Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure

S.Q. Xie[1], Ni Dong[2], S.C. Wong[1], Helai Huang[3], Pengpeng Xu[1*]

[1] Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China
[2] School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China
[3] School of Traffic and Transportation Engineering, Central South University, Changsha, China
*Corresponding author: pengpengxu@yeah.net

ABSTRACT

This study intended to identify the potential factors contributing to the occurrence of pedestrian crashes at signalized intersections in a densely populated city, based on a comprehensive dataset of 898 pedestrian crashes at 262 signalized intersections during 2010-2012 in Hong Kong. The detailed geometric design, traffic characteristics, signal control, built environment, along with the vehicle and pedestrian volumes were elaborately collected. A Bayesian measurement errors model was introduced as an alternative method to explicitly account for the uncertainties in volume data. To highlight the role played by exposure, models with and without pedestrian volume were estimated and compared. The results indicated that the omission of pedestrian volume in pedestrian crash frequency models would lead to reduced goodness-of-fit, biased parameter estimates, and incorrect inferences. Our empirical analysis demonstrated the existence of moderate uncertainties in pedestrian and vehicle volumes. Six variables were found to have a significant association with the number of pedestrian crashes at signalized intersections. The number of crossing pedestrians, the number of passing vehicles, the presence of curb parking, and the presence of ground-floor shops were positively related with pedestrian crash frequency, whereas the presence of playgrounds near intersections had a negative effect on pedestrian crash occurrences. Specifically, the presence of exclusive pedestrian signals for all crosswalks was found to significantly reduce the risk of pedestrian crashes by 43%. The present study is expected to shed more light on a deeper understanding of the environmental determinants of pedestrian crashes.

Keywords: Pedestrian crash frequency; Signalized intersections; Pedestrian exposure; Measurement errors

1. INTRODUCTION

Pedestrian safety continues to be a considerable public health concern worldwide (Naci et al., 2009; Zegeer and Bushell, 2012; Stoker et al., 2015). Although pedestrian casualties due to traffic crashes in Hong Kong have dropped by 19.3% over the past decade, approximately 3,500 pedestrians are still injured each year (HKTD, 2017). Pedestrians also accounted for more than half of the road traffic fatalities, a proportion much higher than that in other high-income areas. To improve the safety of these vulnerable road users, effective interventions are urgently required to be formulated and implemented.

With the rapid progress of urbanization, a growing number of intersections in cities are controlled by traffic signals. The inadequate accommodation of pedestrians' needs makes them difficult to cross streets and increases the number of pedestrian injuries (Xu et al., 2016). In 2016, about 1,200 pedestrian injuries occurred at intersections in Hong Kong, among which 50% were under signal control (HKTD, 2017). A better understanding of factors contributing to pedestrian crashes at signalized intersections is therefore imperative if walking is advocated as a safe and attractive travel mode. Such information can also facilitate safety planners and policy makers in the design of appropriate infrastructures to improve pedestrian mobility and safety.

In the past two decades, researchers have attempted to develop different predictive models to explore the effects of different types of factors on pedestrian crash counts. Existing studies have focused primarily on the area-wide level (LaScala et al., 2000; Graham and Glaister, 2003; Noland and Quddus, 2004; Morency and Cloutier, 2006; Wedagama et al., 2006; Loukaitou-Sideris et al., 2007; Dissanayake et al., 2009; Sebert Kuhlmann et al., 2009; Wier et al., 2009; Chakravarthy et al., 2010; Cottrill and Thakuriah, 2010; Ha and Thill, 2011; Delmelle et al., 2012; Rifaat et al., 2012; Ukkusuri et al., 2011, 2012; Siddiqui et al., 2012; Dumbaugh and Zhang, 2013; Graham and McCoy, 2013; Noland et al., 2013; Wang and Kockelman, 2013; Jermprapai and Srinivasan, 2014; Steinbach et al., 2014; DiMaggio, 2015; Lee et al., 2015a, 2015b; Yao et al., 2015; Yu, 2015; Cai et al., 2016, 2017; Chen and Zhou, 2016; Wang et al., 2016; Guo et al., 2017; Osama and Sayed, 2017; Tasic et al., 2017; Xie et al., 2017; Ding et al., 2018; Goel et al., 2018). Relatively little research effort has been devoted to investigating the relationship between the number of motor vehicle–pedestrian crashes and potential risk factors at intersections (See: Table A1 in Appendix; Leden, 2002; Lyon and Persaud, 2002; Lee and Abdel-Aty, 2005; Geyer et al., 2006; Schneider et al., 2010; Torbic et al., 2010; Miranda-Moreno et al., 2011; Pulugurtha and Sambhara, 2011; Elvik et al., 2013, 2016; Strauss et al., 2014; Quistberg et al., 2015a, 2015b; Kröyer, 2016; Mooney et al., 2016; Lee et al., 2017; Thomas et al., 2017; Wang et al., 2017), particularly at signalized intersections in a densely populated city (Leden, 2002; Lyon and Persaud, 2002; Torbic et al., 2010; Pulugurtha and Sambhara, 2011; Miranda-Moreno et al., 2011; Strauss et al., 2014).

Not surprisingly, with the increase in vehicle and pedestrian volumes, the absolute number of pedestrian crashes also increases. A nonlinear relationship has consistently been reported, indicating that as the number of pedestrians increases, the crash risk for each individual pedestrian decreases (Leden, 2002; Lyon and Persaud, 2002; Geyer et al., 2006; Schneider et al., 2010; Torbic et al., 2010; Miranda-Moreno et al., 2011; Elvik et al., 2013; Strauss et al., 2014; Elvik, 2016; Kröyer, 2016; Mooney et al., 2016). This is referred to as "safety in numbers" effects (Jacobsen, 2003; Elvik and Bjørnskau, 2017; Xu et al., 2017b). Although pedestrian volume is crucial in determining pedestrian crash occurrences, few transportation agencies regularly collect these data on a large scale due to limited resources. The number of pedestrians is thus mostly estimated based on a short period of field observations (Leden, 2002; Lyon and Persaud, 2002; Schneider et al., 2010; Torbic et al., 2010; Miranda-Moreno et al., 2011; Pulugurtha and Sambhara, 2011; Elvik et al., 2013, 2016; Strauss et al., 2014; Quistberg et al., 2015b; Kröyer, 2016; Mooney et al., 2016), predicted by pedestrian activity models such as Space Syntax (Geyer et al., 2006) and "Ballpark" method (Thomas et al., 2017), or surrogated as surrounding land use and demographic characteristics (Quistberg et al., 2015a; Lee et al., 2017; Wang et al., 2017). It is noteworthy that either absence or improper representation of pedestrian exposure probably leads to inconsistent results (Steinbach et al., 2014). The measurement errors induced in this process may also bias the parameter estimates (Kröyer, 2016).

In addition to the vehicle and pedestrian volumes, geometric design, such as the number of approaches (Miranda-Moreno et al., 2011; Pulugurtha and Sambhara, 2011; Quistberg et al., 2015a; Lee et al., 2017; Thomas et al., 2017), the number of lanes (Thomas et al., 2017), the number of right-turn-only lanes (Schneider et al., 2010), the maximum number of lanes crossed by pedestrians (Torbic et al., 2010), lane width (Quistberg et al., 2015a), average slope of terrain (Thomas et al., 2017), the presence of raised medians (Schneider et al., 2010), the presence of one-way streets (Quistberg et al., 2015a), the presence of sidewalks (Quistberg et al., 2015b), the presence of pedestrian barriers (Quistberg et al., 2015b), the presence of marked crosswalks (Mooney et al., 2016), and the presence of on-street parking (Quistberg et al., 2015b; Thomas et al., 2017) were found to be closely related to the frequency of pedestrian crashes at intersections. The presence of specific facilities close to intersections, i.e., bus stops (Torbic et al., 2010; Miranda-Moreno et al., 2011; Mooney et al., 2016; Thomas et al., 2017), transit stops (Pulugurtha and Sambhara, 2011), schools (Miranda-Moreno et al., 2011), street vendors (Quistberg et al., 2015b), alcohol sales establishments (Torbic et al., 2010), and billboards (Mooney et al., 2016), was reported to significantly increase pedestrian crashes. Intersections located in neighborhoods with commercial land use (Geyer et al., 2006; Schneider et al., 2010; Torbic et al., 2010; Miranda-Moreno et al., 2011; Thomas et al., 2017), lower income levels (Torbic et al., 2010; Thomas et al., 2017), denser population (Quistberg et al., 2015a; Lee et al., 2017; Wang et al., 2017), higher employment rates (Quistberg et al., 2015a), and a higher proportion of residents under 18 years old (Schneider et al., 2010) were also associated with more pedestrian crashes. However, relative to the geometric and built environment factors, there is potential for further insights regarding the effects of signal timing, although they are usually designed according to the intersection geometry and traffic volume.

Relationships between the aforementioned explanatory variables and pedestrian crash counts can be established using crash prediction models. Traditional Poisson and negative binomial models have a strong assumption that their observations should be mutually independent. This fundamental hypothesis is almost always violated (Mannering and Bhat, 2014). More advanced models, such as the conditional autoregressive (Xu et al., 2014; Guo et al., 2017; Goel et al., 2018; Cai et al., 2018a), random parameters (Anastasopoulos and Mannering, 2009; Hou et al., 2018), geographically weighted regression (Xu and Huang, 2015; Gomes et al., 2017; Cai et al., 2018b), spatially varying coefficients (Xu et al., 2017a), and spatiotemporal mixture models (Cheng et al., 2018) have therefore been introduced to achieve more accurate and reliable estimations. In particular, El-Basyouny and Sayed (2010) proposed an approach to address the measurement errors in traffic volume when modeling freeway crash counts. Their results suggested that the adjustment of measurement errors in traffic volume could significantly improve model performance and result in unbiased inferences.

Based on a comprehensive dataset of 898 pedestrian crashes at 262 signalized intersections over a 3-year period in Hong Kong, this study intends to quantify the effects of various factors, including the geometric design, traffic characteristics, signal controls, and built environment characteristics, on the frequency of motor vehicle–pedestrian crashes at signalized intersections in a densely populated city. A novel Bayesian measurement errors model is elaborately developed to accommodate the uncertainties in vehicle and pedestrian volumes. To illustrate the role played by exposure, the estimated coefficients of models with and without pedestrian volume are presented and compared. The present study is expected to shed more light on a deeper understanding of the environmental determinants of pedestrian crashes.

2. DATA

We sampled the intersections based on a comprehensive set of traffic impact assessment reports for the years 2011 and 2012. As the traffic impact assessment was conducted for planning and design purposes and did not investigate the crash records in advance, we assumed no systemic biases in this sampling process. A total of 262 signalized intersections (77 on Hong Kong island, 130 in Kowloon and 55 in New Territories) with adequate traffic and geometric information were available for analysis, which accounted for 15.8% of all signalized intersections in Hong Kong.

We obtained the crash data from the Traffic Road Accident Database System, which is maintained by the Hong Kong Transport Department and the Hong Kong Police Force. These data were collected by the police officers at the crash scenes. Only crashes resulting in injuries were recorded in the database. In Hong Kong, crashes occurring within 70m of the centerline of an intersection were defined by the police as the intersection crashes. In total, 898 motor vehicle–pedestrian crashes were reported at the selected intersections from 2010 to 2012.

The vehicle volume was estimated based on the peak-hour vehicle flows obtained from the Base District Traffic (BDT) models and the 24-hour vehicle traffic profiles at the counting stations reported in the Annual Traffic Census (ATC). The BDT models were developed by the Hong Kong Transport Department for traffic impact assessments and provided peak-hour traffic flow data. The good coverage of ATC counting stations allowed each intersection to be spatially mapped to the nearest ATC counting station. The proportion of peak-hour traffic extracted from the ATC served as a scaling factor, together with the corresponding peak-hour traffic volumes obtained from BDT models, we computed the average daily traffic for an intersection as:

$$\text{Average daily traffic} = \frac{\text{Peak-hour traffic}_{\text{BDT}}}{\text{Proportion of peak-hour traffic}_{\text{ATC}}} \tag{1}$$

The pedestrian volume was estimated based on the Travel Characteristics Survey 2011 database and was further adjusted according to the onsite survey data. By extracting all walking trips and mapping them to the districts and time slots, the 24-hour pedestrian flow profiles for the 26 board districts were constructed. We then conducted 18-hour onsite surveys (from 06:00 to 24:00 on weekdays) at one core intersection for each district. The 18-hour pedestrian flow profiles observed at 26 core intersections were compared with the 24-hour pedestrian flow profiles of the corresponding districts. A set of hourly adjustment factors was thus computed for each board district. To obtain the pedestrian flow profiles for the studied intersections, we conducted 1-hour field surveys at all 262 selected intersections. The average daily crossing pedestrians could be computed by dividing the pedestrian volume at the sampled hour by the corresponding hourly adjustment factor:

$$\text{Average daily crossing pedestrians} = \frac{\text{Sampled-hour pedestrians}}{\text{Proportion of sampled-hour pedestrians}} \tag{2}$$

The intersections' geometric and built environment characteristics were derived from the Google Street View (Mooney et al., 2016). Most imagery for the intersections of interest was captured by Google during February 2011 and December 2011. We determined the presence of playgrounds and schools by whether these facilities were present in any approach of the studied intersections, whereas other characteristics were measured within 70m of the intersection. The data for the signal phasing scheme were manually measured onsite.

Table 1 lists the characteristics of the 262 selected signalized intersections.

Table 1. Characteristics of the 262 signalized intersections under investigation

Continuous variables	**Range**	**Mean**	**S.D.**
Dependent variable			
Number of pedestrian crashes in 2010	Min: 0; Max: 7	1.23	1.49
Number of pedestrian crashes in 2011	Min: 0; Max: 8	1.11	1.39
Number of pedestrian crashes in 2012	Min: 0; Max: 8	1.09	1.49
Exposure			

Annual average daily traffic in 2010 ($\times 10^3$)	Min: 4.09; Max: 246.90	32.11	22.83
Annual average daily traffic in 2011 ($\times 10^3$)	Min: 4.31; Max: 340.52	31.96	26.45
Annual average daily traffic in 2012 ($\times 10^3$)	Min: 3.99; Max: 442.91	32.75	31.43
Annual average daily crossing pedestrians in 2010 ($\times 10^3$)	Min: 0.28; Max: 393.20	39.63	49.60
Annual average daily crossing pedestrians in 2011 ($\times 10^3$)	Min: 0.27; Max: 316.99	38.37	46.11
Annual average daily crossing pedestrians in 2012 ($\times 10^3$)	Min: 0.31; Max: 310.28	39.00	47.69
Geometric characteristics			
Number of traffic lanes	Min: 5; Max: 28	13.61	5.32
Average lane width (meters)	Min: 2.47; Max: 5.68	3.45	0.44
Maximum number of lanes crossed by pedestrians per crossing maneuver	Min: 2; Max: 11	4.49	1.76
Traffic characteristics			
Ratio of minor-road AADT to major-road AADT	Min: 0.01 Max: 0.99	0.44	0.26
Number of pedestrian-vehicle conflict points	Min: 2; Max: 46	4.49	1.76
Signal phasing scheme			
Cycle time (seconds)	Min: 48; Max: 216	106.03	22.95
Categorical variables	**Attributes**	**Count**	**Proportion**
Geometric characteristics			
Number of legs	3	107	40.84%
	4	143	54.58%
	≥ 5	12	4.58%
Number of legs with crosswalks	1	21	8.02%
	2	89	33.97%
	3	90	34.35%
	≥ 4	62	23.66%
Maximum number of lanes on major legs	2	21	8.02%
	3	51	19.47%
	4	63	24.05%
	5	47	17.94%
	6	47	17.94%
	7 to 9	33	12.58%
Maximum number of lanes on minor legs	1	13	4.96%
	2	85	32.44%
	3	67	25.57%
	4	43	16.41%
	5	27	10.31%
	6 to 7	27	10.31%
Presence of left-turn only lanes on major legs	Yes	99	37.79%
	No	163	62.21%
Presence of right-turn only lanes on major legs	Yes	119	45.42%
	No	143	54.58%
Presence of left-turn only lanes on minor legs	Yes	154	58.78%
	No	107	41.22%
Presence of right-turn only lanes on minor legs	Yes	149	56.87%
	No	113	43.13%
Presence of one-way streets	Yes	157	59.92%
	No	105	40.08%
Presence of raised medians	For all legs	45	17.18%
	For some legs	119	45.42%

	None	98	37.40%
Presence of curb extension	Yes	100	38.17%
	No	162	61.83%
Presence of curb parking	Yes	95	36.26%
	No	167	63.74%
Presence of marked crosswalks	All crosswalks marked*	197	75.19%
	Some crosswalks marked	63	24.05%
	None	2	0.76%
Presence of pedestrian barriers	For all sidewalks	158	60.31%
	For some sidewalks	103	39.31%
	None	1	0.38%
Presence of pedestrian refuge islands	Yes	216	82.44%
	No	46	17.56%
Presence of overpass or underpass	Yes	40	15.27%
	No	222	84.73%
Signal phasing scheme			
Number of signal stages	2	62	23.66%
	3	117	44.66%
	4	72	27.48%
	5 to 7	11	4.20%
Presence of right-turn pocket	Yes	24	9.16%
	No	238	90.84%
Presence of exclusive pedestrian signals	For all crosswalks	216	82.44%
	For some crosswalks	44	16.79%
	None	2	0.77%
Built environment characteristics			
Presence of trees on roadsides	Yes	207	79.01%
	No	55	20.99%
Presence of bus stops	Harbor-shaped	38	14.50%
	Non-harbor-shaped	75	28.63%
	None	149	56.87%
Presence of tram tails	Yes	25	9.54%
	No	237	90.46%
Presence of tram stops	Yes	16	6.11%
	No	246	93.89%
Presence of metro entrances	Yes	25	9.54%
	No	237	90.46%
Presence of metro guiding signs	Yes	114	43.51%
	No	148	56.49%
Presence of ground-floor shops	Yes	158	60.31%
	No	104	39.69%
Presence of parks or playgrounds	Yes	126	48.09%
	No	136	51.91%
Presence of schools	Yes	165	62.98%
	No	97	37.02%

AADT refers to the annual average daily traffic.

3. METHODOLOGY

We modeled the frequency of pedestrian crashes at each intersection consistent with previous studies (Leden, 2002; Lyon and Persaud, 2002; Geyer et al., 2006; Schneider et al., 2010; Miranda-Moreno et al., 2011; Elvik et al., 2013, 2016; Kröyer, 2016; Mooney et al., 2016). Let Y_i denote the reported number of pedestrian crashes at the ith$\left(i=1,2,\ldots,262\right)$ signalized intersection during the years 2010 to 2012. The use of aggregate crash data over 3 years avoids the confounding effects and regression-to-the-mean phenomenon (Cheng and Washington, 2005). E_i is the exposure function, and $\mathbf{X_i}$ refers to the vector of explanatory variables related to site-specific attributes. Given the random, non-negative, and integral nature of crash counts, we have:

$$
\begin{aligned}
&Y_i \sim \text{Possion}(\lambda_i) \\
&\ln(\lambda_i) = \log(E_i) + \mathbf{X_i'\beta} + u_i
\end{aligned}
\tag{3}
$$

where λ_i is the parameter of the Poisson model. $\boldsymbol{\beta}(\beta_1,\beta_2,..,\beta_k)$ represents the vector of regression coefficients to be estimated. u_i accounts for the overdispersion due to unobserved factors and is specified as an exchangeable normal prior distribution:

$$u_i \sim \mathrm{N}(0,\sigma_u^2)$$

(4)

Given the potential non-linear relationship between pedestrian crashes and traffic volumes, the exposure function suggested by Elvik and Bjørnskau (2017) is adopted here:

$$E_i = V_i^{\alpha_1} \times P_i^{\alpha_2}$$

(5)

where V_i and P_i respectively denote the average daily vehicles and crossing pedestrians during the 3-year period, which can be calculated as:

$$
\begin{aligned}
&V_i = \frac{1}{3}\sum_{t=1}^{3}\mathrm{AADT}_{it} \\
&P_i = \frac{1}{3}\sum_{t=1}^{3}\mathrm{AADP}_{it}
\end{aligned}
$$

(6)

in which AADT_{it} and AADP_{it} are the estimated annual average daily traffic and crossing pedestrians for the ith intersection in the tth$\left(t=1,2,3\right)$ year.

Eq. (6) ignores the fact that the volume data are usually measured with errors. If these measurement errors are not well addressed, biased parameters will be produced and incorrect inference can be drawn (El-Basyouny and Sayed, 2010). To this end, a measurement errors model is introduced here, in which V_i and P_i are treated as the latent variables and are approximated by a log-normal distribution (Davis, 2000; Davis and Yang, 2001; El-Basyouny and Sayed, 2010):

$$\begin{aligned} \ln(V_i) &\sim \mathrm{N}(\mu_V, \sigma_V^2) \\ \ln(P_i) &\sim \mathrm{N}(\mu_P, \sigma_P^2) \end{aligned} \tag{7}$$

where μ_V and μ_P refer to the expected values of $\ln(V_i)$ and $\ln(P_i)$. σ_V^2 and σ_P^2 control the variations across intersections.

The observed AADT_{it} and AADP_{it} are linked with V_i and P_i via:

$$\begin{aligned} \ln(\mathrm{AADT}_{it}) &\sim \mathrm{N}(\ln(V_i) + \delta_{t_V}, \sigma_{\varepsilon_V}^2) \\ \ln(\mathrm{AADP}_{it}) &\sim \mathrm{N}(\ln(P_i) + \delta_{t_P}, \sigma_{\varepsilon_P}^2) \end{aligned} \tag{8}$$

where δ_{t_V} and δ_{t_P} are the time effects. This specification recognizes that traffic volumes are likely to change steadily over time. For data observed within a few years, δ_{t_V} and δ_{t_P} can be adequately modeled as the fixed linear trend with coefficients γ_V and γ_P (El-Basyouny and Sayed, 2010):

$$\begin{aligned} \delta_{t_V} &= \gamma_V (t-1) \\ \delta_{t_P} &= \gamma_P (t-1) \end{aligned} \tag{9}$$

Accordingly, Eq. (8) can be rearranged as:

$$\begin{aligned} \mathrm{AADT}_{it} &= V_i e^{\delta_{t_V}} e^{\varepsilon_{i_V}} \\ \mathrm{AADP}_{it} &= P_i e^{\delta_{t_P}} e^{\varepsilon_{i_P}} \end{aligned} \tag{10}$$

in which ε_{i_V} and ε_{i_P} denote the measurement errors with variances of $\sigma_{\varepsilon_V}^2$ and $\sigma_{\varepsilon_P}^2$.

The relative magnitude of the measurement errors can then be calculated by the reliability ratio:

$$\begin{aligned} \mathrm{RR_V} &= \frac{\sigma_{\varepsilon_V}}{\sqrt{\sigma_V^2 + \sigma_{\varepsilon_V}^2}} \\ \mathrm{RR_P} &= \frac{\sigma_{\varepsilon_P}}{\sqrt{\sigma_p^2 + \sigma_{\varepsilon_P}^2}} \end{aligned} \tag{11}$$

A full Bayesian inference using the Markov Chain Monte Carlo algorithm was implemented to construct the model. Obtaining the Bayesian posterior estimates requires the specification of prior distributions. Due to the absence of sufficient prior knowledge, a non-informative prior, i.e., $\mathrm{N}(0,1000)$, was specified for α_1, α_2, β_k, μ_V, μ_P, γ_V, and γ_P. In accordance with Gelman (2006), the variance parameters, i.e., σ_u, σ_V, σ_P, σ_{ε_V}, and σ_{ε_P} were assigned as a uniform $(0,10)$.

The deviance information criterion (DIC) was used here to measure model performance:

$$\mathrm{DIC} = D(\overline{\theta}) + 2p_D = \overline{D} + p_D \tag{12}$$

where $D(\overline{\theta})$ is the deviance evaluated at $\overline{\theta}$, the posterior means of the parameter of interest, p_D is the effective number of parameters in the model, and $\overline{D}$ is the posterior mean of the deviance statistic $D(\overline{\theta})$. The lower the DIC, the better the model fit. Generally, differences in the DIC of more than 10 definitely rule out the models with higher DIC. Differences between 5 and 10 are considered substantial, whereas a difference of less than 5 indicates that the models are not statistically different (Spiegelhalter et al., 2002).

To evaluate the overall explanatory power of pedestrian volume, the proportion of reduction in variance (PRV), also known as the explained variance (Raudenbush and Bryk, 2002; Wang et al., 2017) was used:

$$\mathrm{PRV} = \frac{\sigma_{u_0}^2 - \sigma_{u_1}^2}{\sigma_{u_0}^2} \tag{13}$$

where $\sigma_{u_1}^2$ and $\sigma_{u_0}^2$ are the variance of the error term in the models with and without pedestrian volume, respectively. The PRV is bounded by 0 and 1, with a higher value indicating a stronger explanatory power.

4. RESULTS AND DISCUSSION

The freeware software WinBUGS was used to calibrate the models (Spiegelhalter et al., 2005). Three parallel chains with diverse starting points were tracked. The first 5,000 iterations were discarded as burn-ins, and then 5,000 iterations were performed for each chain, resulting in a sample of distribution of 15,000 for each parameter. The model's convergence was monitored by the Brooks-Gelman-Rubin statistic (Brook and Gelman, 1998), visual examination of the Markov Chain Monte Carlo chains, and the ratios of Monte Carlo errors relative to the respective standard deviations of the estimates. As a rule of thumb, these ratios should be less than 0.05.

The model specifications were developed based on the following principles. A correlation test was first conducted to ensure the non-inclusion of highly correlated variables. The correlation analysis indicated a high correlation between the number of legs and the number of legs with crosswalks, with the Spearman's correlation parameter (Washington et al., 2011) estimated at 0.68. Similarly, the total number of traffic lanes, maximum number of lanes on major legs, maximum number of lanes on minor legs, maximum number of lanes crossed by pedestrians per crossing maneuver, and the number of pedestrian-vehicle conflict points were highly correlated, indicating that these five variables should not been added into the model simultaneously. Other variables showed weak collinearity as their Spearman's correlation parameters were all less than 0.40. In the initial model, we included all of the uncorrelated variables. The DIC was then used to compare alternative models with different covariate subsets. The one producing a lower DIC value was considered superior.

For the purpose of comparison, in addition to the measurement errors model, we developed the Poisson lognormal model. To highlight the role of pedestrian exposure, models with and without pedestrian volume were estimated. As such, four models were eventually calibrated. The performance of these models is compared below, followed by the presentation and interpretation of the parameter estimates.

4.1 Model performance comparison

Table 2 shows the results of goodness-of-fit measures for the calibrated models. The model with pedestrian volume outperformed its counterpart without it according to the DIC statistic. This result suggests that explicitly accounting for pedestrian volume will be conducive to a substantial improvement in goodness-of-fit. Specifically, as measured by the PRV, roughly 16% of variations could be explained by pedestrian volume, further implying the dominant explanatory powers of pedestrian volume in predicting pedestrian crash counts. It is also interesting that although the measurement errors model had comparable performance with the Poisson lognormal model in terms of the DIC, the reliability ratio for vehicle and pedestrian volumes in the measurement errors model was significant at the 95% confidence level with estimates of 0.07 and 0.13, respectively, confirming the existence of a moderate magnitude of uncertainties in volume data, particularly in the pedestrian volume. To some extent, this finding is expected as the scale factors for our vehicle volume were derived directly from the counting stations, whereas the number of crossing pedestrians was estimated based on a limited-hour manual onsite observation.

Table 2. Goodness-of-fit measures for the Poisson lognormal and measurement errors models with and without pedestrian volume

Model type	$\overline{D}$	p_D	DIC	RR_V	RR_P	PRV
PL without pedestrian volume	948.36	119.53	1067.89			
PL with pedestrian volume	952.23	110.74	1062.97			16.13%
ME without pedestrian volume	946.28	119.64	1065.92	0.07^{**}		
ME with pedestrian volume	950.41	111.07	1061.48	0.07^{**}	0.13^{**}	16.13%

** indicates significance at the 95% confidence level.

PL and ME are the abbreviations of Poisson lognormal and measurement errors models, respectively.

4.2 Parameter estimates

Table 3 summarizes the parameter estimates in the Poisson lognormal and measurement errors models with and without pedestrian volume. A 5% level of significance was used as the threshold to determine whether the parameters differed from 0. Variables that were insignificant in all four models were excluded. To determine the impacts of these independent variables, Table 4 presents the corresponding elasticities.

Table 3. Results of the measurement errors and Poisson lognormal models for pedestrian crash frequency

Variables	**PL without pedestrian volume**		**PL with pedestrian volume**		**ME without pedestrian volume**		**ME with pedestrian volume**	
	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI
Intercept	−3.10 (0.86) **	(−4.80, −1.41)	−3.98 (0.83) **	(−5.56, −2.31)	−3.24 (0.89) **	(−5.03, −1.46)	−4.01 (0.90) **	(−5.70, −2.20)
Ln(*V*)	0.35 (0.08) **	(0.19, 0.51)	0.27 (0.08) **	(0.11, 0.41)	0.37 (0.09) **	(0.20, 0.54)	0.27 (0.09) **	(0.09, 0.43)
Ln(*P*)			0.21 (0.05) **	(0.11, 0.31)			0.23 (0.05) **	(0.13, 0.33)
Presence of exclusive pedestrian signals for all crosswalks	−0.28 (0.13) **	(−0.54, −0.03)	−0.36 (0.13) **	(−0.60, −0.11)	−0.28 (0.13) **	(−0.54, −0.03)	−0.36 (0.13) **	(−0.60, −0.11)
Presence of curb parking	0.25 (0.11) **	(0.03, 0.47)	0.22 (0.11) **	(0.01, 0.43)	0.25 (0.11) **	(0.03, 0.47)	0.22 (0.11) **	(0.01, 0.43)
Presence of bus stops								
Harbor-shaped	0.06 (0.16)	(−0.26, 0.37)	−0.01 (0.16)	(−0.32, 0.30)	0.06 (0.16)	(−0.26, 0.37)	−0.01 (0.16)	(−0.32, 0.30)
Non-harbor-shaped	0.24 (0.12) **	(0.01, 0.47)	0.17 (0.11)	(−0.05, 0.39)	0.24 (0.12) **	(0.01, 0.47)	0.17 (0.11)	(−0.05, 0.39)
Presence of metro guiding signs	0.24 (0.11) **	(0.03, 0.45)	0.11 (0.11)	(−0.10, 0.32)	0.24 (0.10) **	(0.04, 0.45)	0.11 (0.11)	(−0.10, 0.32)
Presence of ground-floor shops	0.80 (0.13) **	(0.55, 1.05)	0.54 (0.14) **	(0.27, 0.81)	0.80 (0.13) **	(0.55, 1.05)	0.54 (0.14) **	(0.27, 0.81)
Presence of parks or playgrounds	−0.24 (0.11) **	(−0.46, −0.02)	−0.23 (0.11) **	(−0.44, −0.03)	−0.24 (0.11) **	(−0.46, −0.02)	−0.24 (0.11) **	(−0.44, −0.03)
σ_u^2	0.31 (0.06) **	(0.21, 0.44)	0.26 (0.05) **	(0.17, 0.38)	0.31 (0.06) **	(0.21, 0.44)	0.26 (0.06) **	(0.17, 0.39)
σ_V^2					1.64 (0.15) **	(1.38, 1.95)	1.64 (0.15) **	(1.38, 1.96)
σ_P^2							0.42 (0.04) **	(0.36, 0.50)
$\sigma_{\varepsilon_V}^2$					0.008 (0.001) **	(0.007, 0.009)	0.008 (0.001) **	(0.007, 0.009)
$\sigma_{\varepsilon_P}^2$							0.008 (0.001) **	(0.007, 0.009)

PL and ME are the abbreviations of Poisson lognormal and measurement errors models, respectively.
SD refers to the standard deviation.
CI denotes the confidence level.
** *indicates the significance at the 95% CI.*
The time effect was insignificantly different from zero at the 5% significance level. The measurement error model was therefore re-estimated with $\delta_{t_V} = \delta_{t_P} = 0$.

Table 4. Elasticities of covariates in the measurement errors and Poisson lognormal models with and without pedestrian volume

	PL without pedestrian volume		**PL with pedestrian volume**		**ME without pedestrian volume**		**ME with pedestrian volume**	
Variables	Elasticity	95% CI	Elasticity	95% CI	Elasticity	95% CI	Elasticity	95% CI
V	0.35 **	(0.19, 0.51)	0.27 **	(0.11, 0.41)	0.37 **	(0.20, 0.54)	0.27 **	(0.09, 0.43)
P			0.21 **	(0.11, 0.31)			0.23 **	(0.13, 0.33)
Presence of exclusive pedestrian signals for all crosswalks	−0.33 **	(−0.72, −0.03)	−0.43 **	(−0.83, −0.11)	−0.33 **	(−0.71, −0.03)	−0.43 **	(−0.83, −0.12)
Presence of curb parking	0.22 **	(0.03, 0.38)	0.20 **	(0.01, 0.35)	0.22 **	(0.03, 0.37)	0.20 **	(0.01, 0.35)
Presence of bus stops								
Harbor-shaped	0.06	(−0.30, 0.31)	−0.01	(−0.38, 0.26)	0.06	(−0.30, 0.31)	−0.01	(−0.38, 0.26)
Non-harbor-shaped	0.21 **	(0.01, 0.37)	0.16	(−0.05, 0.33)	0.21 **	(0.01, 0.37)	0.16	(−0.05, 0.33)
Presence of metro guiding signs	0.22 **	(0.03, 0.36)	0.10	(−0.10, 0.27)	0.21 **	(0.01, 0.37)	0.10	(−0.11, 0.27)
Presence of ground-floor shops	0.55 **	(0.42, 0.65)	0.42 **	(0.24, 0.55)	0.55 **	(0.42, 0.65)	0.41 **	(0.24, 0.55)
Presence of parks or playgrounds	−0.27 **	(−0.58, −0.02)	−0.26 **	(−0.56, −0.03)	−0.27 **	(−0.58, −0.02)	−0.27 **	(−0.56, −0.03)

PL and ME are the abbreviations of Poisson lognormal and measurement errors models, respectively.

CI denotes the confidence level.

** *indicates the significance at the 95% CI.*

Elasticities for average daily passing vehicles and crossing pedestrians (i.e., log-linear variables) were equal to their estimated coefficients, i.e., $\widehat{\alpha}_1$ *and* $\widehat{\alpha}_2$ *(Washington et al., 2011).*

Elasticities for indicate variables were computed as $\frac{\exp(\widehat{\beta}_k)-1}{\exp(\widehat{\beta}_k)}$ *(Washington et al., 2011).*

Several general observations deserve mentioning. First, the significant variables were not entirely identical between the models with and without pedestrian volume. For example, the presence of non-harbor-shaped bus stops was statistically significant in the base models without pedestrian volume but became totally insignificant in the fully specified models. The same conclusion held true for the variable of the presence of metro signs. Second, relative to the base models, the effects of several risk factors, i.e., vehicle volume, the presence of exclusive pedestrian signals, the presence of curb parking, and the presence of ground-level shops, changed substantially in the fully specified models. Specifically, the elasticity of vehicle volume in the measurement errors model decreased sharply from 0.37 to 0.27 once pedestrian volume was added, resulting in an overestimation by about 37%. Similar results could also be observed for the effect of the presence of curb parking and the presence of ground-floor shops, as their elasticities were roughly overestimated by 10% and 34%, respectively. Regarding the variable of the presence of exclusive pedestrian signals, its elasticity was underestimated by approximately 23%. These findings raise an alarm on previous studies (Quistberg et al., 2015a; Lee et al., 2017; Wang et al., 2017) that the omission of exposure in pedestrian crash frequency models would lead to biased estimates and inadequate inferences.

More interestingly, a comparison between the Poisson lognormal and measurement errors models indicated that overall these two models produced very similar parameter estimates. Only the coefficient of pedestrian volume increased slightly from 0.21 to 0.23 when measurement errors were taken into account. This result implies that our data are fairly robust to model configuration. El-Basyouny and Sayed (2010) reported a similar conclusion that in the presence of weak measurement errors, the measurement errors model was comparable with the traditional ones.

Given that the measurement errors model with pedestrian volume performed best with the lowest DIC value, we chose it to interpret our results in the subsequent section.

As Tables 3 and 4 shows, six variables had a significant association with the frequency of pedestrian crashes: average daily passing vehicles, average daily crossing pedestrians, the presence of exclusive pedestrian signals, the presence of curb parking, the presence of ground-floor shops, and the presence of playgrounds. The signs of these parameters were generally consistent with empirical judgements and the results of previous studies (Retting et al., 2003; Zegeer and Bushell, 2012; Stoker et al., 2015).

Both pedestrian and vehicle volumes were significant and positive, with coefficients estimated at 0.23 and 0.27, respectively. This nonlinear relationship between the number of crossing pedestrians and the number of pedestrian crashes has been widely confirmed (Leden, 2002; Geyer et al., 2006; Schneider et al., 2010; Torbic et al., 2010; Miranda-Moreno et al., 2011; Elvik et al., 2013; Kröyer, 2016; Mooney et al., 2016). Jacobsen et al. (2015) attributed it to the behavior modifications by motorists when they encountered more people walking. All things being equal, an increase in pedestrian activity would lead to an increase in the total number of pedestrian crashes but a decrease in the crash risk for each individual pedestrian. That is, a motorist is less likely to collide with a pedestrian when more people are walking (Jacobsen, 2003). However, this conclusion based on a cross-sectional research design should be interpreted with great caution, because it is impossible to determine whether this nonlinear association is a causal relationship or merely a statistical artifact (Bhatia and Wier, 2011; Xu et al., 2017b). Further efforts are therefore desired to identify the underlying mechanisms.

Instead of simply encouraging people to walk, an alternative sound measure to improving pedestrian safety would be to restrict the usage of motor vehicles. In addition to the benefits of less congestion, fewer emissions of pollutants, and less traffic noise, the strategies to reduce vehicle volume would lower both the number of pedestrian crashes and the crash risk for pedestrians. According to our results, an intersection cutting its passing vehicle volume to half would expect a 17% decrease in pedestrian crashes ($0.5^{0.27} = 0.83$). Therefore, the promotion of a modal shift from motor vehicles to other travel modes, such as public transit, walking, and cycling, should be highly advocated, especially in a dense urban setting.

Traffic signal has long served as a common control measure at intersections. Despite the wide use, its effects on pedestrian crashes remain under-investigated. In this study, we included the total cycle time, the number of signal stages, the presence of right-turn pocket, and the presence of exclusive pedestrian signals in our models. Our results revealed that only the presence of exclusive pedestrian signals had a significant relationship with the occurrence of pedestrian crashes. Accordingly, the installation of exclusive pedestrian signals for all crosswalks would contribute to a reduction of pedestrian crash risk by 43%. This finding is expected to a large extent, because the exclusive pedestrian phasing stops all vehicles to facilitate people making crossings, which dramatically reduces the conflicts between pedestrians and vehicles.

Consistent with Quistberg et al. (2015b), the presence of curb parking close to intersections was associated with an elevated risk of pedestrian crashes. Based on our results, if curb parking were allowed at a crossing area, the risk of pedestrian crashes would increase by about 20%. Although on-street parking can provide friction to slow vehicles and act as a buffer for pedestrians (Ewing and Dumbaugh, 2009), the vehicles parked on the streets indeed obscure vision between motorists and pedestrians. One direct countermeasure is to restrict on-street parking, especially at peak hours with heavy pedestrian activities. While in areas with limited parking spaces and great parking demands, a replacement of parallel parking as diagonal one would be an effective measure. With the design of vehicles parked at an angle (typically 30 degree) to the curb in the direction of traffic flow, diagonal parking allows more pedestrians to scan for traffic before crossing and greatly reduces the number of pedestrians in front of a parked vehicle (Retting et al., 2003).

With respect to land use, the presence of ground-floor shops was found to significantly increase the likelihood of pedestrian crashes. Similar results were also reported by Quistberg et al. (2015b) and Mooney et al. (2016). One plausible explanation is that the frequent roadside advertising alongside ground-floor shops can distract drivers' attention (Young et al., 2009). As another, pedestrians are also more likely to jaywalk to reach stores without noticing the approaching vehicles.

Furthermore, the negative coefficient of the presence of playgrounds indicated that recreational land use pattern was associated with a lower frequency of pedestrian crashes. Compared with those located in commercial areas, intersections near playgrounds generally permit a broader view for both pedestrians and motorists, thus enhancing their visibility. Motorists may also adjust their behaviors when they expect more pedestrians near playgrounds.

5. CONCLUSION

This study sought to identify the factors that determine the frequency of motor vehicle–pedestrian crashes at signalized intersections, by use of a comprehensive dataset collected at 262 signalized intersections in Hong Kong over a 3-year period. Detailed site conditions, including the geometric design, traffic characteristics, signal control schemes, and built environment characteristics, were integrated with the vehicle and pedestrian volumes to construct our dataset. A Bayesian measurement errors model was developed to explicitly account for the uncertainties in volume data. To highlight the importance of pedestrian volume, models with and without it were calibrated and compared.

Some key findings are worth noting. First, the omission of pedestrian volume in pedestrian crash frequency models would result in reduced goodness-of-fit and biased estimations. Our results indicated substantial inconsistences in the effects of several risk factors in the models with and without pedestrian volume. For example, the presence of non-harbor-shaped bus stops and the presence of metro signage were significant in the base model but became totally insignificant when pedestrian volume was added. The elasticity of vehicle volume, the presence of curb parking and the presence of ground-floor shops was also overestimated respectively by as much as 37%, 10%, and 34%, whereas the effect of the presence of exclusive pedestrian signals was biased downwards by approximately 23%.

Although pedestrian volume is indispensable in determining pedestrian crash occurrences, the major concern lies in the availability of reliable pedestrian volume data for a large number of sites. Unlike the vehicle volume typically obtained from counting stations, the number of pedestrians is mostly estimated based on a short period of manual observations. The fact that volume data are measured with errors is usually overlooked in previous studies. For this purpose, a Bayesian measurement errors model was introduced as a methodological alternative. Our empirical analysis demonstrated the existence of uncertainties in the volume data. We also revealed that in the presence of weak measurement errors, the measurement errors model had comparable performance with the traditional Poisson lognormal model in terms of overall fit and parameter estimates.

Six variables were finally found to have a significant association with the frequency of pedestrian crashes at signalized intersections. The nonlinear relationship between the number of crossing pedestrians and the number of pedestrian crashes was confirmed, with an estimated coefficient of 0.23. Traffic volume, the presence of curb parking, and the presence of ground-floor shops were found to be positively associated with pedestrian crash frequency, whereas the presence of playgrounds in the vicinity of intersections had a negative effect on pedestrian crash occurrence. By including signal control schemes in our models, we provided additional evidence to the existing research that the presence of exclusive pedestrian signals for all crosswalks could significantly reduce the risk of pedestrian crashes by 43%. These findings would be informative to policy makers and urban planners in the design of appropriate facilities to improve the safety and mobility of pedestrians at signalized intersections.

Our study is not without limitations. Although we took advantage of Google Street View to extract more than 20 variables related to pedestrian facilities, several potential risk factors could not be accounted for. For example, we failed to accommodate vehicle speeds and demographic characteristics of the pedestrians who crossed at our sampled intersections. Further efforts are required to explore their effects on pedestrian crash occurrences. As our research is cross-sectional in nature that can provide correlational evidence only (Kim and Mooney, 2016), future studies using a quasi-experimental research deign (Ewing et al., 2013) are strongly advocated to provide more insights into the causation of pedestrian crashes.

ACKNOWLEDGEMENTS

We would like to thank the Hong Kong Police Force and the Hong Kong Transport Department for providing the database used for this study. This work was jointly supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 717512), the Joint Research Scheme of National Nature Science Foundation of China/Research Grants Council of Hong Kong (Project Nos. 71561167001 and N_HKU707/15), and the Natural Science Foundation of China (Project No. 71601163; 71671100). S.Q. Xie, S.C. Wong and Pengpeng Xu were respectively supported by the Post-doctoral Fellow Scheme, Francis S Y Bong Professorship in Engineering, and Research Postgraduate Studentship from the University of Hong Kong.

REFERENCES

Anastasopoulos, P.Ch., Mannering, F.L., 2009. A note on modeling vehicle-accident frequencies with random parameter count models. Accid. Anal. Prev. 41(1), 153-159.

Bhatia, R., Wier, M., 2011. “Safety in Numbers” re-examined: can we make valid or practical inferences from available inference? Accid. Anal. Prev. 43, 235-240.

Brook, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. 7(4), 434-455.

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. Accid. Anal. Prev. 93, 14-22.

Cai, Q., Abdel-Aty, M., Lee, J., 2017. Macro-level vulnerable road users crash analysis: a Bayesian joint modeling approach of frequency and proportion. Accid. Anal. Prev. 107, 11-19.

Cai, Q., Abdel-Aty, M., Lee, J., Huang, H., 2018a. Integrating macro- and micro-level safety analyses: a Bayesian approach incorporating spatial interaction. Transportmetrica A, doi:10.1080/23249935.2018.1471752

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., Wang, X., 2018b. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. Anal. Methods Acc. Res. 19, 1-15.

Chakravarthy, B., Anderson, C.L., Ludlow, J., Lotfipour, S., Vaca, F.E., 2010. The relationship of pedestrian injuries to socioeconomic characteristics in a large southern California county. Traffic Inj. Prev. 11(5), 508-513.

Chen, P., Zhou, J., 2016. Effects of the built environment on automobile-involved pedestrian crash frequency and risk. J. Transp. Health 3, 448-456.

Cheng, W., Washington, S.P., 2005. Experimental evaluation of hotspot identification methods. Accid. Anal. Prev. 37, 870-881.

Cheng, W., Gill, G.S., Zhang, Y., Cao, Z., 2018. Bayesian spatiotemporal crash frequency models with mixture components for space-time interactions. Accid. Anal. Prev. 112, 84-93.

Cottrill, C.D., Thakuriah, P., 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. Accid. Anal. Prev. 42, 1718-1728.

Davis, G.A., 2000. Estimating traffic accident rates while accounting for traffic volume estimation error: a Gibbs sampling approach. Transp. Res. Rec. 1717, 94-101.

Davis, G., Yang, S., 2001. Accounting for uncertainty in estimates of total traffic volume: an Empirical Bayes approach. J. Transp. Stat. 4(1), 27-38.

Delmelle, E.C., Thill, J.C., Ha, H.H., 2012. Spatial epidemiology analysis of relative collision risk factors among urban bicyclists and pedestrians. Transportation 39, 433-448.

DiMaggio, C., 2015. Small area spatiotemporal analysis of pedestrian and bicyclist injuries in New York City. Epidemiology 26(2), 247-254.

Ding, C., Chen, P., Jiao, J., 2018. Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: a machine learning approach. Accid. Anal. Prev. 112, 116-126.

Dissanayake, D., Aryaija, J., Wedagama, D.M.P., 2009. Modelling the effects of land use and temporal factors on child pedestrian casualties. Accid. Anal. Prev. 41, 1016-1024.

Dumbaugh, E., Zhang, Y., 2013. The relationship between community design and crashes involving older drivers and pedestrians. J. Plan. Educ. Res. 33, 83-95.

Elvik, R., Sørensen, M.W.J., Nævestad, T.O., 2013. Factors influencing safety in a sample of marked pedestrian crossing selected for safety inspections in the city of Oslo. Accid. Anal. Prev. 59, 64-70.

Elivk, R., 2016. Safety-in-numbers: estimates based on a sample of pedestrian crossings in Norway. Accid. Anal. Prev. 91, 175-182.

Elvik, R., Bjørnskau, T., 2017. Safety-in numbers: a systematic review and meta-analysis of evidence. Saf. Sci. 92, 274-282.

Ewing, R., Dumbaugh, E., 2009. The built environment and traffic safety: a review of empirical evidence. J. Plan. Lit. 23(4), 347-367.

Ewing, R., Chen, L., and Chen, C., 2013. Quasi-experimental study of traffic calming measures in New York City. Transp. Res. Rec. 2364, 29-35.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. 1, 515-533.

Geyer, J., Raford, N., Pham, T., Ragland, D.R., 2006. Safety in numbers: data from Oakland, California. Transp. Res. Rec. 1982, 150-154.

Goel, R., Jain, P., Tiwari, G., 2018. Correlates of fatality risk of vulnerable road users in Delhi. Accid. Anal. Prev. 111, 86-93.

Gomes, M.J.T.L., Cunto, F., Silva, A.R., 2017. Geographically weighed negative binomial regression applied to zonal level safety performance models. Accid. Anal. Prev. 106, 254-261.

Graham, D.J., Glaister, S., 2003. Spatial variation in road pedestrian casualties: the role of urban scale, density and land-use mix. Urban Studies 40(8), 1591-1607.

Graham, D.J., McCoy, E.J., 2013. Quantifying the effect of area deprivation on child pedestrian casualties by using longitudinal mixed models to adjust for confounding, interference and spatial dependence. J. R. Statist. Soc. A 176(4), 931-950.

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. Accid. Anal. Prev. 99, 114-124.

Ha, H.H., Thill, J.C., 2011. Analysis of traffic hazard intensity: a spatial epidemiology case study of urban pedestrians. Comput. Environ. Urban Syst. 35, 230-240.

Hong Kong Transport Department (HKTD), 2017. Road Traffic Accident Statistics, 2016. https://www.police.gov.hk/info/doc/statistics/traffic_report_2016_en.pdf

Hou, Q., Tarko, A.P., Meng, X., 2018. Analyzing crash frequency in freeway tunnels: a correlated random parameters approach. Accid. Anal. Prev. 111, 94-100.

Jacobsen, P.L., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj. Prev. 9, 205-209.

Jacobsen, P.L., Ragland, D.R., Komanoff, C., 2015. Safety in Numbers for walkers and bicyclists: exploring the mechanisms. Inj. Prev. 21, 217-220.

Jermprapai, K., Srinivasan, S., 2014. Planning-level model for assessing pedestrian safety. Transp. Res. Rec. 2464, 109-117.

Kim, J.H., Mooney, S.J., 2016. The epidemiologic principles underlying traffic safety study designs. Int. J. Epidemiol. 45(5), 1668-1675.

Kröyer, H.R.G., 2016. Pedestrian and bicyclist flows in accident modelling at intersections. Influence of the length of observation period. Saf. Sci. 82, 315-324.

LaScala, E.A., Gerber, D., Gruenewald, P.J., 2000. Demographic and environmental correlated of pedestrian injury collisions: a spatial analysis. Accid. Anal. Prev. 32, 651-658.

Leden, L., 2002. Pedestrian risk decreases with pedestrian flow. A case study based on

data from signalized intersections in Hamilton, Ontario. Accid. Anal. Prev. 34(4), 457-464.
Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. Accid. Anal. Prev. 37(4), 775-786.
Lee, J., Abdel-Aty, M., Choi, K., Huang, H., 2015a. Multi-level hot zone identification for pedestrian safety. Accid. Anal. Prev. 76, 64-73.
Lee, J., Abdel-Aty, M., Jiang, X., 2015b. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accid. Anal. Prev. 78, 146-154.
Lee, J., Abdel-Aty, M., Cai, Q., 2017. Intersection crash prediction modeling with macro-level data from various geographic units. Accid. Anal. Prev. 102, 213-226.
Loukaitou-Sideris, A., Liggett, R., Sung, H., 2007. Death on the crosswalk: a study of pedestrian-automobile collisions in Log Angeles. J. Plan. Educ. Res. 26, 338-351.
Lyon, C., Persaud, B., 2002. Pedestrian collision prediction models for urban intersections. Transp. Res. Rec. 1818, 102-106.
Mannering, F.L., Bhat, C.R., 2014. Analytical methods in accident research: methodological frontier and future directions. Anal. Methods Accid. Res. 1, 1-22.
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. Accid. Anal. Prev. 43, 1624-1634.
Mooney, S.J., DiMaggio, C.J., Lovasi, G.S., Neckerman, K.M., Bader, M.D., Teitler, J.O., Sheehan, D.M., Jack, D.W., Rundle, A.G., 2016. Use of Google Street View to assess environmental contributions to pedestrian injury. Am. J. Public Health 106(3), 462-469.
Morency, P., Cloutier, M.S., 2006. From targeted “black spots” to area-wide pedestrian safety. Inj. Prev. 12, 360-364.
Naci, H., Chisholm, D., Baker, T.D., 2009. Distribution of road traffic deaths by road user group: a global comparison. Inj. Prev. 15(1), 55-59.
Noland, R.B., Quddus, M.A., 2004. Analysis of pedestrian and bicycle casualties with regional panel data. Transp. Res. Rec. 1897, 28-33.
Noland, R.B., Klein, N.J., Tulach, N.K., 2013. Do lower income areas have more pedestrian casualties? Accid. Anal. Prev. 59, 337-345.
Osama, A., Sayed, T., 2017. Evaluating the impact of connectivity, continuity, and topography of sidewalk network on pedestrian safety. Accid. Anal. Prev. 107, 116-125.
Pulugurtha, S.S., Sambhara, V.R., 2011. Pedestrian crash estimation models for signalized intersections. Accid. Anal. Prev. 43, 439-446.
Quistberg, D.A., Howard, E.J., Ebel, B.E., Moudon, A.V., Saelens, B., Hurvitz, P.M., Curtin, J.E., Rivara, F.P., 2015a. Multilevel models for evaluating the risk of pedestrian-motor vehicle collisions at intersections and mid-blocks. Accid. Anal. Prev. 84, 99-111.
Quistberg, D.A., Koepsell, T.D., Miranda, J.J., Boyle, L.N., Johnston, B.D., Ebel, B.E., 2015b. The walking environment in Lima, Peru and pedestrian-motor

vehicle collisions: an exploratory analysis. Traffic Inj. Prev. 16, 314-321.
Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical Linear Models: Applications and Data Analysis Methods (2nd Edition). Sage, Thousand Oaks.
Retting, R.A., Ferguson, S.A., McCartt, A.T., 2003. A review of evidence-based traffic engineering measures designed to reduce pedestrian-motor vehicle crashes. Am. J. Public Health 93, 1456-1463.
Rifaat, S., Tay, R., De Barros, A., 2012. Urban street pattern and pedestrian traffic safety. J. Urb. Des. 17(3), 337-352.
Schneider, R.J., Diogenes, M.C., Arnold, L.S., Attaset, V., Griswold, J., Ragland, D.R., 2010. Association between roadway intersection characteristics and pedestrian crash risk in Alameda County, California. Transp. Res. Rec. 2198, 41-51.
Sebert Kuhlmann, A.K., Brett, J., Thomas, D., Sain, S.R., 2009. Environmental characteristics associated with pedestrian-motor vehicle collisions in Denver, Colorado. Am. J. Public Health 99(9), 1632-1637.
Siddqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid. Anal. Prev. 45, 382-391.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B 64(4), 583-639.
Spiegelhalter, D.J., Thomas, A., Best, N., Lunn, D., 2005. WinBUGS User Manual. MRC Biostatistics Unit, Cambridge, United Kingdom.
Steinbach, R., Edwards, P., Green, J., 2014. Controlling for exposure changes the relationship between ethnicity, deprivation and injury: an observational study of child pedestrian injury rates in London. Inj. Prev. 20, 159-166.
Stoker, P., Garfinkel-Castro, A., Khayesi, M., Odero, W., Mwangi, M.N., Peden, M., Ewing, R., 2015. Pedestrian safety and the built environment: a review of the risk factors. J. Plan. Lit. 30(4), 377-392.
Strauss, J., Miranda-Moreno, L.F., Morency, P., 2014. Multimodal injury risk analysis of road users at signalized and non-signalized intersections. Accid. Anal. Prev. 71, 201-209.
Tasic, I., Elvik, R., Brewer, S., 2017. Exploring the safety in numbers effect for vulnerable road users on a macroscopic scale. Accid. Anal. Prev. 109, 36-46.
Thomas, L., Lan, B., Sanders, R.L., Frackelton, A., Gardner, S., Hintze, M., 2017. Changing the Future? Development and application of pedestrian safety performance functions to prioritize locations in Seattle, Washington. Transp. Res. Rec. 2659, 212-223.
Torbic, D.J., Harwood, D.W., Bokenkroger, C.D., Srinivasan, R., Carter, D., Zegeer, C.V., Lyon, C., 2010. Pedestrian safety prediction methodology for urban signalized intersections. Transp. Res. Rec. 2198, 65-74.
Ukkusuri, S., Hasan, S., Aziz, H.M.A., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Transp. Res. Rec. 2237, 98-106.
Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Saf. Sci. 50, 1141-1151.
Wang, J., Huang, H., Zeng, Q., 2017. The effect of zonal factors in estimating crash

risks by transportation modes: motor vehicle, bicycle and pedestrian. Accid. Anal. Prev. 98, 223-231.

Wang, Y., Kockelman, K.M., 2013. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accid. Anal. Prev. 60, 71-84.

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.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2011. Statistical and Economic Methods for Transportation Data Analysis (2nd Edition). CRC Press, NY.

Wedagama, D.M.P., Bird, R.N., Metcalfe, A.V., 2006. The influence of urban land-use on non-motorised transport casualties. Accid. Anal. Prev. 38, 1049-1057.

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. Accid. Anal. Prev. 41(1), 137-145.

Xie, K., Ozbay, K., Kurkcu, A., Yang, H., 2017. Analysis of traffic crashes involving pedestrians using big data: investigation of contributing factors and identification of hotspots. Risk Anal. 37(8), 1459-1467.

Xu, P., Huang, H., Dong, N., Abdel-Aty, M., 2014. Sensitivity analysis in the context of regional safety modeling: identifying and assessing the modifiable areal unit problem. Accid. Anal. Prev. 70, 110-120.

Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting. Accid. Anal. Prev. 75, 16-25.

Xu, P., Huang, H., Dong, N., Wong, S.C., 2017a. Revisiting crash spatial heterogeneity: a Bayesian spatially varying coefficients approach. Accid. Anal. Prev. 98, 330-337.

Xu, P., Xie, S., Dong, N., Wong, S.C., Huang, H., 2017b. Rethinking safety in numbers: are intersections with more crossing pedestrians really safer? Inj. Prev. doi:10.1136/injuryprev-2017-042469.

Xu, X., Xie, S., Wong, S.C., Xu, P., Huang, H., Pei, X., 2016. Severity of pedestrian injuries due to traffic crashes at signalized intersections in Hong Kong: a Bayesian spatial logit model. J. Adv. Transp. 50(8), 2015-2028.

Yao, S., Loo, B.Y.P., 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. Accid. Anal. Prev. 75, 320-332.

Young, M.S., Mahfound, J.M., Stanton, N.A., Salmon, P.M., Jenkins, D.P., Walker, G.H., 2009. Conflicts of interest: the implications of roadside advertising for driver attention. Transp. Res. Part F Traffic Psychol. Behav. 12(5), 381-388.

Yu, C., 2015. Built environmental designs in promoting pedestrian safety. Sustainability 7, 9444-9460.

Zegeer, C.V., Bushell, M., 2012. Pedestrian crash trends and potential countermeasures from around the world. Accid. Anal. Prev. 44, 3-11.

Appendix Table A1. Studies investigating the influential factors on the number of pedestrian-motor vehicle crashes at intersections in the past two decades

Authors	Study region	Study period	Observations	Research methods	Exposure measures		Risk factors included			
					Motor vehicles	Pedestrians	Geometric design	Traffic control	Land use	Demographic patterns
Leden (2002)	Hamilton, Ontario, Canada	1983-1986	749 signalized intersections 126 signalized intersections	Multiple regression model with logarithmic transform	8-hour traffic volumes	8-hour pedestrian volumes	×	×	×	×
Lyon and Persaud (2002)	Toronto, Canada	1985-1995	684 four-legged signalized intersections 263 four-Legged signalized intersections 122 three-legged unsignalized intersections	Negative binomial model	Annual average daily traffic	8-hour pedestrian volumes	×	×	×	×
Geyer et al. (2006)	Oakland, California, United States	2000-2002	247 intersections	Poisson model	Annual vehicle volumes	Estimated by Space Syntax	✓	✓	✓	×
Schneider et al. (2010)	Alameda, United States	1998-2007	81 intersections	Negative binomial model	4-hour vehicle volumes	4-hour pedestrian volumes	✓	✓	✓	✓
Tobic et al. (2010)	Charlotte, United States	1997-2005	267 signalized intersections	Negative binomial model	Annual average daily traffic	12-hour pedestrian volumes	✓	×	✓	✓
Pulugurtha and Sambhara (2010)	Charlotte, United States	2003-2007	176 signalized intersections	Negative binomial model	12-hour vehicle volumes	12-hour pedestrian volumes	✓	✓	✓	✓
Miranda-Moreno et al. (2011)	Montreal, Quebec, Canada	1999-2003	519 signalized intersections	Negative binomial, generalized negative binomial, and latent-class negative binomial models	3-hour traffic volumes	3-hour pedestrian volumes	×	✓	✓	✓

Elvik et al. (2013)	Oslo, Norway	2004-2008 2006-2010	159 marked pedestrian crossings	Negative binomial model	Annual average daily traffic	Number of crossing pedestrians	✓	✓	×	×
Strauss et al. (2014)	Montreal, Quebec, Canada	2003-2008	647 signalized intersections 435 non-signalized intersections	Bivariate Poisson model	8-hour traffic volumes	8-hour pedestrian volumes	✓	✓	✓	✓
Quistberg et al. (2015a)	Seattle, United States	2007-2013	37,360 intersections and mid-blocks	Multilevel mixed effects Poisson model	Annual average daily traffic	×	✓	✓	×	✓
Quistberg et al. (2015b)	Lima, Peru	2006	137 intersections and mid-blocks	Matched case-control design	10-minute passing vehicles	10-minute crossing pedestrians	✓	✓	×	×
Elvik (2016)	Oslo, Norway	2006-2010	389 marked pedestrian crossings	Negative binomial model	6-hour traffic volumes	6-hour pedestrian volumes	✓	✓	×	×
röyer (2016)	Sweden	2008-2012	113 intersections	Negative binomial model	3-hour traffic volumes	3-hour pedestrian volumes	✓	×	×	×
Mooney et al. (2016)	New York, United States	2007-2011	532 intersections	Negative binomial model	×	10-minute pedestrian volumes	✓	×	×	×
Lee et al. (2017)	Florida, United States	2010-2012	8,347 intersections	Mixed effects negative binomial model	Annual average daily traffic	×	✓	✓	×	✓
Thomas et al. (2017)	Seattle, United States	2007-2014	12,266 intersections	Negative binomial model	×	Estimated by “Ballpark” model	✓	✓	✓	✓
Wang et al. (2017)	Hillsborough, Florida, United States	2005-2009	279 intersections	Fixed and random parameters negative binomial	Annual average daily traffic	×	✓	✓	×	✓