https://doi.org/10.1016/j.aap.2020.105898

A joint probability model for pedestrian crashes at macroscopic level: Roles of environment, traffic, and population characteristics

Junbiao Su
Department of Civil and Environmental Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
Email: Junbiao.su@connect.polyu.hk

N.N. Sze (Corresponding author)
Department of Civil and Environmental Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
Tel: +852 2766-6062; Email: tony.nn.sze@polyu.edu.hk

Lu Bai
Jiangsu Key Laboratory of Urban ITS, Southeast University
Si Pai Lou #2, Nanjing, China, 210096
and
Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies
Si Pai Lou #2, Nanjing, China, 210096
Email: xinyuesther@126.com

ABSTRACT

Road safety is a major public health issue, with road crashes accounting for one-fourth of all documented injuries. In these crashes, pedestrians are more vulnerable to fatal and/or severe injuries than car occupants. Therefore, it is necessary to have a better understanding of the relationship between pedestrian crashes and possible influencing factors, including road environment, traffic conditions, and population characteristics. In conventional studies, separate prediction models were established for pedestrian crashes and other crash types, which could have ignored possible correlations among the different crash types. Additionally, these influencing factors can contribute to pedestrian crashes in two manners, i.e., contributing to crash occurrence and propensity of pedestrian involvement. Furthermore, extensive pedestrian count data were generally not available, affecting the estimation of pedestrian crash exposure. In this study, a joint probability model is adopted for the simultaneous modeling of crash occurrence and pedestrian involvement in crashes; effects of possible influencing factors, including land use, road networks, traffic flow, population demographics and socioeconomics, public transport facilities, and trip attraction attributes, are considered. Additionally, trip generation and pedestrian activity data, based on a comprehensive household travel survey, are used to determine pedestrian crash exposure. Markov chain Monte Carlo full Bayesian approach is then applied to estimate the parameters. Results indicate that crash occurrence is correlated to traffic flow, number of non-signalized intersections, and points of interest such as restaurants and hotels. By contrast, population age, ethnicity, education, household size, road density, and number of public transit stations could affect the propensity of pedestrian involvement in crashes. These findings indicate that better design and planning of built environments are necessary for safe and efficient access for pedestrians and for the long-term improvement of walkability in a high-density city such as Hong Kong.

Keywords: Pedestrian safety, joint probability model, crash prediction model, walkability, accessibility

1. INTRODUCTION

Walking is increasingly promoted for short-distance travel and access to public transport services (Miranda-Moreno *et al.,* 2011; Transport Department, 2012a; Bhat *et al.,* 2017) because it not only increases physical activity and improves health, but also alleviates traffic congestion and reduces vehicle emissions (Castro *et al.,* 2013; Osama and Sayed, 2017a; Ferrer and Ruiz, 2018; Woo *et al.,* 2019). However, pedestrian safety has been of great concern. Despite many jurisdictions adopting a "Vision Zero" policy aiming to achieve remarkable reduction of road traffic-related injuries (Bhat *et al.* 2017), pedestrians remain prone to fatal and/or severe injuries on roads (Zegeer and Bushell, 2012; Wang and Kockelman, 2013; Zhang *et al.,* 2015; Wang *et al.,* 2016). In Hong Kong, pedestrians constituted 59% of overall road fatalities in 2010, which increased to 67% in 2015 (Transport Department, 2015). In addition, poor access to destination sites, such as offices, shops, and community facilities, frequent conflicts between pedestrians and vehicular traffic, and inadequate pedestrian facilities can all discourage walking (Weinstein Agrawal *et al.,* 2008; Guo and Loo, 2013; Lindelöw *et al.,* 2014; Kamargianni *et al.,* 2015; Cai *et al.,* 2016). Pedestrian safety is an important indicator of walkability. It is therefore necessary to improve the understanding of the relationship between pedestrian safety and possible risk factors, to allow the implementation of effective measures that can improve pedestrian safety and, in turn, enhance the walkability in Hong Kong (Guo and Loo, 2013; Sze and Christensen, 2017; Chen et al., 2020b).

Many studies have revealed explanatory factors that affect pedestrian safety based on historical crash data. Possible influencing factors, including road environment, traffic attributes, and population profiles, have been investigated (Miles-Doan and Thompson, 1999; Sze and Wong, 2007; Lord and Mannering, 2010; Castro *et al.,* 2012; Wang and Kockelman, 2013; Wang *et al.,* 2016; Cheng *et al.,* 2018b). Researchers have developed univariate models to estimate counts for pedestrian crashes and other crash types (Miranda-Moreno *et al.,* 2011; Lam *et al.,* 2014; Lee *et al.,* 2015a; Yao *et al.,* 2015;

Amoh-Gyimah *et al.,* 2016; Wang *et al.,* 2016). However, correlations may exist among the counts for different crash types. Such correlations could then interfere with the estimates produced by these univariate models (Mannering *et al.,* 2016). Furthermore, contradictory effects have been demonstrated for different influencing factors, including transit accessibility and numbers of intersections, among pedestrian crashes and other crash types (Clifton *et al.,* 2009; Cottrill and Thakuriah, 2010; Merlin *et al.,* 2020). To properly account for these correlations, advanced modeling approaches such as multi-level hierarchical structures and multivariate method have been proposed (Huang and Abdel-Aty, 2010; Castro *et al.,* 2013; Wang and Kockelman, 2013; Lee *et al.,* 2015b; Bhat *et al.,* 2017; Huang *et al.,* 2017; Ma *et al.,* 2017; Alarifi *et al.,* 2018; Liu and Sharma, 2018). These models improve prediction performance by explicitly considering the correlations among different types of crashes and simultaneously estimating the effects of possible explanatory factors.

In conventional multivariate models, the response variables are generally mutually exclusive (i.e., injury versus no-injury crashes, and motorist versus non-motorist crashes). Moderating effects on the relationship between crash risk and road environment, i.e., transport mode and injury severity, with respect to different crash types should then be explored (Lee *et al.,* 2015b; Huang *et al.,* 2017; Cheng *et al.,* 2018a). It is also possible to model pedestrian crashes using a hierarchical approach, wherein pedestrian crashes are considered as a subset of total crashes to estimate the propensity of pedestrian involvement in crashes. Built environments and other relevant factors can affect the propensity of pedestrian crashes in two ways, i.e., contributing to crash incidence and pedestrian involvement in crashes. For example, areas with higher population densities tend to have lower total crash rates (Qin *et al.,* 2004; Cai *et al.,* 2017b) but higher numbers of pedestrian crashes (Siddiqui *et al.,* 2012a; Cai *et al.,* 2016; Cai *et al.,* 2017a; Cai *et al.,* 2017b). Although several studies have already explored the roles of built environments in pedestrian crashes (Clifton *et al.,* 2009; Dai and Jaworski, 2016; Osama and Sayed, 2017b; Ding *et al.,* 2018), the differences in the effects of explanatory factors on crash occurrence and propensity of pedestrian involvement in these crashes remain uninvestigated, which, if investigated properly, can lead to the development of targeted countermeasures to address issues relevant to general crash occurrence and those that are specific to pedestrian behaviors.

The tendency of a person to be involved in a road crash can increase with their amount of travel, i.e., frequency, travel distance, and travel time. This is referred to as exposure. It is necessary to control the exposure in the estimation of crash risk (i.e., probability of crash per unit travel) and in the measurement of the relationship between crash incidence and possible risk factors (Pei *et al.,* 2012; Sze *et al.,* 2019). A common approach to estimating the vehicular crash exposure of a road segment is to calculate the traffic flow or vehicle kilometers traveled using traffic count data, which are often readily available from transport agencies. However, accurate and extensive pedestrian counts are seldom available. To estimate pedestrian crash exposure, indirect measures such as population and job opportunities may be adopted (Huang and Abdel-Aty, 2010). However, these proxies are correlated with other crash-explanatory factors such as land use and development density. This correlation may interfere with the estimation of crash risk and determination of associated risk factors.

In this study, we apply a joint probability modeling approach, which was proposed in an earlier study (Pei *et al.,* 2011), to analyze crash occurrence and pedestrian involvement simultaneously in one model. This approach is efficient at addressing the correlations between total and pedestrian crash counts, with a formulation that is less complicated than the multivariate approach. When the relationship between crash occurrence, land use, and road network characteristics is modeled, vehicular exposure based on traffic count data is incorporated into the modeling. When the crash involvement of pedestrians is estimated, walking trip frequency, based on comprehensive household travel survey data, is used as proxy for pedestrian exposure. Furthermore, other possible influencing factors, such as points of interest, public transport facilities, demographics, and socioeconomics, that characterize pedestrian accessibility and walking mode shares are also considered, and the similarities and differences in the effects of explanatory factors between total and pedestrian crash frequencies are investigated.

The remainder of the paper is organized as follows. Section 2 reviews literature on pedestrian crash analysis. Section 3 describes the data used in this study. Section 4 presents the formulation of the joint probability model. Section 5 shows the results of model estimation. Section 6 discusses the implications of the estimation results. Section 7 summarizes the findings and provides suggestions for future research.

2. LITERATURE REVIEW

Several studies have explored factors such as population characteristics, road network attributes, and built environments that affect pedestrian crashes at a macroscopic level. With regard to the effect of population profiles, demographic and socioeconomic characteristics such as population (Kim *et al.,* 2006; Loukaitou-Sideris *et al.,* 2007; Wier *et al.,* 2009; Siddiqui *et al.,* 2012a; Wang and Kockelman, 2013; Lee *et al.,* 2015b; Cai *et al.,* 2016; Wang *et al.,* 2017a), number of employed people (Kim *et al.,* 2006; Loukaitou-Sideris *et al.,* 2007; Wier *et al.,* 2009; Lee *et al.,* 2015b), proportions of children (Hummel, 1998; Siddiqui *et al.,* 2012a; Zegeer and Bushell, 2012) and elderly (Hummel, 1998; Noland and Quddus, 2004), household income (Chimba et al., 2018), and car ownership (Noland and Quddus, 2004; Kim *et al.,* 2006; Lee *et al.,* 2015b) have all been correlated with pedestrian crashes. However, the effects of pedestrian behaviors, i.e., reckless crossing, jaywalking, and red-light running violation, which can be amended by education, are rarely considered in these studies.

With regard to the effect of road network characteristics and traffic conditions, the increases in traffic volume and vehicle miles traveled are both correlated with the increase in total crash counts

(Loukaitou-Sideris *et al.,* 2007; Wier *et al.,* 2009; Lee *et al.,* 2015b). In particular, the increase in traffic volume on urban streets is correlated with the increase in pedestrian crash risk (Wang and Kockelman, 2013; Lee *et al.,* 2015a; Wang *et al.,* 2017a). Additionally, road density (road length per unit area) is positively correlated with crash rates (Cai *et al.,* 2017b; Wang *et al.,* 2017a). The number of intersections can also affect crash occurrence (Siddiqui *et al.,* 2012a; Bao *et al.,* 2017; Wang *et al.,* 2017a); the density of signal intersections is positively correlated with crashes involving motor vehicles, bicycles, and pedestrians (Lee *et al.,* 2015b). Additionally, environmental factors such as lighting conditions and the presence of sidewalks and crosswalks are correlated with both perceived and actual pedestrian injury risk (Dai, 2012; Rankavat and Tiwari, 2016). However, studies seldom consider the effect of pedestrian exposure when measuring the relationship between pedestrian crashes, road networks, and traffic characteristics. Even though crash occurrence may increase with vehicular traffic flow, safety in numbers can have a favorable effect on pedestrian safety because drivers who are exposed to large groups of pedestrians may drive more safely (Bhatia and Wier, 2011; Elvik, 2013, 2016; Osama and Sayed, 2017b; Xu *et al.,* 2019).

Walking is the primary access method to public transport services, such as rail and bus transits, especially when parking spaces near public transit stations are limited, and car ownership rate is low. Pedestrians also tend to gather in catchment areas around railway stations and bus stops (Osama and Sayed, 2017b). Hence, conflicts between vehicles and pedestrians tend to be frequent in these locations (Lee *et al.*, 2015a). With regard to the effect of public transport access, Yao et al. (2015) suggested that an increase in the number of light bus stops is correlated with an increase in total crash count. Nonetheless, even though the total crash risk may increase, risks of more severe injury could be reduced with increase in the number of bus stops, because drivers tend to drive more cautiously through bus stops (Wang and Kockelman, 2013). However, the effects of the presence of pedestrian facilities, such as pedestrian signal crosswalks, footbridges, and underpasses (especially those that are built with rail transit stations), on pedestrian crashes are rarely considered in these studies.

Lastly, land use characteristics can also affect crash occurrence (Kim *et al.,* 2006; Loukaitou-Sideris *et al.,* 2007; Wier *et al.,* 2009; Lee *et al.,* 2015b; Yao *et al.,* 2015; Zhang *et al.,* 2015; Wang *et al.,* 2016). Pedestrian crashes are more profound in commercial and residential areas and in areas with mixed land use (Loukaitou-Sideris *et al.,* 2007; Ding *et al.,* 2018). Pedestrian safety in school areas is of great concern (Clifton and Kreamer-Fults, 2007), and increases in the number and density of dwelling units are correlated with increase in pedestrian crashes (Zhang *et al.,* 2015). Although these findings can be attributed to increase in pedestrian activity in commercial and residential areas, these may also imply that such land uses are more hazardous to pedestrians for the same amount of travel (Wang *et al.,* 2016). Therefore, it is necessary to understand the role of pedestrian exposure in the relationship between pedestrian crashes, land use, and development density.

To properly account for the aforementioned possible correlations, multivariate approaches have been employed to simultaneously model the crash counts for different crash severity levels (Castro *et al.,* 2013; Wang and Kockelman, 2013; Bhat *et al.,* 2017; Ma *et al.,* 2017; Liu and Sharma, 2018), collision types (Cheng *et al.,* 2017; Wang *et al.,* 2017b), and transport modes (Lee *et al.,* 2015b; Huang *et al.,* 2017; Cheng *et al.,* 2018a). Additionally, joint model structures for the count data model and logit model have been proposed to simultaneously examine the factors that affect the total number of crashes and proportions of particular crash types, i.e. fatal and seriously injured crashes, and non-motorized crashes (Pei *et al.,* 2011; Cai *et al.*, 2017a). This study demonstrates the modeling of crashes involving different transport modes in one single structure, in which indicators of exposure that are relevant to pedestrian involvement in crashes (i.e., walking trips, traffic flow, and population) and occurrence of crashes (i.e., traffic flow and trip generation) are included, based on the data from a comprehensive household travel survey. Moreover, the effects of public transport facilities, points of interest, and land use, which may attract more pedestrians, are considered.

3. DATA

In this study, crash data are obtained from the Traffic Information System, which is maintained by the Hong Kong Transport Department and Hong Kong Police Force. Data on crash location, crash severity, vehicle type, and pedestrian involvement are available. To estimate pedestrian exposure, traffic flow data (average annual daily traffic of all major roads) are obtained from the Annual Traffic Census report, whereas trip data are obtained from the Travel Characteristic Survey (TCS) report. The traffic census was conducted at 1,649 stations and covered 88% of the roads in Hong Kong. Data on the origin, destination, time, and transport mode of every trip leg of the sampled trips are also available in the TCS report (Guo *et al.,* 2017; Sze *et al.,* 2019). Additionally, data on demographics and socioeconomics are obtained from the population census database, which is maintained by the Hong Kong Census and Statistics Department, whereas data on land use and zone boundaries are obtained from the database of the Hong Kong Planning Department.

For the proposed macro-level analysis, data on crash incidence, traffic flow, trips, land use, and population profile are mapped into corresponding geographical units (zones) using a geographical information system (i.e., ArcGIS). In this study, the observation unit is the Small Tertiary Planning Unit Group (STPUG) (Planning Department, 2011), of which there are a total of 209 in Hong Kong. However, because traffic count data are not available for some STPUGs, only 179 STPUGs are included in the subsequent analysis. The model is estimated using the crash, traffic flow, trip, and population profile data from 2011, in which the latest population census and household travel survey were conducted.

<Insert Table 1 here>

Table 1 summarizes the sample data. Of the 14,884 crashes in the 179 STPUGs included in this study, 24.2% (i.e., 3,604) involved pedestrians. According to the TCS report, 13.8 million trips on average were made per day, of which 1.2 million were walk-only trips.

With regard to the effects of demographics and socioeconomics, relevant factors including age, ethnicity, education, household income, and household size are considered. As summarized in Table 1, 14% of the population of a zone, on average, were of ages above 64. Similar to other developed societies, Hong Kong is facing a problem with aging population; the proportion of the population aged above 64 is expected to increase to 25% by 2035 (Sze and Christensen, 2017). Majority of the Hong Kong citizens surveyed were Chinese; approximately 12% of the population of a zone, on average, were non-Chinese. Meanwhile, "education" refers to the highest education level attained, and of the population surveyed, 73% had attained secondary education or above. The average household size was 2.9 persons, and approximately one-third of households had more than 3 members (Transport Department, 2012b).

In addition to the road density (length per unit area) and number of intersections, the effects of the numbers of metro exits and bus stops on pedestrian crashes are also investigated. In Hong Kong, 88% of trips were made by public transport, of which 30% were made by rail and 27% were made by bus. As shown in Table 1, each zone had an average of 3 metro exits and 44 bus stops. The number of conflicts between pedestrians and vehicles is considerable, especially in commercial and mixed-use areas.

With regard to land use and built environments, the numbers of restaurants, schools, shopping malls, and hotels, and the dominating land use (i.e., residential, industrial, commercial, and government & community) of a zone are considered. Tourism has been an important economic driver in Hong Kong. Points of interest, which include restaurants, shops, and hotels, can therefore serve as proxy for the travel demand generated by visitors, which is not captured in conventional household travel surveys. As shown in Table 1, each zone had an average of 162 restaurants, 8 shopping malls, and 8 hotels.

Prior to the model estimation, the multicollinearity of the candidate variables is assessed using a correlation test. Variables with variance inflation factor values less than ten would be considered for the subsequent analysis (Tay, 2017).

4. METHOD

To properly account for the correlation among total and pedestrian crash counts, a joint probability approach is proposed to simultaneously estimate the effects of possible influencing factors on crash occurrence and pedestrian involvement in crashes. However, whereas overall exposure to crashes can be estimated using traffic flow data, pedestrian count data are often not available, thus affecting the estimation of pedestrian crash exposure. Instead, certain variables such as walking frequency are used as proxies for pedestrian exposure. Other factors that affect pedestrian activity, including public transport access and points of interest, are also considered.

Let $p(y_i)$ denote the probability of having y_i total crashes and $p(y_i^p|y_i)$ denote the probability of having y_i^p pedestrian crashes conditional on y_i total crashes of unit i. The joint probability of y_i total crashes and y_i^p pedestrian crashes occurring can then be specified as

$$p(y_i^p, y_i) = p(y_i) \cdot p(y_i^p|y_i). \quad (1)$$

The formulations of probability functions $p(y_i)$ and $p(y_i^p|y_i)$ are provided in Subsections 4.1 and 4.2.

4.1 Probability function for total crashes

Generally, count data models such as Poisson regression and negative binomial regression models are applied to model crash frequency (Lord and Mannering, 2010). A basic assumption of the Poisson model is that the mean and variance of the crash count are equal. However, when the data are over-dispersed (i.e., variance is significantly greater than the mean), a negative binomial regression model should be applied instead.

In this study, the results of the over-dispersion test indicate that the variance of total crash count is greater than the mean, at a 1% level of significance. Therefore, a negative binomial regression model is established to estimate the total crash, with the probability function:

$$f(y_i;\mu_i,\alpha)=\frac{\Gamma(y_i+1/\alpha)}{(y_i+1)\Gamma(1/\alpha)}\left(\frac{1}{1+\alpha\mu_i}\right)^{1/\alpha}\left(1-\frac{1}{1+\alpha\mu_i}\right)^{y_i}, \quad (2)$$

Where α is the over-dispersion parameter, and μ_i is the expected number of total crashes of unit i.

The variance of pedestrian crash count is also greater than the mean. The relationship between possible risk factors and expected total crashes μ_i is therefore expressed by the following link function:

$$ln(\mu_i)=\beta_N^T\cdot x_i+\varepsilon_i, \quad (3)$$

where x_i is the vector of possible factors, β_N is the vector of corresponding coefficients, and ε_i is a gamma-distributed with mean equal to 1.

4.2 Probability function for pedestrian crashes

To evaluate the relationship between pedestrian involvement in crashes and possible influencing factors, a binomial approach is applied. The probability function of y_i^p pedestrian crashes conditional on y_i total crashes is then specified as

$$p(y_i^p|y_i)=\binom{y_i}{y_i^p}(\pi_i^p)^{y_i^p}(1-\pi_i^p)^{y_i-y_i^p}, \quad (4)$$

where π_i^p is the binomial probability of pedestrian involvement in the crashes.

The relationship between this probability and possible influencing factors can be measured using a logit function, as follows:

$$log\,i\,t(\pi_i^p) = log(\frac{\pi_i^p}{1-\pi_i^p}) = \beta_B^T \cdot x_i, \quad (5)$$

where x_i denotes the vector of explanatory factors, and β_B denotes the corresponding coefficients that reflect their effects on pedestrian involvement in the crashes.

4.3 Model estimation

Based on the formulations provided in Subsections 4.1 and 4.2, the joint probability function of y_i^p pedestrian crashes and y_i total crashes can be established via substitution of Eq. (2) and Eq. (4) into Eq. (1), resulting in

$$p(y_i^p, y_i) = \frac{\Gamma(y_i+1/\alpha)}{(y_i+1)\Gamma(1/\alpha)}(\frac{1}{1+\alpha\mu_i})^{1/\alpha}(1-\frac{1}{1+\alpha\mu_i})^{y_i} \cdot \binom{y_i}{y_i^p}(\pi_i^p)^{y_i^p}(1-\pi_i^p)^{y_i-y_i^p}, \quad (6)$$

A full Bayesian approach is used to model the joint probability of total crashes and pedestrian crashes specified in Eq. (6). In conventional maximum likelihood estimation, point estimates of β_N and β_B are applied. By contrast, the Bayesian approach is superior because it can provide a posterior distribution (probability density function) of the parameters using simulation-based estimation (Christensen *et al.,* 2011; Gelman *et al.,* 2013).

Given that $\hat{y}$ denotes the observed crash counts and θ denotes the parameter estimate, the posterior probability of parameters, based on the Bayes' rule, can be calculated as

$$p(\theta|\hat{y}) = \frac{p(\theta,\hat{y})}{p(\hat{y})} = \frac{p(\theta)p(\hat{y}|\theta)}{p(\hat{y})}, \quad (7)$$

Where $p(\theta,\hat{y})$ is the probability that the observed outcome is $\hat{y}$ and the parameter estimate is θ. $p(\hat{y})$ is of an observed event and thus serves as a normalization constant.

Therefore, the posterior probability is linearly proportional to the prior distribution of parameters $p(\theta)$ multiplied by the likelihood function of observed outcome $p(\hat{y}|\theta)$:

$$p(\theta|\hat{y}) \propto p(\theta) \cdot p(\hat{y}|\theta). \tag{8}$$

The calculation of the prior distribution of parameters $p(\theta)$, which are unknown, is critical for the Bayesian model. When Eq. (6) is substituted into Eq. (8), the posterior probability of parameters can be expressed as

$$p(\beta|y_i^p, y_i) \propto p(\hat{\beta}) \cdot \frac{\Gamma(y_i+1/\alpha)}{(y_i+1)\Gamma(1/\alpha)} (\frac{1}{1+\alpha\mu_i})^{1/\alpha} (1-\frac{1}{1+\alpha\mu_i})^{y_i} \cdot \binom{y_i}{y_i^p} (\pi_i^p)^{y_i^p} (1-\pi_i^p)^{y_i-y_i^p}, \tag{9}$$

where β is the stack column vector of parameters β_N and β_B, which are ultimately estimated. Furthermore, $p(\hat{\beta})$ is the prior probability of β.

The Markov chain Monto Carlo (MCMC) simulation approach is applied for the estimation of the parameters in the Bayesian model. For instance, Gibbs sampling and Metropolis–Hastings algorithm approaches are adopted to draw a sample from the prior distribution of unknown parameters in an iterative process (Gelman *et al.*, 2013).

The non-informative prior distribution of parameters is set to $\theta \sim Normal(0,1000^2)$. There would be 200,000 iterations for the MCMC simulation and 1,000 additional iterations in the burn-in stage. When the Monte Carlo error is less than 0.05, the distribution of parameter estimates is considered to be converged. Additionally, Gelman–Rubin diagnostics are applied in the assessment of model convergence (Gelman and Rubin, 1992). Thus, the posterior distribution of parameters can be determined (Sinharay, 2003; Barua *et al.*, 2016).

4.4 Goodness-of-fit assessment

The deviance of information criterion (DIC) is employed for the goodness-of-fit assessment of the model (Spiegelhalter *et al.* 2002). DIC is calculated as

$$DIC = 2\overline{D(\theta)} - D(\bar{\theta}), \tag{10}$$

where $\overline{D(\theta)}$ is the mean of posterior deviance $D(\theta)$, and $D(\bar{\theta})$ is the deviance at the mean of the posterior parameters.

Deviance $D(\theta)$ of the model at the parameter θ is determined as

$$D(\theta) = -2\, log(\, P(\hat{y}|\theta)). \tag{11}$$

Moreover, the pseudo ρ^2 is used to assess the goodness of fit of the proposed joint probability model and conventional negative binomial models.

5. RESULTS

For comparison with and evaluation of the proposed joint probability model, estimations are also performed with traditional negative binomial models for total and pedestrian crashes. The results for the negative binomial models are outlined in **Table 2**.

<Insert Table 2 here>

As summarized in Table 2, several factors including trip generation, traffic volume, ethnicity, and numbers of non-signalized intersections, restaurants, and hotels are significantly correlated with the total number of crashes. Furthermore, total population, traffic volume, walking frequency, education level, road density, and number of metro exits, and restaurants are correlated with pedestrian crashes.

The proposed joint probability model (pseudo ρ^2 of 0.29), as shown in Table 3, outperforms the negative binomial model for pedestrian crashes (0.23). Additionally, the joint probability model, as compared to the negative binomial model, involves more significant variables contributing to pedestrian involvement in crashes.

<Insert Table 3 here>

Table 3 presents the results of the joint probability model for crash occurrence and pedestrian involvement in crashes using the MCMC approach. For this model, the mean, standard deviation, and 95% Bayesian credible intervals (BCIs) are estimated. Pedestrian crash exposure, population demographics and socioeconomics, road network and transport facilities, and land use characteristics that contribute to total crash occurrence and pedestrian involvement in crashes are identified. For instance, as summarized in Table 3, crash occurrence is positively correlated with traffic volume [mean 0.16; BCI (0.09, 0.23)], trip generation [mean 0.16; BCI (0.10, 0.22)], number of non-signalized intersection [mean 0.004; BCI (0.003, 0.005)], number of restaurants [mean 0.001; BCI (0.0005, 0.002)], and number of hotels [mean 0.01; BCI (0.0.003, 0.01)], at 5% level of significance. By contrast, the proportion of non-Chinese population [mean −0.02; BCI (−0.03, −0.01)] is negatively correlated with crash occurrence at 5% level of significance.

Furthermore, pedestrian involvement in crashes is positively correlated with population [mean 0.24; BCI (0.18, 0.27)], walking frequency [mean 0.29; BCI (0.23, 0.35)], proportions of population of ages below 15 [mean 0.03; BCI (0.01, 0.05)] and of ages above 64 [mean 0.01; BCI (0.0002, 0.02)], non-Chinese population [mean 0.01; BCI (0.01, 0.02)], road density [mean 0.01; BCI (0.005, 0.01)], number of bus stops [mean 0.002; BCI (0.0006, 0.005)], number of restaurants [mean 0.0009; BCI (0.0006, 0.001)], and residential land use [mean 0.19; BCI (0.09, 0.29)], at 5% level of significance. In contrast, pedestrian involvement is negatively correlated with proportion of people with primary education [mean −0.02; BCI (−0.04, −0.005)], number of households of sizes greater than 3 [mean −0.02; BCI (−0.02, −0.01)], number of non-signalized intersections [mean −0.003; BCI (−0.004, −0.003)], number of metro exits [mean −0.03; BCI (−0.05, −0.02)], and number of schools [mean −0.01; BCI (−0.01, −0.001)], at 5% level of significance.

6. DISCUSSION

In this paper, a joint probability model is proposed to simultaneously model crash occurrence and pedestrian involvement in crashes. **Table 4** summarizes the key factors contributing to total and pedestrian crashes.

<Insert Table 4 here>

Interpretations and policy implications of the results regarding pedestrian exposure, population demographics and socioeconomics, road network characteristics, and land use are provided in the following subsections.

6.1. Exposure

As outlined in Table 4, traffic volume and trip generation were determined to be positively correlated with crash occurrence, which then justified the use of traffic volume and trip generation in the estimation of pedestrian exposure across different zones and their variations in a spatial context (Naderan and Shahi, 2010; Siddiqui *et al.*, 2012b; Haleem *et al.*, 2015; Bao *et al.*, 2017; Wang *et al.*, 2017a). In addition, the parameters of the logarithmically transformed traffic flow and trip generation are both significantly less than one, at 5% level, implying a non-linear relationship between total crash and pedestrian exposure. Meanwhile, the marginal crash rate declined when traffic volume was high, which may be due to reductions in traffic speed and associated crash risk attributed to increase in traffic density (Qin *et al.*, 2004; Pei *et al.*, 2012).

Walking frequency and population were positively correlated with pedestrian crashes, suggesting that pedestrian safety was sensitive to the residential population and pedestrian activity in the area (Wang *et al.,* 2016; Wang *et al.,* 2017a; Zeng *et al.,* 2017). The parameter of logarithmically transformed walking frequency is also significantly less than one. Relative crash risk was lower when pedestrian activity was high, which could be attributed to the "safety-in-numbers" effect (Xu *et al.,* 2019). This is indicative of the necessity to properly design pedestrian facilities. For instance, it is necessary to improve built environments such that pedestrian access to local activities and services is enhanced (Wier *et al.,* 2009; Sze and Christensen, 2017; Chen et al., 2020b), especially in residential areas. However, pedestrian counts for individual road facilities, including footpaths, intersections, and crosswalks, are not available, which affected the estimation of pedestrian–vehicle conflict. The relationship between the number of pedestrians and crash exposure for different sites would definitely be worth exploring when comprehensive information on pedestrian behaviors becomes available in the future.

6.2. Demographics and socioeconomics

Effects of population age, race, education, and household income on total and pedestrian crashes were also investigated. As shown in Table 4, no evidence could be established for the association between crash occurrence and demographics and socioeconomic characteristics (except for ethnicity), which could be due to vehicle crash risk being less sensitive to the characteristics of residents in an area. Indeed, road safety level as a whole should be determined by the behaviors of drivers and commuters, who may be residing somewhere else, traveling in the area. In contrast, the pedestrian crash risk was correlated with the demographic and socioeconomic characteristics of the residents. For instance, proportions of the young population (ages below 15 years), elderly (ages above 64), and non-Chinese were positively correlated with pedestrian crashes, which could be due to the tendency of young and elderly people to walk more frequently (Miranda-Moreno *et al.,* 2011; Bhat *et al.,* 2017). More importantly, it could be attributed to a higher propensity of reckless crossing behaviors among young people, impaired cognitive and physical performance among the elderly (Hummel, 1998; Noland and Quddus, 2004; Zegeer and Bushell, 2012), and differences in social norms among non-local residents (Cottrill and Thakuriah, 2010; Lee *et al.,* 2014; Coughenour *et al.,* 2017). Furthermore, increase in the proportions of the population that attained, at maximum, a primary education and lived in bigger households were correlated with a reduction in pedestrian crash risk. Hence, the effects of education level and household attributes on the perception, attitude, and behaviors of pedestrians, and thus on pedestrian crash risk, would be worth investigating. For instance, propensities for jaywalking and red-light running among pedestrians with respect to possible explanatory factors could have been revealed in an attitudinal survey (Li *et al.,* 2014; Chen *et al.,* 2020a) and empricial observation (Zhu *et al.*, 2020).

6.3. Road network and transport characteristics

With regard to road characteristics, even though no evidence was established for a relationship between road density and crash occurrence, the number of pedestrian crashes increased when the road density was high, which could have been due to an inadequate separation between pedestrian and vehicular traffic (i.e., wider traffic lanes, segregated footpaths, and presence of green zone on the footpath; Transport Department, 2012b). In particular, major arterials, primary distributors, and roads with higher speed limits tended to have higher pedestrian crash risks (Aguero-Valverde and Jovanis, 2006; Wang *et al.,* 2016; Cai *et al.,* 2017b).

As revealed in this study, the number of non-signalized intersections was positively correlated with the total number of crashes. Although yield and stop controls were generally applied at intersections with low traffic volume, other factors such as inattentiveness, traffic sign violation, and speeding behaviors of drivers could increase the risk of traffic conflict and its associated crash risk (Abdel-Aty *et al.,* 2005). However, non-signalized intersections were negatively correlated with pedestrian crashes, which could have been due to a tendency among pedestrians to be more cautious in the absence of signal control (Sze and Wong, 2007)

With regard to the intensity of public transport facilities, pedestrian crash risk was determined to be sensitive to the numbers of bus stops and metro exits. This result could have been expected because walking is the primary mode of access to public transport services (Besser and Dannenberg, 2005; Kim *et al.,* 2010; Miranda-Moreno *et al.,* 2011; Lee *et al.,* 2015a; Dai and Jaworski, 2016). For instance, pedestrian crash risk was reduced by 3.0% for every additional metro exit in an area, which could be attributed to the provision of segregated pedestrian crosswalks, including footbridges and underpasses connecting metro stations. In particular, escalators and movable walkways provided at many metro stations in Hong Kong could have improved the accessibility and safety of pedestrians (Sze and Wong, 2007; Wong *et al.,* 2007; Sze and Christensen, 2017). However, each additional bus stop was associated with a 0.2% increase in pedestrian crash risk, which could be attributed to an increase in roadside activities and reckless crossing behaviors, and thus possible increase in vehicle–pedestrian conflicts near bus stops (Kim *et al.,* 2010; Miranda-Moreno *et al.,* 2011; Wang and Kockelman, 2013; Lee *et al.,* 2015b; Yao *et al.,* 2015; Rhee *et al.* 2016). It is therefore necessary to improve the design and planning for more accessible routes, footpaths, passenger waiting areas, drop-off and pick-up areas, and protection, such as barriers and roadside curbs, at bus stops, as suggested by the Independent Review Committee on Hong Kong's Franchised Bus Service (Lunn *et al.,* 2018). With these developments, safe and efficient access for bus passengers can be achieved (Sze and Christensen, 2017).

6.4 Land use and trip attraction attributes

In conventional studies, increase in vehicle and pedestrian crash rates were determined to be correlated with commercial and mixed land use (Wong *et al.,* 2007). However, as revealed in this study, pedestrian crash risk in residential areas was 20% higher than that in the non-residential areas. For low-activity areas such as residential development areas, high pedestrian crash risk may have been attributed to reckless crossing behaviors and inattentiveness among pedestrians (Loukaitou-Sideris *et al.,* 2007; Wier *et al.,* 2009). Therefore, reduced speed limits, traffic calming, and local area traffic management may be effective at enhancing pedestrian safety in residential areas. However, it remains necessary to explore the relationship between built environments, pedestrian activity, and associated crash risk when information on pedestrian behaviors becomes available in future surveys (Merlin *et al.*, 2020).

With regard to the points of interest and trip attraction attributes, increases in the number of restaurants were associated with increases in both total and pedestrian crashes. Each additional restaurant was correlated with a 0.09–0.1% increase in the crash risk. As previously mentioned, commercial areas were more dangerous to both pedestrians and vehicles. For example, restaurants are key attractions to residents, commuters, and visitors (Siddiqui *et al.,* 2012a; Abdel-Aty *et al.,* 2013; Bao *et al.*, 2017). The modifying effect of travel purpose on the relationship between activity level and pedestrian and vehicle crash risk would therefore be worth investigating in future studies (Sze *et al.,* 2019). Additionally, frequent drop-off and pick-up activities can increase the potential of vehicle–vehicle conflicts near hotels. Therefore, the number of hotels was positively correlated with the total crash risk, which could have been increased by 1% for each additional hotel in an area (Wier *et al.*, 2009; Lee *et al.*, 2015a; Lee *et al.*, 2015b). Furthermore, pedestrian crash was determined to be negatively correlated with the number of schools, which could be attributed to better traffic control (e.g., local area traffic management, traffic calming) in school areas and road safety education, which enhanced safety awareness among drivers and students (Ng *et al.,* 2002). Nonetheless, improving the design of built environments and traffic control remains necessary to protect other vulnerable and disadvantaged pedestrians, including the elderly and individuals with disabilities (Sze and Christensen, 2017).

7. CONCLUSION

For conventional crash prediction models, extensive count data are rarely available for the estimation of pedestrian exposure. Moreover, possible correlations exist among crashes of different types (i.e., pedestrians, motor vehicles), which should be considered during the development of separate crash prediction models. Therefore, a joint probability approach is proposed to simultaneously estimate total and pedestrian crashes. A full Bayesian method using an MCMC approach is adopted to investigate the effects of explanatory factors on crash occurrence and pedestrian involvement in crashes using a single model. Population and walking activity are used as proxies for the exposure of pedestrian involvement in crashes. Moreover, the effects of demographics and socioeconomics, road network and facilities, access to public transport services, land use, and trip attraction attributes are considered.

The proposed joint probability model outperforms separate negative binomial models for total and pedestrian crashes. The results of parameter estimation of crash occurrence and pedestrian involvement are also determined to be consistent with those of previous studies and provide useful recommendations (Lee et al., 2015a; 2015b). For instance, traffic flow, trip generation, road characteristics, and points of interest are correlated with crash occurrence. The pedestrian involvement in crashes is correlated with the demographic and socioeconomic characteristics of residents, access to public transport, presence of schools, and residential land use. Results are indicative of the necessity for proper design and planning of various transport facilities, including footpaths, intersections, pedestrian crosswalks, and bus stops. Moreover, the needs of vulnerable groups (i.e., children, adolescent, elderly, and ethnical minorities) and accessibility to essential urban services and attractions (i.e., hotel, government, community and housing development) should be addressed (Sze and Christensen, 2017). With these proposed improvements, safe and accessible walking environment in built environments can be promoted. Additionally, road safety education can be designed and marketed toward vulnerable pedestrian groups to enhance their safety awareness and deter against reckless crossing behaviors (Zhu et al., 2020). Nonetheless, despite the insights devised in this study, pedestrian safety and travel behaviors with respect to walking distance, time, and path should be investigated when comprehensive walking trip distribution data become available in the future. The effects of development intensity on pedestrian safety should also be analyzed. With this knowledge, sustained improvements on pedestrian safety can be achieved in the long run.

ACKNOWLEDGEMENT

This study was supported by grants from the Research Grants Council of Hong Kong (15209818) and Research Committee of the Hong Kong Polytechnic University (1-ZE5V).

Reference

Abdel-Aty, M., Keller, J., Brady, P.A., 2005. Analysis of types of crashes at signalized intersections by using complete crash data and tree-based regression. Transportation Research Record 1908 (1), 37-45.

Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. Transportation Research Part A: Policy and Practice 49, 62-75.

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in pennsylvania. Accident Analysis & Prevention 38 (3), 618-625.

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.

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2016. Macroscopic modeling of pedestrian and bicycle crashes: A cross-comparison of estimation methods. Accident Analysis & Prevention 93, 147-159.

Bao, J., Liu, P., Yu, H., Xu, C., 2017. Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas. Accident Analysis & Prevention 106, 358-369.

Barua, S., El-Basyouny, K., Islam, M.T., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. Analytic methods in accident research 9, 1-15.

Besser, L.M., Dannenberg, A.L., 2005. Walking to public transit: Steps to help meet physical activity recommendations. American journal of preventive medicine 29 (4), 273-280.

Bhat, C.R., Astroza, S., Lavieri, P.S., 2017. A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. Analytic methods in accident research 16, 1-22.

Bhatia, R., Wier, M., 2011. "Safety in numbers" re-examined: Can we make valid or practical inferences from available evidence? Accident Analysis & Prevention 43 (1), 235-240.

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

Cai, Q., Abdel-Aty, M., Lee, J., Eluru, N., 2017b. Comparative analysis of zonal systems for macro-level crash modeling. Journal of safety research 61, 157-166.

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 & Prevention 93, 14-22.

Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. Transportation research part B: methodological 46 (1), 253-272.

Castro, M., Paleti, R., Bhat, C.R., 2013. A spatial generalized ordered response model to examine highway crash injury severity. Accident Analysis & Prevention 52, 188-203.

Chen, T., Sze, N.N., Saxena, S., Pinjari, A.R., Bhat, C.R., Bai, L., 2020a. Evaluation of penalty and enforcement strategies to combat speeding offences among professional drivers: A hong kong stated preference experiment. Accident Analysis & Prevention 135, 105366.

Chen, T., Sze, N.N., Chen, S., Labi, S, 2020b. Urban road space allocation incorporating the safety and construction cost impacts of lane and footpath widths. Journal of Safety Research, in press.

Cheng, W., Gill, G.S., Dasu, R., Xie, M., Jia, X., Zhou, J., 2017. Comparison of multivariate poisson lognormal spatial and temporal crash models to identify hot spots of intersections based on crash types. Accident Analysis & Prevention 99, 330-341.

Cheng, W., Gill, G.S., Ensch, J.L., Kwong, J., Jia, X., 2018a. Multimodal crash frequency modeling:

Multivariate space-time models with alternate spatiotemporal interactions. Accident Analysis & Prevention 113, 159-170.
Cheng, W., Gill, G.S., Zhang, Y., Cao, Z., 2018b. Bayesian spatiotemporal crash frequency models with mixture components for space-time interactions. Accident Analysis & Prevention 112, 84-93.
Christensen, R., Johnson, W., Branscum, A., Hanson, T.E., 2011. Bayesian ideas and data analysis: An introduction for scientists and statisticians CRC Press.
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.
Clifton, K.J., Kreamer-Fults, K., 2007. An examination of the environmental attributes associated with pedestrian–vehicular crashes near public schools. Accident Analysis & Prevention 39 (4), 708-715.
Cottrill, C.D., Thakuriah, P.V., 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. Accident Analysis & Prevention 42 (6), 1718-1728.
Coughenour, C., Clark, S., Singh, A., Claw, E., Abelar, J., Huebner, J., 2017. Examining racial bias as a potential factor in pedestrian crashes. Accident Analysis & Prevention 98, 96-100.
Dai, D., 2012. Identifying clusters and risk factors of injuries in pedestrian–vehicle crashes in a gis environment. Journal of Transport Geography 24, 206-214.
Dai, D., Jaworski, D., 2016. Influence of built environment on pedestrian crashes: A network-based gis analysis. Applied geography 73, 53-61.
Ding, C., Chen, P., Jiao, J., 2018. Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: A machine learning approach. Accident Analysis & Prevention 112, 116-126.
Elvik, R., 2013. Can a safety-in-numbers effect and a hazard-in-numbers effect co-exist in the same data? Accident Analysis & Prevention 60, 57-63.
Elvik, R., 2016. Safety-in-numbers: Estimates based on a sample of pedestrian crossings in norway. Accident Analysis & Prevention 91, 175-182.
Ferrer, S., Ruiz, T., 2018. The impact of the built environment on the decision to walk for short trips: Evidence from two spanish cities. Transport policy 67, 111-120.
Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian data analysis Chapman and Hall/CRC.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical science 7 (4), 457-472.
Guo, Q., Xu, P., Pei, X., Wong, S.C., Yao, D., 2017. The effect of road network patterns on pedestrian safety: A zone-based bayesian spatial modeling approach. Accident Analysis & Prevention 99, 114-124.
Guo, Z., Loo, B.P., 2013. Pedestrian environment and route choice: Evidence from new york city and hong kong. Journal of transport geography 28, 124-136.
Haleem, K., Alluri, P., Gan, A., 2015. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. Accident Analysis & Prevention 81, 14-23.
Huang, H., Abdel-Aty, M., 2010. Multilevel data and bayesian analysis in traffic safety. Accident Analysis & Prevention 42 (6), 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.
Hummel, T., 1998. Dutch pedestrian research review. A review of the main traffic safety research on pedestrians in the netherlands, 1984-1997.

Kamargianni, M., Dubey, S., Polydoropoulou, A., Bhat, C., 2015. Investigating the subjective and objective factors influencing teenagers' school travel mode choice–an integrated choice and latent variable model. Transportation Research Part A: Policy and Practice 78, 473-488.

Kim, K., Brunner, I.M., Yamashita, E.Y., 2006. Influence of land use, population, employment, and economic activity on accidents. Transportation research record 1953 (1), 56-64.

Kim, K., Punt, P., Yamashita, E., 2010. Measuring influences of demographic and land use variables in honolulu, hawaii. Transportation Research Record 2147, 9-17.

Lam, W.W., Yao, S., Loo, B.P., 2014. Pedestrian exposure measures: A time-space framework. Travel behaviour and society 1 (1), 22-30.

Lee, J., Abdel-Aty, M., Choi, K., Huang, H., 2015a. Multi-level hot zone identification for pedestrian safety. Accident Analysis & Prevention 76, 64-73.

Lee, J., Abdel-Aty, M., Jiang, X., 2014. Development of zone system for macro-level traffic safety analysis. Journal of transport geography 38, 13-21.

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

Li, Y.C., Sze, N.N., Wong, S.C., 2014. Effective measures for combating drink-driving offenses: An attitudinal model for hong kong. Transportmetrica A: transport science 10 (8), 722-739.

Lindelöw, D., Svensson, Å., Sternudd, C., Johansson, M., 2014. What limits the pedestrian? Exploring perceptions of walking in the built environment and in the context of every-day life. Journal of Transport & Health 1 (4), 223-231.

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.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation research part A: policy and practice 44 (5), 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 (3), 338-351.

Lunn, M., Auyeung, P.K., Lo, H.K., 2018. Report of the independent review committee on hong kong's franchished bus service. https://www.irc-bus.gov.hk/pdf/IRC_Report_Eng_Redacted.pdf.

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.

Merlin, L.A., Guerra, E., Dumbaugh, E., 2020. Crash risk, crash exposure, and the built environment: A conceptual review. Accident Analysis & Prevention 134, 105244.

Miles-Doan, R., Thompson, G., 1999. The planning profession and pedestrian safety: Lessons from orlando. Journal of Planning Education and Research 18 (3), 211-220.

Miranda-Moreno, L.F., Morency, P., El-Geneidy, A.M., 2011. The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections. Accident Analysis & Prevention 43 (5), 1624-1634.

Naderan, A., Shahi, J., 2010. Aggregate crash prediction models: Introducing crash generation concept. Accident Analysis & Prevention 42 (1), 339-346.

Ng, K.S., Hung, W.T., Wong, W.G., 2002. An algorithm for assessing the risk of traffic accident. Journal of Safety Research 33 (3), 387-410.

Noland, R., Quddus, M., 2004. Analysis of pedestrian and bicycle casualties with regional panel data.

Transportation Research Record: Journal of the Transportation Research Board (1897), 28-33.
Osama, A., Sayed, T., 2017a. Evaluating the impact of connectivity, continuity, and topography of sidewalk network on pedestrian safety. Accident Analysis & Prevention 107, 117-125.
Osama, A., Sayed, T., 2017b. Macro-spatial approach for evaluating the impact of socio-economics, land use, built environment, and road facility on pedestrian safety. Canadian Journal of Civil Engineering 44 (12), 1036-1044.
Pei, X., Wong, S.C., Sze, N.N., 2011. A joint-probability approach to crash prediction models. Accident Analysis and Prevention 43 (3), 1160-1166.
Pei, X., Wong, S.C., Sze, N.N., 2012. The roles of exposure and speed in road safety analysis. Accident Analysis and Prevention 48, 464-471.
Planning Department, H.K.S.a.R., 2011. Tertiary planning unit street block of hong kong. https://www.pland.gov.hk/pland_en/access/pec/1.html.
Qin, X., Ivan, J.N., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis & Prevention 36 (2), 183-191.
Rankavat, S., Tiwari, G., 2016. Pedestrians risk perception of traffic crash and built environment features–delhi, india. Safety science 87, 1-7.
Rhee, K.A., Kim, J.K., Lee, Y.I., Ulfarsson, G.F., 2016. Spatial regression analysis of traffic crashes in seoul. Accident Analysis and Prevention 91, 190-199.
Siddiqui, C., Abdel-Aty, M., Choi, K., 2012a. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accident Analysis & Prevention 45, 382-391.
Siddiqui, C., Abdel-Aty, M., Huang, H., 2012b. Aggregate nonparametric safety analysis of traffic zones. Accid Anal Prev 45, 317-25.
Sinharay, S., 2003. Assessing convergence of the markov chain monte carlo algorithms: A review. ETS Research Report Series 2003 (1), i-52.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the royal statistical society: Series b (statistical methodology) 64 (4), 583-639.
Sze, N.N., Christensen, K.M., 2017. Access to urban transportation system for individuals with disabilities. IATSS research 41 (2), 66-73.
Sze, N.N., Su, J., Bai, L., 2019. Exposure to pedestrian crash based on household survey data: Effect of trip purpose. Accident Analysis & Prevention 128, 17-24.
Sze, N.N., Wong, S.C., 2007. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. Accident Analysis and Prevention 39 (6), 1267-1278.
Tay, R., 2017. Correlation, variance inflation and multicollinearity in regression model. Journal of the Eastern Asia Society for Transportation Studies 12, 2006-2015.
Transport Department, H.K.S.a.R., 2012a. The annual traffic census. https://www.td.gov.hk/en/publications_and_press_releases/publications/free_publications/the_annual_traffic_census_2011/index.html.
Transport Department, H.K.S.a.R., 2012b. Travel characteristics survey 2011 (final report). https://www.td.gov.hk/en/publications_and_press_releases/publications/free_publications/index.html.
Transport Department, H.K.S.a.R., 2015. Road traffic casualties by class of road user and degree of injury. https://www.td.gov.hk/en/road_safety/road_traffic_accident_statistics/2018/index.html.
Wang, J., Huang, H., Zeng, Q., 2017a. The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian. Accident Analysis & Prevention 98, 223-231.

Wang, K., Ivan, J.N., Ravishanker, N., Jackson, E., 2017b. Multivariate poisson lognormal modeling of crashes by type and severity on rural two lane highways. Accident Analysis & Prevention 99, 6-19.

Wang, X., Yang, J., Lee, C., Ji, Z., You, S., 2016. Macro-level safety analysis of pedestrian crashes in shanghai, china. Accid Anal Prev 96, 12-21.

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.

Weinstein Agrawal, A., Schlossberg, M., Irvin, K., 2008. How far, by which route and why? A spatial analysis of pedestrian preference. Journal of urban design 13 (1), 81-98.

Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accident Analysis & Prevention 41 (1), 137-145.

Wong, S.C., Sze, N.-N., Li, Y.-C., 2007. Contributory factors to traffic crashes at signalized intersections in hong kong. Accident Analysis & Prevention 39 (6), 1107-1113.

Woo, A., Yu, C.Y., Lee, S., 2019. Neighborhood walkability for subsidized households: Revisiting neighborhood environments of housing choice voucher and low-income housing tax credit households. Cities 89, 243-251.

Xu, P., Xie, S., Dong, N., Wong, S.C., Huang, H., 2019. Rethinking safety in numbers: Are intersections with more crossing pedestrians really safer? Injury prevention 25 (1), 20-25.

Yao, S.J., Loo, B.P.Y., Lam, W.W.Y., 2015. Measures of activity-based pedestrian exposure to the risk of vehicle-pedestrian collisions: Space-time path vs. Potential path tree methods. Accident Analysis and Prevention 75, 320-332.

Zegeer, C.V., Bushell, M., 2012. Pedestrian crash trends and potential countermeasures from around the world. Accident Analysis & Prevention 44 (1), 3-11.

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

Zhang, Y., Bigham, J., Ragland, D., Chen, X., 2015. Investigating the associations between road network structure and non-motorist accidents. Journal of transport geography 42, 34-47.

Zhu, D., Sze, N.N., Bai, L., 2020. Roles of personal and environmental factors in the red light running propensity of pedestrian: Case study at the urban crosswalks. Transportation Research Part F, in press.

TABLES

Table 1. Descriptive statistics of sample (*Number of STPUG = 179*)

Variable		**Mean**	**Std.**	**Min.**	**Max.**
Zone area (km^2)		4.90	9.77	0.06	71.23
Crash and exposure					
Total crashes		83.15	71.90	2	483
Pedestrian crashes		20.12	22.54	0	134
Population		38,066.76	42,814.64	1,023	287,901
Average annual daily traffic		31,690.71	27,909.59	790	151,840
Trip generation (per day)		74,762.22	74,108.04	289	465,645
Walking frequency (per day)		154,881.50	159,058.40	973	925,433
Demographic and socioeconomic characteristics					
Age	Proportion of age below 15	0.12	0.03	0.04	0.21
	Proportion of age above 64	0.14	0.05	0.03	0.44
	Proportion of other age group	0.74	0.05	0.49	0.92
Ethnicity	Proportion of Chinese	0.88	0.12	0.43	0.99
	Proportion of non-Chinese	0.12	0.12	0.01	0.57
Education	Proportion attained primary education	0.17	0.05	0.01	0.28
	Proportion attained secondary education	0.45	0.07	0.17	0.66
	Proportion attained tertiary education	0.28	0.12	0.03	0.75
Household income	Proportion below HK$10,000 per month	0.22	0.10	0.01	0.52
	Proportion of HK$10,000-39,999 per month	0.46	0.15	0.03	0.71
	Proportion of ≥ HK$40,000 per month	0.32	0.22	0.00	0.96
Household size	Proportion of ≤ 3 members	0.66	0.10	0.35	0.99
	Proportion of more than 3 members	0.34	0.10	0.01	0.65
Road network and transport facilities					
Road density (km/km^2)		17.60	15.77	0.17	90.43
Number of non-signalized intersections		110.69	97.18	7	636
Number of signalized intersections		8.15	8.20	0	41
Number of metro exits		2.68	3.76	0	18
Number of bus stops		44.39	35.79	1	187
Land use and points of interest					
Number of restaurants		161.65	195.17	2	1,137
Number of schools		19.36	17.53	0	95
Number of shopping malls		7.55	9.51	0	55
Number of hotels		7.78	19.63	0	164
Residential land use (Yes = 1)		0.34	0.47	0	1

Industrial land use (Yes = 1)	0.03	0.17	0	1
Commercial land use (Yes = 1)	0.02	0.15	0	1
Government, institutional & community (Yes = 1)	0.04	0.21	0	1
Other land use (Yes = 1)	0.57	0.50	0	1

Table 2: Estimation results of negative binomial regression models

Variable		Total crash			Pedestrian crash		
		Coef.	**Std.**	**95% CI**	**Coef.**	**Std.**	**95% CI**
Constant		IS			-4.37	0.89	(-6.11, -2.63)
ln (Total population)		IS			0.18	0.06	(0.06, 0.29)
ln (Annual average daily traffic)		0.15	0.04	(0.07, 0.23)	0.10	0.04	(0.03, 0.18)
ln (Trip generation)		0.20	0.06	(0.08, 0.33)	N/A		
ln (walking frequency)		N/A			0.44	0.06	(0.32, 0.57)
Demographic and socioeconomic characteristics							
Age	Age below 15	IS			IS		
	Age over 64	IS			IS		
Ethnicity	Non-Chinese	-0.02	0.005	(-0.03, -0.01)	IS		
Education level	Primary	IS			-0.05	0.03	(-0.08, -0.02)
	Tertiary	IS			-0.02	0.01	(-0.04, -0.004)
Household size	More than 3 members	IS			IS		
Road network and transport facilities							
Road density		IS			0.01	0.003	(0.01, 0.02)
Number of non-signalized intersections		0.003	0.0007	(0.002, 0.005)	IS		
Number of metro exits		IS			-0.03	0.01	(-0.05, -0.01)
Number of bus stops		IS			IS		
Land use and points of interest							
Number of restaurants		0.001	0.0003	(0.0003, 0.002)	0.002	0.0003	(0.001, 0.002)
Number of schools		IS			IS		
Number of hotels		0.01	0.002	(0.002, 0.01)	IS		
Residential land use		IS			IS		
Over-dispersion		0.17	0.02	(0.14, 0.22)	0.09	0.02	(0.06, 0.13)
Log likelihood		-843.43			-552.60		
Pseudo ρ^2		0.12			0.23		

IS: Statistically insignificant; N/A: Not applicable

Table 3. Estimation results of Bayesian joint probability model

Variable		**Total crash**			**Pedestrian involvement**		
		Coef.	**Std.**	**95% BCI**	**Coef.**	**Std.**	**95% BCI**
Constant		IS			-6.78	0.73	(-7.80, -5.14)
ln (Population)		IS			0.24	0.03	(0.18, 0.27)
ln (Annual average daily traffic)		0.16	0.04	(0.09, 0.23)	IS		
ln (Trip generation)		0.16	0.03	(0.10, 0.22)	N/A		
ln (Walking frequency)		N/A			0.29	0.04	(0.23, 0.35)
Demographic and socioeconomic characteristics							
Age	Age below 15	IS			0.03	0.01	(0.01, 0.05)
	Age above 64	IS			0.01	0.006	(0.0002, 0.02)
Ethnicity	Non-Chinese	-0.02	0.004	(-0.03, -0.01)	0.01	0.003	(0.01, 0.02)
Education	Primary education	IS			-0.02	0.008	(-0.04, -0.005)
Household size	More than 3 members	IS			-0.02	0.004	(-0.02, -0.01)
Road network and transport facilities							
Road density		IS			0.01	0.002	(0.005, 0.01)
Number of non-signalized intersections		0.004	0.0005	(0.003, 0.005)	-0.003	0.0002	(-0.004, -0.003)
Number of metro exits		IS			-0.03	0.007	(-0.05, -0.02)
Number of bus stops		IS			0.002	0.001	(0.0006, 0.005)
Land use and points of interest							
Number of restaurants		0.001	0.0004	(0.0005, 0.002)	0.0009	0.0002	(0.0006, 0.001)
Number of schools		IS			-0.01	0.002	(-0.01, -0.001)
Number of hotels		0.01	0.003	(0.003, 0.01)	IS		
Residential land use		IS			0.19	0.05	(0.09, 0.29)
Over-dispersion parameter							
α		0.19	0.007	(0.18, 0.21)	N/A		
Goodness-of-fit measure							
DIC		3005.55					
Pseudo ρ^2		0.29					

IS: Statistically insignificant; N/A: Not applicable

Table 4. Effect of significant factors on total and pedestrian crashes

Category	Total crash		Pedestrian crash	
	Increase	Decrease	Increase	Decrease
Exposure	- Average annual daily traffic (AADT) - Trip generation		- Population - Walking frequency	
Demographic & socioeconomics		- Non-Chinese	- Age below 15 - Age above 64 - Non-Chinese	- Primary education - Household size > 3
Road network & transport	- Non-signalized intersection		- Road density - Bus stop	- Non-signalized intersection - Metro exit
Land use	- Restaurant - Hotel		- Restaurant - Residential	- School