The following publication Li, H., Wu, D., Graham, D. J., & Sze, N. N. (2020). Comparison of exposure in pedestrian crash analyses: A study based on zonal origin-destination survey data. Safety science, 131, 104926 is available at https://doi.org/10.1016/j.ssci.2020.104926.

Highlights

- An exposure metric based on pedestrian road crossing behavior is proposed using an integrated trip assignment approach.
- The proposed exposures and those using conventional approaches are assessed and compared.
- The road crossing based exposure approach provides the best model fit.
- Road safety measures are suggested to be widely implemented on low-grade roads.

Abstract

In recent years, the importance of exposure for pedestrian crash analysis has received increasing attention. Unlike the case of motor vehicle crashes, the definition of exposure for pedestrian crash analysis is sometimes vague and the mechanism behind the association between exposure and pedestrian crash is often ambiguously defined. In this study, the number of roads crossed and walking distance is estimated using an integrated trip assignment method at the aggregate level. The number of walking trips are also considered and compared with the distance travelled and road crossing based exposure using joint probability models. Results show that models using the road crossing based exposure approach provides the best model fit. It is found that the number of roads crossed is the most sensitive to vehicle-pedestrian collisions, as it is more strongly correlated to potential vehicle-pedestrian interactions. The results also indicate that the sensitivity of number of road crossed could vary with road types, and road safety measures for pedestrian protection should be widely implemented on low-grade roads.

Keywords: Pedestrian safety; Exposure; Road crossing; Origin-Destination data

Comparison of exposure in pedestrian crash analyses: a study based on zonal origin-destination survey data

Haojie Li[a,b,c,*], Dongyu Wu[a,b,c], Daniel J. Graham[d], N.N. Sze[e]

a School of Transportation, Southeast University, China

b Jiangsu Key Laboratory of Urban ITS, China

c Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, China

d Transport Strategy Centre, Imperial College London, UK

e Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong

*Corresponding author Email address: h.li@seu.edu.cn

Tel.: +86 13645172527

1 Introduction

Over the last few decades, considerable research have been conducted to identify risk factors related to the occurrence and severity of road crashes involving different types of victims (Miranda-Moreno et al., 2011; Palamara and Broughton, 2013; Kim, 2019; Wang et al., 2013; Lee et al., 2019; Siddiqui et al., 2012; Guo et al., 2018; Wang et al., 2017). Pedestrians are particularly vulnerable as they are directly exposed to collisions. In London, there are around 30,000 road casualties every year, 18% of which are pedestrian. The pedestrian share of total road deaths in London was more than 30% over the period 2005-2014. Furthermore, the injury severity level of pedestrians is relatively high, compared to car occupants. For example, killed and seriously injured casualties (KSIs) accounted for 17.8% and 9.2% of pedestrian casualties and overall road casualties in London respectively in 2011.

In recent years, a number of studies have attempted to measure the relevance of exposure for pedestrian safety (e.g. Siddiqui et al., 2012; Zegeer and Bushell, 2012; Yao et al., 2015; Palamara, 2016; Bao et al., 2017; Elvik and Bjørnskau, 2017; Sze et al., 2019). Unlike the case of motor vehicle crashes, the definition of exposure for pedestrian crash analysis is often vague and the data used for the exposure measurement is lacking in content.

Several aggregated level exposure measures have been proposed for pedestrian crash analysis, such as population, number of trips, and walking distance. However, the linkage between these exposures and pedestrian crashes remains unclear. In addition, the majority of studies only consider the walk-only trips. Even if data is available, the walking trip legs of the multi-modal trips are often ignored because detailed travel data are often not available. This study aims to develop appropriate surrogate measures for the exposure of pedestrian crashes at the aggregate level (i.e. traffic analysis zone, ZIP code, and county, etc.). An integrated trip assignment approach based on gravity model is proposed to estimate the pedestrian exposure at the aggregated level For instance, by integrating the land use and point of interests (POI) data, the potential origin and destination of a walking trip can be identified, and then the walking path is assigned for every (zonal) origin-destination (OD) trip. Hence, the exposure to pedestrian crash can be estimated by computing the walking distance and number of road crossings of the assigned walking path. Separate joint probability models for KSI crashes are then developed using different exposure measures.

This paper is organized as follows. A review of literature on pedestrian safety is provided in the next section. The method and data used for analysis are described in Sections 3. The results are presented in Section 4. Discussion and conclusions are provided in the final section.

2 Literature review

In previous studies, a number of explanatory variables (as shown in Table 1) have been investigated in pedestrian safety analysis models. For example, significant correlations have been found between the occurrence of pedestrian casualties and traffic volume (Mooney et al., 2016; Pulugurtha and Sambhara, 2011; Strauss et al., 2014). An increase in pedestrian and vehicular volume is related to more pedestrian casualties. In terms of the safety impacts of road network features, road density is found to be positively correlated with crash frequency due to the increased potential for vehicular-pedestrian interactions (Wang et al., 2016). The number of pedestrian crossing facilities are also associated with crash frequency. In particular, signalized crossings have lower risk of pedestrian crash, and the zebra crossings without any traffic calming measure are the most hazardous (Lam et al., 2014). In addition, it was found that the meshedness coefficient could better capture the characteristics of traffic analysis zone (TAZ) network than traditional network structure variables (Wang et al., 2013). Studies have also attempted to explore the relationship between pedestrian crashes and socio-economic characteristics. It has been found that higher population, population density and employment rates are associated with more frequent pedestrian casualties (Quistberg et al., 2015; Wang et al., 2017; Yao et al., 2015). Also, number of vehicle-pedestrian collisions for commercial and residential land use were found higher than other land uses (Wier et al., 2009). The safety impacts of age have long been investigated, and results consistently indicate that older people and children are more prone to road casualties because of their slow walking speed, less physical capacity and poor risk perception (Lee and Abdel-Aty, 2005; Sze et al., 2019).

Table 1 Previous studies on explanatory variables for pedestrian crashes

Category	Variables	References
Pedestrian exposure	Trip number, pedestrian volume	Lam et al., 2014; Lee et al., 2019; Mooney et al., 2016; Pulugurtha and Sambhara, 2011;
Vehicular exposure	Average Annual Daily Traffic (AADT), vehicle kilometers travelled (VKT)	Huang et al., 2010; Osama and Sayed, 2017; Pei et al., 2012; Strauss et al., 2014
Road network features	Road density, node density, road network structure	Abdel-Aty et al., 2013; Li et al., 2019; Wang et al., 2013
Environmental characteristics	Number of intersections, intersection density, number of roundabouts, signal density	Lam et al., 2014; Wang et al., 2016; Sze et al., 2019
Socio-demographic characteristics	Population, population density, employment rate, age structure, land use	Lee et al., 2019; Papadimitriou et al., 2016; Quistberg et al., 2015; Wang et al., 2017; Yao et al., 2015

In pedestrian crash analyses, exposure should feature as a crucial metric to reflect the potential for a pedestrian to be involved in a harmful collision a priori. A variety of exposure measures for pedestrian crash analysis have been proposed, both at aggregate and disaggregate levels. At the aggregate level, population, as well as population density are widely adopted as the surrogate for pedestrian exposure because the data can be easily obtained (Chimba et al., 2018; Wang et al., 2016). Such measures, however, do not consider the variability of travel activities among different individuals and geographical locations, and thus may lead to bias in estimation. Another typical aggregate approach is to use trip-based measures, such as the number of trips or trip frequency (Bao et al., 2017). However, the majority of conventional trip-based studies only consider walk-only trips and the walking trip legs in the multi-modal trips are ignored. Thus, pedestrian exposure is typically under-estimated when using such measure. In micro-level studies (Elvik and Bjørnskau, 2017; Xie et al., 2018), point-based pedestrian exposure can be estimated by counting the frequency of pedestrian crossings at designated measurement points (i.e. intersections and zebra crossing). This approach can provide more precise estimates of pedestrian exposure. In macro-level studies, however, there are indefinite locations, where vehicle-pedestrian collisions can occur, in the road network. Therefore, it may not be technically and financially feasible to adopt point-based approach for zonal level studies. Nevertheless, conventional vehicle exposures, such as VKT and AADT, are also employed in pedestrian crash studies (Lee et al., 2017; Osama and Sayed, 2017).

In earlier studies, walking time and distance are adopted as exposure directly (Keall, 1995; Lee and Abdel-Aty, 2005). In recent years, individual-based measurements of pedestrian exposure have been developed based on household travel survey data (e.g. Lee and Abdel-Aty, 2005; Sze et al., 2019;

Yao et al., 2015), that consists of the detailed trip data of each individual, including walking time, walking distance, trip ODs, and personal characteristics, i.e. age, gender, education and occupation. By integrating the trip and road network data, several recent researches (Lam et al., 2014; Yao et al., 2015) have adopted space-time path (STP) and potential path tree (PPT) approaches to predict walking path, and thus estimate the pedestrian exposure. The STP approach is confined to the shortest path, while the PPT approach has the capability to consider all possible paths of a pedestrian. Results indicated that both the STP and PPT approaches are well-performed in predicting the pedestrian exposure. While the STP approach requires less data and computation time, the PPT approach outperforms the STP approach in describing the underlying vehicle-pedestrian collision pattern (Lam et al., 2014).

In terms of the models for road safety analysis, count data models, including the Poisson and negative binomial regression models, have been frequently used and developed in previous studies to estimate crash occurrence (i.e., Guo et al., 2018). Crash severity is also an important topic on road safety researches. To investigate the relationship between crash severity and the explanatory factors, a common practice is to estimate crash occurrence for different severity levels. However, this method may suffer from the homogeneity between crash count and severity. To address this issue, researchers have proposed a series of advanced modeling techniques, such as multivariate model (song et al., 2006) and joint probability model (Pei et al., 2011).

To sum up, in the absence of appropriate pedestrian exposure data, proxies such as population, employment and the number of trips are still used in pedestrian crash studies. For example, Ferenchak and Marshall (2019) indicate that they did not account for exposure as reliable pedestrian count data was not available. Compared to traditional exposure metrics, although individual-based exposures can better reflect the potential in involving pedestrian-vehicle collisions, the data required is often difficult and costly to collect. Therefore, it is important to make efficient use of available aggregate-level data to develop reliable and robust pedestrian exposure metrics. Furthermore, another issue is the choice of formulation used to describe the relationship between pedestrian activities, exposures and occurrence of vehicle-pedestrian collisions. In this study, an integrated trip assignment approach is proposed to calculate the exposure metrics, and the sensitivity of different pedestrian exposure metrics to pedestrian crash frequency is also compared.

3 Methodology

3.1 Data preparation

Aggregate pedestrian crash prediction models used in this study are established at the Middle Super Output Area (MSOA) level in London. An MSOA is a spatial aggregation of census blocks with an average population of 8,000 people. Data on pedestrian crashes involving three types of severity (slight injured, serious injured and killed) in 2011 were obtained from STATS19. Specifically, slight injured crash includes sprains, whiplash, bruises, slight cuts and slight shock not necessarily requiring medical treatment. Serious injured crash includes fractures, severe cuts, internal injury, concussion and where a person is detained in hospital or dies in 30 or more days. Fatal injured crash includes only those cases where death occurs in less than 30 days as a result of the accident (TfL, 2003). Since serious injured crash and fatal injured crash only share a proportion of 1.4% and 16.6% respectively, they are classified as KSI crash in this paper. Socio-demographic data, such as population, employment, land use and age structure were obtained from Office for National

Statistics (ONS). The land use data include the proportion of domestic land (e.g. residential area), non-domestic land (e.g. business and office district area) and greenspace (e.g. park and garden). The road network data were obtained from Ordnance Survey (OS) Meridian^{TM} 2. The road network features considered in this study include road density and meshedness coefficient. Meshedness coefficient is used to measure the structure of closed path in a road network. Greater meshedness coefficient is associated with higher degree of clustering and more intersections. Road network data were aggregated at the MSOA level and matched with the crash and socio-demographics data using a Geographical Information System (GIS), e.g. ArcGIS.

To estimate pedestrian exposure, the principal data source used is the 2011 census OD data. The OD data (also known as flow data or interaction data) consists of counts of flows between two locations at different spatial scales (e.g. MSOA). In the case of census OD data, the moves within an area can be also identified. The most common flow data relate to commuting flows. Land use information is obtained from the OpenStreetMap. It provides the Point of Interest (POI), e.g. restaurants, shopping malls, hotels, schools and gyms, etc., block polygon and the locations of public transport stations. Each block polygon and POI is labeled with a land use classification (e.g. residential, industrial, commercial, government, and green area etc.), those blocks labeled with residential class would be treated as the potential origin of a commuter trip, and those blocks or POIs labeled with land use classes related to various travel purposes, i.e. school, work and others, would be treated as the potential destination of a commuter trip. Table 2 summarizes the variables included in this study.

Table 2 Descriptive statistics of variables

Variables	Min	Max	Mean	S.D.
Dependent variables				
Total pedestrian casualties count	0	116	5.51	7.25
Slightly injured pedestrian casualties count	0	95	4.53	6.06
KSI pedestrian casualties count	0	21	0.98	1.54
Exposure variables				
Trip number	80	17254	592.19	819.74
Distance travelled (km)	125.22	21736.98	786.71	15129.83
Number of road crossed	580.51	101733.81	8479.66	9408.30
Road network variables				
Number of Class A Road crossed (10^3)	0	7.40	0.69	0.85
Number of Class B Road crossed (10^3)	0	5.87	0.22	0.44
Number of Minor Road crossed (10^3)	0.048	43.45	3.96	3.66
Class A Road density (km/km^2)	0	6.48	1.39	1.15
Class B Road density (km/km^2)	0	5.33	0.45	0.65
Minor Road density (km/km^2)	1.31	23.30	8.83	3.31
Meshedness Coefficient	0	0.19	0.08	0.03
Socio-demographic variables				
Population density (persons/hectare)	2.9	251.6	84.7	48.66
Employment rate (%)	82.6	97.5	92.6	7.99
Domestic	0.011	0.313	0.125	0.05
Non-domestic	0.004	0.380	0.063	0.05
Greenspace	0.013	0.871	0.251	0.17
People aged below 15 years (%)	6.04	35.84	19.81	4.10
People aged above 64 years (%)	3.09	27.18	11.23	4.10

3.2 Pedestrian exposure estimation

In this study, trips involving walking activities are used for the computation of the pedestrian exposure for each MSOA. Three metrics are adopted in this study to reflect the pedestrian exposure, (1) trip frequency (2) distance travelled (DT), and (3) number of road crossed (NRC). We first assign the trips to all possible walking paths between origin MSOA to destination MSOA. Then the pedestrian exposures of every walking path in term of DT and NRC are estimated. Finally, the exposures are aggregated at the MSOA level.

3.2.1 Estimation of the trip number for each walking path

In order to obtain the potential walking paths from origin MSOA to destination MSOA, the geometry centroid of a residence block is defined as the potential origin of a commuter trip, and the concerned POI is regarded as the potential destination. As discussed before, the majority of trip-based studies ignore the walking trip legs to and from public transport stations, and thus pedestrian exposure is underestimated. To address this problem, both the (a) walk-only trip and (b) walking trip legs of multi-modal trip are considered here. The latter one includes at least two walking components, one from the origin to public transport station and another from the public transport station to the destination. It is assumed that pedestrians are willing to walk up to 1500 meters to and from the public transport stations for multi-modal trips, and 5000 meters for walk-only trips respectively.

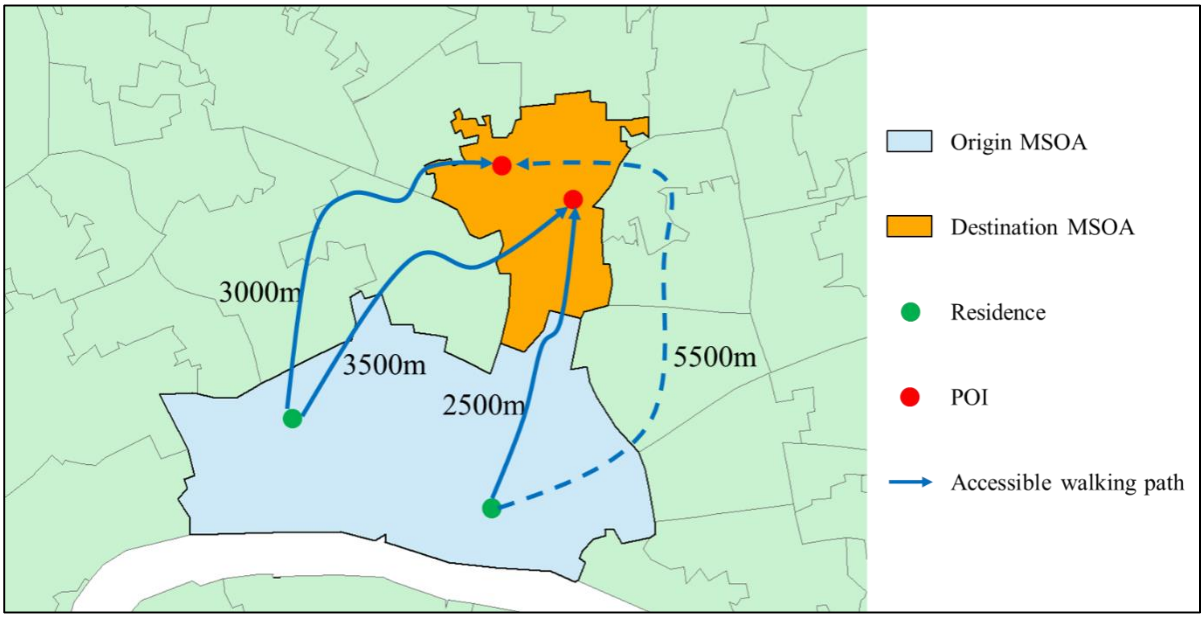

Fig. 1 Accessible walking paths between MSOAs

Assuming that there are b_1 potential origins in the origin MSOA and b_2 potential destinations in the destination MSOA. The number of possible OD pairs is therefore $b_1 \times b_2$. For each OD pair, the walking path is defined based on the shortest path principle and estimated using ArcGIS, and those paths of distance less than 5000 meters are defined as accessible (see Fig.1). Since the total walking trips among MSOAs are known, the number of trips t_i for each path *i* (*i*=1,2,3, …, *n*) can be calculated using the gravity model:

$$t_i = T \times \frac{\frac{p_i e_i}{d_i}}{\sum_{i=1}^{n} \frac{p_i e_i}{d_i}} \tag{1}$$

where T is the total number of walking trips from origin MSOA to destination MSOA, n is the number of accessible walking paths, d_i is the distance of path *i*, p_i is the population density of the Lower Super Output Area (LSOA, a lower-level unit of British administrative with an average population of 1500) at which the origin of path *i* is located, e_i is the employment rate of the LSOA at which the destination of path *i* is located.

As for multi-modal trips, all public transport stations located in the area that is less than 1500 meters from a particular OD will be considered and matched using the GIS approach. The number of trips for each accessible walking path *k* can be calculated using the same approach.

3.2.2 Pedestrian exposure measures

Several studies (Lee and Abdel-Aty, 2005; Papadimitriou, 2016) have indicated that walking time could be a better exposure than walking distance. However, in order to obtain a reliable estimate of walking time, it is necessary to obtain the information on individual characteristics that are closely related to his/her own walking speed (i.e. age, gender). Such information is often not available in conventional macro-level studies. For the estimation of DT, the distance of the shortest walking path of an OD is adopted as the pedestrian exposure. It is often the case that the origin and/or destination are not located on a road when calculating the distance using GIS tools. Then, the ODs are projected to the nearest road. Therefore, the shortest walking distance can be calculated if the road network is coded.

Traditional distance-based methods assume that pedestrian-vehicle collisions could happen anywhere on the road network. However, unlike vehicles, pedestrians are relatively safe when walking along the sidewalk. As shown in Fig. 2, the shortest walking distances from location O to D1 is longer than that from O to D2, but pedestrians walking from O to D1 could be safer as they do not need to cross any road. This suggests that the pedestrian-vehicle collision risk might not necessarily correlate to the walking distance or walking time. In reality, pedestrians may not be exposed to any road crash unless they enter the motor vehicle lanes. In previous studies (Lam et al., 2013; Lam et al., 2014), the walking distance is usually calculated in a network-constrained environment, where the pedestrians are assumed to enter the motor vehicle lanes only when passing through road junctions, and the crossing behaviors occurring at mid-block locations or outside of a crosswalk (i.e. jaywalking) cannot be captured.

In order to address the above issues, the NRC approach is proposed to estimate the least possible number of road crossed in a particular path. Two assumptions are made when counting the number of road crossed in this study. (1) Diagonal crosswalks are not taken into account. Pedestrians can cross the intersection obliquely to complete the crossing of two streets using diagonal crosswalk. Therefore, the number of road crossed will be one rather than two in such case. (2) We did not consider the underpass or overpass in this study. Both assumptions may lead to the over-estimation of the number of road crossed. However, since diagonal crosswalks and underpass/overpass are very rare in London, such influence is not significant. As shown in Fig. 2, line AB presents the shortest path form location A to B without network constraints. The red dots indicate the intersections between the road and line AB, thus possible location (and number) of road crossings. The number of such intersection points on the shortest path is therefore the least possible number of road crossed. In Fig. 2, line AB intersects with two roads at point G, therefore pedestrians need to cross twice (G1 and G2) to get destination B. Hence, the least possible number of road crossed for a walking trip from A to B is six. In addition, crash involvement rate may vary with road types. Therefore, the number of crossing behavior in three road classes (Class A Road, Class B Road and Minor Road) are also computed respectively.

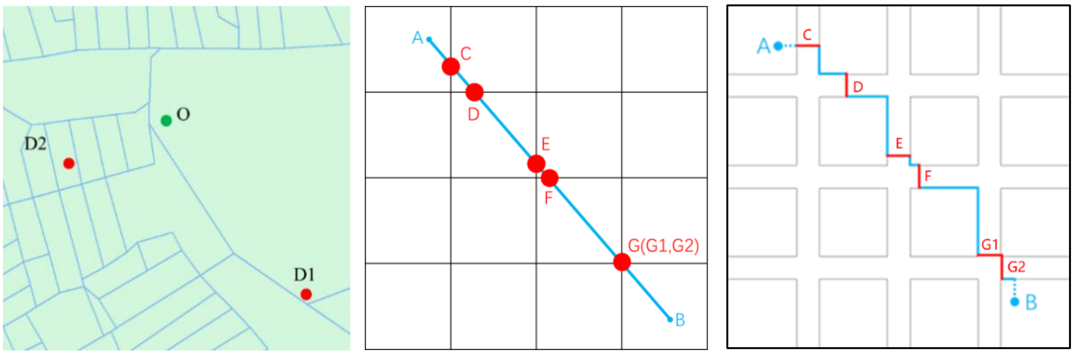

Fig. 2 An illustration of the NRC method

As shown in Fig. 3, the path from A to B passes through three MSOAs. The distances travelled in each of the 3 MSOAs are 600m, 2000m, and 500m, while the number of road crossed are 1, 4, and 1 respectively. Assuming that 50 trips are assigned to this path, then the pedestrian exposures for this path are 600×50, 2000×50 and 500×50 for each MSOA when using the DT approach, and 1×50, 4×50 and 1×50 when using the NRC approach. The total pedestrian exposure for MSOA*p* (*p=1,2,...,n*, where *n* is the total number of MSOA) can be calculated as:

$$E_p = \sum_{i=1}^{n} e_{ip} \times t_i \tag{2}$$

where *n* is the number of walking paths starting from, passing through, and ending at MSOA*p*, e_{ip} presents the walking distance or number of road crossings for path *i* in MSOA*p*, and t_i is the number of trips assigned to walking path *i*.

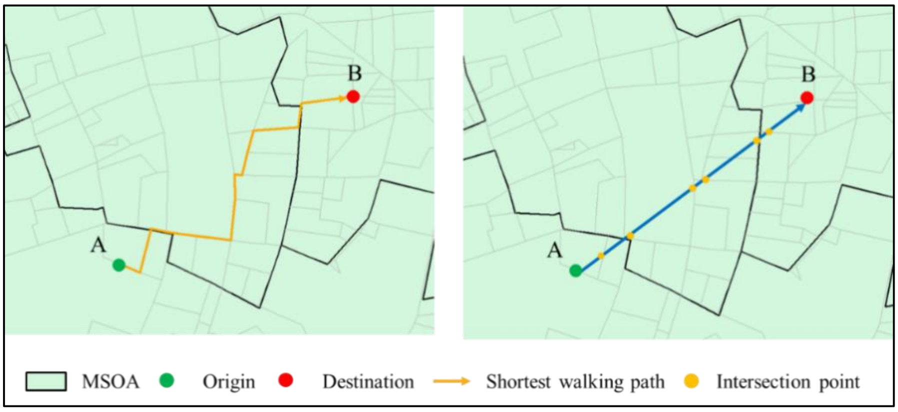

Fig. 3 An illustration of the calculation of exposure at MSOA-level

3.3 Model

To deal with the issue of the homogeneity between crash severity and frequency, the joint-probability approach proposed by Pei et al. (2011) was adopted.

3.3.1 Probability function for crash occurrence

Poisson model is one of the most prevalent models applied to measure the association between crash frequency and risk factors. The traditional Poisson regression model is defined as follows:

$$ln(\lambda_i) = \alpha + ln(E_i) + \beta X_i \quad (3)$$

where λ_i is the expected number of crashes in $MSOA_i$, E_i is the exposure in $MSOA_i$, α is the intercept, β represent the parameters to be estimated and X_i is the vector of explanatory variables. The basic assumption of Poisson model is that the variance of crash data equals to the mean. We used the test suggested by Cameron and Trevedi (1990) to investigate the over-dispersion in the data. The test results suggest significant over-dispersion in our data. To address this issue, a negative binomial regression model with a Gamma distributed error term is employed (Hilbe, 2007). The negative binomial regression model used in this research can be written:

$$ln(\lambda_i) = \alpha + ln(E_i) + \beta X_i + \varepsilon_i \quad (4)$$

$$\varepsilon_i \sim \text{Normal}\ (0, \tau^2) \quad (5)$$

where ε_i is the expectation of unobserved heterogeneity error component, and τ^2 is assigned as gamma distribution.

3.3.2 Probability function for crash severity

In this paper, crash severity is separated into KSI crashes and slight injured crashes. For crash severity upon occurrence, the hierarchical binomial-logistic approach is considered to model the relationship between crash severity and explanatory factors. As proposed by Pei et al. (2011), the number of KSI crashes k_i follows a binomial distribution

$$k_i \sim \text{Binomial}(p_i, \lambda_i) \tag{6}$$

where λ_i denotes the total crash number, p_i is the probability of KSI crash. Then a logit function can be established as

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \theta X_i \tag{7}$$

where θ is the vector of coefficients.

3.3.3 Joint probability

Integrating the probability function for crash occurrence (Eq. (4)) and crash severity (Eq. (7)), the joint probability function of having λ_i total crashes and k_i KSI crashes can be defined as

$$\text{P}(\lambda_i, k_i) = f(\beta X_i) \times f(\theta X_i) \tag{8}$$

In this study, three joint probability models are established to measure the association between pedestrian crash frequency and possible risk factors. Trip number, walking distance, and the number of roads crossed are adopted as the exposure in the three models respectively. The performances of proposed models are assessed using Akaike information criterion (AIC) and Bayesian information criterion (BIC).

4 Results

4.1 Comparison of exposures

The correlation between involved explanatory variables has been checked using Pearson correlation coefficient test. The results are shown in Appendix A. It is found that variables domestic and greenspace exist significant correlation with other variables, so we delete them from the final models. The model results for three joint probability models are presented in Table 3-5 respectively. The pedestrian exposure was proxied by the trip number in Model 1, distance travelled in Model 2 and the number of road crossed in Model 3. The results show that Model 3 has the best fit with the lowest AIC and BIC, indicating that models using the number of road crossed as the proxy for pedestrian exposure have the best performance. In addition, both Model 2 and Model 3 are superior to Model 1, suggesting that the proposed integrated trip assignment approach can provide more efficient exposure than simple trip number. All three proposed pedestrian exposures are significant and positively correlated to pedestrian crashes. The strength of association between exposures and pedestrian crashes are assessed using z values (Yao et al., 2015). Regarding the results of crash occurrence, the NRC based exposure has the highest z value (6.6), while the z values for DT and trip number are 6.48 and 3.84 respectively, indicating that pedestrian crash is the most sensitive to number of roads crossed.

Besides, the prediction performance of three models are also evaluated. The original dataset was randomly separated into a training set and a testing set with a ratio of 4:1. Three evaluation indicators, mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE), are calculated as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{9}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{11}$$

Where *N* is the predicted number, y_i is the real value of MSOA$_i$, and $\hat{y}_i$ is the predicted value of MSOA$_i$. The results of the three indicators are showed in Table 6. It is indicated that models using the exposures estimated based on zonal OD data (distance travelled and number of road crossed) perform better than the model using the trip number as the exposure. And Model 3 using the number of road crossed outperforms other models in terms of the prediction accuracy.

In this study, we further investigate the association between pedestrian casualties and road classes. In particular, numbers of Minor Road crossed were found positively correlated to crash frequency. It can be speculated that the high pedestrian-vehicle collision risk on Minor Road is attributed to the absence of central medians and pedestrian barriers. Additionally, no evidence could be established for the association between number of high-grade roads (Class A Road and Class B Road) crossed and pedestrian crash, though high-grade roads are usually considered as hazardous to pedestrian for higher vehicular speed. Yet, illegal and unsafe pedestrian crossing behavior can be deterred because of the prevalent of central median. Moreover, traffic signal and segregated crossing (i.e. underpass and footbridge) are prevalent on high-grade roads, therefore, providing more protections for pedestrians.

4.2 Estimation Results

As shown in Table 3-5, the parameter estimates of other explanatory variables are similar across all models. Regarding crash severity, the results showed that all pedestrian exposure variables was negatively associated with the proportion of KSI crashes in total crashes. This is similar with the findings from Pei et al. (2011), which indicated that the occurrence of fatal crashes increases at a lower rate than that of total crashes as exposure increase. Another significant factor is Class A Road density. It can be speculated that the crashes on high-grade roads are more severe due to higher vehicle speed.

In terms of crash occurrence, the results indicate that densities of Class A Road and Class B Road are positively correlated with crash frequency, while density of Minor Road is negatively correlated with crash frequency. This is consistent with the findings of previous study (Graham and Glaister, 2003). Meshedness is also found to be positively correlated with crash frequency. As a topological feature of road network, meshedness coefficient is associated with the number of meshes. It can be speculated that road networks with higher meshedness coefficient may imply more conflicts (Wang et al., 2013). As for socio-economic variables, population density and employment are positively correlated with the occurrence of vehicle-pedestrian collisions in all models. This is again consistent with that of previous findings (Khondakar et al., 2009; Quddus, 2008). Increases in population density and employment can imply more overall travel activities. It is suggested that higher level of travel activities may imply higher likelihood of crash involvement (Marshall and Garrick, 2011). This could eventually increase the probability of pedestrian-vehicle crashes. Age structure was also found significantly correlated with crashes. Consistent with the findings of previous studies (Lee and Abdel-Aty, 2005; Sze et al., 2019), higher proportions of children and elderly people were correlated to more pedestrian crashes. It is because age can affect the mean walking speed, risk perception and risk-taking behavior. They all in turn affect the crash risk of pedestrian. Especially, children have a higher tendency to engage in reckless behavior on roads. For the elderly pedestrians, the higher crash risk could be attributed to the lower cognitive performance and impaired mobility (Palamara and Broughton, 2013; Palamara, 2016). For instance, elderly pedestrians have slower reaction and require longer time gap for crossing. Non-domestic land use are found to be positively related to the number of vehicle-pedestrian collisions in all models. This may be due to the higher pedestrian flow and frequent jaywalking behavior in non-domestic areas.

Table 3 Estimation results of the model using the trip number as exposure

	Crash occurrence			Crash severity		
Variable	Coef.	Std. Err.	z	Coef.	Std. Err.	z
log(Trip)	0.001	0.000	3.43	-0.213	0.106	-2.01
Population density	0.004	0.001	4.35	-0.123	0.111	-1.11
Employment rate	0.077	0.012	6.47	-0.003	0.042	-0.08
Class A Road density	0.210	0.026	8.13	7.715	2.182	3.54
Class B Road density	0.106	0.041	2.57	0.058	0.158	0.37
Minor Road density	-0.052	0.011	-4.65	0.043	0.041	1.05
Number of Class A Road crossed	0.091	0.481	0.19	1.177	0.836	1.41
Number of Class B Road crossed	0.302	0.368	0.82	-1.502	1.482	-1.01
Number of Minor Road crossed	0.670	0.219	3.06	1.448	1.629	0.89
Meshedness coefficient	0.048	0.011	4.33	-2.189	2.529	-0.87
People aged below 15 years (%)	0.021	0.007	2.96	-0.025	0.028	-0.87
People aged above 64 years (%)	0.047	0.008	6.15	-0.001	0.025	-0.03
Non-Domestic	9.398	1.207	7.79	7.259	1.245	5.83
Constant	0.004	0.001	4.35	-0.123	0.111	-1.11
AIC=4255.073 BIC=4326.789						

Table 4 Estimation results of the model using the distance travelled as exposure

	Crash occurrence			Crash severity		
Variable	Coef.	Std. Err.	z	Coef.	Std. Err.	z
log(Distance travelled)	0.227	0.032	7.15	-0.215	0.105	-2.04
Population density	0.003	0.001	3.3	0.106	0.109	0.97
Employment rate	0.076	0.011	6.64	-0.012	0.042	-0.28
Class A Road density	0.199	0.025	8	7.097	2.240	3.17
Class B Road density	0.103	0.040	2.61	0.062	0.158	0.39
Minor Road density	-0.050	0.011	-4.67	0.044	0.041	1.07
Number of Class A Road crossed	0.455	0.473	0.96	1.099	0.831	1.32
Number of Class B Road crossed	0.449	0.356	1.26	-1.359	1.477	-0.92
Number of Minor Road crossed	0.531	0.212	2.5	1.708	1.642	1.04
Meshedness coefficient	0.048	0.011	4.45	-1.999	2.523	-0.79
People aged below 15 years (%)	0.017	0.007	2.54	-0.028	0.028	-1
People aged above 64 years (%)	0.047	0.007	6.47	0.001	0.025	0.06
Non-Domestic	6.647	1.221	5.44	5.591	1.244	4.5
Constant	0.003	0.001	3.3	0.106	0.109	0.97
AIC=4220.102 BIC=4291.818						

Table 5 Estimation results of the model using the number of road crossed as exposure

	Crash occurrence			Crash severity		
Variable	Coef.	Std. Err.	z	Coef.	Std. Err.	z
log(Number of road crossed)	0.221	0.030	7.34	-0.210	0.105	-1.99
Population density	0.003	0.001	3.47	-0.097	0.107	-0.9
Employment rate	0.071	0.011	6.21	-0.011	0.042	-0.26
Class A Road density	0.182	0.025	7.26	6.992	2.283	3.06
Class B Road density	0.089	0.040	2.25	0.057	0.158	0.36
Minor Road density	-0.049	0.011	-4.52	0.043	0.041	1.07
Number of Class A Road crossed	0.452	0.472	0.96	1.090	0.832	1.31
Number of Class B Road crossed	0.475	0.356	1.33	-1.365	1.477	-0.92
Number of Minor Road crossed	0.505	0.212	2.38	1.701	1.641	1.04
Meshedness coefficient	0.047	0.011	4.35	-1.994	2.523	-0.79
People aged below 15 years (%)	0.014	0.007	2.04	-0.027	0.028	-0.96
People aged above 64 years (%)	0.043	0.007	5.88	0.003	0.025	0.11
Non-Domestic	1.684	0.521	3.23	-0.007	0.003	-1.9
Constant	7.186	1.198	6	6.253	1.226	5.1
AIC=4206.685 BIC=4278.772						

Table 6 Predictive performance of three models

Dataset	Model	MAE	MAPE(%)	RMSE
Training set	Model 1	1.098	17.473	1.532
	Model 2	1.068	17.071	1.466
	Model 3	1.051	16.791	1.445
Testing set	Model 1	1.093	17.599	1.511
	Model 2	1.065	17.080	1.428
	Model 3	1.052	16.990	1.395

5 Discussion and conclusion

Measurement of pedestrian exposure has long been an important research topic in pedestrian safety studies. In the absence of appropriate macro-level pedestrian exposure data, many previous studies have to use population, employment, and the number of trips as proxies. However, such measures may lead to bias in estimation results. In addition, the mechanism in terms of how pedestrian exposure influence is also unclear. Therefore, we propose an integrated trip assignment approach to compute macro-level pedestrian exposure (distance travelled and number of road crossed) based on available zonal survey data. This paper develops an aggregated level pedestrian crash model based on population, road network, traffic and crash data in London. We hope that this contributes to the literature as a more efficient surrogate measure for pedestrian exposure.

Three pedestrian crash prediction models using different exposure metrics were established and estimated. Among these exposure metrics, trip number is available from 2011 census OD data, while distance travelled and number of road crossed are computed using an integrated trip assignment approach based on gravity model. The model using distance travelled (Model 2) and number of road crossed (Model 3) as exposure are superior to that using number of trip (Model 1). This suggests that the proposed integrated trip assignment approach can provide more efficient exposure and higher prediction accuracy than simple trip number exposure. We also find that the vehicle-pedestrian collision is most sensitive to the number of road crossed, as it directly reflects the opportunities of vehicle-pedestrian conflicts. For the effect of road type, results indicate that it is more hazardous to cross Minor Roads. A possible reason is that pedestrians tend to opt for a protected crossing (e.g. signalized crossing) on principal roads because of various engineering and physical constraints, i.e. prevalent of median separation, high vehicular speed, wider road. This is consistent with the results of previous studies that signalized crossings with the absence of traffic calming measures could have higher pedestrian crash risk (Lam et al., 2014). Also, reckless crossing is less prevalent (although still observable) on principal roads, but more frequent on Minor Roads (e.g. Papadimitriou, 2016).

Our findings are informative for the planning of road infrastructure and traffic control measures for pedestrian safety. Since the number of road crossed is more correlated to potential vehicle-pedestrian interactions than the distance travelled. It is suggested that the transport infrastructures (e.g. public transport stations) should be constructed in a way that can decrease the number of road crossed or guide pedestrians to cross the roads with traffic calming measures (even if this may increase walking distance). Besides, road infrastructures (e.g. median separations), traffic signs and other traffic calming measures should be widely implemented on low-grade roads to reduce unsafe crossing behaviors (e.g. jaywalking). In addition, road user education and enforcement measures, especially for pedestrians, may also help to enhance road safety (Huang et al., 2010).

There are several limitations of our study that can be addressed in future work. When assigning trips to potential paths, only the population density of the origin, the employment rate of destination and the path distance between ODs were considered in the gravity model. Given the availability of data, other factors should be also considered in future studies. For example, since the capability to attract walking trips may vary across different POI, the attraction coefficient of POI can be included in the model. In addition, the potential walking paths are selected using only the shortest path principle for simplicity. The route choice behavior of pedestrians can of course be more complex and this is worth exploring in future work.

Acknowledgements

This work was supported by National Key R&D Program of China [Grant No.2018YFE0102700], the National Natural Science Foundation of China [Grant No. 71701042], the Key Project of National Natural Science Foundation of China [Grant No. 51638004], and the grants from the Research Grants Council of Hong Kong (15209818).

References

Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. Transp. Res. Part A: Policy Pract. 49, 62–75.

Bao, J., Liu, P., Yu, H., Xu, C., 2017. Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas. Accid. Anal. Prev. 106, 358–369.

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.

Cameron, A., Trevedi, P., 1990. Regression based tests for overdispersion in the Poisson model. J. Econ. 46, 347–364.

Chimba, D., Musinguzi, A., Kidando, E., 2018. Associating pedestrian crashes with demographic and socioeconomic factors. Case Stud. Transp. Policy 6, 11-16.

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

Ferenchak, N., Marshall, W., 2019. Suppressed child pedestrian and bicycle trips as an indicator of safety: Adopting a proactive safety approach. Transp. Res. Part A: Policy Pract. 124, 128–144.

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

Guo, Y., Li, Z., Wu, Y., Xu, C., 2018. Exploring unobserved heterogeneity in bicyclists' red-light running behaviors at different crossing facilities. Accid. Anal. Prev.115, 118-127.

Hilbe, J.M., 2007. Negative Binomial Regression. Cambridge University Press, Cambridge.

Huang, H., Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. Accid. Anal. Prev. 42, 1556–1565.

Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-level crash risk analysis in Florida: bayesian spatial modeling. Transp. Res. Rec.: J. Transp. Res. Board 27–37.

Keall, M.D., 1995. Pedestrian exposure to risk of road accident in New Zealand. Accid. Anal. Prev. 27 (5), 729–740.

Khondakar, B., Sayed, T., Lovegrove, G., 2009. Transferability of community based macro-level collision prediction models between different time-space regions for use in road safety planning application. In: Compendium of Papers CD-ROM. Transportation Research Board 88th Annual

Meeting, Washington, DC.

Kim, D., 2019. The transportation safety of elderly pedestrians: Modeling contributing factors to elderly pedestrian collisions. Accid. Anal. Prev. 131, 268-274.

Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. Accid. Anal. Prev. 37, 775–786.

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.

Lee, J., Abdel-Aty, M., Shah, I., 2019. Evaluation of surrogate measures for pedestrian trips at intersections and crash modeling. Accid. Anal. Prev. 130, 91-98.

Lee, J., Abdel-Aty, M., Xu, P., Gong, Y., 2019. Is the safety-in-numbers effect still observed in areas with low pedestrian activities? A case study of a suburban area in the United States. Accid. Anal. Prev. 125, 116-123.

Lam, W.W.Y., Loo, B.P., Yao, S., 2013. Towards exposure-based time-space pedestrian crash analysis in facing the challenges of ageing societies in Asia. Asian Geogr. 30, 105–125.

Lam, W.W.Y., Yao, S., Loo, B.P.Y., 2014. Pedestrian exposure measures: A time-space framework. Travel Behav. Soc. 1, 22-30.

Marshall, W.E., Garrick, N.W., 2011. Does street network design affect traffic safety? Accid. Anal. Prev. 43 (3), 769–781.

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 (5), 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.

Osama, A., Sayed, T., 2017. Evaluating the impact of connectivity, continuity, and topography of s idewalk network on pedestrian safety. Accid. Anal. Prev. 107, 117-125.

Palamara, P., Broughton, M., 2013. An investigation of pedestrian crashes at traffic intersections in the Perth Central Business. J. Public Health 93, 1456–1463.

Papadimitriou, E., 2016. Towards an integrated approach of pedestrian behaviour and exposure. Accid. Anal. Prev. 92, 139-152.

Pei, X., Wong, S.C., Sze, N.N., 2012. A joint-probability approach to crash prediction models. Accid. Anal. Prev. 43, 1160–1166.

Pei, X., Wong, S.C., Sze, N.N., 2012. The roles of exposure and speed in road safety analysis. Accid. Anal. Prev. 48, 464–471.

Pulugurtha, S.S., Sambhara, V.R., 2011. Pedestrian crash estimation models for signalized intersections. Accid. Anal. Prev. 43, 439–446.

Quddus, M.A., 2008. Modeling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accid. Anal. Prev. 40(4), 1486–1497.

Quistberg, D.A., Howard, E.J., Ebel, B.E., Moudon, A.V., Saelens, B., Hurvitz, P.M., Curtin, J.E., Rivara, F.P., 2015. Multilevel models for evaluating the risk of pedestrian-motor vehicle collisions at intersections and mid-blocks. Accid. Anal. Prev. 84, 99–111.

Siddiqui, C., Abdel-Aty M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid. Anal. Prev. 45, 382–391.

Song, J.J., Ghosph, M.,Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. J. Multivariate Anal. 97 (1), 246–273.

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.

Sze, N.N., Su, J., Bai, L., 2019. Exposure to pedestrian crash based on household survey data: Effect of trip purpose. Accid. Anal. Prev. 128, 17-24.

Transport for London, 2003. Accidents and casualties on London's roads 2002. December.

Wang, C., Xu, C., Xia, J., Qian, Z., 2017. Modelling faults among e-bike related fatal crashes in china. Traffic Inj. Prev. 18(2), 175-181.

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, X., Wu, X., Abdel-Aty, M., Tremont, P.J., 2013. Investigation of road network features and safety performance. Accid. Anal. Prev. 56, 22–31.

Wang, X., Yang, J., Lee, C., Ji, Z., Y, S., 2016. Macro-level safety analysis of pedestrian crashes in

Shanghai, China. Accid. Anal. Prev. 96, 12-21.

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, 137–145.

Xie, S.Q., Dong, N., Wong, S.C., Huang, H., Xu, P., 2018. Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure. Accid. Anal. Prev. 121, 285-294.

Yao, S., 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. Accid. Anal. Prev. 75, 320-332.

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

Appendix A

Table A1 Results of Pearson correlation coefficient test for the model using the trip number as exposure

	Trip number	Popu -lation density	Employ -ment rate	Class A Road density	Class B Road density	Minor Road density	Number of Class A Road crossed	Number of Class B Road crossed	Number of Minor Road crossed	Meshed -ness coefficient	People aged below 15 years	People aged above 64 years	Non -Domestic
Trip number	1												
Population density	0.103	1											
Employment rate	0.077	-0.333	1										
Class A Road density	0.068	0.419	-0.076	1									
Class B Road density	0.090	0.336	-0.022	-0.031	1								
Minor Road density	0.071	0.572	-0.113	0.226	0.167	1							
Number of Class A Road crossed	0.071	0.002	0.004	0.500	-0.068	-0.084	1						
Number of Class B Road crossed	-0.067	-0.022	-0.058	0.034	-0.559	0.085	-0.189	1					
Number of Minor Road crossed	-0.071	0.063	-0.022	-0.001	0.021	-0.010	-0.554	0.292	1				
Meshedness coefficient	-0.055	-0.015	0.121	0.013	0.022	0.221	-0.009	-0.001	-0.055	1			
People aged below 15 years (%)	-0.104	-0.164	-0.613	-0.277	-0.176	-0.159	-0.132	0.093	0.026	-0.040	1		
People aged above 64 years (%)	-0.069	-0.600	0.507	-0.305	-0.180	-0.391	-0.042	0.011	-0.113	0.051	-0.165	1	
Non-Domestic	0.085	0.420	-0.241	0.539	0.266	0.301	0.199	-0.097	0.025	-0.027	-0.251	-0.465	1

Table A2 Results of Pearson correlation coefficient test for the model using the distance travelled as exposure

	Distance travelled	Popu -lation density	Employ -ment rate	Class A Road density	Class B Road density	Minor Road density	Number of Class A Road crossed	Number of Class B Road crossed	Number of Minor Road crossed	Meshed -ness coefficient	People aged below 15 years	People aged above 64 years	Non -Domestic
Distance travelled	1												
Population density	-0.179	1											
Employment rate	0.027	-0.333	1										
Class A Road density	0.047	0.419	-0.076	1									
Class B Road density	-0.044	0.336	-0.022	-0.031	1								
Minor Road density	-0.167	0.572	-0.113	0.226	0.167	1							
Number of Class A Road crossed	0.251	0.002	0.004	0.500	-0.068	-0.084	1						
Number of Class B Road crossed	-0.130	-0.022	-0.058	0.034	-0.559	0.085	-0.189	1					
Number of Minor Road crossed	-0.237	0.063	-0.022	-0.001	0.021	-0.010	-0.554	0.292	1				
Meshedness coefficient	0.003	-0.015	0.121	0.013	0.022	0.221	-0.009	-0.001	-0.055	1			
People aged below 15 years (%)	0.043	-0.164	-0.613	-0.277	-0.176	-0.159	-0.132	0.093	0.026	-0.040	1		
People aged above 64 years (%)	0.035	-0.600	0.507	-0.305	-0.180	-0.391	-0.042	0.011	-0.113	0.051	-0.165	1	
Non-Domestic	0.140	0.420	-0.241	0.539	0.266	0.301	0.199	-0.097	0.025	-0.027	-0.251	-0.465	1

Table A3 Results of Pearson correlation coefficient test for the model using the number of road crossed as exposure

	Number of road crossed	Popu -lation density	Employ -ment rate	Class A Road density	Class B Road density	Minor Road density	Number of Class A Road crossed	Number of Class B Road crossed	Number of Minor Road crossed	Meshed -ness coefficient	People aged below 15 years	People aged above 64 years	Non -Domestic
Number of road crossed	1												
Population density	0.055	1											
Employment rate	-0.070	-0.333	1										
Class A Road density	0.250	0.419	-0.076	1									
Class B Road density	0.070	0.336	-0.022	-0.031	1								
Minor Road density	-0.012	0.572	-0.113	0.226	0.167	1							
Number of Class A Road crossed	0.333	0.002	0.004	0.500	-0.068	-0.084	1						
Number of Class B Road crossed	-0.170	-0.022	-0.058	0.034	-0.559	0.085	-0.189	1					
Number of Minor Road crossed	-0.231	0.063	-0.022	-0.001	0.021	-0.010	-0.554	0.292	1				
Meshedness coefficient	-0.001	-0.015	0.121	0.013	0.022	0.221	-0.009	-0.001	-0.055	1			
People aged below 15 years (%)	-0.038	-0.164	-0.613	-0.277	-0.176	-0.159	-0.132	0.093	0.026	-0.040	1		
People aged above 64 years (%)	-0.164	-0.600	0.507	-0.305	-0.180	-0.391	-0.042	0.011	-0.113	0.051	-0.165	1	
Non-Domestic	0.374	0.420	-0.241	0.539	0.266	0.301	0.199	-0.097	0.025	-0.027	-0.251	-0.465	1