in vehicle-pedestrian crashes: Simple threshold model. *Transportation Research Record*, *1773*, 108–113.

El-Basyouny, K., Barua, S., & Islam, M. T. (2014). Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. *Accident Analysis & Prevention*, *73*, 91–99.

El-Basyouny, K., & Sayed, T. (2009). Collision prediction models using multivariate Poisson-lognormal regression. *Accident Analysis & Prevention*, *41*, 820–828.

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Pacific Grove, CA: Thomson Brooks/Cole.

Fitzpatrick, K., Avelar, R., & Turner, S. (2018). *Guidebook on identification of high pedestrian crash locations* (Report No. FHWA-HRT-17-106). College Station, TX: Texas A&M Transportation Institute.

Gelman, A. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515–533.

Governors Highway Safety Association (GHSA). (2019). *Pedestrian traffic fatalities by state: 2018 preliminary data*. Retrieved from https://www.ghsa.org/sites/default/files/2019-02/FINAL_Pedestrians19.pdf

Hankey, S., & Lindsey, G. (2016). Facility-demand models of peak period pedestrian and bicycle traffic: Comparison of fully specified and reduced-form models. *Transportation Research Record, 2586*, 48–58.

Hankey, S., Lu, T., Mondschein, A., & Buehler, R. (2017). *Merging traffic monitoring and direct-demand modeling to assess spatial patterns of annual average daily bicycle and pedestrian traffic*. Paper presented at the Transportation Research Board 2017 Annual Meeting, Washington, DC.

Harwood, D. W., Bauer, K. M., Richard, K. R., Gilmore, D. K., Graham, J. L., Potts, I. B., . . . Hauer, E. (2008). *Pedestrian safety prediction methodology.* Washington, DC: Transportation Research Board.

Hasani, M., Jahangiri, A., Sener, I. N., Munira, S., Owens, J. M., Appleyard, B., . . . Ghanipoor Machiani, S. (2019). Identifying high-risk intersections for walking and bicycling using multiple data sources in the city of San Diego. *Journal of Advanced Transportation, 2019*. doi:10.1155/2019/9072358

Heydari, S., Fu, L., Miranda-Moreno, L. F., & Joseph, L. (2017). Using a flexible multivariate latent class approach to model correlated outcomes: A joint analysis of pedestrian and cyclist injuries. *Analytic Methods in Accident Research, 13*, 16–27.

Huang, H., & Abdel-Aty, M. (2010). Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis & Prevention, 42*, 1556–1565.

Huang, H., Zhou, H., Wang, J., Chang, F., & Ma, M. (2017). A multivariate spatial model of crash frequency by transportation modes for urban intersections. *Analytic Methods in Accident Research, 14*, 10–21.

Jacobsen, P. L. (2003). Safety in numbers: More walkers and bicyclists, safer walking and biking. *Injury Prevention, 9*, 205–209.

Jensen, S. U. (1998). *Pedestrian safety analyses and safety measures* (Report No. 148). Copenhagen, Denmark: Danish Road Directorate, Division of Traffic Safety and Environment.

Jones, M. G., Ryan, S., Donlon, J., Ledbetter, L., Ragland, D. R., & Arnold, L. S. (2010). *Seamless travel: Measuring bicycle and pedestrian activity in San Diego County and its relationship to land use, transportation, safety, and facility type*. Berkeley, CA: University of California, Berkeley, Safe Transportation Research and Education Center.

LaScala, E. A., Gerber, D., & Gruenewald, P. J. (2000). Demographic and environmental correlates of pedestrian injury collisions: A spatial analysis. *Accident Analysis & Prevention, 32*, 651–658.

Lawson, A. B. (2013). *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology*. London, England: Chapman & Hall/CRC.

Leden, L. (2002). Pedestrian risk decrease with pedestrian flow: A case study based on data from signalized intersections in Hamilton, Ontario. *Accident Analysis & Prevention, 34*, 457–464.

Lee, J., Abdel-Aty, M., & Jiang, X. (2015). Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. *Accident Analysis & Prevention, 78*, 146–154.

Lee, J., Abdel-Aty, M., & Shah, I. (2019). Evaluation of surrogate measures for pedestrian trips at intersections and crash modeling. *Accident Analysis & Prevention, 130*, 91–98.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. London, England: Chapman & Hall/CRC.

Liu, C., & Sharma, A. (2017). Exploring spatio-temporal effects in traffic crash trend analysis. *Analytic Methods in Accident Research, 16*, 104–116.

Liu, C., & Sharma, A. (2018). Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. *Analytic Methods in Accident Research, 17*, 14–31.

Liu, H., & Zhu, X. (2017). Joint modeling of multiple crimes: A Bayesian spatial approach. *ISPRS International Journal of Geo-Information, 6*(1), 16.

Liu, X., & Griswold, J. (2009). Pedestrian volume modeling: A case study of San Francisco. *Association of Pacific Coast Geographers Yearbook, 71*, 164–181.

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A, 44*, 291–305.

Loukaitou-Sideris, A., Liggett, R., & Sung, H. G. (2007). Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles. *Journal of Planning Education and Research, 26*, 338–351.

Ma, J., Kockelman, K. M., & Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention, 40*, 964–975.

Ma, X., Chen, S., & Chen, F. (2017). Multivariate space-time modeling of crash frequencies by injury severity levels. *Analytic Methods in Accident Research, 15*, 29–40.

Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research, 11*, 1–16.

Miaou, S. P., & Lord, D. (2003). Modeling traffic crash-flow relationships for intersections: Dispersion parameter, functional form, and Bayes versus empirical Bayes. *Transportation Research Record, 1840*, 31–40.

Miranda-Moreno, L., & Fernandes, D. (2011). Modeling of pedestrian activity at signalized intersections: Land use, urban form, weather, and spatiotemporal patterns. *Transportation Research Record, 2264*, 74–82.

Mitra, S., & Washington, S. (2012). On the significance of omitted variables in intersection crash modeling. *Accident Analysis & Prevention, 49*, 439–448.

MRC Biostatistics Unit. (2004). DIC: Deviance information criteria. Retrieved from https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-dic/

Munira, S., & Sener, I. N. (2017). *Use of the direct-demand modeling in estimating nonmotorized activity: A meta-analysis*. Texas A&M Transportation Institute.

Narayanamoorthy, S., Paleti, R., & Bhat, C. R. (2013). On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. *Transportation Research Part B: Methodological, 55*, 245–264.

Nashad, T., Yasmin, S., Eluru, N., Lee, J., & Abdel-Aty, M. A. (2016). Joint modeling of pedestrian and bicycle crashes: Copula-based approach. *Transportation Research Record, 2601*, 119–127.

Nordback, K., Marshall, W. E., Janson, B. N., & Stolz, E. (2013). Estimating annual average daily bicyclists: Error and accuracy. *Transportation Research Record, 2339*, 90–97.

Ortuzar, J. D. D., & Willumsen, L. G. (2011). *Modelling transport* (4th ed.). Chichester, UK: John Wiley & Sons.

Osama, A., & Sayed, T. (2017). Evaluating the impact of connectivity, continuity, and topography of sidewalk network on pedestrian safety. *Accident Analysis & Prevention, 107*, 117–125.

Park, E. S., & Lord, D. (2007). Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transportation Research Record, 2019*(1), 1–6.

Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering—ASCE, 122*(2), 105–113.

Pulugurtha, S., & Repaka, S. (2008). Assessment of models to measure pedestrian activity at signalized intersections. *Transportation Research Record, 2073*, 39–48.

Pulugurtha, S. S., & Sambhara, V. R. (2011). Pedestrian crash estimation models for signalized intersections. *Accident Analysis & Prevention, 43*(1), 439–446.

Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accident Analysis & Prevention, 40*, 1486–1497.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rana, T. A., Sikder, S., & Pinjari, A. R. (2010). Copula-based method for addressing endogeneity in models of severity of traffic crash injuries: Application to two-vehicle crashes. *Transportation Research Record, 2147*, 75–87.

Song, J. J., Ghosh, M., Miaou, S., & Mallick, B. (2006). Bayesian multivariate spatial models for roadway traffic crash mapping. *Journal of Multivariate Analysis, 97*, 246–273.

Speigelhalter, D., Best, N., Carlin, B., & Linde, V. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B, 64*, 583–616.

Strauss, J., Miranda-Moreno, L. F., & Morency, P. (2014). Multimodal injury risk analysis of road users at signalized and non-signalized intersections. *Accident Analysis & Prevention, 71*, 201–209.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software, 12*(3), 1–16.

Tabeshian, M., & Kattan, L. (2014). Modeling nonmotorized travel demand at intersections in Calgary, Canada: Use of traffic counts and geographic information system data. *Transportation Research Record, 2430*, 38–46.

Texas Department of Transportation (TxDOT). (2016a). *Texas intersection safety implementation plan: Preliminary findings for Texas's Capital Area Metropolitan Planning Organization.* Retrieved from https://www.texasshsp.com/wp-content/uploads/2017/02/Preliminary-Findings_CAMPO_2016-06-15.pdf

Texas Department of Transportation (TxDOT). (2016b). *State of Texas classification of motor vehicle traffic crashes in Texas: 2016 edition*. Retrieved from https://ftp.dot.state.tx.us/pub/txdot-info/library/forms/cit/crash102_0616.pdf

Turner, S., Sener, I. N., Martin, M., Das, S., Shipp, E., Hampshire, R., . . . Robinson, S. (2017). *Synthesis of methods for estimating pedestrian and bicyclist exposure to risk at areawide levels and on specific transportation facilities*. Washington, DC: U.S. Department of Transportation.

U.S. Census Bureau. (2017). *Commuting (journey to work).* Retrieved from https://www.census.gov/topics/employment/commuting/guidance/home-based-workers.html

Wali, B., Khattak, A. J., & Xu, J. (2018). Contributory fault and level of personal injury to drivers involved in head-on collisions: Application of copula-based bivariate ordinal models. *Accident Analysis & Prevention, 110*, 101–114.

Wang, J., & Huang, H. (2016). Road network safety evaluation using Bayesian hierarchical joint model. *Accident Analysis & Prevention, 90*, 152–158.

Wang, K., Bhowmik, T., Yasmin, S., Zhao, S., Eluru, N., & Jackson, E. (2019a). Multivariate copula temporal modeling of intersection crash consequence metrics: A joint estimation of injury severity, crash type, vehicle damage and driver error. *Accident Analysis & Prevention, 125*, 188–197.

Wang, K., Yasmin, S., Konduri, K. C., Eluru, N., & Ivan, J. N. (2015). Copula-based joint model of injury severity and vehicle damage in two-vehicle crashes. *Transportation Research Record, 2514*, 158–166.

Wang, W., Yuan, Z., Yang, Y., Yang, X., & Liu, Y. (2019b). Factors influencing traffic accident frequencies on urban roads: A spatial panel time-fixed effects error model. *PLoS One, 14*(4), e0214539.

Wang, X., Song, Y., Yu, R., & Schultz, G. G. (2014). Safety modeling of suburban arterials in Shanghai, China. *Accident Analysis & Prevention, 70*, 215–224.

Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention, 60*, 71–84.

Wang, Y., Kockelman, K., & Jamali, A. (2017). A synthesis of spatial models for multivariate count responses. In R. Jackson & P. Schaeffer (Eds.), *Regional research frontiers* (Vol. 2, pp. 221–237). Cham, Switzerland: Springer.

Xie, K., Ozbay, K., & Yang, H. (2019). A multivariate spatial approach to model crash counts by injury severity. *Accident Analysis & Prevention*, *122*, 189–198.

Xie, K., Wang, X., Huang, H., & Chen, X. (2013). Corridor-level signalized intersection safety analysis in Shanghai, China using Bayesian hierarchical models. *Accident Analysis & Prevention*, *50*, 25–33.

Xie, K., Wang, X., Ozbay, K., & Yang, H. (2014). Crash frequency modeling for signalized intersections in a high-density urban road network. *Analytic Methods in Accident Research*, *2*, 39–51.

Yasmin, S., & Eluru, N. (2018). A joint econometric framework for modeling crash counts by severity. *Transportmetrica A: Transport Science*, *14*(3), 230–255.

Yasmin, S., Eluru, N., Lee, J., & Abdel-Aty, M. (2016). Ordered fractional split approach for aggregate injury severity modeling. *Transportation Research Record*, *2583*, 119–126.

Yasmin, S., Eluru, N., Pinjari, A. R., & Tay R. (2014). Examining driver injury severity in two vehicle crashes—A copula based approach. *Accident Analysis & Prevention*, *66*, 120–135.

Yasmin, S., Momtaz, S. U., Nashad, T., & Eluru, N. (2018). A multivariate copula-based macro-level crash count model. *Transportation Research Record*, *2672*, 64–75.

Ye, X., Pendyala, R. M., Washington, S. P., Konduri, K., & Oh, J. (2009). A simultaneous equations model of crash frequency by collision type for rural intersections. *Safety Science*, *47*, 443–452.

Zegeer, C. V., Carter, D. L., Hunter, W. W., Stewart, J. R., Huang, H., Do, A., & Sandt, L. (2006). Index for assessing pedestrian safety at intersections. *Transportation Research Record*, *1982*, 76–83.

Zegeer, C. V., Stewart, J. R., Huang, H., & Lagerwey, P. (2001). Safety effects of marked versus unmarked crosswalks at uncontrolled locations: Analysis of pedestrian crashes in 30 cities. *Transportation Research Record*, *1773*, 56–68.

Zeng, Q., & Huang, H. (2014). Bayesian spatial joint modeling of traffic crashes on an urban road network. *Accident Analysis & Prevention*, *67*, 105–112.

Zeng, Q., Wen, H., Huang, H., Pei, X., & Wong, S. C. (2017). A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity. *Accident Analysis & Prevention*, *99*, 184–191.

Zhan, X., Aziz, H. A., & Ukkusuri, S. V. (2015). An efficient parallel sampling technique for multivariate Poisson-lognormal model: Analysis with two crash count datasets. *Analytic Methods in Accident Research*, *8*, 45–60.

Zhao, M., Liu, C., Li, W., & Sharma, A. (2018). Multivariate Poisson-lognormal model for analysis of crashes on urban signalized intersections approach. *Journal of Transportation Safety & Security*, *10*(3), 251–265.