# MODELING OF PEDESTRIAN SAFETY AT THE MACROSCOPIC LEVEL

JUNBIAO SU

PhD

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University

Department of Civil and Environmental Engineering

# Modeling of Pedestrian Safety at the Macroscopic Level

**Junbiao SU**

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

December 2020

## Certificate of originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

________________________(Signed)

______SU Junbiao________(Name of student)

## Abstract

Pedestrians are vulnerable to severe injury and mortality in road crashes. They account for more than 60% of total road fatalities in Hong Kong. It is necessary to promote the understanding of the relationship between pedestrian crashes and possible influencing factors. Therefore, effective countermeasures can be developed to improve pedestrian safety, and more importantly, the well-being of society. The tendency of a person to be involved in a road crash can increase with their amount of travel. This is referred to as exposure. It is necessary to control for the exposure in the estimation of crash risk. A common approach to estimating the vehicular crash exposure is to calculate the traffic flow or vehicle kilometers traveled using traffic count data, which are often readily available. However, accurate and extensive pedestrian counts are seldom available for the estimation of pedestrian crash exposure. In this study, comprehensive population, traffic, and travel data, obtained from multiple data sources, are used to determine the pedestrian crash exposure. In addition, the effects of influencing factors, including travel purposes and transport modes, on pedestrian crash exposure are considered. Then, the association between land use, built environment, population characteristics, traffic conditions, and pedestrian crash risk at the macroscopic level is measured using the Bayesian spatial model.

First, the efficient exposure measures for pedestrian crashes are explored, based on the comprehensive travel survey data in Hong Kong. For instance, zonal population, and frequency and time of walking trips are adopted to represent the pedestrian exposure to road crashes. The random-parameter negative binomial regression approach is then applied to measure the relationship between pedestrian crash frequency, exposure, and possible influencing factors. Results indicate that model that applies the frequency of walking trips as the proxy for pedestrian exposure is superior to that using zonal population and walking time.

Second, the effects of travel purposes on pedestrian crash risk are examined. Pedestrian crash exposures, represented by the frequency and time of walking trips, are discretized by time of the day and travel purposes. For instance, the trips are stratified into six types, including home, work, school, shopping, dining, and others. Results indicate that the

crash risk of “back home” walking trips is amongst the highest. This phenomenon aligns with the distribution of pedestrian crash rate by time of the days, i.e., pedestrian crash rates at noon and in the late afternoon are amongst the highest. This could shed light on the implementation of effective policy strategies that can improve the safety of vulnerable pedestrian groups in specific time periods.

In Hong Kong, 90% of total trips are made by public transport. Walking is the primary mode to get access to public transport services. Safe access to transport facilities is an important issue in sustainable urban development. In this part, pedestrian crash exposures, are categorized by the different transport modes, i.e., metro, bus, light bus, taxi, and private car. Results indicate that pedestrian crash risk is positively correlated to the frequency of walking trips for roadway transport services. However, the association between pedestrian crash risk and frequency of walking trips accessing to metro is not significant. This is indicative to the design and planning of road facilities, i.e., pedestrian crossings and traffic signals, that can enhance the accessibility and safety of public transport.

Fourth, a joint probability approach is developed for simultaneous modeling of crash occurrence and pedestrian involvement in crashes, with which the possible correlations among different crash types, i.e., crashes involving and not involving pedestrians, are accounted. Possible influencing factors, including land use, road network, traffic flow, population demographics and socioeconomics, public transport facilities, and trip attraction attributes are considered. Markov chain Monte Carlo and full Bayesian approach then applied to estimate the parameters. Results indicate that crash occurrence is are correlated to traffic flow, the number of non-signalized intersections, and points of interest such as restaurants and hotels. By contrast, population age, ethnicity, education, household size, and road density could affect the propensity of pedestrian involvement in crashes.

In conclusion, this study has addressed several fundamental problems, including the estimation of pedestrian crash exposure, effects of travel purposes and transport mode involvement in pedestrian crash exposure, and possible correlations among crashes of different types, for the development of pedestrian crash prediction models. Findings are useful for the design of various transport facilities, road safety education for vulnerable pedestrian groups, and more importantly, accessibility to essential urban services and attractions. With these proposed improvements, a safe and accessible walking environment can be promoted in Hong Kong.

# List of publications

## Journal papers:

**Su, J.**, Lam, W.H.K., Lai, X., Tam, M.L., Cheng, L., 2017. Modeling the relationships between home-work-home activity durations and travel times of workers in Hong Kong. *Journal of the Eastern Asia Society for Transportation Studies*, 12, 672-689.

Sze, N.N., **Su, J.**, Bai, L., 2019. Exposure to pedestrian crash based on household survey data: Effect of trip purpose. *Accident Analysis and Prevention*, 128, 17-24.

**Su, J.**, Sze, N.N., Bai, L., 2021. A joint probability model for pedestrian crashes at macroscopic level: Roles of environment, traffic, and population characteristics. *Accident Analysis and Prevention*, 150, 105898.

**Su, J.**, Sze, N.N. 2021. Safety of walking trips accessing to different transport modes: A macro-level study incorporating pedestrian exposure and spatial correlation. *Accident Analysis and Prevention*, to be submitted

## Conference papers:

**Su, J.**, Sze N.N., Bai L. Exposure to pedestrian-related crash based on household survey data: Safety effect of trip purpose. Presenting on the *Transportation Research Board (TRB) 98th Annual Meeting*, 2019.

**Su J.**, Lam W.H.K., Lai X., Zhong R., M.L. Tam. Modeling Activity Duration and Travel Time Uncertainty on Departure Time Choice of Students. Proceedings of the 22nd *Conference of Hong Kong Society for Transportation Studies: Transport and Society*, 9 December 2017 through 11 December 2017, 224-232.

# Acknowledgement

First of all, I would like to express my great gratitude to my supervisor Dr. Nang-ngai Sze, for his endless supervision and kind encouragement to let me stand at this location. It was Dr. Sze who gave me a hand when I was in a hard time and led me to my research today in the past years. I have been benefited a lot from the discussions with my superior and great thanks for his constructive comments, from major comments to minor collections to mistakes. I also learned a lot from his guidance, wisdom, and expertise, which will be greatly important for my future career.

I also want to express my sincere thanks to my co-supervisor, Prof. William H.K. Lam, for his supervision in my Ph.D. study. Prof. Lam had delivered a lot of critical but illuminating comments at my early stage of research time, which was important for help forming a strict attitude on research and the development of independent ability for work. It is of great opportunity to learn from Prof. Lam to layout the foundation for my research.

Great appreciation is also given to my second co-supervisor, Dr. Renxin Zhong from Sun Yat-sen University for his support and help ever since my bachelor's study. Whenever I got into trouble, Dr. Zhong was always there for enlightening me and offering help for a better solution, including not only the technical supervision in research but also the mentor guidance in life. It has been my great luck to be under the supervision of him and to learn from him.

Special thanks are expressed to Prof. Anthony Chen for his kind discussions with me, showing the guidelines and useful rules for the improvement of my research writing. It is appreciated that Prof. Chen offered extra courses with his spare time to help me layout the foundation of methodology and created opportunities for social academic activities.

Also, I would like to thank Dr. Xinjun Lai, from Guangdong University of Technology. The time was just like yesterday that when Dr. Lai guided me on how to read academic papers and how to develop academic techniques, spending her precious time listening to me and explaining to me in the late afternoon time. Many thanks are also given to Dr. Lai, currently at the University of Hong Kong for her discussion with me, guiding me on academic writing, and always encouraging me to move on.

Many thanks also go to the past and present members of transportation groups of the Hong Kong Polytechnic University for their accompany and friendship in the past years, including Dr. Karen Tam, Dr. Tiantian Chen, Dr. Yuan Zhang, Mr. Guoyuan Li, Mr. Hao Fu, Mr. Can Chen, Miss Yunping Huang, Mr. Dianchen Zhu, Miss Yingying Xu, and Mr. Hongliang Ding. I would like to thank them for always being there for discussions and sharing minds including but not limited to the research topics, which had been unforgettable memories throughout my Ph.D. study.

Finally, I would like to express my deepest gratitude to my parents for their endless, unconditional love, and solid support that had given to me. They are always giving me warmth, comfort, and confidence, which support me going over all those hard times.

# Table of Contents

# List of Figures

# List of Tables

# Notations

The lists of abbreviations and nomenclatures in the thesis for easy reference.

## *Abbreviations*

| **Abbreviations** | **Full descriptions** |
|---|---|
| WHO | World Health Organization |
| VMT | Vehicle Miles Travel |
| VKT | Vehicle Kilometers Travel |
| FLM | First Mile and Last Mile |
| MCMC | Markov Chain Monte Carlo |
| TAZ | Traffic Analysis Zone |
| AADT | Annual Average Daily Traffic |
| CAR | Conditional AutoRegressive |
| SB_VC | Street Block and Village Cluster |
| TPU | Tertiary Planning Unit |
| SPU | Secondary Planning Unit |
| PPU | Primary Planning Unit |
| PDD | Planning Data Districts |
| TIS | Traffic Information System |
| TCS | Travel Characteristics Survey |
| ATC | Annual Traffic Census |
| HIS | Household Interview Survey |
| POI | Points-Of-Interest |
| AIC | Akaike Information Criterion |
| BIC | Bayesian Information Criterion |
| CBD | Central Business Districts |
| GIS | Geographical Information System |
| VIF | Variance Inflation Factor |
| BCI | Bayesian Criterion Interval |
| DIC | Deviance Information Criterion |
| STPUG | Small Tertiary Planning Unit Group |
| MAUP | Modifiable Areal Unit Problem |

## *Nomenclatures*

| Elements | Descriptions |
|---|---|
| $p_i^j$ | The proportion of land-use type $j$ in Zone (or entity) $i$ |
| $P(y;\lambda)$ | Poisson probability of observing $y$ events given parameter $\lambda$ |
| $y_i$ | The number of crashes (pedestrian crash/total crash) occurred in Zone $i$ |
| $\mu_i$ | The expected number of crashes occurred in Zone $i$ |
| $H_0$ | The null hypothesis: the mean of the number of crashes equals the variance of the number of variance of crashes |
| $g(\cdot)$ | The function of the mean of the number of crashes across zone units ( $g(E(y_i))$ ) |
| $z_i, w_i$ | Two indicators for over-dispersion test |
| $u_i$ | Error term following the gamma distribution |
| $v$ | Scale parameter of the gamma distribution |
| $\alpha$ | Invert of $v$ |
| $\varepsilon$ | Error term |
| $P(y_i)$ | The probability of $y_i$ crashes (total crashes/pedestrian crashes) occurred in Zone $i$ |
| $y_i^p$ | The number of pedestrian crashes occurred in Zone $i$ |
| $p(y_i^p, y_i)$ | The joint probability of occurring $y_i^P$ pedestrian crashes and $y_i$ total crashes in Zone $i$ |
| $\pi_i^P$ | The binomial probability of pedestrian involvement in crashes of Zone $i$ |
| $I$ | Moran's statistics |
| $w_{ij}$ | Spatial weights indicating the spatial dependence between Zone $i$ and Zone $j$ |
| $w_{i+}$ | The sum of spatial weights connecting Zone $i$ |
| $W_{sum}$ | The sum of all spatial weights |
| $S_i$ | The random spatial effects of Zone $i$ |

| | |
|---|---|
| $S_{-i}$ | The random spatial effects of all the zones expect Zone $i$ |
| $\tau_i$ | The standard deviation of the random spatial effect $S_i$ |
| $\omega^2$ | The precision parameter of CAR prior |
| $L$ | Likelihood of the crash occurrences in all zones |
| $LL$ | Log-likelihood: the logarithm value of $L$ |
| $LL(conv)$ | Log-likelihood value at the convergence of parameter estimation ( $\boldsymbol{\beta}$ ) |
| $LL(cons)$ | Log-likelihood value at the convergence of parameter estimation |
| $D(\cdot)$ | The deviance value |

# Chapter 1 Introduction

## 1.1 Background

The vehicle industry has enjoyed the surge of economic development in recent decades, and an increasing number of vehicles are now running on roads. Although this development has brought great convenience to daily life and facilitates work and study, the increasing usage of motor vehicles has raised significant concerns regarding public health. According to the report by the World Health Organization (World Health Organization, 2018), though the death rate of road crashes had slightly lowered from 18.8 per population in 2000 to 18.2 per 100,000 population in 2016, the number of deaths has continued to rise, from 1.15 million to more than 1.35 million. Road traffic crash has been the eighth-leading cause of deaths across all ages worldwide. In the specific context of injury deaths, road traffic crashes account for the largest portion (23%), followed by suicide (15%) and homicide (11%) (World Health Organization, 2010). Regarding the number of deaths, the number of injuries owing to road traffic crashes is even larger, i.e., between 20 and 50 million, according to the estimation of the World Health Organization (2018). Traffic crashes not only take lives; they also cause indelible suffering to survivors, both physically and mentally (Loo et al., 2007). Road crashes incur direct and indirect costs, including death and human suffering from injury, emergency treatments, medical costs, rehabilitation costs, long-term care and treatment, insurance administration expenses, legal costs, workplace costs, lost productivity, property damage, travel delays, psychosocial impacts, and losses of functional capacity. All of these costs can consequently lead a family into poverty, especially in developing countries (Hijar et al., 2003). In terms of economic terms, the estimated cost as a direct or indirect result of road traffic crashes lies between 1% and 2% of the total gross national product, and varies between low-income countries and high-income countries (Jacobs et al., 2000). Therefore, road safety is not a single problem for transportation, but is a problem concerning both public health and society, and is involved in various sectors.

Among all of the deaths resulting from road traffic crashes, vulnerable road users (including pedestrians, cyclists, and motorcyclists) constitute a major component, i.e., up to more than 50% according to the WHO report (World Health Organization, 2018). Pedestrians accounted for 11% of road deaths in the United States between 2003 to 2006, 13% in Canada from 2002 to 2006, around 25% in mainland China, and over 50% in Hong Kong in the past 10 years (Miranda-Moreno et al., 2011; Lam et al., 2013; Transport Department, 2019). When comparing the number of injuries within each road user group based on injury severity, it is found that 2.45% of pedestrians involved in traffic crashes died, whereas only 0.14% and 0.38% of car passengers and drivers died, respectively. In addition, the seriously injured rate of pedestrians involved in crashes was 21.47%, whereas it was 7.46% and 12.28% for passengers and drivers, respectively (Transport Department, 2016). In some studies, the fatality rate of pedestrians in traffic crashes increased up to 75% or fell within a range between 41% and 75% (Odero et al., 1997; Peden et al., 2004).

However, compared with the larger economic and social costs owing to road traffic crashes, the investments put into road safety research and improvement are far less (Allsop, 2001; Koornstra et al., 2002; Waters et al., 2004). For example, the disability-adjusted life year[1] for road traffic crashes was expected to rank third among a set of causes for deaths and injuries; HIV/AIDS ranked tenth. However, the funding for research on road traffic crashes was 24 to 33 million dollars, whereas the funding for HIV/AIDS was over 900 million dollars (World Health Organization, 1996). Road safety should have drawn more attention than it has to make roads safer and guarantee the safety of vulnerable road users, especially pedestrians. This study focuses on pedestrian safety analysis and modeling at the macroscopic level, addressing several relevant key issues and concerns, as explained below.

[1] Disability-adjusted life year (DALY): A primary metric used by World Health Organization Road to provide an integrated index for the combination of the burden of mortality and morbidity (non-fatal health problems).

## 1.2 Motivation

To advance the study of pedestrian safety and promote the understanding of the occurrence of pedestrian crashes, this thesis explores four questions concerning pedestrian crash prediction and the improvement of road safety.

Pedestrian exposure is mostly not available owing to the difficulty in measuring complex movements of pedestrians on roads (Qin and Ivan, 2001; Greene-Roesel et al., 2007; Jiménez-Mejías et al., 2016; Wang et al., 2016b). However, the measurement of exposure is critical for the quantification of crash risk and providing valid crash predictions, as pedestrian crashes only happen when pedestrians are exposed to conflicts with vehicles. This raises the first research question of this thesis: can we develop an efficient and reliable approach to measure pedestrian exposure?

Travel is driven by participating in different kinds of human activities. As such, increasing attention from transportation researchers had been paid to activity-based modeling for travel demand prediction, which takes into account the individuals' activity-travel behaviors. In addition, people with different travel purposes might have different emotions, which could lead to various pedestrian behaviors on the road. This raises the second question: will the purposes of trips impact the occurrence of pedestrian crashes and how to investigate these effects through crash prediction models?

Beyond walk-only trips, most of the walking connected to motorized modes, especially public transport, in a transit-oriented society. Different types of equipment and associated facilities, as well as the service quality of different transport modes, will result in different pedestrians' actions on roads. Thus, the question is raised: what are the effects of transport mode involvement in pedestrian exposure (walking connecting to different transport modes) on the crash occurrence?

In addition, correlation effects of various factors may exist among multiple crash outcomes. Thus, another question is, how do the various zonal factors impact different types of crashes, especially total crashes and pedestrian crashes? What might be the appropriate methodology to address this concern?

Besides, spatial and temporal effects are also integrated into the modeling in this thesis. Further elaborations of the motivations of this study are discussed in flowing four subsections in this Chapter.

### 1.2.1 Pedestrian exposure measurement

Efficient exposure measurement is necessary for the development of crash prediction modeling and for a better understanding of safety assessments among different (space and time) entities. The concept of exposure has been explored for decades. Chapman (1973) summarized the definitions of exposure based on the number of opportunities with the potential to develop into road crashes under specific conditions (time, location, and crash type). Later, Wolfe (1982) furthered discussed the concept of exposure. Several metrics for exposure were reviewed regarding their features and limitations, including the total population, the number of registered vehicles, distance traveled, travel time spent, and the number of conflicts with vehicles. For a better understanding and modeling of system safety, Hauer (1982) attempted to clarify the concepts of "conflict" and "exposure". The distinction between these two terminologies was explained clearly: a conflict count is an estimation of system safety, whereas the measurement of exposure enables the estimation of the risk of crashes. Besides, the exposure was explained as the trails that could lead to the occurrence of traffic accidents (Hauer, 1982).

Although there are few controversial opinions on the concept of exposure, a consistent agreement on the appropriate measure for crash exposure has far not reached (Wolfe, 1982; Lam et al., 2013). At the macroscopic level, various measures have been adopted for the calculation of crash risk, such as the total population, number of trips, traffic volume (Lam et al., 2014; Bao et al., 2017; Bhowmik et al., 2018), vehicle miles/kilometers traveled (VMT/VKT) (Lee et al., 2014; Dong et al., 2015; Amoh-Gyimah et al., 2016; Cai et al., 2016; Dong et al., 2016; Liu and Sharma, 2017; Xu et al., 2017), travel time spent (Pei et al., 2012; Guo et al., 2017), number of registered vehicles, number of licensed drivers, and number of road crossings (Keall, 1995).

Exposure reflects the number of road users under the risk to be involved in vehicle conflicts. The vehicular traffic volume and its derivatives (such as vehicle-kilometers and vehicle-hours) are commonly adopted as exposure measures for vehicular crash modeling, given the advantages of automatic vehicle detection techniques (Qin et al., 2006; Pei et al., 2012). As for pedestrian crash analysis (beyond analyzing the vehicular traffic), pedestrian exposure is critical for modeling and understanding the mechanisms of pedestrian crash occurrence at the macroscopic level (Mukherjee and Mitra, 2019). However, it is difficult to obtain accurate pedestrian counts, especially for long-term and extensive monitoring and evaluation. Although the collection of high-quality pedestrian count data may be possible at specified locations (applicable for local studies), in the context of macroscopic pedestrian crash modeling, robust measures of exposure, in terms of the walking frequency and walking time still remain rare. Therefore, in order to achieve reliable pedestrian crash modeling, it is important to begin with more reliable measures of pedestrian exposure, which motivates the first objective of the thesis study.

### 1.2.2 Travel behavior and pedestrian safety

Travel is indeed the derived demand directly or indirectly of economic activities including work, education, trade, and recreation, etc. A work-related activity involves travel between the place of residence and the workplace. There is a supply of labor in one location (residence) and a demand for labor in another (workplace). Therefore, travel (commuting) is directly derived from the relationship between the two locations. Towards this end, not only the time, location, and mode of movements of people are of interest, but also the need of traveling for economic activities (e.g., work, education, and recreation, etc.) should be of concern. Travel purpose affects pedestrian behaviors and therefore the pedestrian crash incidence (Bao et al., 2017). Pedestrians with different travel purposes will behave differently on-road. For example, individuals who walk out for leisure recreation will wait for the red light patiently while people are hurrying for work will take more risk-taking actions, i.e., rushing to cross the roads. Activity-based transport demand modeling has received more and more attention in the past decades (Bowman, 1998; Bhat and Koppelman, 1999; Bhat and Singh, 2000; Bowman and Ben-Akiva, 2001; Recker, 2001; Miller and Roorda, 2003; Ettema et al., 2004; Bhat et al., 2013). In conventional pedestrian crash prediction models, the population has been widely used as the proxy of exposure (Wier et al., 2009; Cottrill and Thakuriah, 2010). However, travel demand could vary among different population groups, and more specifically by time, location, and activity purposes. Therefore, it is of importance to estimate the pedestrian exposure to crashes based on the derived transport demand by different travel purposes (e.g., economic activities) (Chliaoutakis et al., 1999; Chliaoutakis et al., 2005; Abdel-Aty et al., 2013; Elias and Shiftan, 2014; Lee et al., 2014; Lee et al., 2015a). Furthermore, trip generation factors could affect the accessibility to different economic activities, and then the pattern of travel activities by purposes and thus the crash risk (Naderan and Shahi, 2010; Siddiqui et al., 2012a; Siddiqui et al., 2012b; Blazquez and Celis, 2013; Bao et al., 2017; Zou et al., 2017).Therefore, the investigation of the effects of travel purposes on the occurrence of pedestrian crashes will help improve the understanding of the essence of the mechanism of crash occurrence, which leads to the second objective of the thesis.

### 1.2.3 Transport mode usage and pedestrian safety

Owing to the time cost and physical exhaustion, the walk-only mode is impractical for long-distance travel, which causes a “first mile and last mile (FLM)” problem. The FLM refers to the first/last trip legs to/from the main motorized modes without door-to-door service. The FLM problem is one of the principal problems for promoting public transport usage as better accessibility for the FLM will bring transit users much convenience and thus attract higher transit demands. From a road safety perspective, the FLM problem is also an important issue for the analysis of pedestrian safety. As promoting the usage of public transport modes has become a strategic policy for many metropolitan cities, increased pedestrians’ walking can be expected as walking is the primary option for their access to/egress from public transport stations, such as metro stations and bus stops (Mohanty et al., 2017). It means that there would be higher pedestrian exposure to vehicle conflicts. Factors that affect the accessibility to public transport services are well studied (Paez et al., 2012; Cheng and Chen, 2015; Van Wee, 2016; Sze and Christensen, 2017), however, the safety of FLM walking trips access to public transport services is not well studied. On the other hand, pedestrians are vulnerable road users, which are easily severely injured or killed once they are involved in vehicle collisions (Cafiso et al., 2013; Transport Department, 2016). Therefore, regarding the gap between the vulnerability of pedestrians and the expected increase of pedestrian walking for daily travel, it is important to improve road safety for pedestrians (Guo and Loo, 2013).

In addition, the vehicle involvement in crashes might be quite different across modes. According to the accident in Hong Kong, the involvement rates in road crashes by vehicle classes (the number of vehicles involved in crashes per one million vehicle-kilometers) for major transport modes in 2019 were 6.82, 3.39, 2.50, 1.95, and 1.46 for motorcycles, public buses, public light buses, taxis, and private cars, respectively. Among these rates, the rate of public buses ranked second, which is 1.7 and 2.3 times the rates of taxis and private cars, respectively (Transport Department, 2019). The high ranking of bus involvement rate in crashes should raise the concerns of government managers and scholars about the crash risk of the roadway public transport mode usage, especially when there is continue advocating for promoting public transport mode usage.

The gap between pedestrian vulnerability and expected increasing pedestrian exposure due to the FLM problem, and the high involvement rate of buses in crashes motivates our study on transport mode involvement in pedestrian exposure and pedestrian safety. Although the accessibility to public transport has been studied in a number of studies (Paez et al., 2012; Cheng and Chen, 2015; Van Wee, 2016; Sze and Christensen, 2017), the influencing factors of safe access walking to/from different transport modes on pedestrian crash occurrence are rarely investigated. Exposure is a critical metric for quantifying the crash risk for safety analysis. Regarding that walking connecting to motorized modes plays a major role in citizens' total daily walking, especially in Hong Kong, a city with over 90% of public transport usage, it is necessary to consider the roles of pedestrians' walking by different connecting transport modes in pedestrian safety. What's more, pedestrians might behave differently when accessing different motorized travel modes, especially public transport modes vs private modes (Delbosc and Currie, 2012). Different facilities are equipped for the access of different motorized modes, i.e., fixed bus stations on roadsides, metro stations underground, flexible loading and boarding for taxis. All these differences together with different buffer times for catching different modes will cause pedestrians to behave differently on their walk, e.g., catching or leaving the bus, boarding on and alighting from roads.

Understanding the impacts of the involvement of motorized modes in pedestrian walking on the risk of crash occurrences will provide valuable insights for pedestrian safety improvements. Motivated by this, the third objective focuses on the incorporation of access/egress walking as the pedestrian exposure in pedestrian crash prediction models; to the best of our knowledge, this is the first time that the problem is being studied.

### 1.2.4 Modeling issues

#### *1.2.4.1 Correlation between different types of crashes*

The effects of contributors can vary across different frequency outcomes (Lee et al., 2015b; Bao et al., 2017; Cheng et al., 2017). For instance, some attributes majorly influence the occurrence of pedestrian crashes. Many pedestrian-vehicle conflicts can be avoided by building a footbridge/pedestrian tunnel. In the meantime, some factors may have a similar influence on the occurrence risk of non-pedestrian crashes and pedestrian crashes. Drunk driving will increase the risk of both total crash and pedestrian crash frequency. However, there might exist some factors that influence the occurrences of these two types of crashes in an opposite way (Bhowmik et al., 2018). A wider sidewalk may leave less space for vehicle running. It may decrease the pedestrian crash risk but increase the risk of a vehicle-vehicle crash due to the reduction of vehicle running space. These cases show that the risk factors might have correlated effects on different crash outcomes, especially those with hierarchical relationships, such as total crash frequency and pedestrian crash involvement. Due to the existence of the correlated effects, separate univariate models might provide biased estimations of count models with multiple outcomes and thus misleading implications for policy making and long-term safety planning (Mannering et al., 2016). The goal for a comprehensive improvement of pedestrian safety should call for crash prediction models with these correlated effects within one single structure. Multivariate approaches to simultaneously model multiple roadway crash frequencies have been applied in many studies, including frequencies by collision types (Wang and Kockelman, 2013; Alarifi et al., 2018) and by severity types (Castro et al., 2013). In most studies of multivariate modeling for multiple crash outcomes, dependent variables are at the same attribute levels or mutually exclusive from each other, for example, fatality crash frequency and severe injury crash frequency. Therefore, the fourth objective in the thesis is a joint probability model considering hierarchical crash data, i.e., total crashes and pedestrian crashes involvement simultaneously.

#### *1.2.4.2 Spatial and temporal effects*

Owing to the difficulty in observing all influencing factors on crash occurrence, the heterogeneity of unobserved factors can cause the spatial correlation effects on crash occurrence (Mannering and Bhat, 2014; Bhat et al., 2017). Therefore, spatial correlation effects are worth investigating in crash prediction models, so as to account for the correlations resulting from unobserved factors.

As public transit services are not as flexible as private car service, which offers door-to-door service, passengers will have chances to walk from home to a neighboring street/block to get on buses or metro. Similarly, at the alighting ends of the transit trips, some passengers may need to walk from the alighting zones to neighboring zones for their final destinations or transfer purposes (Verzosa and Miles, 2016). Inter-zonal travel can be more possible for travels by metro as a result of fewer metro stations compared with the number of bus stops. In that, the public transport facilities in one zone can generate or attract outer travel demands to/from neighboring zones; this might lead to potential spatial correlations of pedestrian exposure and thus the crash occurrence. In addition, unobserved heterogeneities across the spatial areas may also be the reason for correlations among pedestrian crashes of different zones. Ignoring these spatially correlated effects in a zonal crash frequency analysis might lead to biased results and inaccurate interpretations. Therefore, the spatial correlation effects are investigated in this thesis for a better understanding of the essence of pedestrian crash occurrence, especially the role of motorized transport mode involvement in pedestrian walking (Noland and Quddus, 2004; Aguero-Valverde and Jovanis, 2006).

In addition to the spatial correlation effects, the heterogeneities resulting from the unobservable factors also cause temporal insatiability. As road traffic accidents are rare events, aggregated data of crashes over time are most likely to be used in the safety analysis of crash frequency. However, there has been widely acknowledged that both crash occurrence and the influencing factors to crashes are not temporally stable as the fundamental change of human behaviors, road infrastructure, and built environment over time (Mannering, 2018). In that, the temporal instability of crash frequency analysis should be studied to reduce possible biased inference and implications to safety improvement due to the highly aggregated crash data (Castro et al., 2012). The analysis or the modeling of road crashes had been conducted under different time dimensions across the year (Castro et al., 2012; Dong et al., 2016; Cheng et al., 2018a; Cheng et al., 2018b), season (Olszewski et al., 2015), month (Quddus, 2008b), week (Sukhai et al., 2011), day (Mason, 1979; Ma et al., 2017). Evidence from these studies has proved the variations of traffic crashes, which should not be ignored in crash prediction models. For within-day variation, the crash analysis is mostly aggregated over several hours, i.e., 07:00-11:00 (Guo et al., 2017). However, the average journey time for citizens daily travel is within one hour, around 50 minutes, in Hong Kong (Transport Department, 2012b). Thus, it will be an appeal to the pedestrians for their daily travel schedules or arrangements if the information of risk hours to pedestrian-vehicle crashes by time of day is available. In addition, as either traffic characteristics or crash frequencies vary along the day (Mason, 1979), important implications can be derived when investigating temporal time-of-day effects. However, scant studies have focused on the time of day effects (Chiou et al., 2017; Zou et al., 2017). The study in the thesis will make effort to narrow down this research gap on the investigation of time-of-day patterns of pedestrian crash occurrence.

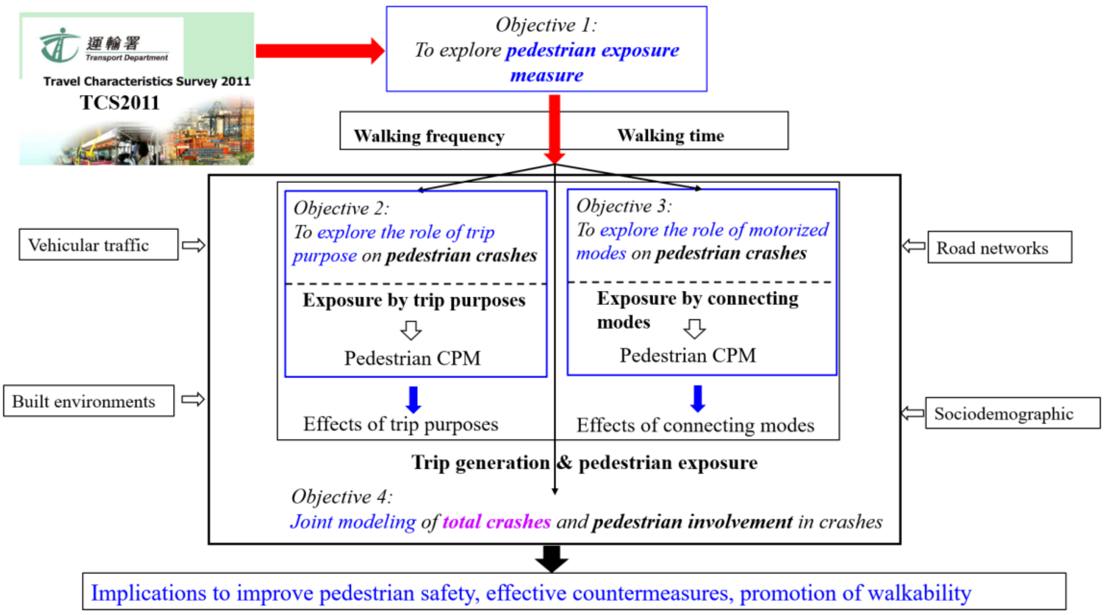
## 1.3 Objectives

Generally, pedestrian safety analysis can be carried out at two levels, i.e., microscopic level and macroscopic level. The microscopic analysis looks at the pedestrian safety performance at specific locations, such as intersections and road segments, or the severity of a specific crash. Microscopic studies provide a close eye on the crash occurrence and will come up with engineering solutions for safety improvement. On the other hand, the macroscopic analysis looks at the safety problem from a broad perspective, paying attention to the safety evaluation at larger areas, i.e., traffic analysis zones (TAZs), census blocks, and counties. Macroscopic studies offer implications of safety improvement and it is a core issue for long-term transportation safety planning purposes (Aguero-Valverde, 2013).

In response to the existing concerns elaborated in Section 1.2, the aim of the thesis is to advance the pedestrian crash modeling at the macroscopic level with proposed pedestrian exposure measures and study the essence of pedestrian crash occurrence, i.e., the investigation of influencing contributors to pedestrian crash frequencies at zonal level. Efficient measures for pedestrian exposure in crash prediction modeling are developed and various influencing factors, including trip purposes, connecting transport modes, traffic characteristics, sociodemographic, road network features, and built environment factors are investigated. Better understanding the effects of these factors on pedestrian crash occurrences at larger areas will provide valuable implications for the development of effective safety countermeasures, the promotion of walkability in society, and the improvement of social well-being in the long run.

To fulfill these targets, the pedestrian safety study at the macroscopic level in the thesis sets forth the objectives from four perspectives, i.e., pedestrian exposure measures, travel purposes, the involvement of motorized transport modes, and the joint modeling of hierarchical crash outcomes. The specific objectives regarding the research gaps are summarized below and **Figure 1.1** outlines the interconnections of the objectives of this dissertation:

- **Objective 1**: To explore a way to derive appropriate pedestrian exposure measures by making use of a large-scale traffic characteristics survey for the pedestrian crash prediction model.
- **Objective 2**: To examine the effects of trip purposes on pedestrian crash occurrence with crash prediction models.
- **Objective 3**: To investigate the association between pedestrian crash occurrence and connected transport modes with a spatial crash prediction model.
- **Objective 4**: To develop a joint probability model for delivering an overall understanding of the crash occurrence and pedestrian crash involvement in crashes simultaneously.




CPM: Crash prediction model

Figure 1.1 Objectives of the thesis

As presented in **Figure 1.1**, the four objectives constitute the major components of this thesis. Objective 1 exploits an efficient and reasonable approach to derive pedestrian exposure measures from the large-scale travel characteristic survey. The survey revealed every daily trip of each member over two years old in the interviewed households. Further calibration would be carried out on the trips from interviewed surveys with statistics from independent third parties, including the traffic census and public transport ridership (Transport Department, 2012b). This adjustment to the survey statistics improved the survey's validity and representativeness of the trip makings. As

Hong Kong is a transit-oriented society with up to 90% of the daily passenger trips are made by public transport mode, walking plays the primary role in these trips. Thus, the walking trip legs within motorized modes should be ignored in the estimation of pedestrian exposure. To achieve better exposure measures for pedestrians, the daily trip dairies were processed to extract the walking trip legs from motorized trips. The walking trip legs include the walking to get boarding on motorized modes, alighting from motorized mode to final destinations, and the possible walking for mode transferring. All these walking trip legs in motorized trips, together with the walk-only trips, serve as the base for pedestrian exposure measures. Therefore, the final output of the travel characteristics survey offers a reliable dataset for the calculation of pedestrian exposure at the macroscopic level. The achievement of Objective 1 will lay the foundation for following Objectives 2, 3, and 4 by supporting pedestrian exposure measures for crash prediction models.

Based on the approach for deriving the pedestrian exposure measures of Objective 1, Objective 2 targets the investigation of the effects of trip purposes on pedestrian crash occurrences. The trip records from the survey also provide information on trip purposes, which enables the categorization of pedestrian exposure measures by different travel purposes. Along with the complement of Objective 2, hourly variation in pedestrian crash occurrence is also taken into account when constructing the crash prediction model. On top of the methodology for pedestrian exposure measure achieved in Objective 1, the pedestrian exposure is further disaggregated by six main travel purposes according to the reported trip purposes of surveyed trips. The random-parameter negative binomial regression model is applied to investigate the effects of trip purpose.

While Objective 2 focuses on the trip purposes of pedestrians, Objective 3 targets the involvement of transport mode in pedestrian walking to investigate the impacts of access trips by different transport modes on crash occurrence, especially public transport modes. To achieve this objective, the frequency of access and egress trip legs to and from the motorized modes are categorized by six major transport modes that the walking trip legs are connected to. Bayesian inference approach is applied for the estimation of the crash prediction model. A conditional autoregressive model is also introduced for examining the spatial correlation effects.

The final objective, i.e., Objective 4, extends the study of pedestrian safety from focusing on pedestrian crashes only in Objective 1, 2, and 3 to a hierarchical perspective and models the zonal pedestrian involvement in crashes and total crash frequency simultaneously. Different effects of potential influencing factors across crash occurrence at different levels will be investigated and discussed. A joint probability modeling structure, i.e., a negative binomial regression model for total crashes modeling and a binomial regression model for pedestrian involvement in crashes will be applied to achieve the hierarchical crash prediction modeling.

## 1.4 Thesis organization

The thesis comprises four parts. The first part, including Chapter 1 and Chapter 2, gives a brief introduction about the background of road safety and pedestrian safety and the literature review on the relevant topics on pedestrian crash frequency modeling at the macroscopic level.

The second part is the methodology section. Within this chapter, the geographic unit system in Hong Kong is described followed by the introduction of various sources of data supporting the study. The mathematical formulations of crash prediction models are presented.

The third part includes three studies on pedestrian crash frequency modeling as in Chapter 4, 5, and 6. Chapter 4 carries out Study 1, a crash prediction model on the crash data during the period from 2011 to 2015. Total population, walking frequency, and walking time are adopted to represent pedestrian exposure to a road crash. The effects of trip purposes on the pedestrian crash are evaluated by disaggregating the pedestrian exposure proxies by purpose. Three random-parameter negative binomial regression models are developed to compare the performances of the three pedestrian exposure proxies. Chapter 5 processes the access/egress walking to and from different motorized modes for the derivation of pedestrian exposure considering the involvement of different transport modes, which is Study 2 of the thesis. Investigation on the association between pedestrian walking and public transport usage is studied, which is believed to be the first for pedestrian safety analysis. The spatial correlation effects across areas are also included to account for the impacts of unobserved measurements. The models are formulated with a full Bayesian approach and estimated with Markov chain Monte Carlo (MCMC) methods. Furthermore, the consideration of spatial correlation improves the model performance. Finally, the last study of the thesis, Study 3 is presented in Chapter 6, modeling the total crash occurrence and pedestrian involvement in crashes in one structure simultaneously with a joint probability model, tackling the subset relationship between the two crash counts. The effects of factors including land use, road network, traffic flow, population demographic and socioeconomics, public transport facilities, and trip attraction attributes on both total crash occurrence and pedestrian involvement in crashes are considered.

Chapter 7 summarizes the study of the thesis and suggests recommendations for further study in the future.

# Chapter 2 Literature review

## 2.1 Influencing factors to road safety

Starting from a broad perspective, macroscopic analysis of road safety concentrates on investigating and understanding the effects of zonal factors on crash occurrences. The factors associated with crash occurrences are commonly categorized into four groups, i.e., traffic characteristics, socio-economic and demographic characteristics, and road network features (Jiang et al., 2016; Cai et al., 2017). Each set of associated factors has been investigated by a number of scholars to analyze the influence on crash occurrence.

### 2.1.1 Traffic characteristics

As one of the most outstanding features of the transportation system, traffic volume (commonly observed as annual average daily traffic (AADT)) plays an important role in the occurrence of a crash. The positive effect of traffic volume has been examined in many studies (LaScala et al., 2000; Lee and Abdel-Aty, 2005; Loukaitou-Sideris et al., 2007; Ma et al., 2008; Wier et al., 2009; Dumbaugh and Li, 2010; Miranda-Moreno et al., 2011; Barua et al., 2014; Lam et al., 2014; Barua et al., 2016; Bao et al., 2017; Bhowmik et al., 2018; Xie et al., 2018). It is quite reasonable to observe a positive association between traffic volume and crash occurrence because crashes will not happen without running vehicles. A higher traffic volume with possible interactions with pedestrians means higher opportunities to result in pedestrian−vehicle conflicts. The speed limit is another factor that influences pedestrian fatalities (Kim et al., 2010a; Lee et al., 2015a; Lee et al., 2015b). Finch et al. (1994) reported that an additional increase in the risk of fatality and serious injury would be up to 4% and 3%, respectively, as a result of a 1% increase in the mean vehicle speed. Inversely, a significant reduction (30%) in fatalities has been seen when the travel speed is reduced by 5% (World Health Organization, 2017). Especially, if pedestrians are hit by the front of a moving vehicle, the death risk is 4.5 times at the speed of 65 km/h than that at a speed of 50 km/h (Martin and Wu, 2018). The high speeds leave less time for either drivers or pedestrians to take action to avoid collisions. What is more, the momentum and the energy of a high-speed vehicle will easily lead to serious injuries and even fatalities of pedestrians. In contrast to the higher risk of fatality when hit by a vehicle, some studies observed that it is more likely to witness the occurrence of pedestrian crashes within areas where local roads are under low-speed limit restrictions (Lee et al., 2015a; Lee et al., 2015b). This might be attributed to the narrowness of the roads and crowded pedestrians in local networks. In terms of the traffic components, a higher proportion of heavy vehicles, i.e., trucks, also brings greater risk to pedestrians (Hadayeghi et al., 2007; Pulugurtha et al., 2013; Amoh-Gyimah et al., 2016). Pedestrians can be unseen as the result of reduced visibility of the surrounding environment due to the large size of these vehicles.

### 2.1.2 Sociodemographic

Socio-economic and demographic characteristics are important factors for area-wide analysis of crash occurrence. Population-related factors, e.g., population density, the proportion of the population for specific groups, are common predictors in safety prediction models at macroscopic level studies (Kim et al., 2006; Ukkusuri et al., 2011; Lee et al., 2015a; Lee et al., 2015b; Amoh-Gyimah et al., 2016; Wang et al., 2016b; Cai et al., 2017; Xu et al., 2017). A higher population indicates more human activities and travel demand, either by motorized vehicles or walking. The increased pedestrian and vehicle volume will increase the opportunities for getting into conflicts. However, regarding the “safety in number” effect, although the absolute number of crash counts increases with the vehicle and pedestrian volume, the risk to each individual might drop (Xu et al., 2019). Regarding the difficulty of measuring pedestrian exposure, population statistics have commonly been adopted as the surrogate of pedestrian exposure, which is further reviewed in Chapter 2.2.

Young teenagers and children have been examined to be risk groups in many studies (Keall, 1995; Fontaine and Gourlet, 1997; Hummel, 1998; Johnson et al., 2004; Ponnaluri and Nagar, 2010; Siddiqui et al., 2012a; Zegeer and Bushell, 2012; Abdel-Aty et al., 2013; Elias and Shiftan, 2014; Amoh-Gyimah et al., 2016). Children may perform risky behaviors because of their unawareness of road safety owing to the lack of education and undeveloped ability to fully understand the moving intention of drivers.

As for youngsters, they are more likely to be risk-taking travelers and tend to ignore potential dangers to some extent. Similarly, Abdel-Aty et al. (2005) reported that male drivers have higher chances to be involved in more pedestrian crashes than female drivers due to their risk-taking behaviors.

Elderly people have been widely explored to have a positive influence on pedestrian crash occurrence (Lee et al., 2014, 2015b; Amoh-Gyimah et al., 2016; Dong et al., 2016; Cai et al., 2017; Guo et al., 2017; Xu et al., 2017). Retiring from work, the elderly have more time for daily leisure activities within short distances, such as taking a stroll after meals around their neighboring districts. Due to their reduced physical ability, elderly people will take a longer time to observe road conditions, perceive potential dangers, and take actions to protect themselves from unexpected emergencies (Keall, 1995). Therefore, it is reasonable to have more pedestrian crashes observed within areas with a higher proportion of elderly people.

Income level and education level are also zonal factors that have been examined to have an influence on the crash occurrences within the zone areas. Households with higher income levels enjoy higher opportunities to travel in their private vehicles and walk less. Thus, lower pedestrian exposure will be found in these areas and thus fewer pedestrian crashes. People with higher educational level will also have higher opportunities to find a job with a higher salary and thus to own their private cars. Moreover, higher educated people have a stronger awareness of obeying the traffic rules on roads, either as drivers or pedestrians, and behave themselves. Therefore, it is easy to find that those zone areas with a higher proportion of the population with high income and education levels will experience lower pedestrian crash counts.

In addition, some other socioeconomic and demographic factors are also investigated in the literature, including employment/employment density (Miranda-Moreno et al., 2011; Siddiqui et al., 2012a; Wang and Kockelman, 2013; Cai et al., 2016; Dong et al., 2016; Liu and Sharma, 2017), median household income (Siddiqui et al., 2012a; Dong et al., 2015; Lee et al., 2015a; Bao et al., 2017; Bhat et al., 2017; Cai et al., 2017; Wang et al., 2017a).

### 2.1.3 Road network features

Roadway infrastructure and network features are important factors that influence crash occurrence. First and foremost, road length/density is the one to be concerned in most of the studies. Many studies have found that increased road length/road density is associated with more crash occurrence (Aguero-Valverde and Jovanis, 2006; Zhang et al., 2015; Huang et al., 2016; Tasic and Porter, 2016; Cai et al., 2017). The increased road length/road length density reflects the need for higher travel demand, which results in higher exposure on roads for both vehicles and pedestrians and increases the chance of more pedestrian-vehicle conflicts.

The number of intersections/amounts of intersection density is another feature that has been widely investigated in previous studies (Siddiqui et al., 2012a; Siddiqui et al., 2012b; Bao et al., 2017; Guo et al., 2017). Intersections are possible locations where pedestrians run into conflicts with a vehicle. It is more likely to observe pedestrian-vehicle crashes around intersections as the result of pedestrians' crossings and complex driver maneuvers (Siddiqui et al., 2012a). Carter and Council (2007) reported in their study that 39% of the pedestrian crashes in rural areas were correlated with intersections. Similar effects are also found for crosswalks (Barua et al., 2014; Sun et al., 2014) in that more crosswalks indicate higher demand for pedestrian crossings, thus increasing the exposure and possible conflicts. In addition, other factors including the proportions of road types (local roads (Cai et al., 2016) and minor roads (Quddus, 2008a)), have also been investigated in the literature.

### 2.1.4 Built environments

Built environment features are important factors that influence crash occurrence at the macroscopic level, especially in long-term transport planning. Land-use types are the most widely investigated factors. Residential land use has been examined to have positive associations with crash occurrence (Hadayeghi et al., 2007; Loukaitou-Sideris et al., 2007; Wier et al., 2009; Siddiqui et al., 2012a; Amoh-Gyimah et al., 2016). An increase in the areas/proportion of residential land-use type within a zone will properly increase the residential activities surrounding the neighborhoods, which are more likely to be conducted on foot due to the short distances. Thus, the neighboring activities increase pedestrian exposure. In the meantime, roads in residential areas are more likely to be narrow local roads with crowded pedestrian flows, especially in high-density developed societies like Hong Kong. As a result, the frequency of crash occurrence may increase.

Commercial land use is another interested factor that is widely examined with a positive influence on pedestrian crash occurrence (Kim et al., 2006; Wedagama et al., 2006; Wier et al., 2009; Pulugurtha et al., 2013). Prosperous commercial activities attract both vehicular traffic and pedestrian flows, which adds to the probability of pedestrian-vehicle crashes. However, the analysis of Guo et al. (2017) showed that there was a reduction in pedestrian crashes in commercial land use areas compared to other land use areas. The authors explained that pedestrian facilities are better equipped and in good quality in commercial areas, such as overpasses to separate the pedestrians and vehicles and avoid some road-crossing conflicts.

Beyond the residential and commercial land use, other land-use types including industrial areas and institutional areas, have also been studied (Miranda-Moreno et al., 2011; Amoh-Gyimah et al., 2016; Bhowmik et al., 2018). In addition to the effects of specific land-use types on crash occurrence, many researchers begin to pay attention to the relationship between land use diversity and crash occurrence. Mixed land use indices have been widely used to describe the diversity of land use in the analyzed zone (Loo et al., 2010; Wang and Kockelman, 2013; Lam et al., 2014; Yao et al., 2015). Amoh-Gyimah et al. (2016) noted that as the mixed land use areas provide more opportunities for different types of activities, citizens can fulfill their daily needs within the zones and avoid many inter-zonal travels but generate more intra-zonal travel. These travels within zones are more likely to be on foot as the result of short distances and thus increase the number of pedestrians on local roads. Besides, vehicular traffic might also be complex as a result of various travel demands. Therefore, although mixed land use help reduce motorized travel as a whole because the demand for long-distance travel is reduced, the risk of pedestrians being involved in crashes might still increase in these areas. Pedestrian safety in school areas is of great concern (Clifton and Kreamer-Fults, 2007). Also, the increases in the number/density of dwelling units are correlated with the increase in pedestrian crashes (Zhang et al., 2015). Although these findings can be attributed to the increase of pedestrian activities in commercial and residential areas, these may also imply that such land uses are more hazardous to pedestrians (Wang et al., 2016b).

Public transport facility is another group of built environment factors that has been reported to have a significant influence on pedestrian crash occurrence. Either the presence of bus stations or metro stations increase the accessibility of an area (Loo et al., 2010; Wang and Kockelman, 2013; Lam et al., 2014; Kamargianni et al., 2015; Lee et al., 2015a; Tasic and Porter, 2016; Bhat et al., 2017). Pedestrians also tend to gather in catchment areas around railway stations and bus stops (Osama and Sayed, 2017). Hence, it is expected to witness more conflicts between vehicles and pedestrians in these locations (Lee et al., 2015a). Concerning the effect of public transport access, Yao et al. (2015) suggested that an increase in the number of light bus stops is correlated with an increase in total crash count. Nonetheless, even though the total crash risk may increase, risks of crashes with severe injuries could be reduced with an increase in the number of bus stops, because drivers tend to drive more cautiously through bus stops (Wang and Kockelman, 2013). However, the effects of the presence of pedestrian facilities, such as pedestrian signals, crosswalks, footbridges, and underpasses (especially those connected with rail transit stations), on pedestrian crashes are rarely considered in these studies. In addition, although the usage of public transport helps reduce private vehicle traffic and thus the number of crashes as a whole, the higher usage of public transport means more pedestrian exposure and possibly more pedestrian crashes, which is detailed reviewed in Chapter 2.3.2.

Additionally, environmental factors such as lighting conditions and the presence of sidewalks and crosswalks are correlated with both perceived and actual pedestrian injury risk (Dai, 2012; Rankavat and Tiwari, 2016).

### 2.1.5 Summary

In summary, numerous efforts have been put into investigating the effects of various zonal factors on pedestrian crash occurrence for the improvement of pedestrian safety. Although not any single study can have incorporated all possible factors in their research, a variety of factors have been considered in various studies subject to the different situations in different places and the data availability. The findings of previous research have provided important implications and offered useful references for subsequent studies. However, a comprehensive picture clearly delivering the effects of factors on pedestrian crash occurrence has far not been achieved due to the limitation of data availability. More importantly, without a reliable measure as the premise for safety performance function, the results estimated for the factors can be biased. The exposure measures incorporated within a model, to some extent, can take over the effects of the factors to some extent. Therefore, continuous effort should be devoted to the analysis of the association of risk factors and pedestrian crash occurrence, especially to look for credible measures of the exposure to reduce biased estimation.

## 2.2 Pedestrian exposures

Although various data collection approaches and exposure proxies have been explored to determine exposures in crash frequency analysis, it is necessary to carefully consider what kind of exposure measures to be adopted. One study might adopt a measure for exposure derived with much effort, whereas only very simple indicators are adopted in other studies (Van den Bossche et al., 2005). Chipman et al. (1993) warned that different exposure measures might lead to different results when making comparisons of crash risks. The authors had compared and discussed two direct, easily accessed measurements, i.e., the travel time spent and distance traveled. Their results indicated that in many comparisons, these two measures cannot be viewed as equivalent to each other. The travel time spent can be easily affected by weather or the environment, whereas the distance traveled cannot reflect conflicts owing to intersections and crossings. Accordingly, researchers should be careful when deciding which exposure measurement to be adopted as the proxy of exposure in road safety analysis (Hauer, 1982).

In pedestrian crash analysis, especially at the macroscopic level, it is more difficult to obtain valid exposure measures for pedestrians than other road users. For vehicle-vehicle crashes, the VMT/VKT and average traffic volume are the most common surrogates for exposure measurement (Liu and Sharma, 2017; Ma et al., 2017; Liu and Sharma, 2018), and can easily be accessed from traffic census; this process has also been aided by the fast development of vehicular detection techniques. However, measuring pedestrian exposure is far more difficult because the behavior of pedestrians is much more complex than that of vehicles, and it is impractical to track all of the movements of pedestrians (Wang et al., 2016b).

Accurate pedestrian volume is a desirable measure for the pedestrian crash prediction model (Lam et al., 2013; Xie et al., 2018; Lee et al., 2019). Site-specific observational counting does provide pedestrian volume at specific sites such as intersections. However, the data collection is time-consuming and expensive, and therefore it is not practical to apply to a large-scale analysis (Qin and Ivan, 2001). Moreover, this method may not be able to incorporate the effect of pedestrian characteristics, including demographics and travel purposes, when evaluating the relationship between pedestrian behavior and safety (Lam et al., 2014).

Due to the difficulty in measuring accurate pedestrian volume, several alternative measures are adopted as pedestrian exposure surrogates in crash prediction models. Population-based data, including total population and population density, are the most prevalent surrogates for pedestrian exposure (Wier et al., 2009; Cottrill and Thakuriah, 2010; Siddiqui et al., 2012a; Lam et al., 2013; Wang et al., 2017a). The information is easily accessed from existing census statistics, which is the cheapest way without too much additional effort. However, the limitations of using population-based data for exposure measures are obvious. Firstly, not every member within the household will conduct trips, specifically trips on foot (Qin and Ivan, 2001). Therefore, the population in the area does not necessarily account for the number of people walking on-road as it is less likely to share the same distribution of traffic, especially pedestrian traffic (Greene-Roesel et al., 2007). What is more, using population as a proxy does not consider the habits of travelers and behavior among different population groups. Therefore, the population is only a rough estimate of pedestrian exposure (Wundersitz and Hutchinson, 2008).

Due to the difficulty in observing pedestrian volume at a large-scale area, one alternative method of pedestrian exposure most adopted is to use trip estimations from four-step models as the number of pedestrian trips (Naderan and Shahi, 2010; Siddiqui et al., 2012b; Pulugurtha et al., 2013; Bao et al., 2017; Wang et al., 2017a). However, the transport mode of every trip leg within a single trip had been ignored (Allsop, 2005). For instance, a multi-modal trip (home – metro – bus – office) could generate three walking trip legs, taking into account the access to the public transport station and the transfer between modes. Therefore, the exposure could be underestimated when only the number of Origin−Destination trips is used as a proxy (Lam et al., 2014).

Travel characteristics survey through household interviews can be made use of estimating pedestrian exposure. It is possible to deduce the origin, destination, and timings of each walking trip (and trip leg) from the household travel survey data. More importantly, the attributes of every trip could be linked to the personal and household characteristics of the commuter. In addition, the socio-economic characteristics of the traveler, such as sex, age, education level, occupation and income information, travel costs, and trip purposes can be available from the survey (Wundersitz and Hutchinson, 2008). This could help improve the understanding of the association between pedestrian crash occurrence and personal factors (Allsop, 2005; Lam et al., 2013; Papadimitriou, 2016). Some examples of pedestrian exposure measures used in previous studies are summarized in **Table 2.1**.

Table 2.1 Some examples of pedestrian crash exposure measures

| Exposures | References | Advantages | Limitations |
|---|---|---|---|
| Population / population density | Wier et al., 2009; Cottrill and Thakuriah, 2010; Siddiqui et al., 2012a; Lee et al., 2015b; Wang et al., 2016 | ➢ Easy to extract | ➢ Not considering the difference in the amount of travel between individual |
| Pedestrian volume from observational surveys | Lam et al., 2014 | ➢ Considering the spatial and temporal distributions of pedestrian exposure | ➢ Not considering the effects of demographics and other personal characteristics |
| Trip generation estimates from the four-step model | Bao et al., 2017 | ➢ Considering the effects of land use, demographics, and socio-economic characteristics on travel demand | ➢ Not considering the multi-modal (walking and non-walking) trip legs |
| Walking frequency from household travel surveys | Elias et al., 2010; Naderan and Shahi, 2010; Lam et al., 2013 | ➢ Considering the effects of demographics and travel purpose on safety<br>➢ Considering the temporal effect on safety | ➢ Should consider multi-modal (vehicle and pedestrian) exposures |

## 2.3 Effects of travel behavior

### 2.3.1 Trip purpose

The effects of the trip purposes on the risks of vehicular crashes have been reported in the literature (Elias et al., 2010; Naderan and Shahi, 2010; Lam et al., 2013), and been documented to have a close link with road safety, but has not received sufficient attention (Papadimitriou et al., 2012; Papadimitriou, 2016). Indeed, trip purposes may affect pedestrian behaviors, in terms of walking path choices, walking speeds, crossing locations and timings, and the propensity of convicted crossing behavior (Sze and Wong, 2007; Lavieri and Bhat, 2019). For instance, distraction by mobile phone use and social interactions were correlated to the prevalence of pedestrian crashes at the crosswalk (Elias and Shiftan, 2014). A pedestrian might behave differently between hurrying to work and going for leisure activities. One might tend to take more risky actions or even disobey traffic rules (e.g., for crossing an intersection) when he/she is going to be late for work (Lam et al., 2014; Lee et al., 2019).

Elias et al. (2010) called for the consideration of daily-activity patterns in the study of crashes as few research studies had fully accounted for the roles of trip attributes in the occurrence of crashes. The authors recommended further investigation into activity-based exposure measures. An analysis of the risk of exposure for children walking to and from school was conducted by Elias and Shiftan (2014) and considered the influences of daily activity patterns on the occurrence of crashes. The research found that additional trips for leisure purposes (e.g., beyond walk trips to and from school) were more likely to lead to be involved in vehicle crashes. And the crash rate regarding injuries was found to be much higher for trips for shopping and leisure than that for other trip purposes. The authors noted that the ignorance of the trip attributes, especially the trip purposes, weakens the explanatory power of understanding the reasons for the crash occurrence. The roles of trip productions and attractions for different purposes in the context of crash occurrence were also investigated in Siddiqui et al. (2012b) using a random forest technique, a nonparametric method. The authors pointed out that trips with social-recreation purposes played a more important role in crash occurrence than other purposes.

Further studies on the relationships between activity-travel behaviors and crashes have been conducted (Lam et al., 2013; Lam et al., 2014; Yao et al., 2015). Space-time activity-travel patterns were considered for exposure estimates in Lam et al. (2013) followed by the proposal of a space-time path and potential path tree for pedestrian exposure measurement based on the pedestrian activity-travel patterns (Lam et al., 2014; Yao et al., 2015). Bao et al. (2017) investigated the relationship between human activities and the occurrences of crashes under a spatial framework. Human activities for different purposes (e.g., working, eating, entertainment, recreation, shopping, social, and education) were accessed from Twitter check-in websites. It was noted that these activities, with certain trip purposes, could be a good surrogate for the measure of the exposure for crash count modeling at the zonal level. The authors found that the occurrence of pedestrian crashes increased with exposure to trip purposes such as eating, recreation, and education.

In summary, the effects of the trip purposes on the associations between pedestrian crashes, exposures, and other possible factors should not be ignored. In previous studies on safety analysis at the macroscopic level, the exposure was primarily approximated by population, or on conventional travel demands at the zonal level. However, exposure at the zonal level can be better approached by the amount of travel regarding certain travel purposes. Accident risks for different exposure measures such as the exposure categorized by different purposes should have different impacts on the occurrence of crashes (Chliaoutakis et al., 1999; Chliaoutakis et al., 2005; Elias and Shiftan, 2014).

### 2.3.2 Transport mode

Travel amounts by different modes have been investigated in many studies of roadway crash analyses (Moeinaddini et al., 2015; Tasic and Porter, 2016; Cai et al., 2017; Wang et al., 2017b; Xu et al., 2017; Truong and Currie, 2019). The research reveals that travel amounts by different modes pose different effects on the occurrence of crashes. Xu et al. (2017) used a Bayesian model to explore the impacts of mode usage of commuting trips on crash occurrence. The authors found that higher bicycle usage lowered the crash occurrence. The impacts of public transport on the total crash occurrence and severe crashes in Melbourne had been studied in Truong and Currie (2019). Their study investigated the effects of the mode shares of commuting trips on crash occurrences and pointed that pubic transport is a possible solution for road safety. A mode shift from private cars was found to reduce both total crashes and severe crashes. However, the authors had also pointed out the safety issues that the usage of public transport, which included active trip legs, might lead to higher casualty rate.

Although many existing studies of transport modes focused on the total number of crashes, less attention has been paid to pedestrian crashes. What’s more, the investigated travel amounts by different modes in most existing studies are limited to commuting trips. Moeinaddini et al. (2015) studied the effects of different transport modes for commuting trips on accident-related fatalities at the city level. The authors included the percentages of travel by different modes: public transport, motorcycle, bicycle, on foot, and private car, as independent variables. The use of public transport reduces the number of trips made by private cars, and thus reduces vehicular traffic and total crash numbers (Part, 2010; Cafiso et al., 2013; Tasic and Porter, 2016). However, from the perspective of bus users, it does not necessarily imply that it is safer for them to travel by bus than by other modes (Pulugurtha and Penkey, 2010). Mohan (2001) noted that non-motorized modes are inevitable for travel by public transport, owing to the required access/egress travel at the origin and destination ends. Therefore, with the increase of public transport usage, the exposure of pedestrians also increases. Bus users tend to walk more when using public transport modes than private modes as a result of the increased access/egress walking (Mohanty et al., 2017). Therefore, the opportunities for pedestrians to get caught in vehicle collisions also increase (Lakhotia et al., 2019).

The travel behaviors of pedestrians can also be different between walking access to/egress from public transport modes and private modes, which can result in different risks of crash occurrence (Delbosc and Currie, 2012). Mukherjee and Mitra (2019) found that pedestrians who are catching buses are more likely to disobey traffic rules to avoid waiting a long time for the next bus. In addition, every simple round trip by bus has high probabilities to cross streets, which increases the chance of conflicts between pedestrians and motorized vehicles (Mohan, 2001). Walking to public transport stations and waiting for public transport vehicles also places pedestrians in situations with a relatively higher risk of crashes or collisions (Evans, 1994; Kharola et al., 2010). What's more, regarding pedestrians' safety perceptions, Delbosc and Currie (2012) found that pedestrians tend to feel less safe when getting access to/egress from public stations. Close interactions between buses and pedestrians had been documented as one reason that led to the vulnerability of pedestrians in Cafiso et al. (2013), in which the authors pointed out that the most dangerous actions for these vulnerable road users might be passengers boarding, aligning, and crossing roads near bus stops.

However, in the field of pedestrian safety analysis at the macroscopic level, few studies have paid sufficient attention to the effects of walking trips accessing to/egressing from different transport modes on the occurrences of pedestrian crashes. Study 3 of the thesis explores these FLM walking trip legs as pedestrian exposure measures for the investigation of their effects on the occurrence of pedestrian crashes.

### 2.3.3 Summary

In addition to exposure measures, human behavior, i.e., the activity-travel pattern also plays an important role in crash occurrence. Travel purpose is the starting point of activity-travel modeling under the wide agreement that travel is derived from the fulfillment of participating in particular activities. The trip purposes of individuals for their travels may influence their behaviors on roads. For instance, pedestrians hurrying to work might behave more riskily on roads and are more likely to disobey the traffic rules to cross the roads during the red light signals. This will obviously increase the risk of crash occurrence.

Mode choice is another factor within the activity-travel pattern that influences the behaviors of pedestrians. People walk more for accessing to or regressing from public transport stations with comparison to choosing to use private cars or taxis. People experience closer distances with moving vehicles when they are waiting for, boarding on, and alighting from buses. Pedestrians may also speed up or become risk-taking travelers to catch up with a bus to avoid waiting for the next. Therefore, although the amount of pedestrian exposure is the same, the risk of crash occurrence can vary across transport modes.

Although the activity-travel behaviors have been recognized to have impacts on pedestrian crash occurrences at the macroscopic level, the efforts that have been put into such studies are not enough. An investigation into the effects of trip purposes and the classification of pedestrian exposure measures by transport modes for pedestrian safety analysis at the macroscopic level will help to provide a better understanding of pedestrian crash occurrence.

## 2.4 Crash prediction model

In crash frequency modeling, the count model framework is the most prevalent methodology in response to the count outcome of crash occurrences, which is nonnegative integers. Among the family of count models, a number of models have been well applied for the formulations of crash prediction models. A brief review is given in this chapter and more details can be seen in (Lord and Mannering, 2010; Washington et al., 2010; Hilbe, 2011).

### 2.4.1 Conventional models

In this subsection, three count models that are mostly used for crash frequency modeling are briefly introduced, i.e., the Poisson regression model, the negative binomial model, and the Poisson-lognormal model.

*2.4.1.1 Poisson regression model*

The Poisson regression model provides a starting point to fulfill the modeling requirement. The model is developed from binomial trails. A number of studies have adopted the methodology in the early decades (Jovanis and Chang, 1986; Joshua and Garber, 1990; Jones et al., 1991). One central property of the Poisson regression model is that the Poisson mean is restricted to be equal to the variance. However, this strong assumption might not be met by all of the crash data, which are rarely happened events. Therefore, the adaptation of the Poisson regression model for modeling crash frequency might be somewhat problematic if the crash data are overdispersed or under-dispersed. Biased estimation and inconsistent estimated parameters can possibly lead to erroneous inferences for influencing factors with these data (Lord and Mannering, 2010).

*2.4.1.2 Negative binomial regression model*

Due to the restriction that the mean should equal the variance of the fitting data, Poisson regression is not applicable for modeling with overdispersed data with its variance deviating from the mean. For fitting the overdispersed data, the negative binomial regression model has been formulated for crash count analysis.

Negative binomial regression is also known as the Poisson-gamma model, which can be derived from the Poisson model. A gamma heterogeneity error, $\varepsilon$, is introduced for the expression of the mean of the Poisson distribution, $\ln(\mu_i) = \boldsymbol{\beta}^T \cdot \mathbf{x}_i + \varepsilon$. $\varepsilon$ is an error term following gamma distribution with a mean of 1 and variance $\alpha$, which is referred to as the over-dispersion parameter. The addition error term allows the variance of the modeling data to deviate from the mean as $\mathrm{var}(y) = \mu + \alpha\mu^2$. When $\alpha$ approaches zero, it indicates that the variance equals the mean. In this case, the negative binomial regression model collapses to the Poisson regression model.

As a majority of the crash data have a high probability of being overdispersed, the negative binomial model has been more frequently adopted in crash frequency analysis (Siddiqui et al., 2012a; Lam et al., 2014; Yao et al., 2015; Cai et al., 2016; Cai et al., 2017; Lee et al., 2019).

*2.4.1.3 Poisson-lognormal model*

Though the negative binomial model is applicable for the overdispersed data set, it still does not fit well the data that is under-dispersed. The Poisson-lognormal model, which is capable of dealing with over-dispersion and under-dispersion data, has been proposed for the modeling of crash occurrence (Bulmer, 1974; Izsák, 2008; Siddiqui et al., 2012a; Williams and Ebel, 2012; Wang and Kockelman, 2013; Lee et al., 2014; Lee et al., 2015a; Lee et al., 2015b; Cai et al., 2017; Guo et al., 2017; Huang et al., 2017; Cai et al., 2019). Starting from the Poisson regression model, and different from the negative binomial regression model, the mean of the Poisson-lognormal is associated with the heterogeneous error term, the log value of which follows the normal distribution.

### 2.4.2 Advanced models

Although the Poisson model, negative binomial models, and the Poisson/negative binomial lognormal models have solved part of the issues in the domain of crash frequency modeling, there is still room for improvement in the methodology. This section reviews the methodologies that have helped address some widely concerned issues for crash frequency modeling in recent years, i.e., heterogeneity and the correlation between multiple crash outcomes.

#### *2.4.2.1 Unobserved heterogeneity and random-parameter model*

Considering the existence of heterogeneity of exogenous variables on the observed outcomes and the varying effects across the entities, a fixed parameter might not be able to capture such unobserved heterogeneity on crash occurrence (Savolainen et al., 2011; Xiong and Mannering, 2013; Mannering et al., 2016). In order to catch the potential variation of the effects on crash counts, the random-parameter modeling has been applied for crash count prediction (Anastasopoulos and Mannering, 2009; Mannering and Bhat, 2014; Xiong et al., 2014; Mannering et al., 2016; Wang et al., 2017a).

With the comparison of fixed-parameter modeling, the random-parameter models allow the coefficients of exogenous variables to vary across the individual observations. To better present the varying coefficients estimated for variables, it is usually assumed that these values follow a specific distribution. The most frequently adopted distribution is the normal distribution. Given the random distribution of estimated coefficients, the probability of zonal crash frequency becomes conditional probabilities given the distribution of parameters to be estimated.

As the coefficients are modeled to follow some distributions which can vary across the individuals, it is believed that they better fit the data than a fixed value for the variable. Therefore, better goodness-of-fit are observed for random-parameter models with the comparison of fixed-parameter models. However, due to the formulation of a specific random distribution for each interested variable, the computation of the estimation procedure becomes much more complex, especially for the estimation of large data sets.

*2.4.2.2 Multivariate model*

In the field of the analysis for crash occurrence, either the crash data or the crash outcome can be classified into different groups, i.e., vehicle types (modes) and severity levels. For groups by modes, motorized and non-motorized modes are two of the most common categories for crash classification by involved modes. For crash severity, crashes can be categorized into four groups, i.e., fatal, serious/disabling injury, slight/non-disabling injury, potential injuries, and no injury. However, due to a wide acknowledgement fact that the crash frequencies across these multivariate groups are not independent, there might be correlated effects of potential factors on different groups of crash frequencies. Therefore, separate modeling of univariate crash counts has its limitations and lacks a comprehensive view of understanding crash occurrence. In light of this concern, multivariate models have been introduced for the modeling of crash counts by different crash types simultaneously within one structure (Bhat, 2005; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009; Lee et al., 2015b; Wang et al., 2016b; Cheng et al., 2017; Huang et al., 2017; Ma et al., 2017). The multivariate Poison lognormal model is one of the most frequently used models for multivariate crash analysis (Ma et al., 2008; Aguero-Valverde and Jovanis, 2009; Wang and Kockelman, 2013; Cheng et al., 2017; Huang et al., 2017; Wang et al., 2017b).

*2.4.2.3 Spatial and temporal effects*

Spatial correlation effects on crash count data have been widely modeled in the literature, aiming to consider unobserved heterogeneities. (Aguero-Valverde and Jovanis, 2006; Wang and Abdel-Aty, 2006; Quddus, 2008a). For example, neighboring zone areas are likely to share similar characteristics (i.e., unobserved factors), which can lead to potential correlations in the crash counts among these zones. For example, pedestrians are more likely to walk across neighboring zones between locations of public transport stations and their origins/destinations for public transport services

(Verzosa and Miles, 2016). The situation should have received more attention in a transit-oriented society with high public transport usage, as a result of the increased inter-zonal travel by passengers for access to/egress from transit stations.

Various models have been proposed to account for spatial effects in crash prediction modeling (Noland and Quddus, 2004; Aguero-Valverde and Jovanis, 2006; Quddus, 2008a; Castro et al., 2013; Wang and Kockelman, 2013; Jonathan et al., 2016; Bhat et al., 2017; Ma et al., 2017; Liu and Sharma, 2018). One of the possible methods to address this issue is the random-parameter model, as reviewed in Chapter 2.4.2.1. Geographical weight regression method is one of the widely applied methodologies accounting for spatial correlation effects (Fotheringham et al., 2003; Hadayeghi et al., 2010; Xu and Huang, 2015; Gomes et al., 2017; Liu et al., 2017; Hezaveh et al., 2019). The maximum approximate composite marginal likelihood model proposed by Bhat (2011) was adopted in Castro et al. (2013) and Bhat et al. (2017) to model spatial effects and heterogeneities. Huang et al. (2017) developed a multivariate spatial model and adopted a Wishart distribution and Bayesian approach for estimation. The intrinsic conditional autoregressive (CAR) model proposed by Besag et al. (1991) is a prevalent approach for addressing spatial correlation effects (Wang and Kockelman, 2013; Ma et al., 2017; Liu and Sharma, 2018).

However, relatively few studies have fully considered spatial correlation effects in pedestrian crash prediction models. Study 2 adopted the CAR model to investigate the relationships between the usage of different transport modes and the occurrences of pedestrian crashes in Chapter 5.

The patterns of accident occurrence over different hours within a day and different days of the week have continuously attracted attention since the late 1970s (Mason, 1979). In the past decades, a number of studies have examined the temporal effect on crash counts under different time dimensions, for example: broad time hours (Lam et al., 2013; Chiou et al., 2017; Zou et al., 2017; Wang et al., 2018); over one day (Mason, 1979; Liu and Sharma, 2017; Ma et al., 2017; Zou et al., 2017); or over one year. Zegeer et al. (1993) observed that the fatality rate was higher during the daytime, on weekdays, and in wintertime for elderly people. Most of the existing crash frequency analyses have explored the temporal effect to investigate the changing tendency of crashes year by year (Castro et al., 2012; Dong et al., 2016; Cheng et al., 2017; Park et al., 2017; Cheng et al., 2018a; Cheng et al., 2018b). A 10-year temporal effect was studied to shed light on the long-term development of reducing traffic crash frequencies (Liu and Sharma, 2017; Ma et al., 2017; Liu and Sharma, 2018). A higher revolution analysis of temporal effects was left for future study due to data limitations in their studies. Pahukula et al. (2015) highlighted the importance of the time-of-day effect on truck crashes. Other research had focused their attention on investigating or modeling the daily/week variation impacts of traffic crashes (Mason, 1979; Blazquez and Celis, 2013; Xiong et al., 2014; Wang et al., 2018).

Though the importance and the necessities have been emphasized in a number of descriptive analyses (Zhou and Sisiopiku, 1997; Clarke et al., 2006; Blazquez and Celis, 2013; Chiou et al., 2017), few pieces of literature have ever modeled the time-of-day effect on traffic crash frequency. On the other hand, the average journey time in Hong Kong is around 50 minutes according to the report of the transportation department (Transport Department, 2016). In response to the concern of passengers who travel daily, it is an appeal to pedestrians that which hour of a day is at higher risk of pedestrian–vehicle crashes when they are scheduling their daily trips. In addition, as witnessed from the hourly variation of pedestrian crash occurrence, obvious temporal features can be found in Hong Kong’s crash data. Given the same level of exposure, higher crash occurrences were observed in the late afternoon in our data set. This temporal variation is worth more attention and will be considered in the study in Chapter 4.

### 2.4.3 Summary

This section delivers a brief review of the methodologies within the domain of crash frequency modeling. Conventional models, i.e., Poisson models, negative binomial models, and Poisson lognormal models, layout the methodological foundation for crash prediction models. Advance models are further developed and adopted for addressing complex issues, e.g., heterogeneity and correlation between crash types. Random parameter models and multivariate models are frequently used to incorporate these issues into the concern.

Spatial and temporal characteristics also form as parts of travelers' activity-travel patterns that people tend to participate in particular activities at particular places at particular times within a day. Therefore, the consideration of temporal effects will help provide a better picture of the influence of activity-travel patterns on crash occurrence.

## 2.5 Remarks

This chapter reviews several issues concerning road safety, especially pedestrian crash frequency analysis. It reviews the effects of a set of various risk factors, including traffic characteristics, socio-economic and demographic data, road network features, and built environment characteristics. All these variables might have possible effects on road crash occurrence, which have been examined in the previous studies. However, these effects can be different from site to site and have not yielded a full picture of the understanding of pedestrian safety partially due to the limitation of data or the lack of valid pedestrian exposure measures. Further investigation of the effects of the influencing factors on pedestrian safety in Hong Kong will provide valuable information for the improvement of pedestrian safety in the high-density development city.

Conventional pedestrian exposure measures to pedestrian-vehicle collisions, including population-based measures, forecasting trips, and observing pedestrian counts have also been reviewed. However, these measures have their limitations in representing the exposure of pedestrians to vehicle crashes due to the unavailability of representative data or the high cost of observing high-quality data. More reliable pedestrian measures are worth further exploration. An approach for estimating pedestrian exposure is proposed in this study by making use of a large-scale travel characteristics survey. The survey provides rich information on trip information, which is well calibrated with third parties, including the vehicular traffic census and bus company statistics. The developed exposure measure takes into account not only the walk-only trips but also the walking trip legs of motorized trips that constitute the major component of pedestrians walking in their daily lives in a transit-oriented city. Therefore, the measures will better present the movements of pedestrians for the quantification of the potential risk of crash occurrence.

Although the more reliable measure of pedestrian exposure advances the crash prediction models for pedestrians, a single and simple sum of zonal exposure may still overlook some valuable features, i.e., activity-travel patterns. Pedestrians traveling for different purposes will perform differently on roads. With different behaviors on roads, the risks of pedestrian−vehicle crashes might not be the same for pedestrian movements with different travel purposes. Further investigation into exposure measures by considering pedestrians' activity participation, which is of interest for travel planning, will benefit the understanding of the pedestrian crash occurrence and improve pedestrian safety. The relation between the usage of motorized modes, especially public transport, and pedestrian safety has also been reviewed. It is necessary to identify the involvement of transport modes in pedestrian exposure on the occurrences of pedestrian crashes.

A brief introduction to the methodologies on crash frequency modeling is reviewed. Different modeling methodologies fit different data sets and address different issues of concern. In addition, spatial and temporal effects will be considered in the studies of the thesis. As a result of the space−time attribute of pedestrian activities and traveling, there exist spatial correlations and hourly variations of pedestrian crash occurrence, which should not be ignored in the analysis.

# Chapter 3 Methodology

This chapter introduces the methodologies for data processing and model formulations of the crash prediction models. The data are obtained from various sources, covering the information from different domains of society. Regarding the data from various sources in different formats, techniques are needed to aggregate the data into a consistent format for model estimation. Section 3.1 elaborates the geographical unit systems in Hong Kong. Section 3.2 introduces the data from various sources supporting the study in the thesis. Section 3.3 presents the formulations of the crash prediction model at the macroscopic level in this thesis. Section 3.4 is the summary of this chapter.

## 3.1 Geographical unit

Hong Kong is a high-density city, with approximately 7.5 million people living on approximately 1,110 square kilometers of land area; only 24% of this area is reserved for urban development. For planning purposes (e.g., to provide a reference for other departments), the Hong Kong Planning Department has demarcated different geographic units covering the whole territory of Hong Kong. The boundaries of the geographic units are delineated by major development sites, road centerlines, coastal lines, natural ridges, village clusters, zoning boundaries in outline zoning plans, and boundaries of district councils. The zone system is updated every five years. Regarding the consistency of data from various sources, the geographic units from 2011 were adopted herein. The zone system is referred to as Tertiary Planning Unit (TPU) system and comprises four hierarchy levels, which from the bottom to top are the street block and village cluster (SB_VC) level, the TPU level, the secondary planning units (SPU) level, and the primary planning unit (PPU) level. In addition, a compatible system with TPU is also introduced, named the planning data districts (PDD).

### 3.1.1 Street Block and Village Cluster

In the system, the SB_VC is the lowest level, with the smallest zone size. SBs are divided in urban areas with streets as boundaries, whereas VCs are divided in rural areas by using roads/streets and streams as boundaries. The average zone size at this level is 0.22 $km^2$. The SB_VC level serves as the geographical reference basis for many surveys, including the population census and travel diary survey. The outlines of SB_VC zones are presented in **Figure 3.1**. From the figure, we can see that the zone size is smaller in urban areas, especially Kowloon and the north shore of Hong Kong Island, whereas the zone size is larger in rural or mountainous areas. However, regarding small zone sizes with sparse populations, the release of sociodemographic statistics to the public concerning these units can potentially release the private information of the citizens within these zones. Therefore, no sociodemographic statistics are released on the SB_VC level.

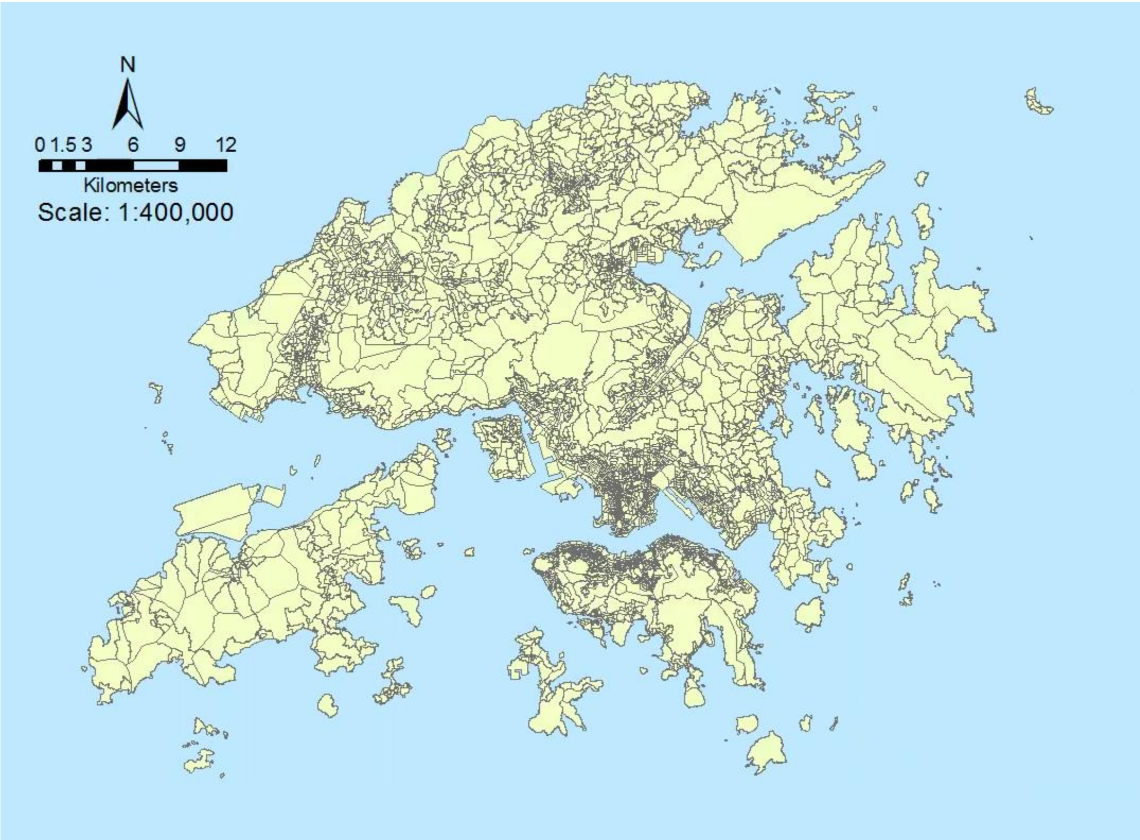
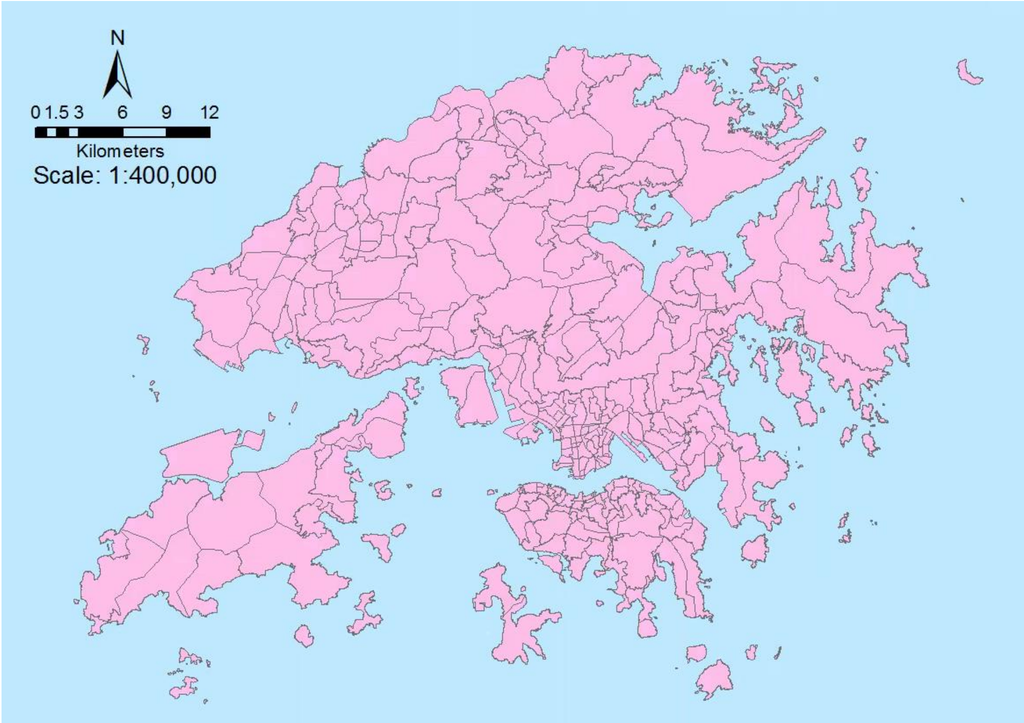



Figure 3.1 Boundaries of street block/village center (SB_VC) units

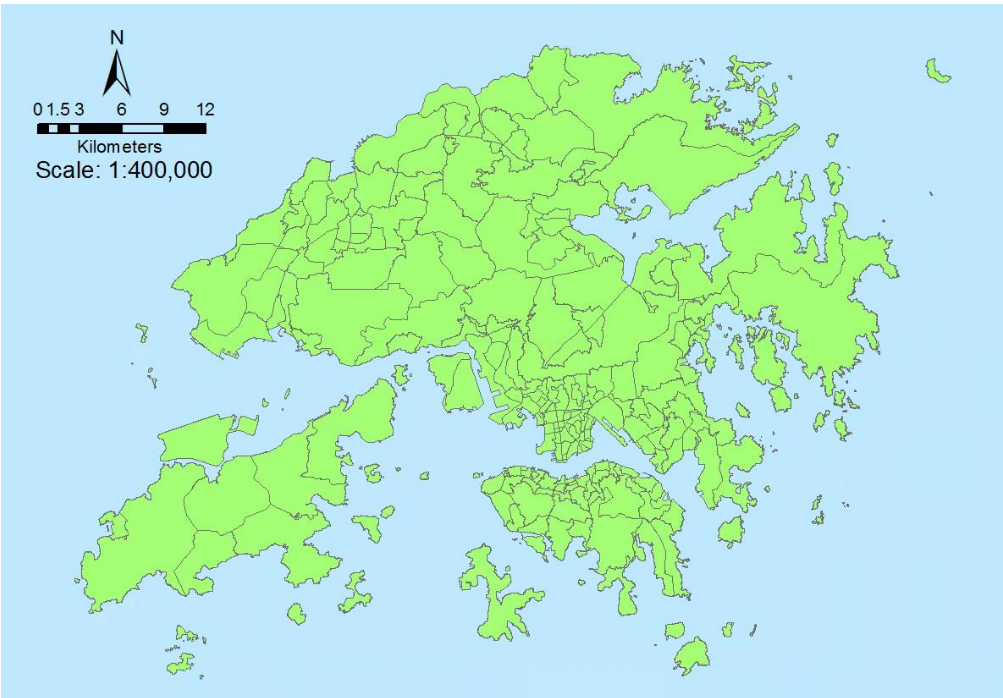
### 3.1.2 Tertiary Planning Units

A number of SB_VC zones are aggregated into a TPU, i.e., the third level of the geographic units. The boundaries are determined by the nature of geographic features in the area, such as roads, railway lines, coastlines, contours, waterways, lot boundaries, and zoning boundaries. There are a total of 289 TPUs at this level, and the average zone area is 3.84 $km^2$. TPUs are the most widely referenced zones in Hong Kong for planning purposes, especially for population projections.

Census statistics are released at the level of TPUs. In response to privacy laws, TPUs with less than 1,000 persons are merged with adjacent zones when the census data are released to the public. The census statistics are ultimately released on 209 zones after the zone merging process. The average zone size of the TPUs is 5.31 $km^2$. The boundary outlines of the total 289 TPUs and the final 209 TPUs are presented in **Figure 3.2**.




(a) Boundaries of 289 TPUs

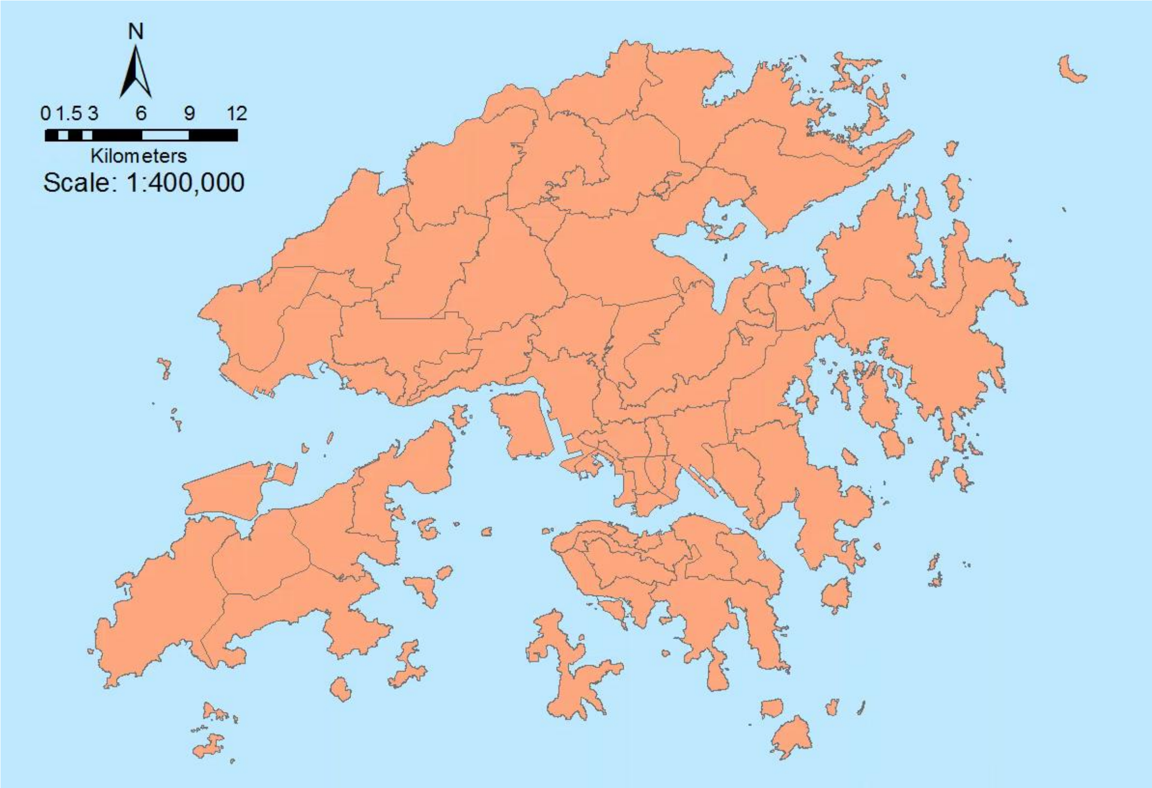



(b) Boundaries of 209 TPUs

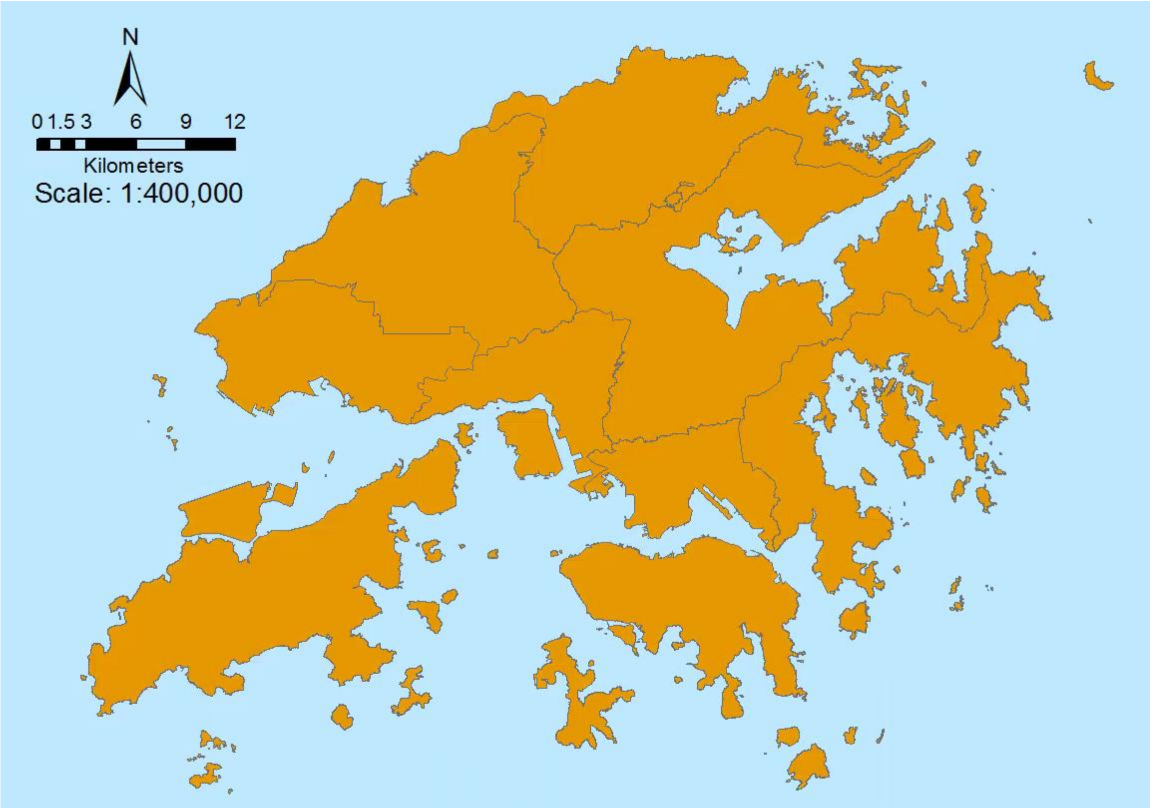
Figure 3.2 Boundaries of geographic units: 289 tertiary planning units (TPUs) and 209 TPUs

### 3.1.3 Secondary Planning Units and Primary Planning Units

Above the TPU level, there are two higher levels, with larger geographic units. Secondary planning units (SPUs) are at the second level and are aggregated with TPUs; there are a total of 52 SPUs. Primary planning units (PPUs) are at the top level of the TPU geographic unit system. There are nine PPUs at this level; these can also be aggregated with TPUs, and are larger than SPUs. The boundaries of the SPUs and PPUs are presented in **Figure 3.3**.




(a) Boundaries of SPUs




(b) Boundaries of PPUs

Figure 3.3 Boundaries of geographic units: 52 secondary planning units (SPUs) and 9 primary planning units (PPUs)

### 3.1.4 Planning Data Districts

In addition to the TPU geographic unit system, a board district reference system is adopted in the travel characteristics survey. It employs planning data districts (PDDs) to generate a territorial population and employment data matrix. To make use of the population census statistics released at the TPU level, the PDDs are compatible with the TPU zone system. In addition, the PDDs also consider the accessibility of roads and public transportation. The boundaries of the PDDs are presented in **Figure 3.4**. In view of the need to disaggregate the pedestrian crash data over 24 hours to consider the hourly variation in pedestrian crashes (discussed in Chapter 4), the TPU zone system would cause the crash counts in zone $i$ at time $t$ to have excess zeros. Therefore, the PDD zone system is adopted in Chapter 4 as the geographic unit system for pedestrian crash modeling by time of day.

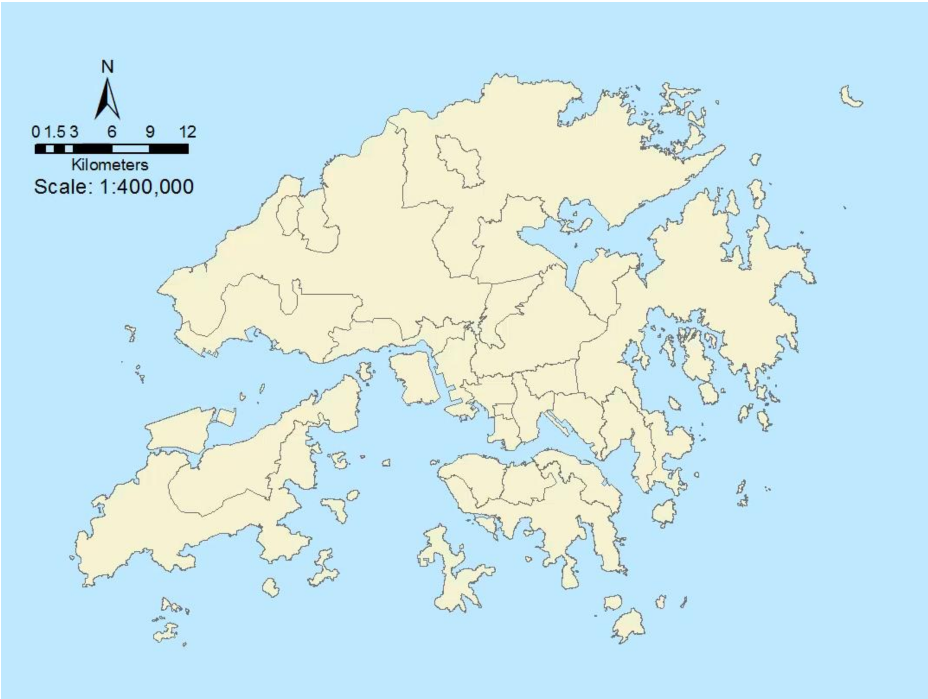



Figure 3.4 Boundaries of geographic units: planning data districts (PDDs)

### 3.1.5 Summary

Section 3.1 describes the geographic unit systems in Hong Kong. The TPU system has a hierarchy with four levels, namely, SB_VCs, TPUs, SPUs, and PPUs from the bottom to top level; the number of zones at each level are 4993, 289, 52, 9 respectively according to the demarcation method in 2011, with an increasing average land area. To incorporate the accessibility of public transport and land use changes into the presentation of population and employment distributions over the territory, the PDD zone system is also demarcated, and is compatible with the TPU system. Throughout this thesis, the PDD and TPU zone systems are the two geographic unit systems for data analysis, model formulation, and estimation.

## 3.2 Data

This section introduces the data supporting the study of the thesis. The data are obtained from various sources, covering the information from different domains of society. The crash data are obtained from the Traffic Information System (TIS) maintained by the Hong Kong Transport Department. The 2011 Travel Characteristics Survey (TCS) is also from the same department. The survey was conducted from September 2011 to early 2012 (hereinafter referred to as 'TCS2011'). The department also releases the data set of the road network and transport facilities, vehicular annual traffic census (ATC). Population census statistics from the Census and Statistics Department, provide socio-economic and demographic information. The land use information is obtained from the Hong Kong Planning Department. And finally, the points of interest is grabbed from Amap.

### 3.2.1 Police crash record

The crash data were obtained from the TIS, as maintained by the Hong Kong Transport Department. The crash data set is police-reported data and recorded the crashes with injuries. The crash database summarizes crash count data using three categories: information on crash events, information on involved vehicles, and information on casualties. Pedestrian crashes were sorted out from the whole data set (i.e., from casualties). Five-year pedestrian crash records are aggregated in Chapter 4, with consideration of the hourly variation of pedestrian crashes. One-year pedestrian crash data are used for estimations in Chapter 5 and Chapter 6. In addition to using pedestrian crashes only, the total number of crashes is incorporated into a joint probability model in Chapter 6. The crash distributions are presented in the corresponding chapters.

### 3.2.2 Traffic data

#### *3.2.2.1 Travel Characteristics Survey*

The TCS2011 offered valid support for measuring pedestrian exposure. The survey was conducted between September 2011 and January 2012, aiming to collect the latest travel characteristics data and to build a database. The personal trip data were collected in one of the three main surveys, i.e., the household interview survey (HIS). The HIS covered 35,401 households from randomly sampled quarters, with a 1.5% sampling rate. All household members aged 2 or above were interviewed in regard to their trip information and trip-making characteristics on normal weekdays (excluding weekends and public holidays). The HIS recorded both household and personal characteristics. The revealed trip records provided detailed information on every trip, including boarding and lighting locations, activities being traveled for, departure and arrival times, and chosen modes. Instead of a single record for a trip, the details of the trip legs and interchange locations were also collected, thereby supporting more valid measures for pedestrian exposure.

Although a sampling rate of 1.5% is not very high, a two-stage expansion of the recorded trip data improves the quality of the survey for the representation of the travel characteristics over the territory. First, the sample characteristics at the household and personal levels were expanded based on a territory population census conducted in the same year. The household data was stratified by district, housing type, and household income group for household levels, whereas the individual data was by districts, housing type, gender, and age group. Finally, demographic expansion factors were derived for households and individuals to represent the distribution of the population, using the statistics of the population census as a control.

The adjustment in the second stage was conducted on revealed trips. The individual expansion factors derived at the previous stage were initially applied to the trips. In view of the possible under-reporting of trips by the interviewed household members, independent transport statistics from the third parties were introduced to calibrate the sampled trips, for a more accurate representation of the trips made by the entire population. The control statistics included an annual traffic census survey from over 1,600 traffic count stations distributed over the entire road network in Hong Kong, along with occupancy data and ridership statistics from public transport. Trip expansion factors were derived for adjusting each trip. As a result, the processed trip records from the HIS were of good quality for representing the daily trips in Hong Kong.

*3.2.2.2 Annual Traffic Census*

The Transport Department also releases a data set concerning the road network and transport facilities, i.e., ATC. The 2011 census was conducted based on a total of 1,649 traffic count stations, distributed on roads over the entire territory of Hong Kong. The counting stations were designed into three categories: core (A) stations, coverage (B) stations, and coverage (C) stations. The three types of stations followed different work patterns. Core (A) stations were set to cover the entirety of the major links in the road network whereas the coverage stations were set for minor links. For these types of stations, the traffic was surveyed one week in a month throughout the year. Continuous data were used for the establishment of the hourly, daily, and monthly variation patterns. The surveyed periods for coverage (B) and coverage (C) stations were one week and one weekday, respectively.

Notably, owing to the general limitation of resource availability, a rotating census plan has to be accepted for coverage counting stations for surveying the minor links. For instance, all of the minor links were grouped into five groups, and traffic counts of each link would be obtained twice every five years. Especially, between two successive years, one group of stations was designed to remain as overlapping. If five station groups were classified as a, b, c, d, and e, a circle surveying plan for the five years would be a and b for the first year, followed by b and c for the next year, c and d for the third year, d and e for the fourth year, and e and a for the fifth year. Over half of the stations (i.e., 844 stations) were surveyed to determine the vehicular traffic in 2011, covering 87.1% of the trafficable road length in Hong Kong.

Although the coverage stations were surveyed for a short time, the surveyed data from the core stations can be used to develop scaling factors for the estimation of the AADT at the coverage stations. The scaling factors include hourly factors, daily factors, and monthly factors. For the stations working in the surveyed year, the AADT of the stations is derived from the product of the surveyed traffic and corresponding scaling factors. For sites that were not being surveyed during the year, the AADT at the current year were determined as the product of the AADT in the previous year and proper scaling factors.

### 3.2.3 Sociodemographic

Population census statistics were obtained from the Census and Statistics Department. The census is conducted every ten years with a by-census in the middle of the interdental period, which is vital to government planning and policy formulation. The statistics adopted in this thesis are from the 2011 census. The census was based on household surveys. A simple enumeration was conducted on nine-tenths of the households for basic information such as sex and age, whereas further collections of detailed demographic and socioeconomic characteristics of the household members were conducted for the remaining one-tenth of the households. The 2011 Census was the first time that the Hong Kong government adopted multiple modal approaches for data collection. More than half of the households completed the survey through their self-enumeration modes, using either postal or electronic questionnaires.

With regard to privacy protection, the population statistics are released at the TPU level. The total number of persons presented in Hong Kong during the surveyed year was up to 7 million; approximately 6.6 million were usual residents, living in a total of 2.36 billion domestic households. The dataset provides socioeconomic and demographic information, including information on population sizes and structures, nationalities, ethnicities, education, labor forces, and household characteristics.

### 3.2.4 Road network characteristics

The road network data available from the Transport Department includes the road centerlines and a set of roadway features. The data are updated monthly. The length of the road centerlines is calculated as the road length. The road density is derived as the ratio of the road length in a zone to the zone area. The roadway features include intersections (signalized and non-signalized), on-street parking lots, zebra crossings, yellow boxes, tolls, cul-de-sacs, and roundabouts. All of these features are released in a well-geocoded database format, enabling their presentation in a geographical information survey map, and can be summarized according to the zone.

### 3.2.5 Built environments

The built environment data includes the land-use information from the Planning Department, and a points-of-interest (POI) dataset obtained from Amap. The land use information is coded in three ways: area, proportion, and a dummy indicator for the major land-use type. Five land use types are classified according to the land use categories from the Hong Kong Planning Department: commercial, governmental and institutional, industrial, residential, and other (grassland, woodland, shrubland, etc.). The areas of each land-use type are counted in the PDD zones and TPU zones. The proportion of each land-use type is calculated as the ratio of the area of the land-use type to the total area of the zone. The land-use type with the highest proportion among all types within a zone is coded with a dummy variable 1, and all other cases are coded with a 0. As the majority of land in Hong Kong is mountainous, only 24.9% of the land is available for urban development. Therefore, when coding the dummy variable for the major land-use type in a larger zone system, i.e., a PDD zone system, other land-use types make up the highest proportion of the land use in all 26 zones, which might weaken the explanatory power of such a factor. Thus, the dummy status of the land use is more appropriate to be applied in the TPU zone system instead of the PDD zone system.

In addition to using the areas and proportions of land-use types to represent the land-use intensity, the entropy index has also been introduced to quantify the diversity of land use, and is formatted as follows:

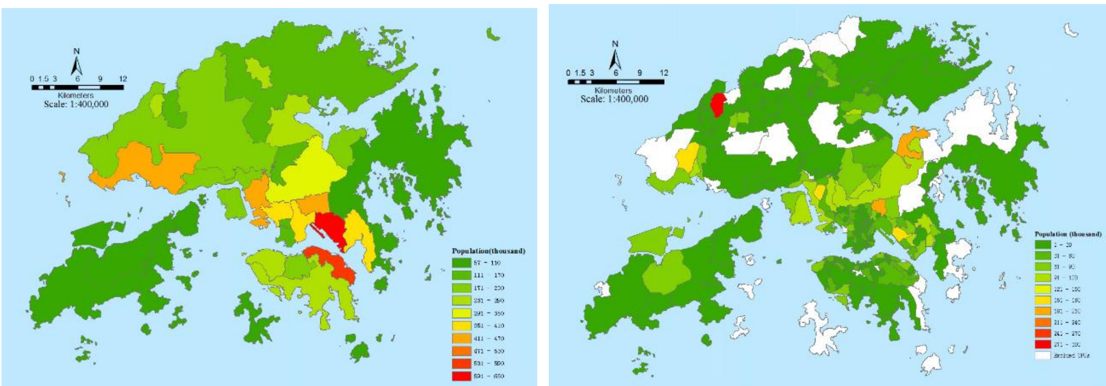
$$Entropy_i = -\frac{\sum_{j=1}^{J} p_i^j \bullet \ln(p_i^j)}{\ln(J)} \tag{3.1}$$

In the above, $p_i^j$ is the proportion of land-use type $j$ in zone $i$, and $J$ is the total number of land-use types ( $J$=5 in our study). The entropy index in Equation (3.1) ranges between 0 and 1. Given that land uses are categorized in the same way and the number of the types is the same over all zones, a higher value of the entropy index indicates a higher level of mixed land use.

Finally, the points-of-interest data, including the number of metro exits, bus stops, restaurants, hotels, schools, and shopping malls, is also considered in the crash prediction model.

### 3.2.6 Summary

As mentioned before, throughout the study of the thesis, the data have been summarized into two zone systems, i.e., the TPU system and the PDD system. The census statistics are released based on the TPU zone system. In the case of the analysis for hourly variation in Chapter 4, to avoid the excess zero counts, the PDD zone system is also adopted in crash prediction models beyond the TPU zone system. **Figure 3.5** shows an example of the distribution of potential factors between the two zone systems. As seen in the figure, the distribution of the same factor can be quite different under different geographic demarcations. Being divided into a number of smaller zones, the outstanding hot zones under PDDs, such as the Kwun Tong and Kowloon areas, are no longer distinguished under TPUs.




3.5(a) Distribution of population on PDDs 3.5(b) Distribution of population on TPUs

Figure 3.5 Distribution of zonal population between PDDs and TPUs

## 3.3 Model formulation

### 3.3.1 Poisson regression model

The Poisson regression model is developed from the binomial trails. Given that within one trial, an event occurs at a probability of $p$, then the probability of occurring $y$ events within $n$ trails is calculated as

$$P(y;n,p) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y} \qquad (3.2)$$

$\lambda$ is introduced as the mean of the number of the events occurring:

$$\begin{aligned} &\lambda = np \Rightarrow \\ &p = \frac{\lambda}{n} \end{aligned} \qquad (3.3)$$

When substituting $p = \frac{\lambda}{n}$ in Equation (3.2), the probability of occurring $y$ events is expressed as:

$$\begin{aligned} P(y;\lambda,n) &= \frac{n!}{y!(n-y)!} (\frac{\lambda}{n})^y (1-\frac{\lambda}{n})^{n-y} \\ &= \frac{n(n-1)...(n-y+1)}{y!} \frac{\lambda^y}{n^y} (1-\frac{\lambda}{n})^{n-y} \\ &= \frac{n(n-1)...(n-y+1)}{n^y} \frac{\lambda^y}{y!} (1-\frac{\lambda}{n})^n (1-\frac{\lambda}{n})^{-y} \end{aligned} \qquad (3.4)$$

If the number of trials goes larger enough and approaches infinite, recall that $\lim_{x\to\infty}(1+1/x)^x = e$ , $\lim_{x\to\infty}(1+k/x)^x = e^k$ , $\lim_{n\to\infty}\frac{n(n-1)...(n-y+1)}{n^y} = 1$ , and $\lim_{n\to\infty}(1-\frac{\lambda}{n}) = 1$ , then the Equation (3.4) is expressed as

$$P(y;\lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad (3.5)$$

The standard form of the Poisson regression model is derived in Equation (3.5).

In line of the application for crash prediction analysis, let $y_i$ indicate the number of the crash counts of a specific entity $i$; $\mu_i$ ($\lambda_i$ in Equation (3.5)) indicates the mean or expected crash counts of the corresponding entity. The probability function of the Poisson regression model with regard to the entity $i$ is as below:

$$P(y_i;\mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \tag{3.6}$$

A link function is then formulated to associated with the crash counts and the potential explanatory factors, as follows

$$\ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + ... + \beta_K x_{iK} \tag{3.7}$$

where $\beta_0$ is the intercept and $\beta_k (k = 1, 2, ..., K)$ are the coefficients to be estimated for corresponding exogenous variables $x_{ik} (k = 1, 2, ..., K)$.

### 3.3.2 Negative binomial regression model

The negative binomial regression is also known as the Poisson-gamma model, which can be derived from the Poisson model. Owing to the distributional assumption, the Poisson model is only applicable to the data with equal mean and variance. Therefore, the Poisson regression model is not capable of modeling overdispersed data.

An overdispersion test to the data should be carried out for the decision on appropriate the model (the Poisson or the negative binomial model) to be used for the model estimation. The hypotheses, $H_0$ and $H_1$, assuming that the relationships between the mean of the data and the variance are examined as follows (Cameron and Trivedi, 1990).

$$H_0 : Var(y_i) = E(y_i) \tag{3.8}$$

$$H_1 : Var(y_i) = E(y_i) + a \cdot g(E(y_i)) \tag{3.9}$$

In the above, $H_0$ is the null hypothesis that the mean of the data equals the variance. $y_i$ indicates the number of crashes in zone entity $i$. $E(y_i)$ and $Var(y_i)$ are the mean and variance of $y_i$, respectively. In contrast to the hypothesis $H_0$, the hypothesis $H_1$ assumes that the variance of data does not simply equal the mean. $g(\cdot)$ is the function of data mean, $E(y_i)$, and $a$ is the scaling parameter of $g(\cdot)$. Two possible functions for $g(\cdot)$ have been tested are tested as follows:

$$g(E(y_i)) = E(y_i) \tag{3.10}$$

and

$$g(E(y_i)) = E^2(y_i) \tag{3.11}$$

The statistics of ovedispersion test are then derived using a linear regression model as follows.

$$z_i = b \cdot w_i \tag{3.12}$$

where $z_i$ and $w_i$ are two indicators calculated for each entity $i$ in Equation (3.13) and Equation (3.14), respectively. $b$ in Equation (3.12) indicates the regression parameter between $z_i$ and $w_i$.

$$z_i = \frac{(y_i - E(y_i))^2 - y_i}{\sqrt{2}E(y_i)} \tag{3.13}$$

$$w_i = \frac{g(E(y_i))}{\sqrt{2}E(y_i)} \tag{3.14}$$

To better fit the overdispersed data, a heterogeneity error is introduced to the mean of the Poisson distribution, given by

$$P(y_i;\lambda_i,u_i) = \frac{(\lambda_i u_i)^{y_i} e^{-(\lambda_i u_i)}}{y_i!} \tag{3.15}$$

Equation (3.15) can be treated as a Poisson model, in which the mean, $\mu_i$ is expressed as $\mu_i = \lambda_i u_i$. $u_i$ is an error term following gamma distribution. The unconditional form of Equation (3.15) is calculated as:

$$P(y_i;\lambda_i) = \int_0^\infty \frac{(\lambda_i u_i)^{y_i} e^{-(\lambda_i u_i)}}{y_i!} f(u_i) du_i \tag{3.16}$$

where $f(u_i)$ is the gamma distribution given by $u_i = e^{\varepsilon}$ with mean 1. Equation (3.16) is then expanded as below:

$$
\begin{aligned}
P(y_i;\lambda_i,\nu) &= \int_0^\infty \frac{(\lambda_i u_i)^{y_i} e^{-(\lambda_i u_i)}}{y_i\,!} \frac{\nu^\nu}{\Gamma(\nu)} u_i^{\nu-1} e^{-(\nu u_i)} du_i \\
&= \frac{\lambda_i^{y_i}}{\Gamma(y_i+1)} \frac{\nu^\nu}{\Gamma(\nu)} \int_0^\infty e^{-(\lambda_i+\nu)u_i} u_i^{(y+\nu)-1} du_i
\end{aligned}
\tag{3.17}
$$

where $\nu$ is the parameter of the gamma distribution for $u$. Recall the nature of gamma distribution, Equation (3.17) is derived as:

$$
\begin{aligned}
P(y;\lambda,\nu) &= \frac{\lambda^y}{\Gamma(y+1)} \frac{\nu^\nu}{\Gamma(\nu)} \frac{\Gamma(y+\nu)}{(\lambda+\nu)^{y+\nu}} \\
&= \frac{\lambda^y}{\Gamma(y+1)} \frac{\Gamma(y+\nu)}{\Gamma(\nu)} \frac{\nu^\nu}{(\lambda+\nu)^{y+\nu}} \\
&= \frac{\Gamma(y+\nu)}{\Gamma(y+1)\Gamma(\nu)} \left(\frac{\nu}{\lambda+\nu}\right)^\nu \left(\frac{\lambda}{\lambda+\nu}\right)^y \\
&= \frac{\Gamma(y+\nu)}{\Gamma(y+1)\Gamma(\nu)} \left(\frac{\nu}{\lambda+\nu}\right)^\nu \left(1-\frac{\nu}{\lambda+\nu}\right)^y \\
&= \frac{\Gamma(y+\nu)}{\Gamma(y+1)\Gamma(\nu)} \left(\frac{1}{1+\lambda/\nu}\right)^\nu \left(1-\frac{1}{1+\lambda/\nu}\right)^y
\end{aligned}
\tag{3.18}
$$

Inverting the scale parameter of the gamma distribution, $\nu$, yields $\alpha = 1/\nu$, which is known as the negative binomial heterogeneity or the overdispersion parameter. A standard form of the negative binomial regression model is shown below.

$$
P(y_i;\lambda_i,\alpha) = \frac{\Gamma(y_i+1/\alpha)}{\Gamma(y_i+1)\Gamma(1/\alpha)} \left(\frac{1}{1+\alpha\lambda_i}\right)^{1/\alpha} \left(1-\frac{1}{1+\alpha\lambda_i}\right)^{y_i} \tag{3.19}
$$

Regarding the properties of Gamma function, $\Gamma(y_i+1/\alpha) = (y_i+1/\alpha-1)!$ , $\Gamma(y_i+1) = y_i\,!$ and $\Gamma(1/\alpha) = (1/\alpha-1)!$, Equation (3.19) is usually written as:

$$
P(y_i;\lambda_i,\alpha) = \begin{pmatrix} y_i+1/\alpha-1 \\ 1/\alpha-1 \end{pmatrix} \left(\frac{1}{1+\alpha\lambda_i}\right)^{1/\alpha} \left(1-\frac{1}{1+\alpha\lambda_i}\right)^{y_i} \tag{3.20}
$$

The link function associating the expected mean and the influencing factors is given below:

$$
\ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK} + \varepsilon_i \tag{3.21}
$$

For the crash data with variance larger than the mean, the negative binomial regression model is a preferable approach for the modeling.

### 3.3.3 Random-parameter model

Since not all influencing factors to crash occurrence are available, the unobservable heterogeneity can result in biased estimators based on the limited observable factors (Mannering and Bhat, 2014). To capture the variations of the effects of influencing factors on crash occurrence, the random-parameter approach has been introduced for crash prediction modeling (Anastasopoulos and Mannering, 2009). The parameters associated with the factors are no longer fixed but they can vary across the studied entities (e.g., individuals, time, spatial units) (Train, 2009). Assumed that the set of parameters $\boldsymbol{\beta}$ are following some kinds of distributions and their probability density function is $f(\boldsymbol{\beta})$, the probability of crash occurrence in Equation (3.6) or (3.20) is calculated as:

$$P(y_i) = \int p(y_i \mid \boldsymbol{\beta}) f(\boldsymbol{\beta}) d\boldsymbol{\beta} \tag{3.22}$$

If the random parameters, $\boldsymbol{\beta}$, collapse to the case that only the intercept is found to be randomly distributed, the random-parameter model is equivalent to the random-effect model. Halton draws simulation method has been used to solve random-parameter models and provided stable results (Train, 2001).

### 3.3.4 Joint probability model

In response to the modeling of potential correlation among total and pedestrian crash counts and their hierarchy subset relationship, a conditional joint probability approach (Pei et al., 2011) is applied to estimate the possible correlation of factors on the occurrence of total crashes and pedestrian crashes simultaneously. Let $p(y_i)$ denote the probability of $y_i$ total crashes occurred and $p(y_i^P \mid y_i)$ denote the probability of $y_i^P$ pedestrian crashes occurred conditional on $y_i$ total crashes in zone $i$. The joint probability of $y_i$ total crashes and $y_i^P$ pedestrian crashes occurring can then be specified as

$$p(y_i^P, y_i) = p(y_i) \cdot p(y_i^P \mid y_i) \tag{3.23}$$

The formulations of probability functions $p(y_i)$ and $p(y_i^P \mid y_i)$ are given in the following paragraphs.

In the study of the thesis, the results of the over-dispersion test indicated that the variance of the crash counts is greater than the mean, at a 1% level of significance. Therefore, a negative binomial regression model is formulated to model the total crash frequency as in Equation (3.20).

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

$$p(y_i^P \mid y_i) = \begin{pmatrix} y_i \\ y_i^P \end{pmatrix} (\pi_i^P)^{y_i^P} (1-\pi_i^P)^{y_i - y_i^P} \quad (3.24)$$

where $\pi_i^P$ is the binomial probability of pedestrian involvement in crashes. The relationship between this probability and possible influencing factors can be measured using a logit function specified, as follows

$$\log it(\pi_i^P) = \log(\frac{\pi_i^P}{1-\pi_i^P}) = \boldsymbol{\beta}_P^T \cdot \boldsymbol{x}_i \quad (3.25)$$

where $\boldsymbol{x}_i$ denotes the vector of explanatory factors and $\boldsymbol{\beta}_P$ denotes the corresponding coefficients that reflect their effects on pedestrian involvement in crashes.

Based on the formulations provided in Section 3.3.1 and Section 3.3.2, the joint probability function of observing $y_i^P$ pedestrian crashes and $y_i$ total crashes in zone area $i$ can be established via substitution of Equation (3.20) and Equation (3.24) into Equation (3.23), resulting in

$$p(y_i^P, y_i) = \frac{\Gamma(y_i + 1/\alpha)}{(y_i+1)\Gamma(1/\alpha)} (\frac{1}{1+\alpha\mu_i})^{1/\alpha} (1-\frac{1}{1+\alpha\mu_i})^{y_i} \cdot \begin{pmatrix} y_i \\ y_i^P \end{pmatrix} (\pi_i^P)^{y_i^P} (1-\pi_i^P)^{y_i - y_i^P} \quad (3.26)$$

### 3.3.5 Bayesian inference

A full Bayesian approach is used to model the joint probability of total crashes and pedestrian crashes as specified in Equation (3.26). In conventional maximum likelihood estimation, point estimates of $\beta_N$ and $\beta_B$ are applied. By contrast, the Bayesian approach is superior because it can provide a posterior distribution (probability density function) of the parameters using simulation-based estimation (Christensen et al., 2011; Gelman et al., 2013).

Given that $y$ denotes the observed crash counts and $\theta$ denotes the parameters to be estimated, the posterior probability of parameters, based on the Bayes' rule, can be calculated as

$$p(\theta \mid y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)\, p(y \mid \theta)}{p(y)} \qquad (3.27)$$

where $p(\theta, y)$ is the probability that the observed outcome is $y$ and the parameters to be estimated is $\theta$. $p(y)$ is of an observed event and thus serves as a normalization constant. Therefore, the posterior probability is linearly proportional to the prior distribution of parameters $p(\theta)$ multiplied by the likelihood function of the observed outcome $p(y \mid \theta)$ as

$$p(\theta|y) \propto p(\theta) \cdot p(y|\theta) \qquad (3.28)$$

The calculation of the prior distribution of parameters, $p(\theta)$, which are unknown, is critical for the Bayesian model.

The MCMC simulation approach is applied for the estimation of the parameters in the Bayesian model. For instance, the Gibbs sampling and Metropolis-Hastings algorithm approaches are adopted to draw a sample from the prior distribution of unknown parameters in an iterative process (Gelman et al., 2013). The non-informative prior distributions are set as for the parameters as the prior knowledge. During the estimation, a number of iterations will be run as the burn-in stage before the achievement of proper distributions of the values of parameters interested.

### 3.3.6 Spatial random effect with conditional autoregressive (CAR) model

Considering the potential heterogeneous effects from unobserved factors, spatially correlated effects might exist concerning the crashes (Bhat et al., 2017). Global Moran's I statistic, as proposed by Moran (1950) are introduced to examine the existence of such spatial correlation. The statistics are defined as shown in Equation (3.29).

$$I = \frac{n\sum_{i}\sum_{j}\omega_{ij}(y_i - \bar{y})(y_j - \bar{y})}{W_{sum}\sum_{i}(y_i - \bar{y})^2} \tag{3.29}$$

Here, $n$ denotes the number of zones; $y_i$ and $y_j$ follows the definition in Equation (1) as the pedestrian crash counts in zone area $i$ and $j$ respectively; and $\bar{y}$ is the average number of the pedestrian crash counts. $\omega_{ij}$ denotes a spatial weight that indicates the spatial dependence between zone $i$ and zone $j$. $W_{sum}$ denotes the sum of all the spatial weights. A zero value for statistics $I$ indicates that there is no spatial correlation across the spatial areas. The $I$ test results for the pedestrian crashes over the analyzed zone area in this thesis is 0.12 with p-value of 0.004. The statistical significance indicates that the spatial correlation should not be ignored in the analysis.

Regarding the consistency of spatial correlation effect in the occurrence of crashes in these areas, a CAR is constructed in the model formulation, to model crash occurrence. Equation (2) can then be further expressed as shown in Equation (3.21).

$$\ln(\mu_i) = \boldsymbol{\beta}^T \cdot \boldsymbol{x}_i + \mathrm{S}_i + \varepsilon_i \tag{3.30}$$

In the above, $S_i$ stands for the random spatial effects (for capturing potential correlations), and is subject to a conditional Gaussian distribution given the value of the other zone areas denoted by $\mathrm{S}_{-i}$ . The distribution is formulated as shown in Equation (3.31)

$$S_i \mid S_{-i} \sim Normal(\overline{S_i}, \tau_i^2) \tag{3.31}$$

$$\overline{\mathrm{S}_i} = \sum_{j \neq i} \frac{\omega_{ij}\mathrm{S}_j}{\omega_{i+}} \tag{3.32}$$

$$\tau_i^2 = \frac{\omega^2}{\omega_{i+}} \tag{3.33}$$

In the above, $\tau_i$ is the standard deviation of the conditional normal distribution of $S_i$. $\omega_{ij}$ follows the same definition as in Equation (3.29), with $\omega_{ij} = \omega_{ji}$, $\omega_{ii}$ = 0 and $\omega_{i+}$ = $\sum_j \omega_{ij}$. For precision, $\omega_{ij}$ is set to be 1. And $\omega^2$ is the precision parameter in the CAR prior.

### 3.3.7 Assessment of the model performance

The likelihood of the data set to be observed over all the observations given the probability density functions as in Equation (3.6), (3.20), and (3.26) is

$$L{=}P(Y) = \prod_{i=1}^{N} p(y_i) \tag{3.34}$$

The parameters that enable $L$ reach the maximum value are the optimal parameters for a model. For ease of calculation, logarithm value is taken to $L$ for transforming multiply operation to additive operation as below:

$$LL = \log(L) = \sum_i \log(p(y_i)) \tag{3.35}$$

In line of this operation, the parameters that enable $LL$ reach its maximum value are the optimal estimators. To indicate that the achievement of optimal estimators, Monte Carlo (MC) error is a commonly used criterion for the assessment of convergence of the simulations. When the MC error is getting less than 0.05, the distributions of parameters estimated are considered to be converged. In addition, Gelman-Rubin diagnostics is preferable as the criterion for the convergence assessment which required Gelman-Rubin statistics, $R$, to be less than 1.1 (Gelman and Rubin, 1992). Then, the posterior distribution of parameters can be given (Sinharay, 2003; Barua et al., 2016).

To make comparisons of the performances between different models, three indicators are commonly used for evaluating the goodness-of-fit of the random-parameter models, they are Pseudo $\rho^2$, Akaike information criterion (AIC), and Bayesian information criterion (BIC). They can be calculated as follows:

$$\rho^2 = 1 - \frac{LL(conv)}{LL(cons)} \tag{3.36}$$

$$AIC = -2\ln(L^*) + 2k \tag{3.37}$$

$$BIC = -2\ln(L^*) + \ln(n)k \tag{3.38}$$

In the above, $k$ is the number of parameters in the model and $n$ is the sample size. $L^*$ is the final likelihood at convergence while $LL(conv)$ and $LL(cons)$ represent the log-likelihood values for the estimated and constant only models, respectively. The model with the highest value of Pseudo $\rho^2$ and the lowest AIC and BIC values is considered to be best fitted (Washington et al., 2010; Hilbe, 2011).

The deviance of information criterion (DIC) was introduced for model assessment and comparison for Bayesian inference (Spiegelhalter et al., 2002), which is calculated as

$$DIC = 2\overline{D(\theta)} - D(\bar{\theta}) \tag{3.39}$$

where $\overline{D(\theta)}$ is the mean of the posterior deviance $D(\theta)$ and $D(\bar{\theta})$ is the deviance at the mean of posterior parameters. The deviance $D(\theta)$ of the model at the values of the parameter $\theta$ is defined as the minus twice of the log-likelihood:

$$D(\theta) = -2\log(P(\hat{y} \mid \theta)) \tag{3.40}$$

## 3.4 Summary

This chapter introduces the geographic demarcation system. The Planning Department develops the TPU zone system for planning, population projection, and providing a reference system for other departments. The TPU system divides the whole territory of Hong Kong into fourth levels, i.e., the PPUs, the SPUs, the TPUs, and the SB_VCs. A large set of zonal factors, i.e., sociodemographic characteristics, are released at the TPU level. Therefore, the TPU system serves as the major geographical unit system for carrying out the crash prediction models. On the other hand, when taking into account the within-day hourly variations of crash occurrence, the PDD reference system is applied for the analysis of hourly crashes. By doing so, the excess zeros of the crash counts will be avoided. Data from various sources, including traffic characteristics, sociodemographic census information, road network characteristics, POI information, built environment, and land use data are processed and matched into TPU and PDD zone systems via geographic information system (GIS) technique.

Methods have been elaborated on how to estimate pedestrian exposure at the macroscopic level from a larger-scale travel characteristics survey by making use of the surveyed trip diaries. The estimation approach pays great attention to pedestrians' walking trip legs in a multimodal trip, especially the trips made by using public transit modes. In a transit-oriented society like Hong Kong, walking to and from public transport stations contributes to the majority of the total amount of pedestrian walking. Therefore, taking into account walking trip legs among motorized trips to estimate pedestrian exposure can offer a more reasonable estimation of pedestrian exposure for pedestrian safety analysis.

Finally, the formulations of crash prediction models, including random-parameter negative binomial regression, the joint probability model, and Bayesian inference have been presented. Overdispersion tests and various assessment criteria for model performance have also been introduced. The methodologies lay out the theoretical foundation of the investigation into the pedestrian crash occurrence.

# Chapter 4 Role of exposure in pedestrian safety analysis

## 4.1 Introduction

An exposure measure is inevitable for the evaluation of crash risk. For pedestrian exposure measures at the macroscopic level, pedestrian volumes are rarely available due to data limitations (Qin and Ivan, 2001; Lam et al., 2014; Wang et al., 2016b; Lee et al., 2019). A number of the metrics have been examined as pedestrian exposure surrogates, including population/population density, the number of workers, predicted trips, and site counting (Jonah and Engel, 1983; Davis and Braaksma, 1988; Greene-Roesel et al., 2007; Wier et al., 2009; Chakravarthy et al., 2010; Cottrill and Thakuriah, 2010; Siddiqui et al., 2012a). However, none of the existing measures for pedestrians have been widely acknowledged and there is still a call for efforts to explore more reliable exposure measures for crash prediction models.

On the other hand, it has been noted that crash involvement is the by-product of activity participation (Elias et al., 2010; Elias and Shiftan, 2014). The trip purposes can influence the performances of on-road travelers, which play important roles in the occurrence of crash occurrence. Under similar conditions, people in a hurry to get to work have higher opportunities to disobey traffic rules, i.e., crossing roads during the red light period. Moreover, particular activity participation at particular times may also lead to hourly variations of crash occurrence.

This chapter examines the efficiency of the proposed measurement of pedestrian exposure in crash prediction modeling at the macroscopic level. The exposure measures are estimated from the revealed traffic diaries recorded during the travel characteristics survey. Three types of pedestrian exposure measures are compared, including conventional zonal population, pedestrian walking frequency, and walking time. In addition to the examination of the efficiency of pedestrian exposure derived from the travel characteristics survey, the exposure measure for pedestrians is disaggregated by the purposes of pedestrians' trip makings. The explanatory power of zonal factors, such as sociodemographic characteristics and road networks, has been tested. Three hypotheses to be examined in this chapter are: 1). the proposed pedestrian exposure measures estimated from revealed travel diaries improve the performance of crash prediction models with better goodness-of-fit; 2). different trip purposes play different roles in pedestrian crash occurrence; 3) there exist hourly variations of pedestrian crash occurrences.

Section 4.2 describes and summarizes the data used in this chapter. Section 4.3 shows the results of the crash prediction models. Following that, discussions are presented in Section 4.4. Lastly, Section 4.5 summarizes Study 1.

## 4.2 Data

The sources of the data supporting Study 1 were as described in Chapter 3. Regarding the excess zero crash count when the data is highly disaggregated (of small zone size and/or short time), the PDDs system with 26 zone units was adopted as the geographic system. The data, including crash counts and explanatory factors, is aggregated into PDDs using GIS technique.

In Study 1, 12,672 pedestrian crashes that occurred on weekdays during the period from 2011 to 2015 were captured from the TIS. TIS database consists of three profiles, namely crash environment, vehicle attributes, and casualty characteristics. Since the precise information on the location and time of every single crash is available, it is possible to allocate the crashes into units by space (26 zones) and time (24 hours) when evaluating the spatial-temporal variations in pedestrian crash incidence. Finally, there are 624 units for the proposed pedestrian crash prediction models. The distribution of pedestrian crashes (by zone) is depicted in **Figure 4.1**.

As shown in **Figure 4.1**, the pedestrian crashes are concentrated in Central Kowloon and North shore of Hong Kong Island (colored in red). It is not surprising since they are the Central Business Districts (CBDs), and the interactions between pedestrian and vehicular traffic are prevalent. This justifies the need to estimate pedestrian exposure, in terms of the amount of travel by walking.

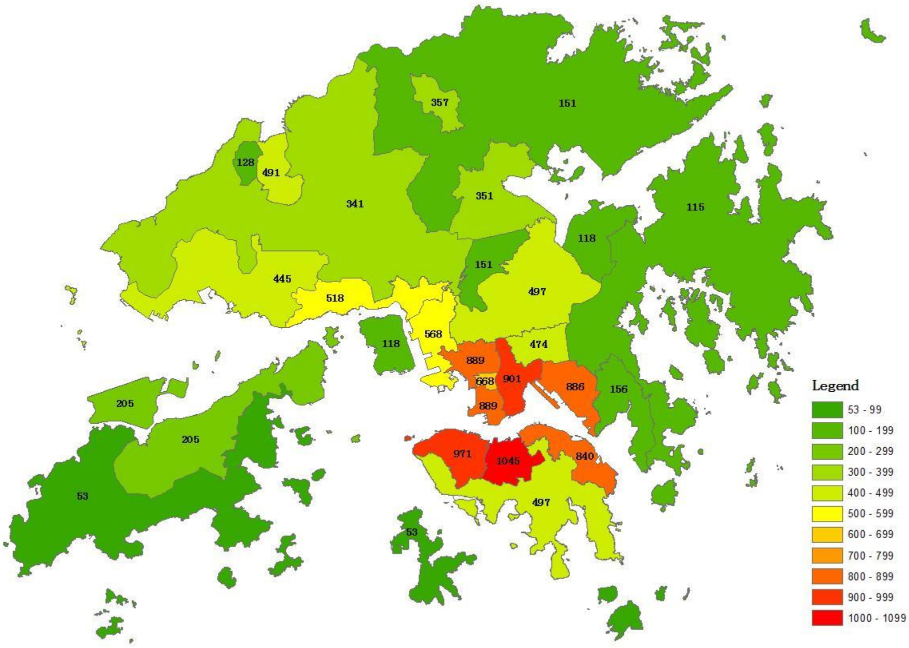



Figure 4.1 Spatial distribution of pedestrian crash

The estimation of pedestrian exposure makes use of the HIS database, mainly the information of revealed motorized trips (involved with one or more motorized modes with a trip) records and walk-only trips. A total of 122,237 motorized trips and 13,885 walk-only trips were recorded in the survey. 20% of the walk-only trips more than 10 minutes were recorded with trip-making time and origin/destination locations. After the adjustment for the trip-making profile of the whole territory, 12.6 million motorized trips and 1.2 million tracked walk-only trips were estimated to be made on a weekday. 84% of the motorized trips involved only one motorized mode while 14% of the trips have two motorized modes and the left 2% comprised more than two motorized modes.

With detailed trip legs recorded in the motorized modes, the walking trip legs accessing to/egressing from a motorized mode can be identified for the calculation of pedestrian exposure. Given the locations of origin/destination, the boarding, and alighting sites, the pedestrian movements of each trip leg were well assigned onto the map. For instance, a trip-making that took the franchised bus and transferred to the metro involved three walking trip legs, shown in **Figure 4.2**. The citizen departed in a zone in Tuen Mun at 07:30 a.m. and arrived at a franchised bus station in the neighboring zone after a 2-minutes walking. He/she alighted from the franchised bus at a zone in Central and walked approximately 4 minutes to a metro station in the neighboring zone. Finally, the citizen alighted from the subway in a zone of North Point and walked for another 15 minutes to the destination. The three walking activities would be ignored without further information about these motorized tips provided and being looked at. However, these walking activities within a motorized trip constitute a large proportion of daily walking in a transit-oriented society. The inclusion of these walking trip legs for area-wide pedestrian exposure measures enhances the representative of more accurate pedestrian movements, which got little studied in the existing studies.

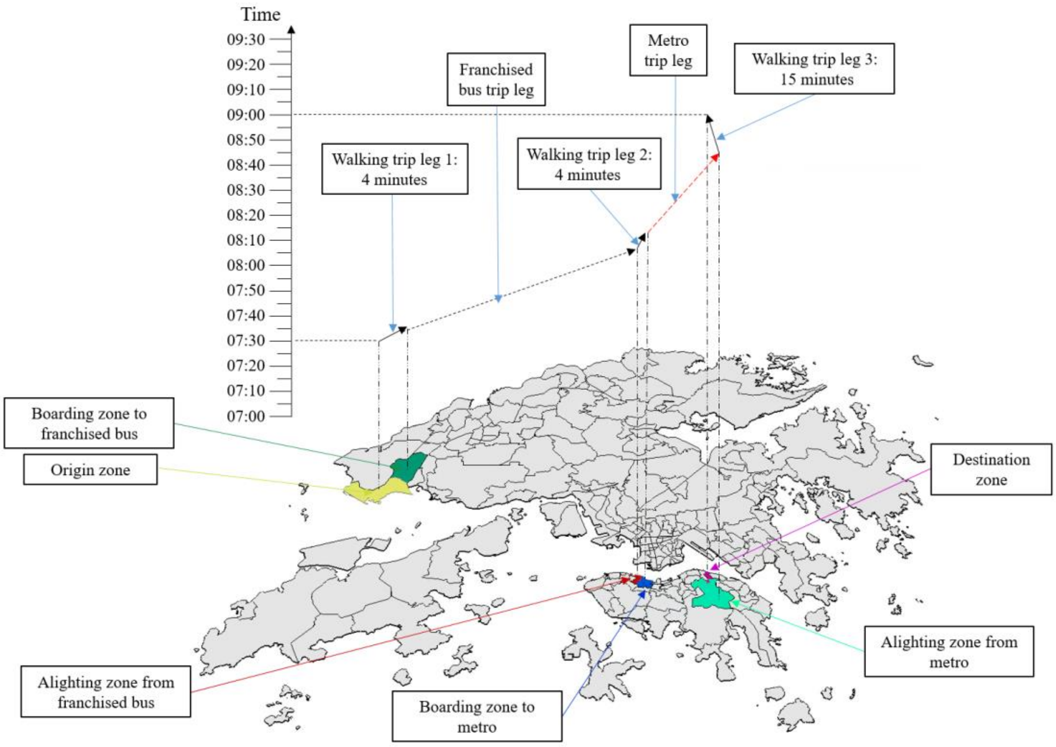



Figure 4.2 A multimodal trip-making with three walking trip legs

A total of 28.71 million walking frequencies (including walk-only trips and trip legs) and 2.40 million walking hours had been estimated for representing pedestrians' movements from the TCS2011 after adjustment and validation. With detailed locations recorded for travelers' origins, destinations, boarding, and alighting during transferring, the estimated walking frequencies and walking-hour can be easily matched into different zones as area-wide pedestrian exposure measures.

In addition, the information provided from the trip records in TCS is more than the trip counts and time spend but also the information related to the activity-travel pattern, i.e., travel purposes and the mode usage of pedestrians. The information enables further investigation on pedestrian exposure by trip purposes and by modes.

The distribution of walking frequency by the 624 analysis units (i.e., 26 zones x 24 hours) is depicted in **Figure 4.3**. As shown in **Figure 4.3**, four analysis units had zero walking trips (leg). They were all in the rural area and before dawn (e.g., from 2.00 am to 4.59 am).

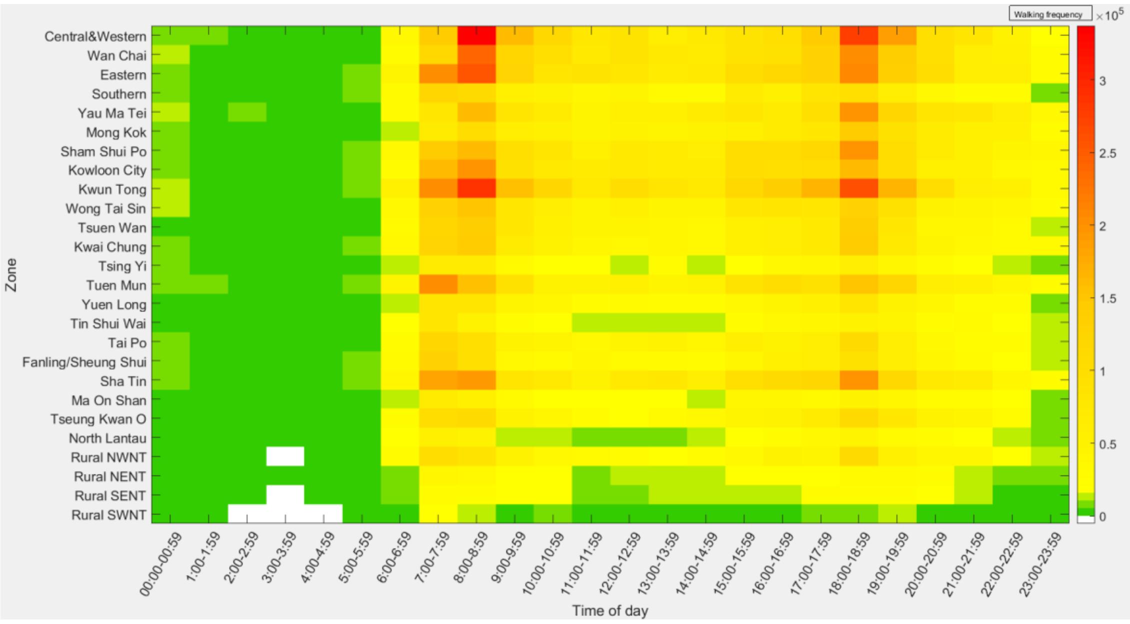



Figure 4.3 Space-time distribution of walking frequency

Variations in the pedestrian crash frequency and walking frequency, by time of the day, are depicted in **Figure 4.4**. The peak walking frequencies occurred during the periods from 08:00 am to 08:59 am and from 18:00 pm to 18:59, respectively. Except during the period from midnight to dawn (e.g., from 23:00 to 05:59), the walking frequency was the lowest in the middle of the day while the pedestrian crash frequency was consistently high throughout the period from 8:00 am to 19:00. Because of the discrepancy between crashes and walking frequency, bias may exist in the association measure when only the aggregated walking frequency was used as the exposure for the pedestrian crash prediction model. In Study 1, the walking frequency and walking time of different trip purposes in different time periods were considered. Specifically, trip purposes could be categorized into six: home, work, school, shopping, dining, and others (Miller and Roorda, 2003; Bhat et al., 2013; Allahviranloo and Recker, 2015).

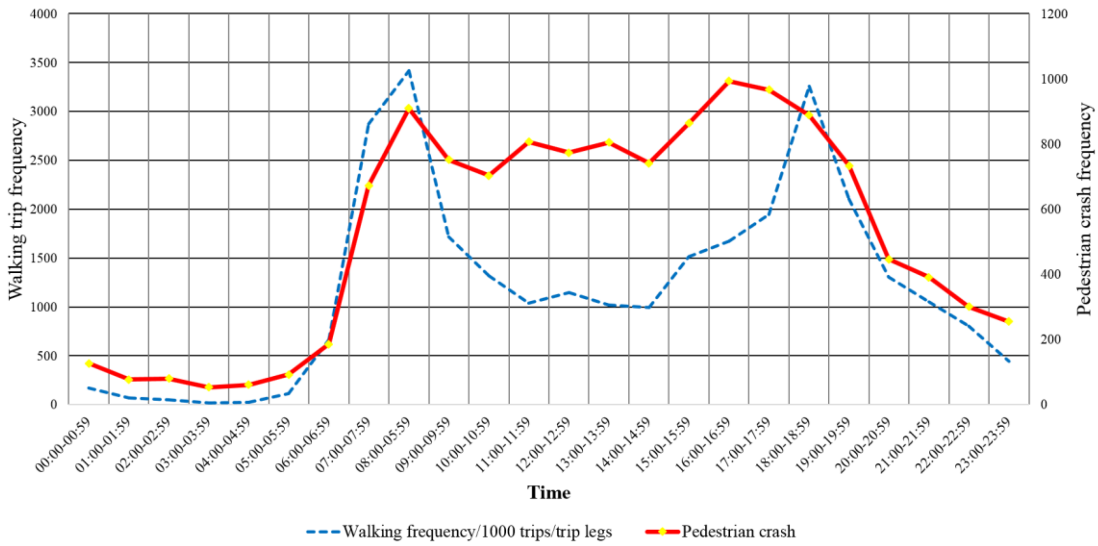



Figure 4.4 Temporal distributions of pedestrian crash frequency and walking frequency

Taking into account the effect of multi-modal (vehicle and pedestrian) exposure, vehicular traffic was also considered in the proposed crash prediction model. For instance, the information on AADT (and hourly count) was obtained from the ATC 2011. Additionally, information on age, employment status, household size, and household income were obtained from the population census. Furthermore, the transport network and traffic control attributes including the number of signalized and non-signalized intersections, and the number of zebra crossings, were also incorporated into the proposed model. The above data was matched with the crash, traffic count, and pedestrian exposure data using the GIS technique. **Table 4.1** summarizes the characteristics of the candidate variables.

Table 4.1 Summary statistics of the variables used in Study 1

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Pedestrian crash count | 20.31 | 20.69 | 0 | 89 |
| Walking frequency | | | | |
| Home | 20,711.00 | 29,755.74 | 0 | 219,524 |
| Work | 11,282.38 | 26,590.24 | 0 | 288,345 |
| School | 3,725.92 | 10,508.50 | 0 | 92,389 |
| Shopping | 3,792.33 | 5,023.76 | 0 | 32,209 |
| Dining | 2,547.89 | 4,251.15 | 0 | 35,439 |
| Others | 3,953.78 | 4,699.43 | 0 | 25,726 |
| Walking time (minute) | | | | |

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Home | 106,281.40 | 156,019.10 | 0 | 1,360,410 |
| Work | 54,653.49 | 135,696.30 | 0 | 1,410,589 |
| School | 18,066.02 | 53,136.65 | 0 | 544,495 |
| Shopping | 21,320.60 | 28,749.95 | 0 | 168,726 |
| Dining | 11,684.00 | 20,491.71 | 0 | 205,374 |
| Others | 19,035.26 | 22,239.84 | 0 | 124,815 |
| Annual average hourly traffic | 1,514 | 1,098 | 10 | 4,646 |
| Zonal population | 271,938 | 146,182.50 | 57,338 | 622,152 |
| Proportion of population | | | | |
| Age below 15 years | 0.12 | 0.01 | 0.10 | 0.15 |
| Age 15-64 years | 0.75 | 0.26 | 0.72 | 0.80 |
| Age above 64 years | 0.13 | 0.03 | 0.07 | 0.18 |
| Unemployed | 0.33 | 0.03 | 0.27 | 0.37 |
| Non-Chinese | 0.08 | 0.06 | 0.02 | 0.21 |
| Proportion of household with more than 3 members | 0.34 | 0.04 | 0.25 | 0.40 |
| Median household income | 23,365 | 5,461.49 | 7,500 | 12,500 |
| Road density (km/km$^2$) | 9.18 | 7.58 | 0.33 | 31.28 |
| Number of signalized intersections | 56.81 | 39.60 | 0 | 141 |
| Number of non-signalized intersections | 798.19 | 342.02 | 156 | 1,544 |
| Number of zebra crossings | 10.23 | 8.89 | 0 | 31 |

As shown in **Table 4.1**, walking frequency and time of back home trip were predominant among the six travel purposes. Back home trips constituted 45% of walking frequency and 46% of the walking time respectively. For the mandatory travel (e.g., commuting), work trips constituted 25% of walking frequency and 24% of the walking time respectively. For the demographic characteristics, vulnerable groups (i.e., elderly and children) constituted one-fourth of the overall population. In Hong Kong, the average household size was 2.8 people, and the proportion of households with more than three members was 34%.

With regard to the effects of demographics and socioeconomics, relevant factors including age, ethnicity, education, household income, and household size are considered. As summarized in **Table 4.1**, 13% of the population of a zone, on average, were of ages above 64. Similar to other developed societies, Hong Kong is facing a problem with the aging population; the proportion of the population aged above 64 is expected to increase to 25% by 2035 (Sze and Christensen, 2017). The majority of the Hong Kong citizens surveyed were Chinese; approximately 8% of the population of a zone, on average, were non-Chinese. The average household size was 2.9 persons, and approximately one-third of households had more than 3 members (Transport Department, 2012b).

Prior to the estimation between pedestrian crash frequency, exposure, and possible factors, the variance inflation factor (VIF) was estimated to identify any possible collinearity among the variables. The candidate variables with VIF greater than five (e.g., proportion of unemployment, the number of signalized intersections) were not considered in the consolidated model. As the main purpose of Study 1 is to identify the significant contributory factors (and their effects) to pedestrian crash, it is necessary to exclude the (theoretically less important) correlated variables, even which the estimates of remaining correlated variables might be biased (Hair et al., 1998; Loo et al., 2010; Tay, 2017).

## 4.3 Estimation results

The over-dispersion test indicated that the pedestrian crash data was subject to significant over-dispersion at the 1% level. Therefore, the negative binomial regression model was preferred to the Poisson model.

Three random-parameter negative binomial regression models were established to measure the association between pedestrian crash frequency and possible risk factors. The exposure to pedestrian crashes was represented by the zonal population in Model 0, walking frequency in Model 1, and walking time in Model 2 respectively. Results of the three estimated models were given in **Table 4.2**.

Table 4.2 Estimation results of random parameter negative binomial models

| Variables | Model 0 | | Model 1 | | Model 2 | |
|---|---|---|---|---|---|---|
| | Coef. | p-value | Coef. | p-value | Coef. | p-value |
| Constant | -0.214 | 0.320 | -3.22 | <0.001 | -3.39 | <0.001 |
| ln(walking frequency) | | | | | | |
| Home | | | 0.25 | <0.001 | | |
| Work | | | 0.11 | <0.001 | | |
| School | | | 0.08 | <0.001 | | |
| Shopping | | | 0.07 | <0.001 | | |
| Dining | | | 0.02 | 0.07 | | |
| Others | | | 0.03 | 0.02 | | |
| ln(Walking time) | | | | | | |
| Home | | | | | 0.22 | <0.001 |
| Work | | | | | 0.09 | <0.001 |
| School | | | | | 0.06 | <0.001 |
| Shopping | | | | | 0.05 | <0.001 |
| Dining | | | | | 0.03 | <0.001 |
| Others | | | | | 0.02 | <0.001 |
| Annual average hourly traffic | 0.09 | 0.06 | -0.03 | 0.01 | -0.02 | 0.03 |
| Zonal population | 1.59 | <0.001 | | | | |
| Proportion of population | | | | | | |
| Age below 15 years | | | 8.02 | <0.001 | 6.78 | <0.001 |
| *std. of age below 15 years* | | | *1.47* | *<0.001* | *2.09* | *<0.001* |
| Age above 64 years | 6.53 | 0.01 | 4.15 | <0.001 | 3.92 | <0.001 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Proportion of household with more than 3 members | -4.77 | 0.003 | -2.44 | <0.001 | -2.33 | <0.001 |
| *std. of household with more than 3 members* | | | *0.45* | *<0.001* | | |
| Median household income | 0.003 | 0.001 | 0.002 | <0.001 | 0.003 | <0.001 |
| *std. median household income* | | | *0.45* | *<0.001* | *0.000* | *<0.001* |
| Road density (km/km$^2$) | 0.03 | <0.001 | 0.02 | <0.001 | 0.03 | <0.001 |
| Number of non-signalized intersection | 0.02 | <0.001 | 0.04 | <0.001 | 0.04 | <0.001 |
| Number of zebra crossing | -0.007 | <0.001 | 0.01 | <0.001 | 0.01 | <0.001 |
| Log likelihood with constant only | -2514.17 | | -2514.17 | | -2513.17 | |
| Log likelihood at convergence | -2408.90 | | -1838.17 | | -1857.56 | |
| Number of observations | 624 | | 624 | | 624 | |
| Pseudo $\rho^2$ | 0.04 | | 0.27 | | 0.26 | |
| AIC | 4837.80 | | 3714.35 | | 3751.12 | |
| BIC | 4875.73 | | 3798.64 | | 3824.54 | |

As shown in **Table 4.2**, Pseudo $\rho^2$ of Model 0, Model 1 and Model 2 were 0.04, 0.27 and 0.26, AICs of Model 0, Model 1 and Model 2 were 4,837.80, 3,714.35 and 3,751.12, and BICs of Model 0, Model 1 and Model 2 were 4,875.73, 3,798.64 and 3,824.54, respectively. Model 1 was superior to the two other counterparts, given the highest Pseudo $\rho^2$ while the lowest AIC and BIC. This indicated that the model using the walking frequency as the proxy for pedestrian exposure had the best prediction performance.

For the model using walking frequency as a proxy for pedestrian exposure (Model 1), as shown in **Table 4.2**, walking frequency (for back home, work, study, and shopping), traffic flow, the proportion of population age below 15, age above 64, proportion of household with more than 3 members, median household income, road density, number of non-signalized intersection and zebra crossing all significantly affected the crash frequency, at the 1% level. In particular, increase in walking frequency (coefficient ranged from 0.03 to 0.25), the proportion of the vulnerable population (age below 15, coefficient = 8.02; and above 64, coefficient = 4.15), median household income (0.002), road density (0.02), and the number of non-signalized intersection (0.04) and zebra crossing (0.01), were correlated to the increase in pedestrian crash frequency. In contrast, increases in traffic flow (-0.03) and households with more than 3 members (-2.44) were correlated to the reduction in pedestrian crash frequency. Additionally, the effects of the proportion of the young population, and households with more than 3 members, and median household income on crash frequency varied across the observations. Similar results could be revealed for the model using walking time as the proxy for pedestrian exposure (Model 2).

Table 4.3 Elasticity of pedestrian crash frequency

| **Variable** | **Elasticity** | | |
|---|---|---|---|
| Walking frequency | Mean | t-statistics | (*p*-value) |
| Home | 0.25 | 26.52 | (<0.001) |
| Work | 0.11 | 11.47 | (<0.001) |
| School | 0.08 | 10.53 | (<0.001) |
| Shopping | 0.07 | 6.89 | (<0.001) |
| Dining | 0.02 | 1.81 | (0.069) |
| Other | 0.03 | 2.28 | (0.023) |

**Table 4.3** illustrates the elasticity estimates of walking frequency by travel purpose. An increase in pedestrian crash frequency is less proportionate to that of walking frequency, as the elasticities are all less than one (0.25, 0.11, 0.08, 0.07, and 0.03 for back home, work, school, shopping, and other trips respectively). Nevertheless, the effect of back-home travel is dominant among other travel purposes.

## 4.4 Discussion

### 4.4.1 Pedestrian exposure

The pedestrian crash prediction model using walking frequency as the proxy for pedestrian exposure was found superior to that using the zonal population and walking time. Though the population is often used as the proxy for exposure as it is readily available in the census dataset, it is not necessary that every individual would make the same number of (walking) trips in a specified period of time. On the other hand, the population could only reflect the number of residents in the area concerned, it should not necessarily reflect the exposure of commuters traveling in the area.

Besides, the model using walking frequency as the proxy better reflected the prevalence of vehicular-pedestrian interaction, as compared to that using walking time. For instances, the increase in the number of the walking trip by ten times was more likely correlated to the increase in the frequency of vehicular-pedestrian interaction to the same extent, however, it might not be the case for increasing the walking time by ten times (it could be attributed to the increase in the time spent on the footpath or waiting at the crosswalk) (Siddiqui et al., 2012a; Siddiqui et al., 2012b; Cai et al., 2017; Wang et al., 2017a). On the other hand, in the household travel survey, the reliability of (self-reported) information on walking frequency could be higher than that of walking time, because of the variation of time perception among individuals and thus the unreliability of self-reported walking time (Chu, 2003; Greene-Roesel et al., 2007). On the other hand, the increase in traffic flow was correlated to the reduction in pedestrian crash frequency. It could be attributed to the increase in safety awareness of the pedestrian due to the increase in traffic volume. Then, the propensity of reckless crossing behavior could be reduced, and therefore, the risk of pedestrian crashes reduced.

The increase in pedestrian exposure was found correlated to the increase in the frequency of pedestrian crashes. However, as shown in **Figure 4.3**, the pedestrian crash rates (per unit walking frequency) were remarkably high during the noon (12:00 – 13:59) and late afternoon (15:00 - 17:59), even though the overall walking frequency was very low during the periods. Specifically, further categorizing the population that made back home trips from 12:00 p.m. to 13:59 by their economic status, students accounted for the majority of trips, up to 40% of the back home trips followed by the employed population at 36%. While turning to the back home trips made during the late afternoon, the employed population accounted for around 55% of the trips followed by around 30% of the back home trips made by students. When further looking at the distributions of the population who made back home trips during noon and late afternoon periods by age. It was found that up to 24% of the back home trips were made by children below 15-yers old, followed by 18% from teenagers between 15-24 years old. The statistics were consistent with the result that students contributed most to the back home trips during noon. Similar statistics were also found in the late afternoon. It could be attributed to the reckless behavior of pedestrians since it is the prime time for back home travel during the periods (Clarke et al., 2006; Anderson, 2008; Devlin et al., 2010; Palamara and Broughton, 2013). Considering the distribution of walking frequency by travel purpose, back home travel was dominating during the noon (about 35%) and late afternoon (about 75%) periods. Indeed, pedestrian inattentiveness was one of the top contributory factors to road crashes in recent years (Transport Department, 2012-2016). These are consistent with the findings of earlier studies that the travel purpose and time of the day can affect the travelers' behavior and thus the safety level on roads (Elias et al., 2010; Naderan and Shahi, 2010). This could shed light on the setting of road safety targets, and then better planning of road safety strategies that can enhance the safety of the vulnerable pedestrian group at specified time periods and locations (Wong et al., 2006; Sze et al., 2014; Sze and Christensen, 2017). In particular, it is necessary to improve road user education and enforcement measures that could enhance the safety awareness of pedestrians on their way back home (Wong et al., 2008). Yet, the association between safety perception and safety risk of an individual pedestrian was blurred when the crash data was aggregated by time and zone. It would be worth exploring the factors affecting the safety perception of pedestrians and thus crash involvement using observational and/or perceptional surveys.

### 4.4.2 Sociodemographic characteristics

Age could affect the frequency of pedestrian crashes. In particular, increases in the proportions of children and elderly people were correlated to the increase in the frequency of pedestrian crashes (Fontaine and Gourlet, 1997; Noland and Quddus, 2004; Palamara and Broughton, 2013; Jiménez-Mejías et al., 2016). It could be attributed to the variations in the obedience of traffic rules, risk perception, and safety awareness across different age groups (Ponnaluri and Nagar, 2010; Siddiqui et al., 2012a; Siddiqui et al., 2012b). In particular, the children and younger teens had a higher propensity to involve in reckless behavior on roads (Tay, 2003). For the elderly pedestrian, the high crash risk could be attributed to the degradation of cognitive performance, gap acceptance, and impaired mobility (Fontaine and Gourlet, 1997; Noland and Quddus, 2004; Palamara and Broughton, 2013; Jiménez-Mejías et al., 2016). Therefore, targeted road user education programs should be conducted in schools and community centers.

An increase in household median income was found correlated to the increase in pedestrian crash incidence. It could be attributed to the difference in safety perception and choice decisions across the population of different income levels. In particular, the trade-off between travel cost and crash risk could vary with the value of time of commuters (Tay, 2003). Such finding could be indicative to the planning and design of built environments, especially for the enhancement of accessibility, and therefore the safety performance, in accordance with the need of different population demographics and socio-economics (AECOM Asia Co.Ltd., 2010; Mannings (Asia) Consultants Ltd., 2010; MVA Hong Kong Limited, 2013). Additionally, an increase in the proportion of households with more than 3 members was found correlated to the reduction in the frequency of pedestrian crashes. It could be because the vulnerable road users (i.e., children and elderly) are usually accompanied by other household members, in particular the domestic helper or health caretaker, for the household with more members, which was commonly seen along the streets in Hong Kong. It could then enhance safety awareness and reduce the crash risk on roads (Huang et al., 2010; Cai et al., 2017). This is indicative to the road safety education and promotional strategies targeted to the caretakers of vulnerable pedestrian groups (Sze and Christensen, 2017).

### 4.4.3 Road network features

Last but not the least, transport system characteristics including road density (road length per unit area), number of non-signalized intersections, and the number of zebra crossing could affect pedestrian safety. Specifically, the increase in road density was found correlated to the increase in pedestrian crashes. It could be attributed to the increase in the potential vehicular-pedestrian interactions, and thus the incidence of pedestrian crashes. This should also be applicable to the association between pedestrian crashes, the number of non-signalized intersections, and the number of zebra crossings (Varhelyi, 1997). This implied that the installation of signal control at the intersection and crosswalk (segregating the conflicting vehicular and pedestrian streams by time and space) might effectively reduce the incidence of pedestrian crashes (Castro et al., 2012; Siddiqui et al., 2012a; Siddiqui et al., 2012b). This is indicative to the planning of pedestrian infrastructure and traffic management & control strategies that could mitigate the vehicular-pedestrian interaction, and therefore the pedestrian crash potential (Sze and Wong, 2007). It was noticed that the signs of annual average hourly traffic and number of zebra crossing were different across the three models. When the amount of walking is not considered (using zonal population as the proxy), it is expected that the crash incidence would increase with traffic volume and decrease with the presence of zebra crossing. Pedestrian perception and behavior are indeed correlated to traffic volume and traffic control. Therefore, when the amount of walking is used as the proxy (effect of pedestrian behavior is incorporated implicitly), the effects of traffic volume and presence of zebra walking could be reversed. Therefore, the increase in pedestrian crashes is correlated to the reduction in annual average hourly traffic and the presence of zebra crossing.

## 4.5 Conclusion

Study 1 contributes to the literature by developing a pedestrian crash prediction model, at the zonal level, that takes into account the effects of pedestrian exposure and travel purpose. Three hypotheses are verified: 1) the pedestrian exposure measures estimated from the travel characteristics survey provided better model performance than traditional zonal population; 2) the crash risks of pedestrian exposure by different trip purposes contribute differently on pedestrian crash occurrence; 3) the hourly variation of crash occurrence was witnessed and help verify the different crash risk of exposure by different trip purposes.

The contribution is twofold. Firstly, we sought an alternative approach for measuring pedestrian exposure based on the comprehensive household travel characteristics survey data. Therefore, the effect of the difference in the amount of walking across population groups on crash incidence can be considered. In particular, the model that applied the walking frequency as the proxy to pedestrian exposure was superior to that using the zonal population and walking time. Secondly, the effect of travel purpose on pedestrian crash risk was examined. The exposure was discretized by hours of the day and travel purposes. Results indicated that the crash risk of back home walking trip was more profound than that of other travel purposes. Additionally, the effects of confounding factors including age, household size, household income, road density, and intersection control on pedestrian safety were revealed. Results should be indicative to the targeted road user education, enforcement, and traffic control measures that could enhance the safety awareness of vulnerable pedestrian groups (i.e., young and elderly population), and therefore improve pedestrian safety in the long run. This is an essence to the promotion of walkability in an aging society like Hong Kong.

# Chapter 5 Effects of connecting transport mode on pedestrian safety

## 5.1 Introduction

Travel mode is one of the important attributes within activity-travel patterns. Different mode choices result in different travel patterns and behavior on roads. Individuals may also encounter different road environments and transport facilities. All those differences will account for differences in the risk of crash occurrence. Therefore, the relation between crash risk and mode selection should be investigated for a better understanding of crash occurrence and improvement in road safety.

A reasonable estimation of pedestrian exposure has been verified in Chapter 4 to advance the modeling of pedestrian crashes. Within a society where more than 90% of travel is by using public transport, walking connected with various motorized modes constitutes the majority of pedestrians' daily walking. The pedestrian exposure measures are classified by motorized modes to investigate the potential effects of walking exposure by different motorized modes.

In addition, due to the existence of unobserved heterogeneity across the spatial areas, there may exist spatial correlated effects on crash occurrence as not all possible factors can be perfectly incorporated in a study. For example, an unobserved factor, such as the number of hospitals in a zone, attracts not only travel within the located zone but also travel demand from neighboring zones. Ignorance of such spatial correlated effects for zonal crash frequency analysis would cause biased results and interpretation. A number of studies have taken into account the spatial correlation issue for crash analysis (Noland and Quddus, 2004; Aguero-Valverde and Jovanis, 2006). The issue will also be considered Study 2 in this chapter.

Apart from Study 1 in Chapter 4 that the effects of trip purposes on pedestrian crash occurrence had been investigated, Study 2 will consider the involvement of motorized modes in pedestrian exposure and its impacts on pedestrian crash occurrence. The hypotheses are: 1) walking by different transport modes contribute differently to pedestrian crash occurrence; 2) walking to and from roadway public transport, such as buses, is more likely than railway transit to lead to involvement with crash occurrence; 3) spatial correlations with crash occurrence does exist.

The rest of this chapter is organized as follows. Section 5.2 presents and summarizes the data aggregated in the TPU system. Section 5.3 shows the results of six pedestrian crash prediction models with different exposure measures. Section 5.4 goes into a detailed discussion about the results of the models. Section 5.5 gives a summary of Study 2.

## 5.2 Data

Multiple types of data were integrated to provide the investigation, including police-reported crash data, traffic volume data (e.g., AADT), daily passenger travel diaries, and sociodemographic information. In addition, roadway characteristics were available in the dataset, including data regarding the road length density, accessibility facilities for public transport (metro entrances/exits and bus stops), and intersections. In addition, built environment factors such as the number of restaurants, schools, shopping malls, and hotels were also included as candidate variables. Finally, the land areas for the different land-use types in each zone were also investigated. The sources of the data were described in Chapter 3.

The statistics from the population census were released based on a geographical unit system from the Planning Department, the TPU, which divided the whole territory of Hong Kong into 289 zones in 2011. However, owing to the sparse populations in some zones, the population census statistics were finally released as 209 TPUs, based on emerging zones with populations of less than 1000. Further, as the traffic count stations for the annual traffic census were only allocated for 179 zones out of the total 209 TPUs, a geographic unit with 179 zones was finally adopted for the data aggregation and model formulation. All of the above information was aggregated using the ArcGIS technique, according to the zone boundaries of the 179 TPUs. A summary of the data is presented in **Table 5.1**.

In view of the year of exposure measurements from TCS2011, the crash data in 2011 were used, with a total of 3,604 pedestrian crashes. The distribution of the pedestrian crashes over the studied areas is presented in **Figure 5.1**, and ranges between 0 and 134.

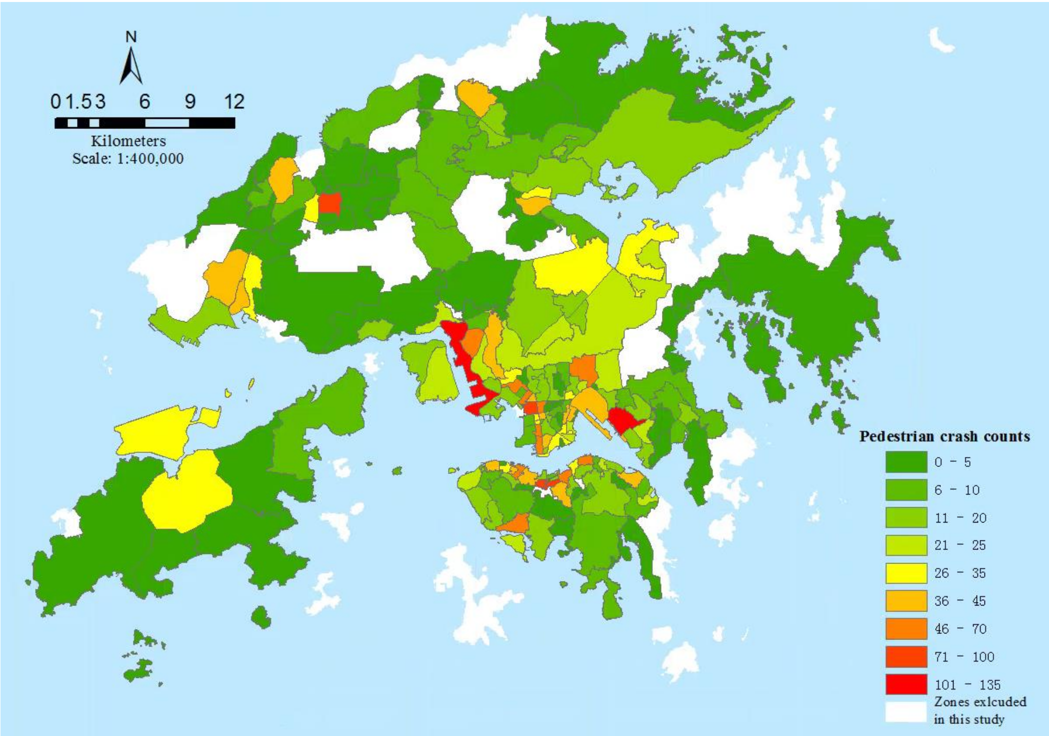



Figure 5.1 Distribution of pedestrian crashes in 2011 in Hong Kong

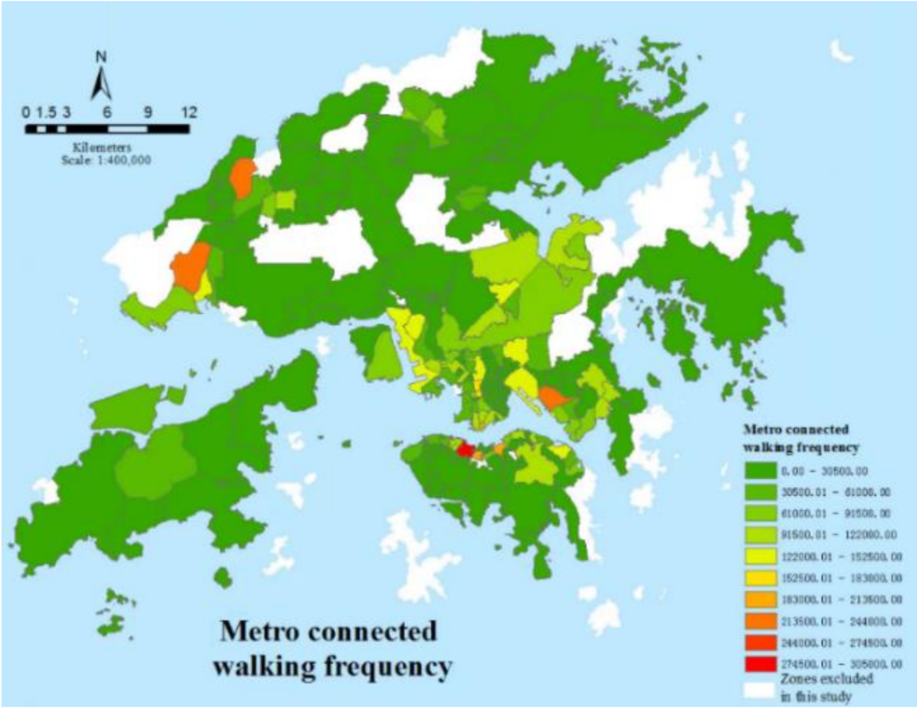
In Study 2, three alternatives of pedestrian exposure are explored for pedestrian crash modeling, i.e., zonal population, the aggregate pedestrian walking frequency, and the pedestrian walking frequency by different transport modes. The zonal population was directly available from the census statistics. Pedestrian walking frequency is derived from the TCS2011 database.

Table 5.1 Summary statistics of data samples

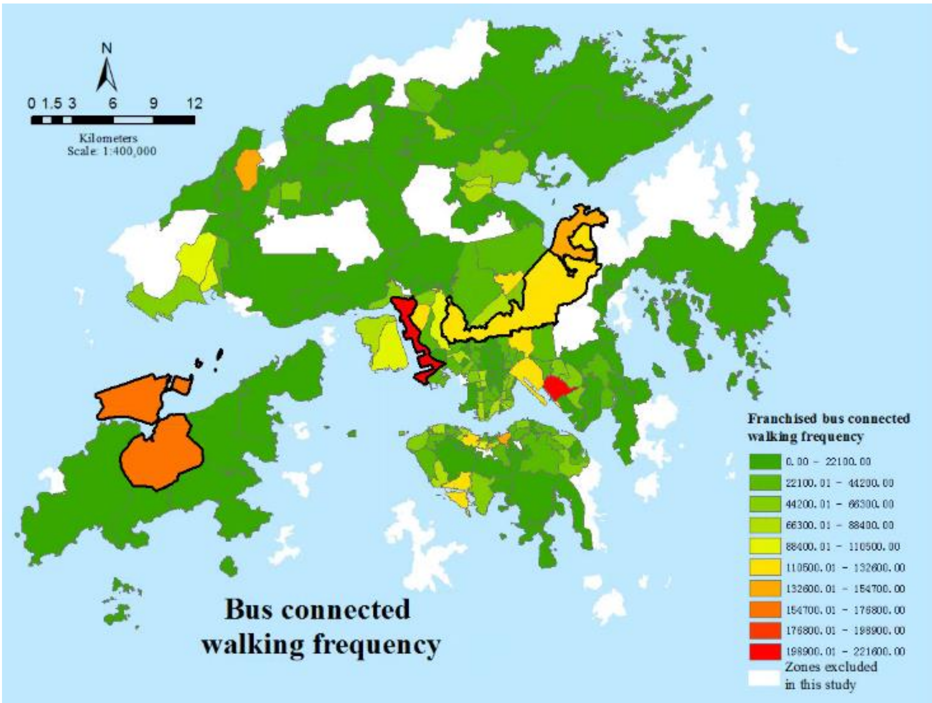
| Scope of work | Variable | Mean | Std. dev. | Min. | Max. |
|---|---|---|---|---|---|
| Crash | Pedestrian crash frequency | 20.13 | 22.54 | 0 | 134 |
| Exposure | Zonal population | 38066.76 | 42814.64 | 1023 | 287901 |
| | Daily walking trip (overall) | 154881.47 | 159058.39 | 973 | 925433 |
| | Daily walking trip by transport mode | | | | |
| | Metro | 42468.70 | 57489.38 | 0 | 304936 |
| | Bus | 36944.77 | 41044.65 | 0 | 221574 |
| | Light bus | 16613.97 | 21845.74 | 0 | 155141 |
| | Taxi | 9580.98 | 11337.50 | 0 | 68198 |
| | Private car | 11844.94 | 11025.63 | 0 | 66282 |
| | Others (include walk-only trip) | 37571.43 | 44155.42 | 0 | 254980 |
| | Annual average daily traffic (AADT) | 31690.71 | 27909.59 | 790 | 151840 |
| Population characteristics | % of population of age below 15 | 11.93 | 2.74 | 4.39 | 20.76 |
| | % of population of age 15-64 | 74.46 | 4.59 | 49.03 | 91.90 |
| | % of population of age over 64 | 13.61 | 5.46 | 3.13 | 43.81 |
| | % of Chinese population | 88.29 | 12.21 | 4.31 | 99.17 |
| | % of non-Chinese population | 11.71 | 12.21 | 0.83 | 56.81 |
| | % of population attained primary education or below | 27.47 | 7.67 | 7.24 | 61.09 |
| | % of population attained secondary education | 44.50 | 6.81 | 17.35 | 65.56 |
| | % of population attained tertiary education | 28.03 | 12.38 | 2.63 | 75.41 |
| Household characteristics | % of households of monthly income below HK$10,000 | 21.52 | 9.88 | 0.71 | 52.48 |
| | % of households of monthly income of HK$10,000-$39,999 | 46.16 | 14.81 | 3.36 | 71.27 |
| | % of households of monthly income over HK$40,000 | 32.32 | 22.19 | 0.00 | 95.94 |
| | % of households of more than 3 members | 34.20 | 9.96 | 1.29 | 65.42 |
| Road network | Road length per unit area (km per km$^2$) | 17.60 | 15.77 | 0.17 | 90.43 |

| | | | | | |
|---|---|---|---|---|---|
| characteristics | Number of metro exits | 2.68 | 3.76 | 0 | 18 |
| | Number of bus stops | 44.39 | 35.79 | 1 | 187 |
| | Number of non-signalized intersections | 110.69 | 97.18 | 7 | 636 |
| | Number of signalized intersections | 8.15 | 8.20 | 0 | 41 |
| Point of interest | Number of restaurants | 161.65 | 195.17 | 2 | 1137 |
| | Number of schools | 19.36 | 17.53 | 0 | 95 |
| | Number of shopping mall | 7.55 | 9.51 | 0 | 55 |
| | Number of hotels | 7.78 | 19.63 | 0 | 164 |
| Land use | Commercial ($km^2$) | 0.03 | 0.10 | 0.00 | 1.26 |
| | Government and institutional ($km^2$) | 0.32 | 0.35 | 0.00 | 2.46 |
| | Industrial ($km^2$) | 0.07 | 0.32 | 0.00 | 3.78 |
| | Residential ($km^2$) | 0.44 | 0.43 | 0.00 | 1.92 |
| | Transportation ($km^2$) | 0.21 | 0.21 | 0.01 | 1.57 |
| | Others ($km^2$) | 3.83 | 9.44 | 0.00 | 70.31 |

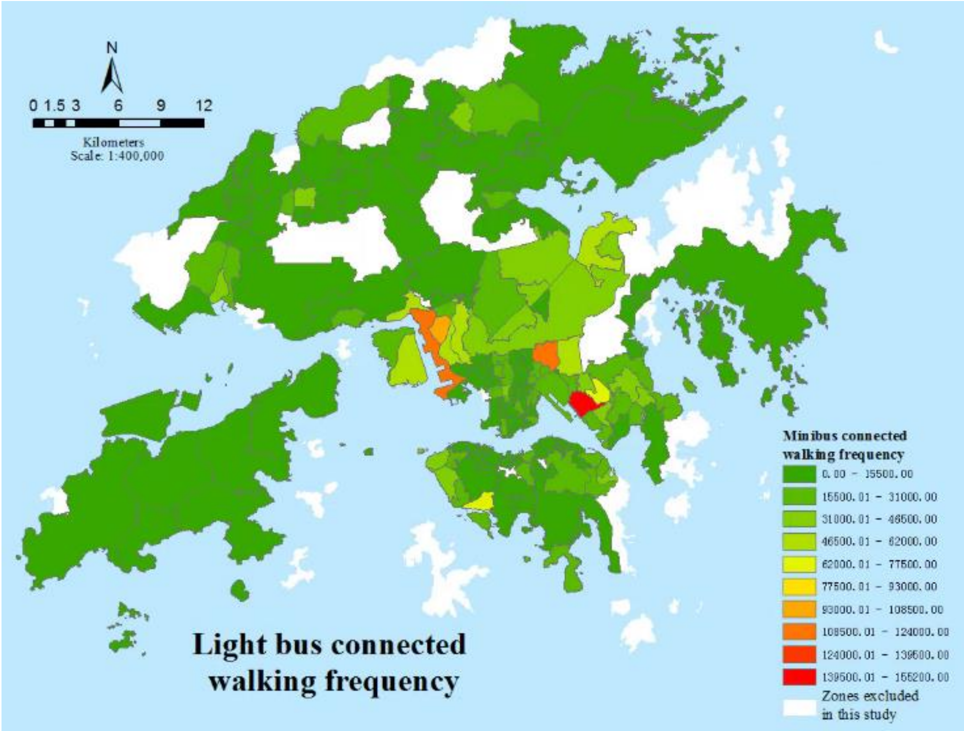
After the adjustment of surveyed motorized trips, 16% of the motorized trips involved more than one motorized mode. Each first/last trip leg of motorized trips (including trips involving more than one motorized mode) was counted as one walking and categorized by specific connected motorized mode when calculating the walking frequencies. Therefore, each motorized trip at least results in two walking trip legs. In particular, each walk-only trip was only counted once. Up to 39 travel modes were recorded throughout the survey. To focus on the connections between pedestrians' walking and major motorized modes, minor modes with less than 5% of the total pedestrian access/egress travel amount were grouped together as *Other*. Finally, six modes were categorized, i.e., metro, bus (franchised bus), light bus (including red and green minibusses), private car, taxi, and *Other* (including walk-only trips with 4%). The transferring trip legs between two motorized modes were counted as transferring trips and were categorized by different combinations of consequent motorized modes. Finally, a total of 2.7 million walking frequencies were derived from the motorized trips. Among the walking frequencies, 8% are transferring trip legs between two motorized modes (based on the aforementioned six categorized modes: 1.7% transferring between metro and light buses, 1.5% transferring between metro and bus, 0.7% transferring between buses, 0.5% transferring between metro lines, and 3.4 % transferring between types other than the four listed types). As seen in **Table 5.1**, the walking connected with public transport (including metro, bus, and light bus) constituted the majority (65%) of pedestrians' daily walking in terms of frequency. The spatial distributions of these exposures over the studied zones are presented in **Figure 5.2**.



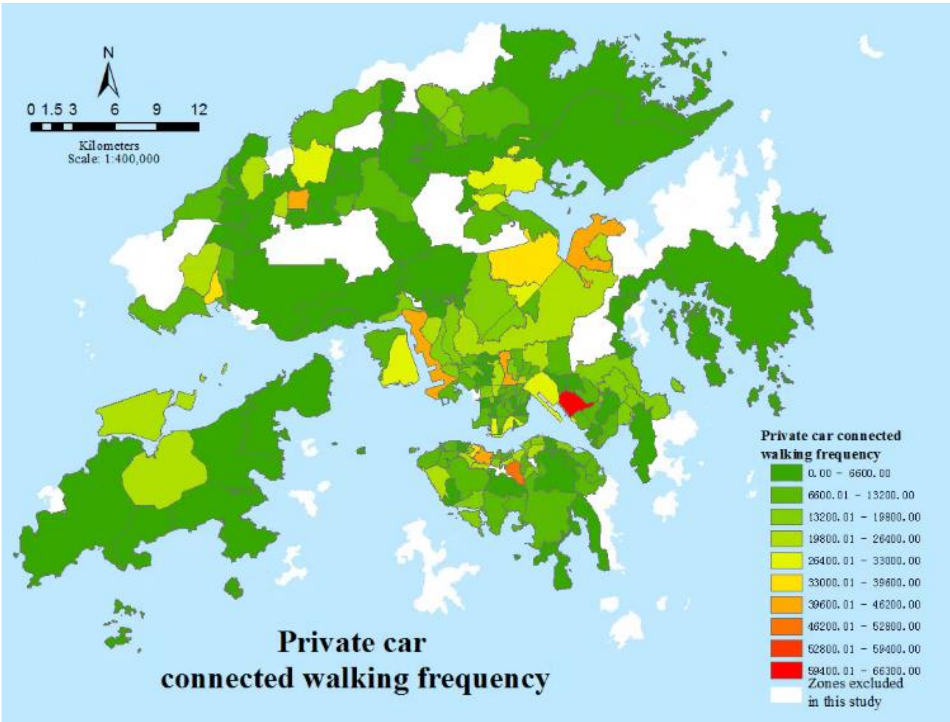

(a) The spatial distribution of pedestrian exposure by metro



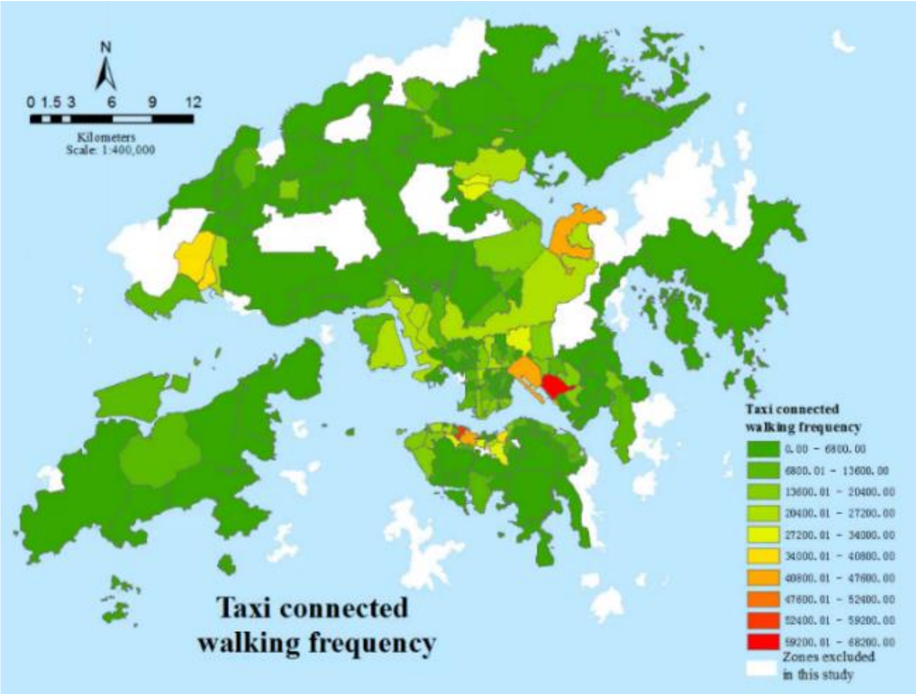

(b) The spatial distribution of pedestrian exposure by franchised bus



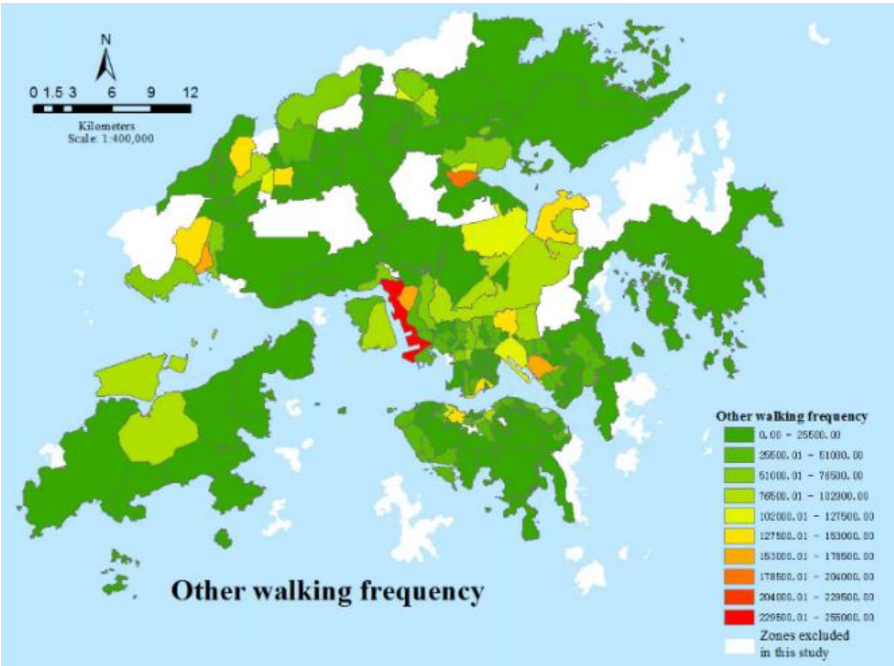

(c) The spatial distribution of pedestrian exposure by light bus




(d) The spatial distribution of pedestrian exposure by private car




(e) The spatial distribution of pedestrian exposure by taxi




(f) The spatial distribution of pedestrian exposure by other mode

Figure 5.2 Spatial distribution of pedestrian exposure by different modes

In **Figure 5.2**, the distribution of pedestrian exposure (walking frequency) connected with the bus is distinguished from the distributions of other exposure measures, with evident higher exposures in the Shatin District, Kwai Tsing District, and Tung Chung areas (thicker black outlines in **Figure 5.2b**). Pedestrian exposures by metro (**Figure 5.2a**), light bus (**Figure 5.2c**), private car (**Figure 5.2d**), and taxi (**Figure 5.2e**) modes, in the referred zones, were not much higher than other zone areas. The difference in the exposure distributions by different transport modes may indicate there would be different influences on the occurrence of pedestrian crashes, which is investigated in Study 2.

## 5.3 Results

As in Chapter 4, VIF had been tested for the potential multicollinearity between the independent variables before running the crash prediction model. Six models were constructed and compared: models with three alternatives for pedestrian exposure (3 models), with and without spatial correlation effects respectively (3*2 models). Model 1 adopted the zonal population as pedestrian exposure surrogate. In Model 2, the total walking frequency was treated as the pedestrian exposure surrogate (instead of the zonal population). The pedestrian exposure surrogate in Model 3 was further categorized by different transport modes. The spatial correlation effect was not considered in Model 1, Model 2, and Model 3. Models 4, 5, and 6 included spatial correlation effects based on the same model specifications of Models 1, 2, and 3, respectively. Significant variables, i.e., those under 90% level of the Bayesian criterion interval (BCI) in any one of the models, were retained for final estimations. The estimated results of six models are presented in **Table 5.2** (non-spatial models) and **Table 5.3** (spatial models). The BCI intervals including zero are marginal significant within 90% level while the others are significant at 95% level.

Table 5.2 Results of parameter estimation of models without considering spatial correlation

| Scope of work | Variable | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Coef. | Std. | 95% BCI | Coef. | Std. | 95% BCI | Coef. | Std. | 95% BCI |
| Constant | | -2.68* | 0.72 | (-4.11, -1.32) | -5.22* | 0.82 | (-6.83, -3.66) | -4.07* | 0.62 | (-5.28, -2.86) |
| Exposure | ln (Population) | 0.32* | 0.05 | (0.22, 0.43) | N/A | | | N/A | | |
| | ln (AADT) | 0.12* | 0.05 | (0.03, 0.21) | 0.13* | 0.04 | (0.05, 0.21) | 0.11* | 0.03 | (0.03, 0.18) |
| | ln (overall walking trip) | N/A | | | 0.50* | 0.06 | (0.38, 0.62) | N/A | | |
| | ln (walking trip to and from bus) | N/A | | | N/A | | | 0.36* | 0.05 | (0.26, 0.45) |
| | ln (walking trip to and from light bus) | N/A | | | N/A | | | 0.06^ | 0.03 | (-0.004, 0.129) |
| | ln (walking trip to and from taxi) | N/A | | | N/A | | | 0.05* | 0.02 | (0.01, 0.10) |
| Population characteristics | % of non-Chinese population | IS | | | -0.006^ | 0.003 | (-0.0133, 0.0006) | IS | | |
| | % of the population of age below 15 | IS | | | IS | | | 0.04* | 0.02 | (0.01, 0.07) |
| Road network characteristics | Road length per unit area | 0.01* | 0.004 | (0.008, 0.022) | 0.01* | 0.003 | (0.007, 0.019) | 0.01* | 0.003 | (0.005, 0.017) |
| | Number of metro exits | IS | | | -0.04* | 0.01 | (-0.07, -0.02) | IS | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Number of bus stops | 0.005* | 0.002 | (0.001, 0.010) | 0.005* | 0.002 | (0.001, 0.008) | IS | | |
| | Number of non-signalized intersections | IS | | | IS | | | -0.001^ | 0.0006 | (-0.00224, 0.00002) |
| Point of interest | Number of restaurants | 0.002* | 0.0003 | (0.0014, 0.0027) | 0.002* | 0.0003 | (0.001, 0.0025) | 0.002* | 0.0003 | (0.0016, 0.0026) |
| | Number of hotels | 0.008* | 0.002 | (0.003, 0.013) | 0.005* | 0.002 | (0.001, 0.009) | 0.004* | 0.002 | (0.0004, 0.0077) |
| Land use | Residential | IS | | | IS | | | 0.29* | 0.12 | (0.06, 0.53) |
| Over-dispersion parameter | | 0.17* | 0.03 | (0.12, 0.24) | 0.12* | 0.02 | (0.08, 0.17) | 0.09* | 0.02 | (0.06, 0.14) |
| DIC | | 1180.57 | | | 1149.47 | | | 1116.35 | | |

**: Statistically significant at the 5% level; ^: marginally significant at the 10% level; IS: statistically insignificant; N/A: Not applicable*

Table 5.3 Results of parameter estimation of models considering spatial correlation

| Scope of work | Variable | Model 4 | | | Model 5 | | | Model 6 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Coef. | Std. | 95% BCI | Coef. | Std. | 95% BCI | Coef. | Std. | 95% BCI |
| Constant | | -3.01* | 0.68 | (-4.35, -1.69) | -5.29* | 0.82 | (-6.95, -3.68) | -4.46* | 0.62 | (-5.63, -3.23) |
| Exposure | ln (population) | 0.32* | 0.05 | (0.21, 0.42) | N/A | | | N/A | | |
| | ln (AADT) | 0.14* | 0.05 | (0.05, 0.23) | 0.15* | 0.04 | (0.07, 0.24) | 0.13* | 0.04 | (0.06, 0.21) |
| | ln (overall walking trip) | N/A | | | 0.47* | 0.06 | (0.35, 0.60) | N/A | | |
| | ln (walking trip to and from bus) | N/A | | | N/A | | | 0.33* | 0.05 | (0.23, 0.42) |
| | ln (walking trip to and from light bus) | N/A | | | N/A | | | 0.10* | 0.04 | (0.01, 0.18) |
| | ln (walking trip to and from taxi) | N/A | | | N/A | | | 0.05* | 0.02 | (0.01, 0.10) |
| Population characteristics | % of non-Chinese population | IS | | | -0.01* | 0.004 | (-0.022, -0.005) | -0.01^ | 0.004 | (-0.0161, 0.0004) |
| | % of the population of aged below 15 | IS | | | 0.04* | 0.02 | (0.01,0.07) | 0.05* | 0.02 | (0.02, 0.09) |
| Road network characteristics | Road length per unit area | 0.01* | 0.04 | (0.006, 0.020) | 0.01* | 0.003 | (0.006, 0.018) | 0.01* | 0.003 | (0.005, 0.018) |
| | Number of metro exits | IS | | | -0.04* | 0.01 | (-0.07, -0.02) | -0.02^ | 0.01 | (-0.038, 0.003) |
| | Number of bus stops | 0.004^ | 0.002 | (-0.0001, 0.0085) | 0.004* | 0.002 | (0.001, 0.008) | IS | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Number of non-signalized intersections | IS | | | IS | | | IS | | |
| Point of interest | Number of restaurants | 0.002* | 0.0003 | (0.001, 0.003) | 0.002* | 0.0003 | (0.001, 0.003) | 0.002* | 0.0002 | (0.0015, 0.0025) |
| | Number of hotels | 0.01* | 0.002 | (0.006, 0.015)) | 0.01* | 0.002 | (0.002, 0.011) | 0.006* | 0.002 | (0.002, 0.010) |
| Land use | Residential | IS | | | IS | | | 0.35* | 0.13 | (0.10, 0.60) |
| S.D. of spatial effects | | 0.22* | 0.06 | (0.12, 0.35) | 0.18* | 0.06 | (0.08, 0.30) | 0.17* | 0.04 | (0.09, 0.26) |
| Over-dispersion parameter | | 0.15* | 0.03 | (0.10, 0.21) | 0.11* | 0.03 | (0.06, 0.16) | 0.08* | 0.02 | (0.05, 0.12) |
| DIC | | 1168.81 | | | 1143.56 | | | 1108.96 | | |

**: Statistically significant at the 5% level; ^: marginally significant at the 10% level; IS: statistically insignificant; N/A: Not applicable*

As shown in **Table 5.2**, the DIC values of the three non-spatial models, namely Model 1, Model 2, and Model 3, are 1180.57, 1149.47, and 1116.35, respectively. The statistics indicate that the exploration of the pedestrian walking frequency (Model 2) as the pedestrian exposure significantly improves the model performance of the pedestrian crash model in comparison to the conventional zonal population (Model 1), as the DIC difference was greater than 10 (31.10 = 1180.57-1149.47). The best performance among these three models assessment was achieved in Model 3 in terms of DIC value, which considered the connections between pedestrian walking and motorized transport modes (33.12 = 1149.47-1116.35)).

The model performances of Model 4, 5, and 6 (in **Table 5.3**) significantly improved with the consideration of spatial correlation effects by CAR structure on top of Model 1, 2, and 3 respectively. As shown in **Table 5.3,** the standard deviations of the structured spatial effects of the three models (0.22, 0.18, and 0.17 respectively) were significant within a 95% credible interval. The statistics confirmed the existence of correlated effects across areas. Model 6 has the best prediction performance among all of the models: (a) pedestrian crash exposure estimated by connecting transport modes is superior, and (b) the model incorporating spatial correlation effects is superior. Therefore, the results of Model 6 are described in the following paragraph.

In Model 6, as expected, pedestrian crash counts increase with vehicular traffic (0.13). Pedestrian exposures connecting three specific motorized modes were examined to have significant effects on pedestrian crash occurrences: bus (0.33), light bus (0.10), and taxi (0.05). A higher proportion of children (aged below 15) (0.05) was also positively associated with crash frequency, whereas the increase in the proportion of non-Chinese people had minor impacts on reducing the occurrence of crashes (-0.01, significant at 90% level). As for road network characteristics, the road length density showed positive effects (0.01). One type of public transport facility, i.e., the number of entrances/exits of the metro stations, was examined to have a negative influence on pedestrian crash occurrence (-0.02) at a 90% BCI significance level. In addition, the number of hotels (0.006), restaurants (0.0002), and residential areas (0.35) in a zone showed increasing effects on the pedestrian crash frequency.

## 5.4 Discussion

### 5.4.1 Effects of pedestrian exposure

Consistent with previous studies, the increase in vehicular traffic (i.e., AADT) would result in more crashes, as there are more vehicles running on the road, which can increase the possibility of crash occurrences (Miranda-Moreno et al., 2011; Barua et al., 2016; Bao et al., 2017; Xie et al., 2018).

Among the six groups of pedestrian exposures by connecting transport modes, three of them are found to have significant influences on the occurrence of pedestrian crashes, i.e., bus, light bus, and taxi. In Hong Kong, buses and light buses are two major roadway public transport modes. Up to 41% of the passengers using motorized modes choose buses and light buses, whereas another 30% of the passengers travel by metro (Transport Department, 2012b). Increased public transport usage undoubtedly generates more pedestrian walking, and thus increased the access/egress walking to/from public transport stations (Mohan, 2001; Lakhotia et al., 2019). Therefore, pedestrians hit by running vehicles would increase as the pedestrian exposure increases (Pulugurtha and Penkey, 2010). Pedestrians have to wait, get on and alight from buses at close distances from running vehicles as bus stops are located at roadsides (Cafiso et al., 2013). What's more, pedestrians are more likely to cross roads to catch buses or get to their destinations after alighting from the buses (Mohan, 2001). The crossing behaviors add to the risk of pedestrian-vehicle collisions. Therefore, to protect pedestrians from conflicts with vehicles while advocating public transport usage and active travel, more attention should be paid to improve pedestrian facilities and strengthen speed controls around places with busy bus routes in response to the increased pedestrian exposure (Mohan, 2001).

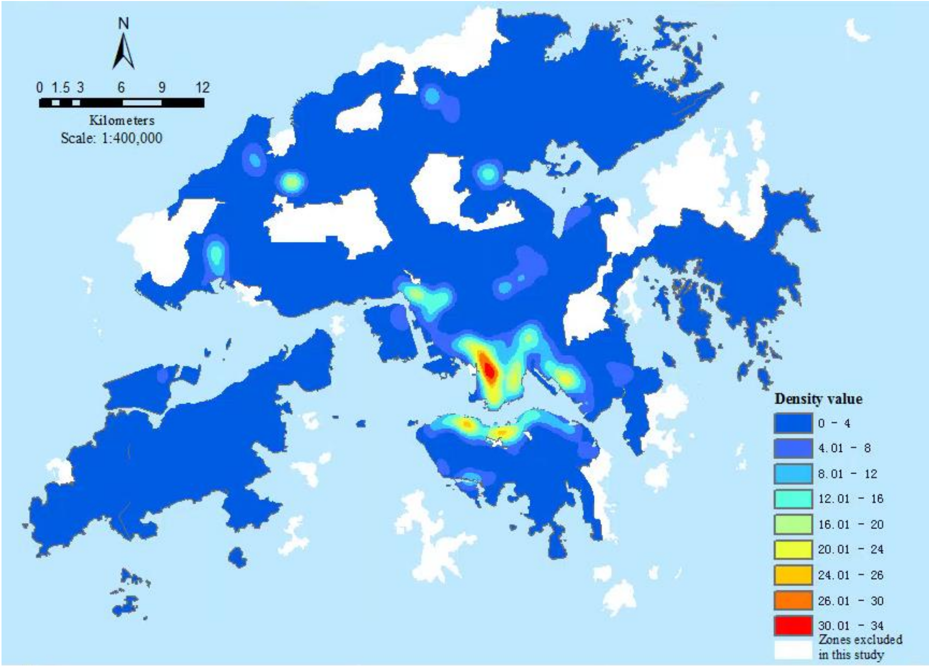
The light bus is an important supplementary mode to the bus for servicing citizens in Hong Kong, providing 15% of the public transport patronage (Transport and Housing Bureau, 2017). Similar to the case of the bus, the increase in the usage of light buses also generates more pedestrians walking on roads, and thus causes higher pedestrian exposure to crashes. In particular, light buses are allocated to service areas with lower passenger demand, where there may be longer distances between consecutive stops (Transport Department, 2001; Planning Department, 2020). As a result, passengers need to walk longer and spend more time on the roads, increasing the chances to be involved with crashes. Furthermore, areas with lower travel demands are more likely to be underdeveloped areas or mountainous areas in Hong Kong, where the road environment will be complex and transport facilities might not be well equipped. As a consequence, the safety level might be lower for both drivers and pedestrians in these areas (Yao et al., 2015). Therefore, drivers/pedestrians should raise their safety awareness for more careful driving/walking at places with poor transport facilities.

Notably, the access/egress walking connected with another major public transport mode, i.e., the metro, was found insensitive to the occurrence of pedestrian crashes. Compared with bus travel, metro traveling offers better travel conditions for passengers. Boarding and lighting of metros usually occur within protected buildings, whereas such activities are performed on roadsides for buses, which accounted for a large portion of fatalities (Kharola et al., 2010). In addition, entrances/exits of metro stations are available in each direction, which provides better accessibility to railway transit (Sung et al., 2014; Li et al., 2017), and reduces the road crossings of pedestrians. What’s more, pedestrians perceived a better traveling experience in walking for metro services than bus service (Prasertsubpakij and Nitivattananon, 2012; Goel and Tiwari, 2016). Therefore, it is not surprising that there is no significant association between the occurrences of pedestrian crashes and the pedestrian exposure connected with metro travel.

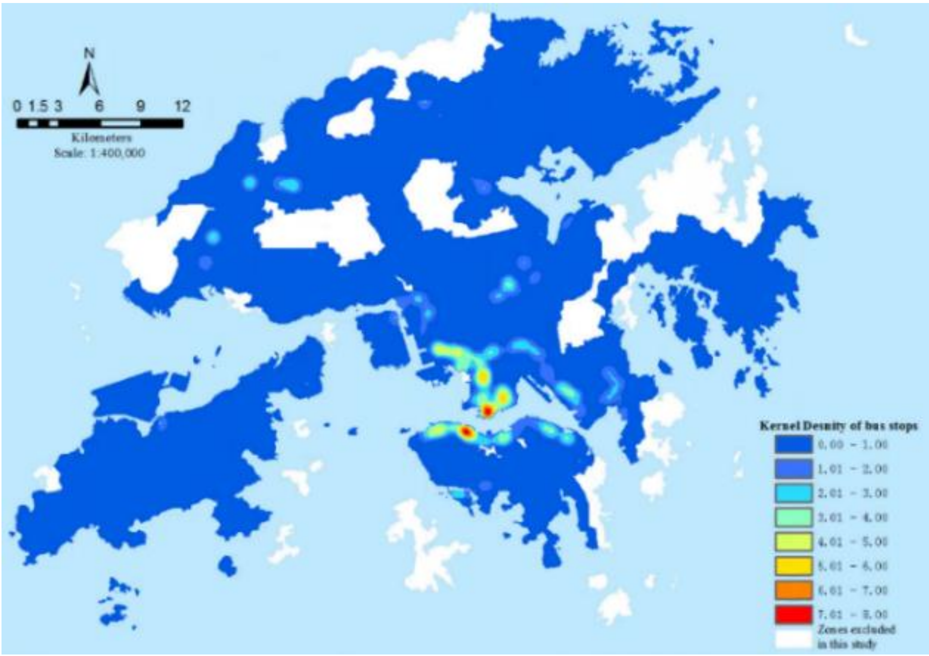
Pedestrians’ walking to take a taxi has significant effects on pedestrian crash occurrence. Although taxi service provides nearly point-to-point service, pedestrians still need to walk to approach taxis at their origin ends. Unlike the case of buses that the routes and frequencies are fixed, resulting in constant vehicle volume regardless of the passenger counts, the increased passenger travel by taxis will increase vehicular traffic, which might also be a reason for increasing crashes (Lam, 2004; Tay and Choi, 2016). Lower speed limits for auto vehicles in high-density areas or the areas with more local street roads might help reduce pedestrian-vehicle conflicts. Finally, the effect of the pedestrian exposure connecting to private cars on pedestrian crash occurrence is found insignificant. It is reasonable that there would be little walking on roads when people travel by their own cars, which offers door-to-door services.

In Models 1, 2, 4, and 5, which did not incorporate the pedestrian exposures categorized by transport modes, the positive effects of bus stops on the occurrence of pedestrian crashes were identified. The results were consistent with the data of Hong Kong presented in **Figure 5.3**, i.e., the density distributions of pedestrian crashes (**Figure 5.3a**), metro entrances/exits (**Figure 5.3b**), and bus stops (**Figure 5.3c**). Referring to the density distributions among pedestrian crashes, metro entrances/exits, and bus stops, it can be easily identified that the density distribution of the bus stops was closer to the pattern of pedestrian crashes than that of the metro exits. This is reasonable because bus stops are the locations where pedestrians walk to and from for bus service, which is highly correlated with pedestrian exposure. The similar patterns of the density distributions between pedestrian crashes and bus stops indicate that pedestrian exposure by bus has a strong relationship with the crash occurrences.

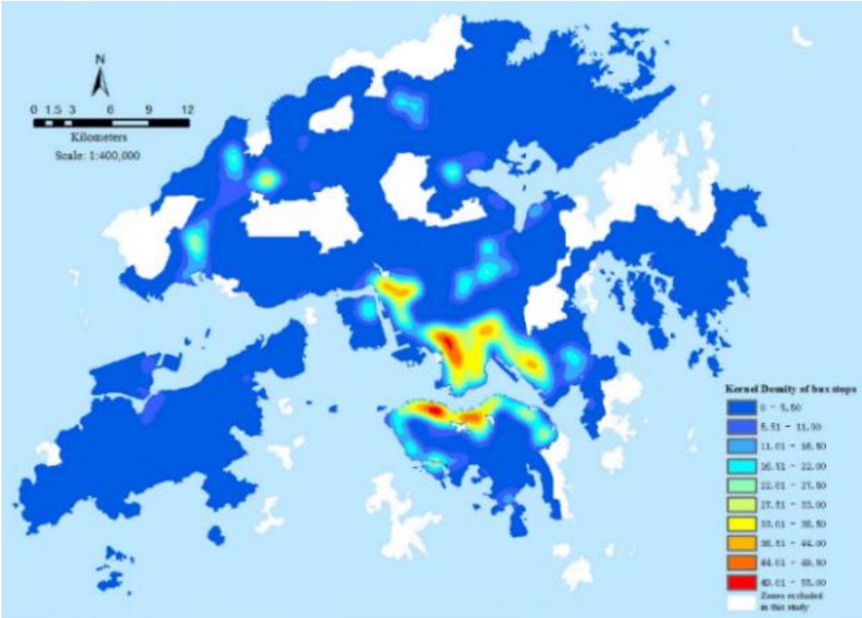
However, when the pedestrian exposure has been further divided by transport modes, the significant effect of bus stops is no longer examined in Models 3 and 6. The existence of bus stops is one of the key attributes in the generation and attraction of travel within such areas (Wedagama et al., 2006). When pedestrian exposure connecting to buses is incorporated as covariates in crash prediction models, the effects of bus stops on crash occurrences might be overtaken by these exposure measures (Yao et al., 2015). The results highlight the need to further investigate and identify the effects of exposure and the associated factors on the occurrences of crashes.




(a) Density distribution of pedestrian crashes




(b) Density distribution of metro entrances/exits




(c) Density distribution of bus stops

Figure 5.3 Density distributions of pedestrian crashes, metro entrances/exits and bus stops

### 5.4.2 Effects of zonal factors

- **Population characteristics**

Among the sociodemographic factors, a few of them show significant effects in Study 2. A higher proportion of children in an area was found to be related to more pedestrian crashes (Hummel, 1998; Elias et al., 2010; Ponnaluri and Nagar, 2010; Siddiqui et al., 2012a). Children might be prone to perform risky behaviors on roads, owing to the lack of road safety awareness and education on traffic rules (Keall, 1995; Tay, 2003; Elias and Shiftan, 2014). In addition, households with children tend to have more out-of-home activities with their children, such as school escorts or necessary recreational outdoor activities (Johnson et al., 2004; Amoh-Gyimah et al., 2016). These travel demands add to either vehicle or pedestrian flows and increase the chance for crash occurrences. Parents and schools can strengthen road safety education to raise the safety awareness of children (Lee et al., 2015b), and parents should pay more attention to the risk behaviors of their children on roads when going outside. Finally, the minor negative effects of the non-Chinese population might be the result of that people in minority groups live together within the same community and thereby reduce long-distance travel (Cottrill and Thakuriah, 2010; Bhat et al., 2017).

- **Transport facilities**

In regards to the road network characteristics, the high road density was found to have positive effects on pedestrian frequency. The results are consistent with the findings in previous studies (Wang et al., 2016a; Cai et al., 2017; Guo et al., 2017). Higher road density reflects higher demands of vehicle running and pedestrian walking. According to the ATC report, local street roads constitute a large proportion of the road network in Hong Kong, and up to 76% of the roads are lack of sufficient segregated pedestrian facilities (Transport Department, 2012b). Therefore, pedestrian-friendly facilities should be built or improved in road-dense areas, to lower the risk of pedestrian-vehicle conflicts in these areas. One factor of public transport facilities, i.e., the number of metro station entrances/exits, was shown to help reduce pedestrian crashes. As discussed in Section 5.4.1, although metro travel may induce pedestrian walking to some extent, the superior facilities for walking to and from the metro stations, including footbridges and underpasses for pedestrians, help reduce pedestrian-vehicle conflicts by separating the pedestrians from vehicles (Guo et al., 2017).

- **Built environments and land use information**

Regarding built environment factors, more restaurants and hotels within a zone tend to increase pedestrian crashes (Yao et al., 2015; Zhang et al., 2015). Hong Kong is a developed high-density society; it is common to see restaurants widely spreading over neighborhood streets. Walking is a reasonable (and primary) means for traveling to and from restaurants within local neighborhoods. Therefore, areas with more restaurants will stimulate more pedestrian walking. Crowded pedestrians on local streets can be easily exposed to conflicts with vehicles, owing to the narrow conditions of local streets. Hotels are key attractions to tourists. The arrival of tourists and their frequent travels induce additional vehicle and pedestrian volume (Lee et al., 2015a). In addition, frequent pick-up and drop-off activities around hotels increase the chances of potential conflicts between vehicles and pedestrians (Verzosa and Miles, 2016). The residential land-use area was also found to be positively correlated with the occurrence of pedestrian crashes (Dong et al., 2015; Amoh-Gyimah et al., 2016). A larger residential area means more domestic households and therefore more active walking activities in surrounding neighborhoods, such as shopping at supermarkets and eating in restaurants (Wier et al., 2009; Bhat et al., 2017; Guo et al., 2017).

### 5.4.3 Effects of spatial correlation

The variances of the spatial correlation effects were found to be significant in all three CAR models. The results indicated the existence of spatial correlations in the occurrence of pedestrian crashes (Barua et al., 2014, 2016). In such cases, the changes in risk factors within a zone will not only impact the occurrence of crashes in the host zones but also the crash occurrences in neighboring zones (Bhat et al., 2017). Notably, the highest value of the variances of the spatial correlation was in Model 4 (treating the conventional population as pedestrian exposure), whereas the value was the smallest in Model 6 (pedestrian exposure measures are categorized by different transport modes). It implies that the exposure measures estimated from the travel characteristics surveys, especially categorizing the pedestrians' walking by different transport modes, showed the capability to capture spatial correlations, which is properly due to the inter-zone travel of pedestrians; this is difficult to achieve with population surrogates.

## 5.5 Conclusion

Study 2 investigated the safety effects of FLM trip legs, i.e., pedestrian walking connecting to different transport modes, on pedestrian crash occurrences. The walking trip legs connecting to different transport modes were explored as pedestrian exposure measures in crash prediction models. Several crash prediction models were developed and adopted different alternatives of pedestrian exposure measures. In the meantime, the effects of various zonal factors were also investigated. The three hypotheses are verified: 1) pedestrian exposure connecting to different transport mode pose different impacts on pedestrian crash occurrence; 2) walking connecting to roadway transport modes is more likely to have a higher risk of crash occurrence than railway modes; 3) the significance of the variance of spatial correlation indicates the existence of spatial correlation of crash occurrence across the zone units.

The contributions to the literature are two-fold. First, to the best of our knowledge, this is the first study to consider the FLM issue in pedestrian crash prediction models. The FLM trip legs to and from different motorized modes, which have been overlooked in previous pedestrian crash analysis, have been explored as the measures of pedestrian exposure in pedestrian crash prediction models. Second, the results showed that the pedestrian exposures by different transport modes posed different impacts on pedestrian crash occurrences. Pedestrian walking connecting to buses, light buses, and taxis was found to have a positive relationship with pedestrian crash frequency. However, pedestrian walking connecting to the metro was found insensitive to pedestrian crash occurrence in this study. In addition, the number of metro station entrances/exits showed negative impacts on pedestrian crashes. It implies that although the increased usage of public transport improves overall road safety, safety issues remain outstanding for the passengers of public transport as a result of the FLM problem, which generates more pedestrian walking to and from transit stations. Therefore, effective countermeasures should be developed to improve the planning, design, and management of different transport facilities. For example, the provision of segregated/elevated walkways and footpaths, segregated pedestrian crossings, rationalization of public transport routes, and optimized relocations of public transport stations will help offer better and safer service for the FLM access to public transport stations, and thus improve pedestrian safety using public transport. This study also showed that consideration of the spatial correlation effects improves the model performance of pedestrian crash prediction models. The lowest variance of the spatial random effects in the model with pedestrian exposure categorized by transport modes indicated the capability of the proposed exposure measures to capture the spatial correlation to some extent, which is properly due to the cross-zone-boundary movements. Beyond the exposure variables, positive effects on crash occurrence were found for the following factors: age (children), road length density, number of restaurants and hotels, and residential land-use area while negative effects were examined for the proportion of the non-Chinese population.

# Chapter 6 Joint probability model for pedestrian crash frequency

## 6.1 Introduction

Study 1 and 2 in Chapter 4 and Chapter 5 respectively have studied the pedestrian crash occurrence with the consideration of the impacts of trip purposes and the involvement of motorized modes that pedestrian walking is connected to. While only pedestrian crashes were analyzed in the above two studies, a better understanding of pedestrian safety can be achieved to incorporate broader types of crash occurrences into the study as crash occurrences of different types can be correlated. Study 3 in this chapter moves on to the investigation of pedestrian crash occurrence under the consideration of total crash incidence.

In conventional studies, separated prediction models were established for pedestrian crashes and other crash types. However, possible correlations among the counts of different crash types could be ignored. Additionally, possible factors can contribute to pedestrian crashes in two different ways, i.e., crash occurrence and the propensity of pedestrian involvement. Furthermore, extensive pedestrian count data were usually not available to estimate the pedestrian crash exposure. In Study 3, a joint probability model is adopted to model crash occurrence and pedestrian involvement in crashes simultaneously in one structure. The effects of the factors including land use, road network, traffic flow, population demographic and socioeconomics, public transport facilities, and trip attraction attributes are considered. Additionally, trip generation and pedestrian activity data, based on a comprehensive household travel characteristics survey, are used to reflect the pedestrian crash exposures. The full Bayesian approach was applied to estimate the parameters.

The hypotheses of Study 3 are: 1) correlation effects do exist between different types of crash occurrences, i.e., total crash frequency and pedestrian crash involvement; 2) Explanatory factors, including environment, traffic, and population characteristics could present different effects across total crash frequency and pedestrian crash frequency; 3) The joint probability modeling outperforms the conventional models for simultaneously modeling total crash frequency and pedestrian crash frequency.

The rest of the paper is organized as follows. Section 6.2 describes the data used in the current study. Section 6.3 shows the results of the model estimation. Implications of estimation results are discussed in Section 6.4. Section 6.5 summarizes the findings.

## 6.2 Data

Similar to the zone system adopted in Chapter 5, 179 TPUs were adopted in Study 3. The data ready for the joint probability model, included crash incidence, traffic flow, trips, land use, and population profile are mapped into corresponding geographical units (zones) with ArcGIS. The summarized statistics are presented in **Table 6.1** (size=179).

Table 6.1 Descriptive statistics of the sample

| **Variable** | | **Mean** | **Std.** | **Min.** | **Max.** |
|---|---|---|---|---|---|
| Zone area ($km^2$) | | 4.90 | 9.77 | 0.06 | 71.23 |
| **Crash and exposure** | | | | | |
| Total crashes | | 83.15 | 71.90 | 2 | 483 |
| Pedestrian crashes | | 20.12 | 22.54 | 0 | 134 |
| Population | | 38,066.76 | 42,814.64 | 1,023 | 287,901 |
| Annual average daily traffic (AADT) | | 31,690.71 | 27,909.59 | 790 | 151,840 |
| Trip generation (per day) | | 74,762.22 | 74,108.04 | 289 | 465,645 |
| Walking frequency (per day) | | 154,881.50 | 159,058.40 | 973 | 925,433 |
| **Demographic and socioeconomic characteristics** | | | | | |
| Age | Proportion of age below 15 | 0.12 | 0.03 | 0.04 | 0.21 |
| | Proportion of age above 64 | 0.14 | 0.05 | 0.03 | 0.44 |
| | Proportion of other age group | 0.74 | 0.05 | 0.49 | 0.92 |
| Ethnicity | Proportion of Chinese | 0.88 | 0.12 | 0.43 | 0.99 |
| | Proportion of non-Chinese | 0.12 | 0.12 | 0.01 | 0.57 |
| Education | Proportion attained primary education | 0.17 | 0.05 | 0.01 | 0.28 |
| | Proportion attained secondary education | 0.45 | 0.07 | 0.17 | 0.66 |
| | Proportion attained tertiary education | 0.28 | 0.12 | 0.03 | 0.75 |
| Household income | Proportion below HK$10,000 per month | 0.22 | 0.10 | 0.01 | 0.52 |
| | Proportion of HK$10,000-39,999 per month | 0.46 | 0.15 | 0.03 | 0.71 |
| | Proportion of ≥ HK$40,000 per month | 0.32 | 0.22 | 0.00 | 0.96 |
| Household size | Proportion of ≤ 3 members | 0.66 | 0.10 | 0.35 | 0.99 |
| | Proportion of more than 3 members | 0.34 | 0.10 | 0.01 | 0.65 |

| | | | | |
|---|---|---|---|---|
| **Road network and transport facilities** | | | | |
| Road density (km/km$^2$) | 17.60 | 15.77 | 0.17 | 90.43 |
| Number of non-signalized intersections | 110.69 | 97.18 | 7 | 636 |
| Number of signalized intersections | 8.15 | 8.20 | 0 | 41 |
| Number of metro exits | 2.68 | 3.76 | 0 | 18 |
| Number of bus stops | 44.39 | 35.79 | 1 | 187 |
| **Land use and points of interest** | | | | |
| Number of restaurants | 161.65 | 195.17 | 2 | 1,137 |
| Number of schools | 19.36 | 17.53 | 0 | 95 |
| Number of shopping malls | 7.55 | 9.51 | 0 | 55 |
| Number of hotels | 7.78 | 19.63 | 0 | 164 |
| Residential land use (Yes = 1) | 0.34 | 0.47 | 0 | 1 |
| Industrial land use (Yes = 1) | 0.03 | 0.17 | 0 | 1 |
| Commercial land use (Yes = 1) | 0.02 | 0.15 | 0 | 1 |
| Government, institutional & community (Yes = 1) | 0.04 | 0.21 | 0 | 1 |
| Other land use (Yes = 1) | 0.57 | 0.50 | 0 | 1 |

**Table 6.1** summarizes the sample data. Of the 14,884 crashes in the 179 STPUGs included in Study 3, 24.2% (i.e., 3,604) involved pedestrians. According to the TCS report, 13.8 million trips on average were made per day, of which 1.2 million were walk-only trips.

In addition to the road density (length per unit area) and the number of intersections, the effects of the numbers of metro exits and bus stops on pedestrian crashes are also investigated. As shown in **Table 6.1**, each zone had an average of 3 metro exits and 44 bus stops. In Hong Kong, 88% of trips were made by public transport, of which 30% were made by rail and 27% were made by franchised buses. The number of conflicts between pedestrians and vehicles is considerable, especially in commercial and mixed-use areas. With regard to land use and built environments, the numbers of restaurants, schools, shopping malls, and hotels, and the dominating land use (i.e., residential, industrial, commercial, and government & community) of a zone are considered. Tourism has been an important economic driver in Hong Kong. POI information, which includes restaurants, shops, and hotels, can therefore serve as a proxy for the travel demand generated by visitors, which is not captured in conventional household travel surveys. As shown in **Table 6.1**, each zone had an average of 162 restaurants, 8 shopping malls, and 8 hotels.

Prior to the model estimation, the multicollinearity of the candidate variables is assessed with VIF and the factors with VIF values less than 10 were kept in the model estimation.

## 6.3 Results

For comparison with and evaluation of the proposed joint probability model, estimations are also performed with traditional negative binomial models for total and pedestrian crashes. The results for the negative binomial models are outlined in **Table 6.2**.

As summarized in **Table 6.2**, several factors including trip generation, traffic volume, ethnicity, and numbers of non-signalized intersections, restaurants, and hotels are significantly correlated with the total number of crashes. Furthermore, total population, traffic volume, walking frequency, education level, road density, and the number of metro exits, and restaurants are correlated with pedestrian crashes.

The proposed joint probability model (pseudo $\rho^2$ of 0.29), as shown in **Table 6.3**, outperforms the negative binomial model for pedestrian crashes (0.23). Additionally, the joint probability model, as compared to the negative binomial model, involves more significant variables contributing to pedestrian involvement in crashes.

**Table 6.3** presents the results of the joint probability model for crash occurrence and pedestrian involvement in crashes using the MCMC approach. For this model, the mean, standard deviation, and 95% BCIs are presented. Pedestrian crash exposure, population demographics and socioeconomics, road network and transport facilities, and land use characteristics that contribute to the total crash occurrence and pedestrian involvement in crashes are identified. For instance, as summarized in **Table 6.3**, crash occurrence is positively correlated with traffic volume (mean 0.16), trip generation (0.16), number of non-signalized intersection (0.004), number of restaurants (0.001), and number of hotels (0.01), at 5% level of significance. By contrast, the proportion of the non-Chinese population (−0.02) is negatively correlated with the crash occurrence at 5% level of significance.

Furthermore, pedestrian involvement in crashes is positively correlated with the population (0.24), walking frequency (0.29), proportions of the population of ages below 15 (0.03) and of ages above 64 (0.01), non-Chinese population (0.01), road density (0.01), number of bus stops ( 0.002), number of restaurants ( 0.0009), and residential land use (0.19), at 5% level of significance. In contrast, pedestrian involvement is negatively correlated with the proportion of people with primary education (−0.02), number of households of sizes greater than 3 (−0.02), number of non-signalized intersections (−0.003), number of metro exits (−0.03), and number of schools (−0.01), at 5% level of significance.

Table 6.2 Estimation results of negative binomial regression models

| Variable | | Total crash | | | Pedestrian crash | | |
|---|---|---|---|---|---|---|---|
| | | Coef. | Std. | 95% BCI | Coef. | Std. | 95% BCI |
| Constant | | 0.47 | 0.73 | (-0.95, 1.90) | -4.37 | 0.89 | (-6.11, -2.63) |
| ln (Total Population) | | | IS | | 0.18 | 0.06 | (0.06, 0.29) |
| ln (Annual average daily traffic) | | 0.15 | 0.04 | (0.07, 0.23) | 0.10 | 0.04 | (0.03, 0.18) |
| ln (Trip generation) | | 0.20 | 0.06 | (0.08, 0.33) | | N/A | |
| ln (walking frequency) | | | N/A | | 0.44 | 0.06 | (0.32, 0.57) |
| **Demographic and socioeconomic characteristics** | | | | | | | |
| Age | Age below 15 | | IS | | | IS | |
| | Age over 64 | | IS | | | IS | |
| Ethnicity | Non-Chinese | -0.02 | 0.005 | (-0.03, -0.01) | | IS | |
| Education level | Primary | | IS | | -0.05 | 0.03 | (-0.08, -0.02) |
| | Tertiary | | IS | | -0.02 | 0.01 | (-0.04, -0.004) |
| Household size | More than 3 members | | IS | | | IS | |
| **Road network and transport facilities** | | | | | | | |
| Road density | | | IS | | 0.01 | 0.003 | (0.01, 0.02) |
| Number of non-signalized intersections | | 0.003 | 0.0007 | (0.002, 0.005) | | IS | |
| Number of metro exits | | | IS | | -0.03 | 0.01 | (-0.05, -0.01) |
| Number of bus stops | | | IS | | | IS | |
| **Land use and points of interest** | | | | | | | |
| Number of restaurants | | 0.001 | 0.0003 | (0.0003, 0.002) | 0.002 | 0.0003 | (0.001, 0.002) |
| Number of schools | | | IS | | | IS | |
| Number of hotels | | 0.01 | 0.002 | (0.002, 0.01) | | IS | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Residential land use | | IS | | | IS | |
| Over-dispersion | 0.17 | 0.02 | (0.14, 0.22) | 0.09 | 0.02 | (0.06, 0.13) |
| Log likelihood | | -843.43 | | | -552.60 | |
| Pseudo $\rho^2$ | | 0.12 | | | 0.23 | |

*IS: Statistically insignificant; N/A: Not applicable*

Table 6.3 Estimation results of Bayesian joint probability model

| Variable | | Total crash | | | Pedestrian involvement | | |
|---|---|---|---|---|---|---|---|
| | | **Coef.** | **Std.** | **95% BCI** | **Coef.** | **Std.** | **95% BCI** |
| Constant | | 0.51 | 0.56 | (-0.62, 1.79) | -6.78 | 0.73 | (-7.80, -5.14) |
| ln (Population) | | | IS | | 0.24 | 0.03 | (0.18, 0.27) |
| ln (Annual average daily traffic) | | 0.16 | 0.04 | (0.09, 0.23) | | IS | |
| ln (Trip generation) | | 0.16 | 0.03 | (0.10, 0.22) | | N/A | |
| ln (Walking frequency) | | | N/A | | 0.29 | 0.04 | (0.23, 0.35) |
| **Demographic and socioeconomic characteristics** | | | | | | | |
| Age | Age below 15 | | IS | | 0.03 | 0.01 | (0.01, 0.05) |
| | Age above 64 | | IS | | 0.01 | 0.006 | (0.0002, 0.02) |
| Ethnicity | Non-Chinese | -0.02 | 0.004 | (-0.03, -0.01) | 0.01 | 0.003 | (0.01, 0.02) |
| Education | Primary education | | IS | | -0.02 | 0.008 | (-0.04, -0.005) |
| Household size | More than 3 members | | IS | | -0.02 | 0.004 | (-0.02, -0.01) |
| **Road network and transport facilities** | | | | | | | |
| Road density | | | IS | | 0.01 | 0.002 | (0.005, 0.01) |
| Number of non-signalized intersections | | 0.004 | 0.0005 | (0.003, 0.005) | -0.003 | 0.0002 | (-0.004, -0.003) |
| Number of metro exits | | | IS | | -0.03 | 0.007 | (-0.05, -0.02) |
| Number of bus stops | | | IS | | 0.002 | 0.001 | (0.0006, 0.005) |
| **Land use and points of interest** | | | | | | | |
| Number of restaurants | | 0.001 | 0.0004 | (0.0005, 0.002) | 0.0009 | 0.0002 | (0.0006, 0.001) |
| Number of schools | | | IS | | -0.01 | 0.002 | (-0.01, -0.001) |
| Number of hotels | | 0.01 | 0.003 | (0.003, 0.01) | | IS | |
| Residential land use | | | IS | | 0.19 | 0.05 | (0.09, 0.29) |

| | | | | |
|---|---|---|---|---|
| **Over-dispersion parameter** | | | | |
| $\alpha$ | 0.19 | 0.007 | (0.18, 0.21) | N/A |
| **Goodness-of-fit measure** | | | | |
| DIC | | | 3005.55 | |
| Pseudo $\rho^2$ | | | 0.29 | |

*IS: Statistically insignificant; N/A: Not applicable*

Table 6.4 Effect of significant factors on total and pedestrian crashes

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

## 6.4 Discussion

A joint probability model is proposed to simultaneously model crash occurrence and pedestrian involvement in crashes. **Table 6.4** summarizes the key factors contributing to total and pedestrian crashes. Interpretations and policy implications of the results regarding pedestrian exposure, population demographics and socioeconomics, road network characteristics, and land use are provided in the following subsections.

### 6.4.1 Exposure

As outlined in **Table 6.4**, traffic volume and trip generation were determined to be positively correlated with the crash occurrence, which then justified the use of traffic volume and trip generation in the estimation of pedestrian exposure across different zones and their variations in a spatial context (Naderan and Shahi, 2010; Siddiqui et al., 2012b; Haleem et al., 2015; Bao et al., 2017; Wang et al., 2017a). In addition, the parameters of the logarithmically transformed traffic flow and trip generation are both significantly less than one, at 5% level, implying a non-linear relationship between total crash and pedestrian exposure. Meanwhile, the marginal crash rate declined when traffic volume was high, which may be due to reductions in traffic speed and associated crash risk attributed to the increase in traffic density (Qin et al., 2004; Pei et al., 2012).

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

### 6.4.2 Demographic and socioeconomics

Effects of population age, race, education, and household income on total and pedestrian crashes were also investigated. As shown in **Table 6.4**, no evidence could be established for the association between crash occurrence and demographics and socioeconomic characteristics (except for ethnicity), which could be due to vehicle crash risk being less sensitive to the characteristics of residents in an area. Indeed, road safety level as a whole should be determined by the behaviors of drivers and commuters, who may be residing somewhere else, traveling in the area. In contrast, the pedestrian crash risk was correlated with the demographic and socioeconomic characteristics of the residents. For instance, proportions of the young population (ages below 15 years), elderly (ages above 64), and non-Chinese were positively correlated with pedestrian crashes, which could be due to the tendency of young and elderly people to walk more frequently (Miranda-Moreno et al., 2011; Bhat et al., 2017). More importantly, it could be attributed to a higher propensity of reckless crossing behaviors among young people, impaired cognitive and physical performance among the elderly (Hummel, 1998; Noland and Quddus, 2004; Zegeer and Bushell, 2012), and differences in social norms among non-local residents (Cottrill and Thakuriah, 2010; Lee et al., 2014; Coughenour et al., 2017). Furthermore, increases in the proportions of the population that attained, at maximum, primary education and lived in bigger households were correlated with a reduction in pedestrian crash risk. Hence, the effects of education level and household attributes on the perception, attitude, and behaviors of pedestrians, and thus on pedestrian crash risk, would be worth investigating. For instance, propensities for jaywalking and red-light running among pedestrians with respect to possible explanatory factors could have been revealed in an attitudinal survey (Li et al., 2014; Chen et al., 2020a) and empirical observation (Zhu et al., 2021).

### 6.4.3 Road network and transport characteristics

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

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

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

### 6.4.4 Land use and trip attraction attributes

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

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

## 6.5 Conclusion

For conventional crash prediction models, extensive count data are rarely available for the estimation of pedestrian exposure. Moreover, possible correlations exist among crashes of different types (i.e., pedestrians, motor vehicles), which should be considered during the development of separate crash prediction models. Therefore, a joint probability approach is proposed to simultaneously estimate total and pedestrian crashes. A full Bayesian method using an MCMC approach is adopted to investigate the effects of explanatory factors on crash occurrence and pedestrian involvement in crashes using a single model. Population and walking activity are used as proxies for the exposure of pedestrian involvement in crashes. Moreover, the effects of demographics and socioeconomics, road network and facilities, access to public transport services, land use, and trip attraction attributes are considered. The validity of the three hypotheses is examined: 1) correlation between total crash occurrence and pedestrian involvement in crashes does exist; 2) risk factors can impact the total crash occurrence and pedestrian involvement in crashes in different directions; 3) the proposed joint probability model outperforms separate negative binomial models for total and pedestrian crashes.

The results of parameter estimation of crash occurrence and pedestrian involvement are also determined to be consistent with those of previous studies and provide useful recommendations (Lee et al., 2015a; Lee et al., 2015b). For instance, traffic flow, trip generation, road characteristics, and points of interest are correlated with crash occurrence. The pedestrian involvement in crashes is correlated with the demographic and socioeconomic characteristics of residents, access to public transport, the presence of schools, and residential land use. Results are indicative to the necessity for proper design and planning of various transport facilities, including footpaths, intersections, pedestrian crosswalks, and bus stops. Moreover, the needs of vulnerable groups (i.e., children, adolescents, elderly, and ethnic minorities) and accessibility to essential urban services and attractions (i.e., hotel, government, community, and housing development) should be addressed (Sze and Christensen, 2017). With these proposed improvements, a safe and accessible walking environment in built environments can be promoted. Additionally, road safety education can be designed and marketed toward vulnerable pedestrian groups to enhance their safety awareness and deter reckless crossing behaviors (Zhu et al., 2021).

# Chapter 7 Conclusion

## 7.1 Summary

This dissertation aims to advance pedestrian safety at the macroscopic level. The study has contributed to the literature by enriching the investigations into four fundamental problems. Firstly, an efficient method was proposed to measure pedestrian exposure at the macroscopic level by making use of the revealed diary passenger trips from the travel characteristics survey. Secondly, the effects of travel behavior, such as trip purposes, on pedestrian crash occurrence were investigated. Thirdly, the relationship between the amount of pedestrian walking stratified by transport modes and pedestrian crashes was examined in a transit-oriented society. Fourthly, the simultaneous modeling of crash occurrence and pedestrian involvement in crashes was developed in a joint probability model for the investigation of the possible correlation among different crash types. Influencing zonal factors, such as population sociodemographic characteristics, household size, road density, traffic controls, public transport facilities, land use, traffic flow, and POI data, have been investigated throughout the study in the dissertation. The findings are useful for the design of various transport facilities, road safety education for vulnerable pedestrian groups, and more importantly, accessibility to essential urban services and attractions. With these proposed improvements, a safe and accessible walking environment can be promoted in Hong Kong.

Chapter 2 reviewed the effort on pedestrian safety at the macroscopic level. Road safety is a complex problem associated with many aspects of society. Risk factors that can influence the occurrence of crashes include traffic characteristics, sociodemographic features, road networks, public transport facilities, and built environment characteristics. However, the influence of risk factors can be different from site to site. It is of importance to have a comprehensive investigation of the risk factors on crash occurrences in Hong Kong, especially for pedestrian crashes in a transit-oriented society where more pedestrian walking occurs.

An exposure measure is an important metric for the quantification of crash risk. It is difficult to measure pedestrian exposure at the macroscopic level. Population proxies, predicted trip demand, and observational pedestrian volumes were served as the pedestrian exposure measures. However, there remain limitations to the existing measures, leaving room for further exploration of more efficient pedestrian exposure measures with better representatives of pedestrian movements.

Activity-travel patterns have caught the attention of scholars for travel demand planning. It has been well acknowledged that travel is driven by participation in various activities. The purposes of travel impact the emotions and behaviors of travelers on roads. In line with this, different trip purposes can also have different effects on crash occurrences, which should be of concern for crash prediction modeling. The ignorance of the possible influence of people's travel purpose might cause the loss of explanatory power and an incomplete understanding of crash occurrence. However, few studies have incorporated trip purposes implicitly into the modeling crash prediction.

Although more or less walking might be needed for access to/egress from different transport modes, with regard to the different features of different modes, the walking characteristics tend to be different across different modes. Within the transit-oriented society in which walking is the primary means to connect with the public transport stations, the crash risk of pedestrians connecting to different transport modes should be identified for better planning and facilitating of the pedestrian walking environment.

Possible correlations among different crash types do exist as the result of possible various effects of risk factors across crash types. Understanding the correlations between crash occurrence and pedestrian involvement in crashes provides implications for improving overall road safety and the pedestrian walking environment simultaneously. The hierarchical structure of overall crash occurrence and pedestrian involvement in crashes calls for specific methodologies.

The research gaps above motivate the work done in Chapter 3 to 6. The geographic unit system in Hong Kong was described in Chapter 3. Given the multilevel hierarchical zone systems and in light of the data availability at different zone levels, the TPUs and the PDDs with an appropriate number of zones for crash prediction modeling were adopted in the study. The data from various sources were introduced, including the crash records, traffic characteristics, population census, and sociodemographic statistics, road network features, land use data, built environment, and points-of-interest information. Methodologies of the models adopted in the dissertation, i.e., the random-parameter negative binomial regression model, the joint probability model, the CAR model, and Bayesian estimation methods were formulated and elaborated for application in the estimations.

An efficient and reasonable method for measuring pedestrian exposure was proposed in Chapter 4. After being well adjusted with multiple statistics, the trip diaries available in the TCS2011 survey were expanded to represent the daily passenger travel. The trip records provided detailed information on motorized trips, including the locations of each trip (motorized and non-motorized). Walk-only trips and walking trip legs of motorized trips were derived for the estimation of pedestrian exposure. Walking frequency (walk-only trips and walking trip legs) and walking time were calculated and examined as pedestrian exposure measures. In addition, the pedestrian exposure was also classified by six trip purposes, i.e., back home, to work, to school, shopping, dining, and others, to explore the effects of trip purposes on pedestrian crash occurrence. A random-parameter negative binomial model was applied for the estimation. The results indicated that the proposed pedestrian exposure measures outperformed the traditional measures with better goodness-of-fit. The walking frequency was found to be a better pedestrian exposure surrogate than walking time as the former showed better model performance. Pedestrian exposure under back-home purposes was tested to be more likely to correlate with pedestrian crashes than other trip purposes. The finding was consistent with the hourly variation of crashes in which more pedestrian crashes were recorded at noon and in the late afternoon even though the exposure value was even higher. This could shed light on the implementation of effective policy strategies that can improve the safety of vulnerable pedestrian groups in specific periods.

Chapter 5 investigated the influence of pedestrian walking associated with different transport modes on pedestrian crash occurrence. Trip records from the survey provided the mode choices of each trip leg, which enabled the stratification of pedestrian exposure by six transport modes, i.e., metro, buses, light buses, taxis, private cars, and others. In the meantime, possible correlations with crash occurrence might exist due to the similarity of the influencing factors of adjacent zones and the heterogeneity of unobserved factors. Therefore, the spatial correlation was incorporated in this study. CAR model used for addressing the spatial correlation was formulated with the negative binomial model. The estimated results showed that walking connected with roadway public transport, i.e., buses, was associated with a higher risk to be involved with pedestrian crashes than other modes. In contrast, walking access to and egress from metro stations was found to have no significant effects on crash occurrences. The incorporation of spatial correlation provided better model performance. The results were indicative to the design and planning of road facilities, i.e., pedestrian crossings and traffic signals, that can enhance the safety and accessibility of public transport.

Integrated modeling of total crash occurrence and pedestrian involvement in crashes was carried out in Chapter 6. Possible correlations of different crash types were examined as the effects of influencing factors on crash occurrence might vary across crash types. A joint probability integrating negative binomial regression model and a binomial regression model were applied to model the hierarchical crash types, i.e., total crash occurrence and pedestrian crashes. The roles of environment, road network, traffic flow, and population characteristics were investigated. The full Bayesian approach was adopted for the estimation of the parameters using the MCMC simulation method. Results indicate that crash occurrence is correlated to traffic flow, the number of non-signalized intersections, and points of interest such as restaurants and hotels. By contrast, population age, ethnicity, education, household size, road density, and the number of public transit stations could affect the propensity of pedestrian involvement in crashes. These findings indicate that better design and planning of built environments are necessary for safer and more efficient access for pedestrians and the long-term improvement of walkability in a high-density city such as Hong Kong.

## 7.2 Key findings and recommendations

In conclusion, with regard to the topic of pedestrian safety at the macroscopic level, the key contributions of the dissertation are summarized in the following aspects:

1) A valid method for measuring pedestrian exposure at the macroscopic level has been proposed by making use of the revealed trip diaries of the travel characteristics survey. The proposed pedestrian measure showed its superiority to the conventional surrogates of the exposures.
2) It is the first time that the influence of travel behavior, i.e., trip purposes, on pedestrian crash occurrence has been explicitly investigated in the pedestrian crash prediction model. The results highlighted that the crash risks of exposure vary with the purpose of people's trip makings.
3) Different crash risks of pedestrian walking connected with different transport modes were evaluated in the study. It was found that the risk opportunities for the crash occurrence of pedestrian exposure connected with different modes are different.
4) The correlations across total crash occurrence and pedestrian involvement in crashes were explored with an integrated joint probability. The varying effects of zonal risk factors on the occurrence of different crash types were evaluated.

Based on the studies of the thesis for the contributions above, the findings are indicative to the improvement of pedestrian safety from various aspects, such as education, planning of pedestrian infrastructure, and traffic management & control strategies. Potential implications for policymaking are suggested to improve pedestrian safety at a macroscopic level.

1) Attention should be given to the better planning of road safety strategies that can enhance the safety of vulnerable pedestrians at specific time periods and locations. In particular, it is necessary to improve road user education and enforcement measures that could enhance the safety awareness of pedestrians on their way back home (Wong et al., 2008).

2) Continuous education in road safety can be enhanced at all education levels from communities to schools, especially for children, to strengthen their safety awareness and to protect themselves from crashes.

3) In order to separate the potential conflicts of pedestrians and vehicles at the intersections, the upgrading of non-signalized intersections into signalized intersections, and the construction of footbridges and underpasses will do a great deal to help reduce the chance of pedestrians running into crashes.

4) Better facilities to improve the walking access to and from local activities and services are a necessity, such as more accessible routes, footpaths, crosswalks.

5) At the same time as advocating public transport usage, it is important to improve the design and planning of the environment for passengers connecting to public transport modes, such as safer passenger waiting areas, drop-off & pick-up areas, and roadside curbs at bus stops.

6) In the case of the growing aging problem in society, a better road crossing infrastructure for the elderly should be facilitated, i.e., elevators for footbridges or underpasses. The elderly population will benefit a lot from these facilities when crossing roads.

## 7.3 Limitations and recommendations for future research

Though the studies in the thesis have put effort into the analysis of pedestrian safety at a macroscopic level, there exist limitations in the thesis. The data used for the model building and analyses were collected a long time ago, i.e., TCS2011 was collected almost 10 years. The validity of the model testing, and then the validity of the findings, and the usefulness of recommendations derived from the analyses will be needed to be verified. The transferability of the results estimated in the thesis remains uninvestigated.

### 7.3.1 Observational study

Throughout the study of this dissertation, the TCS2011 offered important support for the research as the well-recorded trip diaries enabled a superior method of the pedestrian exposure measure compared to the conventional method. However, the study did not directly analyze the actual action of pedestrians per sec. Observational experiments can be carried out to pay close observation to the pedestrian behavior on roads to analyze the pedestrian behavior, such as by video recording the pedestrians at some locations and processing the recorded data later to observe the pedestrian actions (Zhu et al., 2021). In this way, information about the walking speed, attentiveness, group behavior, and crossing behavior at intersections during different times of the day can be possibly identified. The observations will enable close investigation of the relationship between pedestrian behavior and the surrounding built environment, i.e., the different behavior of pedestrians at different land-use areas (residential, commercial, etc.). The results will help the understanding of the effects of trip purposes on pedestrians' behavior on roads, and thus the risk to crash occurrence.

### 7.3.2 Perception survey

The perceptions of pedestrians on road safety can have important impacts on their behavior on roads. Stated preference surveys can be designed for acquiring the perceptions of pedestrians towards the transport facilities, mode selection, and their attitudes on risky behavior under different road conditions and travel purposes. The data will provide more reliable information on pedestrians' travel purposes and motivations at the individual level. The effects of trip purposes on crash risks with an aggregated approach can be validated. Besides, the perceptions of pedestrians help identify the treatments that can create a safe and/or desirable environment for pedestrians themselves. Such studies will give important feedback for the evaluation of the effectiveness of the countermeasures. In addition, further information on the effects of the social norm can be analyzed from the perception survey to investigate the transferability of the results and findings in Hong Kong to other places.

### 7.3.3 Further investigation of spatial dependency

Though spatial correlation has been taken into account in Chapter 5 by adopting the CAR model, the model only captured the heterogeneity effects of spatial dependency from unobserved factors. In fact, spatial dependency includes both the spatial correlation effects from unobserved factors and the spillover effects from observed factors (Bhat et al., 2017). Due to the similarity of zonal characteristics among neighboring spatial units, factors in one zone might possibly have an impact on the neighboring zones, which is referred to as “spillover effects”. The incorporation of such spatial spillover effects will offer more accurate estimates for the influencing factors and reliable implications for the improvement of road safety.

As it has been noted that both crash occurrence and influencing factors might have spatial–temporal heterogeneity, different geographic unit systems might present different statistical inferences and interpretations, which is referred to as the modifiable areal unit problem (MAUP) (Huang et al., 2013; Xu et al., 2014; Xu et al., 2018). Two geographic unit systems, i.e., the TPUs and the PDDs, were adopted as the spatial unit system for data aggregation and model estimation in the thesis. There exist different estimated results of the same factor in different unit systems, e.g., negative effect of AADT was examined in Chapter 4 using the PDD zone system, while the positive effect of AADT was examined in Chapter 5 and Chapter 6 using the TPU zone system. Therefore, it is worth trying to explore the MAUP problem and identifying the optimal zone system for pedestrian safety analysis.

### 7.3.4 Safety consideration for transportation planning

Road crashes are byproducts of transport activities, especially since the start of the automobile stage. For a long time in past decades, road safety analysis was presented as a reactive approach, as the actions could only be taken after crashes had been witnessed, including identification, diagnosis, and remedy suggestions to the existing crash-prone locations (Lovegrove and Sayed, 2006; Pirdavani et al., 2012). However, retrofitting countermeasures to improve the existing problem might be costly as the damage had been done. Therefore, proactive approaches are beginning to catch the attention of researchers in road safety improvement. It is essential to incorporate road safety as an important decision factor when making long-term transportation plans. For instance, a travel demand forecasting model that considers the risk of crash occurrence can be developed. The utility cost of travel will involve not only the financial cost and time but also the risk cost resulting from potential crashes.

# Reference


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

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

Aecom Asia Co.Ltd., 2010. Study on Pedestrian Subways and Related Traffic Improvement Measures in Causeway Bay - Executive Summary. Retrieved from https://www.td.gov.hk/en/publications_and_press_releases/publications/free_publications/index.html

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

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

Aguero-Valverde, J., Jovanis, P.P., 2009. Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. *Transportation Research Record*, 2136 (1), 82-91.

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

Allahviranloo, M., Recker, W., 2015. Mining activity pattern trajectories and allocating activities in the network. *Transportation*, 42 (4), 561-579.

Allsop, R.E., 2001. *Road safety: Britain in Europe*. Parliamentary Advisory Council on Transport Safety.

Allsop, R.E., 2005. Travel survey data required to inform transport safety policy and practice. *Transportmetrica*, 1 (3), 241-245.

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

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis and Prevention*, 41 (1), 153-159.

Anderson, R., 2008. *Pedestrian collisions in South Australia*. Centre for Automotive Safety Research.

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

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

Barua, S., El-Basyouny, K., Islam, M.T., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. *Analytic Methods in Accident Research*, 9, 1-15.

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

Besser, L.M., Dannenberg, A.L., 2005. Walking to public transit: steps to help meet physical activity recommendations. *American Journal of Preventive Medicine*, 29 (4), 273-280.

Bhat, C.R., 2005. A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B: Methodological*, 39 (8), 679-707.

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B: Methodological*, 45 (7), 923-939.

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

Bhat, C.R., Goulias, K.G., Pendyala, R.M., Paleti, R., Sidharthan, R., Schmitt, L., Hu, H.-H., 2013. A household-level activity pattern generation model with an application for Southern California. *Transportation*, 40 (5), 1063-1086.

Bhat, C.R., Koppelman, F.S., 1999. Activity-based modeling of travel demand. *Handbook of Transportation Science.* Springer, pp. 35-61.

Bhat, C.R., Singh, S.K., 2000. A comprehensive daily activity-travel generation model system for workers. *Transportation Research Part A-Policy and Practice*, 34 (1), 1-22.

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

Blazquez, C.A., Celis, M.S., 2013. A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. *Accident Analysis and Prevention*, 50, 304-311.

Bowman, J.L., 1998. The day activity schedule approach to travel demand analysis. Massachusetts Institute of Technology.

Bowman, J.L., Ben-Akiva, M.E., 2001. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A-Policy and Practice*, 35 (1), 1-28.

Bulmer, M., 1974. On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics*, 101-110.

Cafiso, S., Di Graziano, A., Pappalardo, G., 2013. Road safety issues for bus transport management. *Accident Analysis and Prevention*, 60, 324-333.

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

Cai, Q., Abdel-Aty, M., Lee, J., Huang, H., 2019. Integrating macro-and micro-level safety analyses: a Bayesian approach incorporating spatial interaction. *Transportmetrica A: Transport Science*, 15 (2), 285-306.

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

Cameron, A.C., Trivedi, P.K., 1990. Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, 46 (3), 347-364.

Carter, D.L., Council, F.M., 2007. Factors contributing to pedestrian and bicycle crashes on rural highways. In: Proceedings of the *Transportation Research Board 86th Annual Meeting Paper*, pp. 07-2457.

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

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

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 Injury Prevention*, 11 (5), 508-513.

Chapman, R., 1973. The concept of exposure. *Accident Analysis and Prevention*, 5 (2), 95-110.

Chen, T., Sze, N., Chen, S., Labi, S., 2020a. Urban road space allocation incorporating the safety and construction cost impacts of lane and footpath widths. *Journal of safety research*, 75, 222-232.

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

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

Cheng, W., Gill, G.S., Ensch, J.L., Kwong, J., Jia, X., 2018a. Multimodal crash frequency modeling: multivariate space-time models with alternate spatiotemporal interactions. *Accident Analysis and Prevention*, 113, 159-170.

Cheng, W., Gill, G.S., Zhang, Y., Cao, Z., 2018b. Bayesian spatiotemporal crash frequency models with mixture components for space-time interactions. *Accident Analysis and Prevention*, 112, 84-93.

Cheng, Y.-H., Chen, S.-Y., 2015. Perceived accessibility, mobility, and connectivity of public transportation systems. *Transportation Research Part A: Policy and Practice*, 77, 386-403.

Chiou, Y.C., Sheng, Y.C., Fu, C., 2017. Freeway crash frequency modeling under time-of-day distribution. *World Conference on Transport Research 2016*, 25, 664-676.

Chipman, M.L., Macgregor, C.G., Smiley, A.M., Lee-Gosselin, M., 1993. The role of exposure in comparisons of crash risk among different drivers and driving environments*. *Accident Analysis and Prevention*, 25 (2), 207-211.

Chliaoutakis, J.E., Darviri, C., Demakakos, P.T., 1999. The impact of young drivers' lifestyle on their road traffic accident risk in greater Athens area. *Accident Analysis and Prevention*, 31 (6), 771-780.

Chliaoutakis, J.E., Koukouli, S., Lajunen, T., Tzamalouka, G., 2005. Lifestyle traits as predictors of driving behaviour in urban areas of Greece. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8 (6), 413-428.

Christensen, R., Johnson, W., Branscum, A., Hanson, T.E., 2011. *Bayesian ideas and data analysis: an introduction for scientists and statisticians*. CRC Press.

Chu, X., 2003. The fatality risk of walking in America: A time-based comparative approach. In: Proceedings of the *Walk21 conference: Health equity and the environment. Portland, Oregon*.

Clarke, D.D., Ward, P., Bartle, C., Truman, W., 2006. Young driver accidents in the UK: The influence of age, experience, and time of day. *Accident Analysis and Prevention*, 38 (5), 871-878.

Clifton, K.J., Kreamer-Fults, K., 2007. An examination of the environmental attributes associated with pedestrian–vehicular crashes near public schools. *Accident Analysis and Prevention*, 39 (4), 708-715.

Cottrill, C.D., Thakuriah, P.V., 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. *Accident Analysis and Prevention*, 42 (6), 1718-1728.

Coughenour, C., Clark, S., Singh, A., Claw, E., Abelar, J., Huebner, J., 2017. Examining racial bias as a potential factor in pedestrian crashes. *Accident Analysis and Prevention*, 98, 96-100.

Dai, D., 2012. Identifying clusters and risk factors of injuries in pedestrian–vehicle crashes in a GIS environment. *Journal of Transport Geography*, 24, 206-214.

Dai, D., Jaworski, D., 2016. Influence of built environment on pedestrian crashes: A network-based GIS analysis. *Applied Geography*, 73, 53-61.

Davis, D.G., Braaksma, J.P., 1988. Adjusting for luggage-laden pedestrians in airport terminals. *Transportation Research Part A: General*, 22 (5), 375-388.

Delbosc, A., Currie, G., 2012. Modelling the causes and impacts of personal safety perceptions on public transport ridership. *Transport Policy*, 24, 302-309.

Devlin, A., Hoareau, E., Logan, D., Corben, B.F., Oxley, J., 2010. Towards zero pedestrian trauma: Literature review and serious casualty analysis. In: Proceedings of the *Proceedings of the Australasian road safety research, policing and education conference*.

Dong, N., Huang, H., Lee, J., Gao, M., Abdel-Aty, M., 2016. Macroscopic hotspots identification: a Bayesian spatio-temporal interaction approach. *Accident Analysis and Prevention*, 92, 256-264.

Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects. *Accident Analysis and Prevention*, 82, 192-198.

Dumbaugh, E., Li, W., 2010. Designing for the safety of pedestrians, cyclists, and motorists in urban environments. *Journal of the American Planning Association*, 77 (1), 69-88.

Elias, W., Shiftan, Y., 2014. Analyzing and modeling risk exposure of pedestrian children to involvement in car crashes. *Accident Analysis and Prevention*, 62, 397-405.

Elias, W., Toledo, T., Shiftan, Y., 2010. The effect of daily-activity patterns on crash involvement. *Accident Analysis and Prevention*, 42 (6), 1682-1688.

Ettema, D., Ashiru, O., Polak, J.W., 2004. Modeling timing and duration of activities and trips in response to road-pricing policies. *Travel Behavior and Values 2004*, (1894), 1-10.

Evans, A.W., 1994. Evaluating public transport and road safety measures. *Accident Analysis and Prevention*, 26 (4), 411-428.

Finch, D., Kompfner, P., Lockwood, C., Maycock, G., 1994. *Speed, speed limits and accidents*. Safety Resource Centre.

Fontaine, H., Gourlet, Y., 1997. Fatal pedestrian accidents in France: A typological analysis. *Accident Analysis and Prevention*, 29 (3), 303-312.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. *Bayesian Data Analysis*. Chapman and Hall/CRC.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7 (4), 457-472.

Goel, R., Tiwari, G., 2016. Access–egress and other travel characteristics of metro users in Delhi and its satellite cities. *IATSS Research*, 39 (2), 164-172.

Gomes, M.J.T.L., Cunto, F., Da Silva, A.R., 2017. Geographically weighted negative binomial regression applied to zonal level safety performance models. *Accident Analysis and Prevention*, 106, 254-261.

Greene-Roesel, R., Diogenes, M.C., Ragland, D.R., 2007. Estimating pedestrian accident exposure: protocol report. Retrieved from https://safetrec.berkeley.edu/publications/estimating-pedestrian-accident-exposure-protocol-report.

Guo, Q., Xu, P., Pei, X., Wong, S.C., Yao, D., 2017. The effect of road network patterns on pedestrian safety: A zone-based Bayesian spatial modeling approach. *Accident Analysis and Prevention*, 99, 114-124.

Guo, Z., Loo, B.P., 2013. Pedestrian environment and route choice: evidence from New York City and Hong Kong. *Journal of Transport Geography*, 28, 124-136.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2007. Safety prediction models: proactive tool for safety evaluation in urban transportation planning applications. *Transportation Research Record*, 2019 (1), 225-236.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. *Accident Analysis and Prevention*, 42 (2), 676-688.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., 1998. *Multivariate data analysis*. Prentice hall Upper Saddle River, NJ.

Haleem, K., Alluri, P., Gan, A., 2015. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accident Analysis and Prevention*, 81, 14-23.

Hauer, E., 1982. Traffic conflicts and exposure. *Accident Analysis and Prevention*, 14 (5), 359-364.

Hezaveh, A.M., Arvin, R., Cherry, C.R., 2019. A geographically weighted regression to estimate the comprehensive cost of traffic crashes at a zonal level. *Accident Analysis and Prevention*, 131, 15-24.

Hijar, M., Vázquez-Vela, E., Arreola-Risa, C., 2003. Pedestrian traffic injuries in Mexico: a country update. *Injury Control and Safety Promotion*, 10 (1-2), 37-43.

Hilbe, J.M., 2011. *Negative binomial regression*. Cambridge University Press.

Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-level crash risk analysis in Florida: Bayesian spatial modeling. *Transportation Research Record*, (2148), 27-37.

Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., Abdel-Aty, M., 2016. Macro and micro models for zonal crash prediction with application in hot zones identification. *Journal of Transport Geography*, 54, 248-256.

Huang, H., Xu, P., Abdel-Aty, M., 2013. Transportation Safety Planning: A Spatial Analysis Approach. *Transportation*, 2, 3.

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

Hummel, T., 1998. Dutch Pedestrian Research Review: A Review of the Main Traffic Safety Research on Pedestrians in the Netherlands, 1984-1997.

Izsák, R., 2008. Maximum likelihood fitting of the Poisson lognormal distribution. *Environmental and Ecological Statistics*, 15 (2), 143-156.

Jacobs, G., Aeron-Thomas, A., Astrop, A., 2000. Estimating global road fatalities.

Jiang, X., Abdel-Aty, M., Hu, J., Lee, J., 2016. Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach. *Neurocomputing*, 181, 53-63.

Jiménez-Mejías, E., Martínez-Ruiz, V., Amezcua-Prieto, C., Olmedo-Requena, R., De Dios Luna-Del-Castillo, J., Lardelli-Claret, P., 2016. Pedestrian-and driver-related factors associated with the risk of causing collisions involving pedestrians in Spain. *Accident Analysis and Prevention*, 92, 211-218.

Johnson, E., Geyer, J., Rai, N., Ragland, D., 2004. Low income childhood pedestrian injury: understanding the disparate risk. UC Berkeley traffic safety center. Retrieved from https://safetrec.berkeley.edu/publications/low-income-childhood-pedestrian-injury-understanding-disparate-risk.

Jonah, B.A., Engel, G.R., 1983. Measuring the relative risk of pedestrian accidents. *Accident Analysis and Prevention*, 15 (3), 193-206.

Jonathan, A.-V., Wu, K.-F.K., Donnell, E.T., 2016. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. *Accident Analysis and Prevention*, 87, 8-16.

Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. *Accident Analysis and Prevention*, 23 (4), 239-255.

Joshua, S.C., Garber, N.J., 1990. Estimating truck accident rate and involvements using linear and Poisson regression models. *Transportation Planning and Technology*, 15 (1), 41-58.

Jovanis, P.P., Chang, H.-L., 1986. Modeling the relationship of accidents to miles traveled. *Transportation Research Record*, 1068, 42-51.

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

Keall, M.D., 1995. Pedestrian exposure to risk of road accident in New Zealand. *Accident Analysis and Prevention*, 27 (5), 729-740.

Kharola, P., Tiwari, G., Mohan, D., 2010. Traffic safety and city public transport system: case study of Bengaluru, India. *Journal of Public Transportation*, 13 (4), 4.

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

Kim, K., Pant, P., Yamashita, E., 2010a. Accidents and accessibility: Measuring influences of demographic and land use variables in Honolulu, Hawaii. *Transportation Research Record*, 2147 (1), 9-17.

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

Koornstra, M., Lynam, D., Nilsson, G., 2002. SUNflower: a comparative study of the development of road. *Leidschendam: SWOV Institute for Road Safety Research*.

Lakhotia, S., Lassarre, S., Rao, K.R., Tiwari, G., 2019. Pedestrian accessibility and safety around bus stops in Delhi. *IATSS Research*.

Lam, L.T., 2004. Environmental factors associated with crash-related mortality and injury among taxi drivers in New South Wales, Australia. *Accident Analysis and Prevention*, 36 (5), 905-908.

Lam, W.W., Loo, B.P., Yao, S., 2013. Towards exposure-based time-space pedestrian crash analysis in facing the challenges of ageing societies in Asia. *Asian Geographer*, 30 (2), 105-125.

Lam, W.W., Yao, S., Loo, B.P., 2014. Pedestrian exposure measures: a time-space framework. *Travel Behaviour and Society*, 1 (1), 22-30.

Lascala, E.A., Gerber, D., Gruenewald, P.J., 2000. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accident Analysis and Prevention*, 32 (5), 651-658.

Lavieri, P.S., Bhat, C.R., 2019. Modeling individuals' willingness to share trips with strangers in an autonomous vehicle future. *Transportation Research Part A: Policy and Practice*, 124, 242-261.

Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. *Accident Analysis and Prevention*, 37 (4), 775-786.

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

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

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

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

Li, L., Ren, H., Zhao, S., Duan, Z., Zhang, Y., Zhang, A., 2017. Two dimensional accessibility analysis of metro stations in Xi'an, China. *Transportation Research Part A: Policy and Practice*, 106, 414-426.

Li, Y.C., Sze, N.N., Wong, S.C., 2014. Effective measures for combating drink-driving offenses: an attitudinal model for Hong Kong. *Transportmetrica A: Transport Science*, 10 (8), 722-739.

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

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

Liu, J., Khattak, A.J., Wali, B., 2017. Do safety performance functions used for predicting crash frequency vary across space? Applying geographically weighted regressions to account for spatial heterogeneity. *Accident Analysis and Prevention*, 109, 132-142.

Loo, B.P., Chen, C., Chan, E.T., 2010. Rail-based transit-oriented development: lessons from New York City and Hong Kong. *Landscape and Urban Planning*, 97 (3), 202-212.

Loo, B.P., Wong, S.C., Hung, W.T., Lo, H.K., 2007. A review of the road safety strategy in Hong Kong. *Journal of Advanced Transportation*, 41 (1), 3-37.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation research part A: Policy and Practice*, 44 (5), 291-305.

Loukaitou-Sideris, A., Liggett, R., Sung, H.-G., 2007. Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles. *Journal of Planning Education and Research*, 26 (3), 338-351.

Lovegrove, G.R., Sayed, T., 2006. Using macrolevel collision prediction models in road safety planning applications. *Transportation Research Record*, 1950 (1), 73-82.

Lunn, M., Auyeung, P.K., Lo, H.K., 2018. Report of the Independent Review Committee on Hong Kong's Franchised Bus Service. Retrieved from https://www.irc-bus.gov.hk/pdf/IRC_Report_Eng_Redacted.pdf.

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis and Prevention*, 40 (3), 964-975.

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

Mannering, F., 2018. Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research*, 17, 1-13.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research*, 1, 1-22.

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

Mannings (Asia) Consultants Ltd., 2010. Study on improvements to pedestrian links in Mong Kok. Retrieved from https://www.td.gov.hk/filemanager/en/publication/td1712008_executivesummary.pdf

Martin, J.L., Wu, D., 2018. Pedestrian fatality and impact speed squared: Cloglog modeling from French national data. *Traffic Injury Prevention*, 19 (1), 94-101.

Mason, K., 1979. Accident patterns by time-of-day and day-of-week of injury occurrence. *Journal of Occupational Accidents*, 2 (2), 159-176.

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

Miller, E.J., Roorda, M.J., 2003. Prototype model of household activity-travel scheduling. *Transportation Research Record*, (1831), 114-121.

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

Moeinaddini, M., Asadi-Shekari, Z., Sultan, Z., Shah, M.Z., 2015. Analyzing the relationships between the number of deaths in road accidents and the work travel mode choice at the city level. *Safety Science*, 72, 249-254.

Mohan, D., 2001. Planning for public transport: integrating safety, environment and economic issues. Workshop on Transport, Land Use and the Environment', Pune, India. Available at: http://www. iitd. ac. in/tripp/publications/paper/safety/dmpune. pdf.

Mohanty, S., Bansal, S., Bairwa, K., 2017. Effect of integration of bicyclists and pedestrians with transit in New Delhi. *Transport Policy*, 57, 31-40.

Moran, P.A., 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37 (1/2), 17-23.

Mukherjee, D., Mitra, S., 2019. A comparative study of safe and unsafe signalized intersections from the view point of pedestrian behavior and perception. *Accident Analysis and Prevention*, 132, 105218.

Mva Hong Kong Limited, 2013. Traffic Study on Pedestrian Links for the West Kowloon Development Area and its Connections with Surrounding Districts. Retrieved from https://www.td.gov.hk/filemanager/en/publication/final_es.pdf.

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

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

Noland, R., Quddus, M., 2004. Analysis of pedestrian and bicycle casualties with regional panel data. *Transportation Research Record*, (1897), 28-33.

Odero, W., Garner, P., Zwi, A., 1997. Road traffic injuries in developing countries: a comprehensive review of epidemiological studies. *Tropical Medicine International Health*, 2 (5), 445-460.

Olszewski, P., Szagała, P., Wolański, M., Zielińska, A., 2015. Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in Poland. *Accident Analysis and Prevention*, 84, 83-91.

Osama, A., Sayed, T., 2017. Macro-spatial approach for evaluating the impact of socio-economics, land use, built environment, and road facility on pedestrian safety. *Canadian Journal of Civil Engineering*, 44 (12), 1036-1044.

Paez, A., Scott, D.M., Morency, C., 2012. Measuring accessibility: positive and normative implementations of various accessibility indicators. *Journal of Transport Geography*, 25, 141-153.

Pahukula, J., Hernandez, S., Unnikrishnan, A., 2015. A time of day analysis of crashes involving large trucks in urban areas. *Accident Analysis and Prevention*, 75, 155-163.

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

Papadimitriou, E., 2016. Towards an integrated approach of pedestrian behaviour and exposure. *Accident Analysis and Prevention*, 92, 139-152.

Papadimitriou, E., Yannis, G., Golias, J., 2012. Analysis of pedestrian exposure to risk in relation to crossing behavior. *Transportation Research Record*, 2299 (1), 79-90.

Park, J., Abdel-Aty, M., Wang, J.-H., 2017. Time series trends of the safety effects of pavement resurfacing. *Accident Analysis and Prevention*, 101, 78-86.

Part, D., 2010. Highway safety manual. Aashto.

Peden, M., Scurfield, R., Sleet, D., Hyder, A.A., Mathers, C., Jarawan, E., Mohan, D., 2004. *World report on road traffic injury prevention*. World Health Organization.

Pei, X., Wong, S.C., Sze, N.N., 2011. A joint-probability approach to crash prediction models. *Accident Analysis and Prevention*, 43 (3), 1160-1166.

Pei, X., Wong, S.C., Sze, N.N., 2012. The roles of exposure and speed in road safety analysis. *Accident Analysis and Prevention*, 48, 464-471.

Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., 2012. Application of different exposure measures in development of planning-level zonal crash prediction models. *Transportation Research Record*, 2280 (1), 145-153.

Planning Department, H.K.S.A.R., 2020. Internal Transport Facilities.

Ponnaluri, R.V., Nagar, F., 2010. Road crash risk among vulnerable population groups in Andhra Pradesh, India.

Prasertsubpakij, D., Nitivattananon, V., 2012. Evaluating accessibility to Bangkok Metro Systems using multi-dimensional criteria across user groups. *IATSS Research*, 36 (1), 56-65.

Pulugurtha, S.S., Duddu, V.R., Kotagiri, Y., 2013. Traffic analysis zone level crash estimation models based on land use characteristics. *Accident Analysis and Prevention*, 50, 678-687.

Pulugurtha, S.S., Penkey, E.N., 2010. Assessing use of pedestrian crash data to identify unsafe transit service segments for safety improvements. *Transportation Research Record*, 2198 (1), 93-102.

Qin, X., Ivan, J.N., 2001. Estimating pedestrian exposure prediction model in rural areas. *Transportation Research Record*, 1773 (1), 89-96.

Qin, X., Ivan, J.N., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis and Prevention*, 36 (2), 183-191.

Qin, X., Ivan, J.N., Ravishanker, N., Liu, J., Tepas, D., 2006. Bayesian estimation of hourly exposure functions by crash type and time of day. *Accident Analysis and Prevention*, 38 (6), 1071-1080.

Quddus, M.A., 2008a. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. *Accident Analysis and Prevention*, 40 (4), 1486-1497.

Quddus, M.A., 2008b. Time series count data models: An empirical application to traffic accidents. *Accident Analysis and Prevention*, 40 (5), 1732-1741.

Rankavat, S., Tiwari, G., 2016. Pedestrians risk perception of traffic crash and built environment features–Delhi, India. *Safety Science*, 87, 1-7.

Recker, W.W., 2001. A bridge between travel demand modeling and activity-based travel analysis. *Transportation Research Part B: Methodological*, 35 (5), 481-506.

Rhee, K.A., Kim, J.K., Lee, Y.I., Ulfarsson, G.F., 2016. Spatial regression analysis of traffic crashes in Seoul. *Accident Analysis and Prevention*, 91, 190-199.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accident Analysis and Prevention*, 43 (5), 1666-1676.

Siddiqui, C., Abdel-Aty, M., Choi, K., 2012a. Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis and Prevention*, 45, 382-391.

Siddiqui, C., Abdel-Aty, M., Huang, H., 2012b. Aggregate nonparametric safety analysis of traffic zones. *Accident Analysis and Prevention*, 45, 317-25.

Sinharay, S., 2003. Assessing convergence of the Markov chain Monte Carlo algorithms: A review. *ETS Research Report Series*, 2003 (1), i-52.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Rtatistical Rociety: Series B (Statistical Methodology)*, 64 (4), 583-639.

Sukhai, A., Jones, A.P., Love, B.S., Haynes, R., 2011. Temporal variations in road traffic fatalities in South Africa. *Accident Analysis and Prevention*, 43 (1), 421-428.

Sun, G., Oreskovic, N.M., Lin, H., 2014. How do changes to the built environment influence walking behaviors? A longitudinal study within a university campus in Hong Kong. *International Journal of Health Geographics*, 13 (1), 1-10.

Sung, H., Choi, K., Lee, S., Cheon, S., 2014. Exploring the impacts of land use by service coverage and station-level accessibility on rail transit ridership. *Journal of Transport Geography*, 36, 134-140.

Sze, N.N., Christensen, K.M., 2017. Access to urban transportation system for individuals with disabilities. *IATSS Research*, 41 (2), 66-73.

Sze, N.N., Su, J., Bai, L., 2019. Exposure to pedestrian crash based on household survey data: effect of trip purpose. *Accident Analysis and Prevention*, 128, 17-24.

Sze, N.N., Wong, S.C., 2007. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accident Analysis and Prevention*, 39 (6), 1267-1278.

Sze, N.N., Wong, S.C., Lee, C.Y., 2014. The likelihood of achieving quantified road safety targets: A binary logistic regression model for possible factors. *Accident Analysis and Prevention*, 73, 242-251.

Tasic, I., Porter, R.J., 2016. Modeling spatial relationships between multimodal transportation infrastructure and traffic safety outcomes in urban environments. *Safety Science*, 82, 325-337.

Tay, R., 2003. The efficacy of unemployment rate and leading index as predictors of speed and alcohol related crashes in Australia. *International Journal of Transport Economics*, 363-375.

Tay, R., 2017. Correlation, variance inflation and multicollinearity in regression model. *Journal of the Eastern Asia Society for Transportation Studies*, 12, 2006-2015.

Tay, R., Choi, J., 2016. Factors associated with crashes involving taxi owners and non-owners: a case of moral hazard and adverse selection? *Accident Analysis and Prevention*, 87, 78-82.

Train, K., 2009. *Discrete choice methods with simulation*. Cambridge university press.

Transport and Housing Bureau, H.K.S.A.R., 2017. Public transport strategy study.

Transport Department, H.K.S.A.R., 2001. Transport Planning and Design Manual. 9.

Transport Department, H.K.S.A.R., 2012a. The Annual Traffic Census. Retrieved from https://www.td.gov.hk/en/publications_and_press_releases/publications/free_publications/the_annual_traffic_census_2011/index.html.

Transport Department, H.K.S.A.R., 2012b. Travel Characteristics Survey 2011 (final report). Retrieved from https://www.td.gov.hk/en/publications_and_press_releases/publications/free_publications/index.html.

Transport Department, H.K.S.A.R., 2012-2016. Annual Tranport Digest: Major Contributory Factors of Road Traffic Accidents. Retrieved from https://www.td.gov.hk/en/publications_and_press_releases/publications/free_publications/annual_transport_digest/index.html.

Transport Department, H.K.S.A.R., 2016. Road traffic casualties by class of road user and degree of injury 2016.

Transport Department, H.K.S.A.R., 2019. Road Traffic Accident Statistics. Retrieved from https://www.td.gov.hk/en/road_safety/road_traffic_accident_statistics/2019/index.html

Truong, L.T., Currie, G., 2019. Macroscopic road safety impacts of public transport: A case study of Melbourne, Australia. *Accident Analysis and Prevention*, 132, 105270.

Ukkusuri, S., Hasan, S., Aziz, H.A., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. *Transportation Research Record* 2237 (1), 98-106.

Van Den Bossche, F., Wets, G., Brijs, T., 2005. Role of exposure in analysis of road accidents: a Belgian case study. *Transportation Research Record*, 1908 (1), 96-103.

Van Wee, B., 2016. Accessible accessibility research challenges. *Journal of Transport Geography*, 51, 9-16.

Varhelyi, A., 1997. Dynamic speed adaptation based on information technology-a theoretical background. University of Lund.

Verzosa, N., Miles, R., 2016. Severity of road crashes involving pedestrians in Metro Manila, Philippines. *Accident Analysis and Prevention*, 94, 216-226.

Wang, J., Huang, H., 2016. Road network safety evaluation using Bayesian hierarchical joint model. *Accident Analysis and Prevention*, 90, 152-158.

Wang, J., Huang, H., Zeng, Q., 2017a. The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian. *Accident Analysis and Prevention*, 98, 223-231.

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

Wang, L., Abdel-Aty, M., Wang, X., Yu, R., 2018. Analysis and comparison of safety models using average daily, average hourly, and microscopic traffic. *Accident Analysis and Prevention*, 111, 271-279.

Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis and Prevention*, 38 (6), 1137-1150.

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

Wang, X., Yang, J., Lee, C., Ji, Z., You, S., 2016b. Macro-level safety analysis of pedestrian crashes in Shanghai, China. *Accident Analysis and Prevention*, 96, 12-21.

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

Washington, S.P., Karlaftis, M.G., Mannering, F., 2010. *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC.

Waters, H., Hyder, A.A., Phillips, T., 2004. Economic evaluation of interventions to reduce road traffic injuries-a review of the literature with applications to low and middie-income countries. *Asia Pacific Journal of Public Health*, 16 (1), 23-31.

Wedagama, D.P., Bird, R.N., Metcalfe, A.V., 2006. The influence of urban land-use on non-motorised transport casualties. *Accident Analysis and Prevention*, 38 (6), 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. *Accident Analysis and Prevention*, 41 (1), 137-145.

Williams, M.S., Ebel, E.D., 2012. Methods for fitting the Poisson-lognormal distribution to microbial testing data. *Food Control*, 27 (1), 73-80.

Wolfe, A.C., 1982. The concept of exposure to the risk of a road traffic accident and an overview of exposure data collection methods. *Accident Analysis and Prevention*, 14 (5), 337-340.

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

Wong, S.C., Sze, N.N., Yip, H., Loo, B.P., Hung, W., Lo, H.K., 2006. Association between setting quantified road safety targets and road fatality reduction. *Accident Analysis and Prevention*, 38 (5), 997-1005.

Wong, S.C., Wong, C.W., Sze, N.N., 2008. Attitudes of public light bus drivers to penalties to combat red light violations in Hong Kong. *Transport Policy*, 15 (1), 43-54.

World Health Organization, 1996. Investing in health research and development: report of the ad hoc committee on health research relating to future intervention options. World Health Organization, Retrieved from https://apps.who.int/iris/handle/10665/63024.

World Health Organization, 2010. *Injuries and violence: the facts*. World Health Organization.

World Health Organization, 2017. Managing speed. World Health Organization, Retrieved from https://www.who.int/publications/i/item/managing-speed.

World Health Organization, 2018. Global status report on road safety 2018: Summary. World Health Organization.

Wundersitz, L., Hutchinson, P., 2008. *Identifying and improving exposure measures*. Centre for Automotive Safety Research.

Xie, S., Dong, N., Wong, S.C., Huang, H., Xu, P., 2018. Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure. *Accident Analysis and Prevention*, 121, 285-294.

Xiong, Y., Mannering, F.L., 2013. The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach. *Transportation research part B: Methodological*, 49, 39-54.

Xiong, Y.G., Tobias, J.L., Mannering, F.L., 2014. The analysis of vehicle crash injury-severity data: A Markov switching approach with road-segment heterogeneity. *Transportation Research Part B: Methodological*, 67, 109-128.

Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accident Analysis and Prevention*, 75, 16-25.

Xu, P., Huang, H., Dong, N., 2018. The modifiable areal unit problem in traffic safety: Basic issue, potential solutions and future research. *Journal of Traffic and Transportation Engineering-English Edition*, 5 (1), 73-82.

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. *Accident Analysis & Prevention*, 70, 110-120.

Xu, P., Huang, H., Dong, N., Wong, S.C., 2017. Revisiting crash spatial heterogeneity: a Bayesian spatially varying coefficients approach. *Accident Analysis and Prevention*, 98, 330-337.

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

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

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

Zegeer, C.V., Stutts, J.C., Huang, H., Zhou, M., Rodgman, E., 1993. Analysis of elderly pedestrian accidents and recommended countermeasures. *Transportation Research Record*, 1405, 56-63.

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

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

Zhou, M., Sisiopiku, V., 1997. Relationship between volume-to-capacity ratios and accident rates. *Transportation Research Record*, (1581), 47-52.

Zhu, D., Sze, N.N., Bai, L., 2021. Roles of personal and environmental factors in the red light running propensity of pedestrian: Case study at the urban crosswalks. *Transportation research part F: traffic psychology and behaviour*, 76, 47-58.

Zou, W., Wang, X., Zhang, D., 2017. Truck crash severity in New York city: an investigation of the spatial and the time of day effects. *Accident Analysis and Prevention*, 99, 249-261.