The Hong Kong Polytechnic University

Department of Civil and Environmental Engineering

Accident Analyses in Hong Kong: Accident Blackspot Identification, Casualty Injury Severity and Before-After Analysis

YE Daojun

A thesis submitted in partial fulfillment
of the requirements for the
degree of Master of Philosophy

March 2013

Certificate of Originality

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

YE Daojun

Abstract

Road safety is a major concern because of the social and economic costs resulting from traffic crashes. Numerous studies investigating traffic collisions and the resulting costs have been conducted. One important research topic in these studies is accident blackspot identification. The measures of accident frequency and accident rate are commonly adopted as identification criteria. However, identifying accident blackspots based solely on accident frequency or accident rate has been found, in practice, to be inefficient as neither can correctly reveal the extent of accident consequences. For example, high accident frequency does not necessarily result in a large number of fatalities or serious injuries. To supplement the knowledge gained by previous researchers, a new method to rank accident blackspots is proposed in this study.

In the proposed blackspot identification method, instead of accident frequency, accident consequences (in terms of injury or accident costs) are considered when identifying accident blackspots. The merit of the proposed method is that it takes not only number of injuries (or accident frequency) but also injury severity (or accident severity) into the consideration. To illustrate the proposed method, a case study was carried out using Hong Kong traffic accident data. The results indicate that adopting accident consequences, such as injury costs, can identify accident blackspots with higher injury costs but the methods of using accident frequency only or the Hong Kong Transport Department's blackspot definition may not be guaranteed.

In view of the importance of casualty injury costs as regards identification of accident blackspots, this feature is further investigated in this study. Focus is on an analysis of the effects of various contributory factors (categorized by environmental, site, and vehicle factors) on the injury severity of driver, passenger and pedestrian casualties and on accident costs in Hong Kong. Binary logistic regression model is

adopted to quantify the associations between injury severity of casualties and the contributory factors, while linear regression model is used for modeling the effects of contributory factors on accident costs. A Hong Kong traffic accident dataset for the whole territory from 2007 to 2009 is used in this study to estimate the coefficients of the regression models. The results of the regression models reveal that accident time, rain conditions, speed limits, traffic congestion levels, road types, and vehicle types significantly affect casualty injury severity and accident costs. Each contributory factor has a different effect or a different degree of effect on driver/passenger and pedestrian casualties.

A Before-After analysis is also used for investigation of accident effects on traffic speed in this study. Factors affecting the accident effects on traffic speed are firstly identified. Regression models are calibrated with empirical data to quantify the influences of factors on the degree of accident effects. A case study is carried out in which there is a total of 313 accidents occurred on a local urban area of Hong Kong during the study period: from September 2009 to December 2010 together with the corresponding speed data before and after the occurrence of these accidents. From the results of the case study, three factors, namely accident severity, accident time and accident location, were found to significantly affect the degree of accident effects on traffic speed.

Publications Arising from the Thesis

Journal paper:

1. Ye, D.J., Lam, W.H.K. and Tam, M.L., 2011, Automatic incident detection for urban roads in Hong Kong. *Journal of the Eastern Asia Society for Transportation Studies,* Vol. 9, pp. 1897-1912.

Conference paper:

2. Ye, D.J., Lam, W.H.K., Tam, M.L., Chen, W. and Ngai, E.W.T., 2011, Comparing the severity of different types of injuries at junction locations in Hong Kong. *Proceedings of 16th International Conference of the Hong Kong Society for Transportation Studies*, pp. 213-220.

Acknowledgments

I would like to express my deepest gratitude to my chief supervisor and mentor, Prof. William H. K. Lam, for his guidance, advice and patience throughout my study in Hong Kong. This thesis could not be completed without his encouragements and insightful inputs. Beside knowledge, he provides me self-motivated and structured working attitude that substantially influences the rest of my career.

I would like to express my sincere gratitude to my co-supervisor, Dr. Agachai Sumalee, for his guidance and constructive suggestions. I would also like to thank Dr. Mei Lam Tam and Mrs. Elaine Anson for providing comments and advises on my thesis.

Grateful appreciation should be extended to all research staffs in the transportation research group at the Hong Kong Polytechnic University not limited to: Dr. Zhongyi Wu, Dr. Bin Yu, Dr. Jiancheng Long, Dr. Wenzhu Zhou, Dr. Hua Wang. I also would like to acknowledge fellow research students: Dr. Biyu Chen, Dr. Renxin Zhong, Lianqun OuYang, Ding Liu, Yuqing Zhang, Yiliang Xiong, Xiao Fu, Jianting Cong and Lingyun Yuan.

Above all, I am deeply thankful to my family for their endless love, care, and support in no matter what I choose to pursue in life.

Table of Contents

List of Tables

List of Figures

Notation

$AC_{F,T}$	The estimated accident cost per fatal accident during time period T
$AC_{Se,T}$	The estimated accident cost per seriously injured accident during time period T
$AC_{Sl,T}$	The estimated accident cost per slightly injured accident during time period T
$C_{S,T}$	The cost of injury or accident with specific severity S (F represents fatality or fatal accident, Se represents serious injury or seriously injured accident and Sl represents slight injury or slightly injured accident) during time period T
$\overline{C}_{S,T}$	The estimated cost of injury or accident with specific severity S during time period T
$F_{F,T}^{F}$	The number of fatalities in all fatal accidents in Hong Kong during time period T
$F_{Se,T}^{F}$	The number of serious injuries in all fatal accidents in Hong Kong during time period T
$F_{Sl,T}^{F}$	The number of slight injuries in all fatal accidents in Hong Kong during time period T
$F_{Se,T}^{Se}$	The number of serious injuries in all seriously injured accidents in Hong Kong during time period T
$F_{Sl,T}^{Se}$	The number of slight injuries in all seriously injured accidents in Hong Kong during time period T
$F_{Sl,T}^{Sl}$	The number of slight injuries in all slightly injured accidents in Hong Kong during time period T

$F^{J}_{F,T}(d)$	The number of fatalities at junction location d during time period T
$F^{J}_{Se,T}(d)$	The number of serious injuries at junction location d during time period T
$F^{J}_{Sl,T}(d)$	The number of slight injuries at junction location d during time period T
$F^{NJ}_{F,T}(d)$	The number of fatalities at non-junction location d during time period T
$F^{NJ}_{Se,T}(d)$	The number of serious injuries at non-junction location d during time period T
$F^{NJ}_{Sl,T}(d)$	The number of slight injuries at non-junction location d during time period T
G	The number of common blackspots in the two comparison datasets
$g(x)$	The latent dependent variable in binary regression model
H_L	The Hosmer-Lemeshow statistic
$IC_{F,T}$	The estimated injury cost per fatality during time period T
$IC_{Se,T}$	The estimated injury cost per serious injury during time period T
$IC_{Sl,T}$	The estimated injury cost per slight injury during time period T
i	The index of the observation
$L(d)$	The length (in meters) of non-junction location d
L_U	The unit distance adopted (in meters)
m	The number of independent variables or contributory factors in binary regression model and linear regression model
$N_{F,T}$	The number of all fatal accidents in Hong Kong during time period T

$N_{Se,T}$ The number of all seriously injured accidents in Hong Kong during time period T

$N_{Sl,T}$ The number of all slightly injured accidents in Hong Kong during time period T

$N^{J}_{F,T}(d)$ The number of fatal accidents at junction location d during time period T

$N^{J}_{Se,T}(d)$ The number of seriously injured accidents at junction location d during time period T

$N^{J}_{Sl,T}(d)$ The number of slightly injured accidents at junction location d during time period T

$N^{NJ}_{F,T}(d)$ The number of fatal accidents at non-junction location d during time period T

$N^{NJ}_{Se,T}(d)$ The number of seriously injured accidents at non-junction location d during time period T

$N^{NJ}_{Sl,T}(d)$ The number of slightly injured accidents at non-junction location d during time period T

n_k The number of observations in the kth group

p_d The percentage deviation value of blackspots identified based on different methods

T The total number of blackspots in each dataset

$TAC^{J}_{T}(d)$ The total accident costs at junction location d during time period T

$TAC^{NJ}_{T}(d)$ The total accident costs (per unit distance) at road segment d during time period T

$TIC^{J}_{T}(d)$ The total injury costs at junction location d during time period T

$TIC_T^{NJ}(d)$	The total injury costs (per unit distance) at road segment *d* during time period *T*
u	The estimated road speed in the comparison group in the 'after' period
$\mathbf{X_T}$	The vector of factors to be considered in injury or accident cost regression model
x_j	The *j*th independent variable which is the contributory factor
y_i	The *i*th observed outcome
$\boldsymbol{\beta}$	The vector of corresponding coefficients to be calibrated in injury or accident cost regression model
β_j	The corresponding coefficient of the variable (for j = 1, 2, 3, …., *m*) in binary regression model and linear regression model
δ	The difference between "the road speed of the area where the accident did happen" and "what would have been the road speed of the area had the accident not happened" in the 'after' period
θ	The ratio of "the road speed of the area where the accident did happen" to "what would have been the road speed of the area had the accident not happened" in the 'after' period
$\theta_{S,T}$	The weighted factors for injury or accident with specific type of severity *S* (*F* represents fatality or fatal accident, *Se* represents serious injury or seriously injured accident and *Sl* represents slight injury or slightly injured accident) during time period *T*
κ	The estimated road speed in the accident group in the 'before' period
λ	The estimated road speed of the area where the accident did happen in the 'after' period

μ	The estimated road speed in the comparison group in the 'before' period
$\pi(x)$	The conditional probability of the outcome
$\hat{\pi}_i$	The predicted probability for the *i*th observation
π	The predicted road speed of the area in the 'after' period had the accident not happened
$\sigma_{S,T}$	The standard deviation of the cost of injury or accident with specific type of severity S during time period T

The following acronyms are used throughout this thesis:

ATC	Annual Traffic Census
AVI	Automatic vehicle identification
CTS-3	The Third Comprehensive Transport Study
GDP	Gross Domestic Product
JTIS	Hong Kong Journey Time Indication System
PDF	Probability distribution functions
RoCIS	Road Casualty Information System
TAC	Total accident costs
TD	Transport Department
TIC	Total injury costs
TRADS	Hong Kong Traffic Accident Database System

1. Introduction

1.1. Need for the Study

Road safety is of major importance because of the economic and social costs of traffic safety failure in the form of crashes. According to Hong Kong Transport Department road traffic accident statistics, 18,138 casualties resulted from 14,316 police-reported traffic accidents in 2009. In addition, 139 people were killed and the fatality rate per 1,000,000 people was 20 in that year. This Hong Kong rate is comparatively lower than that in most overseas cities of comparable size, for that year (for example, 35 fatality/1,000,000 people in Singapore). However, these statistics still raise awareness that remedial policies or measures should be conducted to improve the road safety. One major reason is that a large number of accidents produce huge economic costs, such as productivity losses, property damage, medical costs, rehabilitation costs and travel delays (Lanrence, 1995).

As indicated above, efficient policies and treatments should be implemented to improve road safety. To begin to achieve this, an accident analysis should be conducted, such as the temporal and spatial distributions of these accidents, to support policy making. In this respect, one valuable study is the identification of traffic accident blackspots. Traffic accident blackspots (accident hot spots, hazardous locations, or sites with promises) are the locations where the level of risk is higher than those in nearby areas with average levels of risk (Geruts et al., 2004; Montella, 2010). The objective of accident blackspot identification is to prioritize locations with a strong need of remedial action (Hauer, 1996). Correct identification of blackspots therefore can be considered a first and essential step in the improvement of accident prone conditions in these locations. Incorrect blackspot identification, however, may result in inefficient use of resources for safety improvements and may reduce global effectiveness of the safety management process (Montella, 2010).

In addition to accident blackspot identification, casualty injury severity is another key issue intensively investigated in the literature (Kockelman and Kweon, 2002; Savolainen and Mannering, 2007; Sze and Wong, 2007). In some studies, injury severity is adopted as a measurement to evaluate the road safety and its improvements after treatments. Although fatal accidents occupy a comparatively small percentage of total accidents, they have more significant impacts than other accidents. Thus, another clear reason is provided for the provision of greater insight into the effects of various contributory factors on casualty injury (or accident) severity so that appropriate countermeasures can be implemented to reduce the occurrence of fatal and seriously injured accidents.

Apart from causing injuries and fatalities, the occurrence of traffic accidents also adversely affects traffic speed and causes prolonged travel delays on roadways (Sethi et al., 1995). When a road accident occurs, the traffic speed would suddenly be decreased due to the reduction in roadway capacity. The degree of traffic speed reduction may depend on several factors, such as the accident severity, accident time and accident duration etc. Thus there is a need to fully investigate in what ways accidents impact traffic conditions particularly in terms of vehicular speed. The identification of factors which have influences on the magnitude of accident effects deserves further research.

1.2. Objectives of the Study

This study focuses on comprehensive statistical investigations of traffic accidents. It is aimed to provide some insightful findings of traffic accidents in Hong Kong based on the statistical analysis of the relevant data. Some possible road safety improvement policies may be proposed based on these insightful findings in order to reduce accident frequency, severity and accident effects on traffic speed. The specific associated objectives of this study are listed as follows:

Objective 1: to propose a traffic accident blackspot identification method with consideration of both injury (or accident) severity and cost.

Objective 2: to propose regression models to model the effects of contributory factors on the injury severity of different casualties and accident costs.

Objective 3: to statistically analyze accident effects on traffic speed and to propose regression models to quantify the associations between accident effects and some contributory factors such as accident severity.

1.3. Structure of the Thesis

The interrelationship of objectives and the structure of the thesis are presented in Figure 1.1. The thesis consists of four sections. The first section (Chapters 1 and 2) gives a brief introduction and relevant literature review of traffic accident blackspot identification, casualty injury severity modeling and accident effects on traffic speed. The second section (Chapter 3) describes the data used in this study.

The third section (Chapters 4, 5 and 6) presents a method to identify accident blackspots based on accident or injury costs, models to quantify the associations between casualty injury severity and accident costs and contributory factors, and a Before-After analysis of accident effects on traffic speed. Case study results or modeling results are also shown and discussed in this section. The forth section (Chapter 7) gives a summary of this study and recommendations for further research.

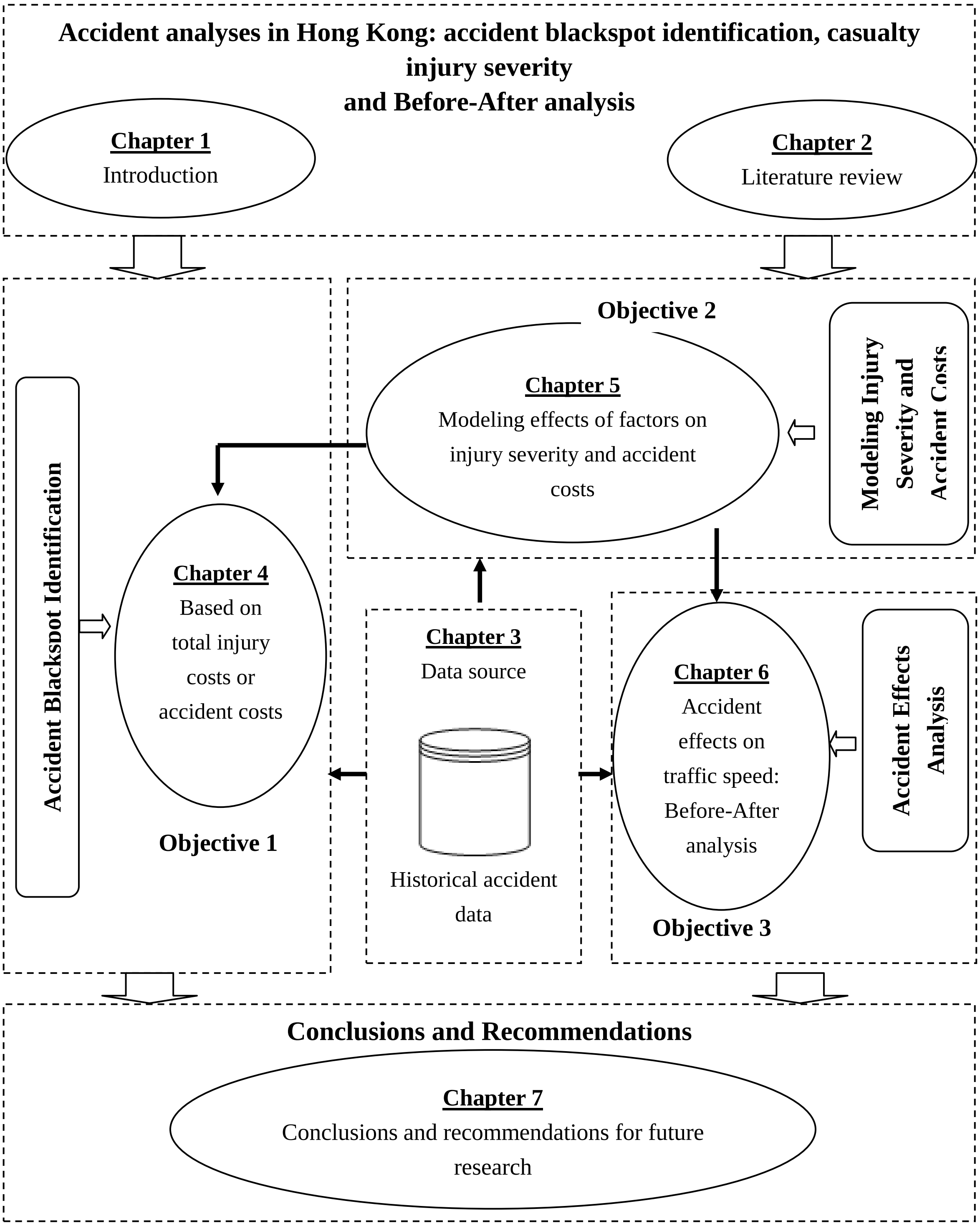


Figure 1.1 The interrelationship of objectives and the structure of the thesis

Specifically, in Chapter 4, an accident blackspot identification method is proposed. In the proposed method, the measure of accident consequences is adopted instead of accident frequency for accident blackspot identification. To illustrate the performance of the proposed method, comparison studies are carried out between the results identified based on accident frequency and the proposed method.

Chapter 5 presents regression models to statistically quantify the associations between injury severity of different casualties (driver, passenger and pedestrian) and accident costs and contributory factors. These contributory factors are classified into three categories, environmental, site, and vehicle factors. The modeling results are also discussed and compared in Chapter 5.

Chapter 6 describes an analysis of accident effects on traffic speed using Before-After analysis method. Factors that would affect the degree of accident effects are identified. Regression models are adopted to quantify the influences of these contributory factors on the accident effects.

2. Literature Review

Literature review on related studies is presented in this chapter. The chapter is structured as follows. Section 2.1 reviews the existing literature about accident blackspot identification and summarizes the methods for identification of accident blackspots. A review of regression models applied to analyze accident or casualty injury severity is given in Section 2.2. The factors that have been considered in these regression models are also summarized and discussed in Section 2.2. Section 2.3 reviews the existing literature related to the investigation of accident effects on traffic conditions such as vehicular speeds, together with the studies on Before-After analysis. A summary of the reviewed literature is given in Section 2.4.

2.1. Accident Blackspot Identification

In view of the importance of correct identification of blackspots, numerous methods have been proposed to identify blackspots in the literature. These methods are divided into two categories according to the accident measures adopted to identify blackspots. The first category is observed accident frequency (or accident rate) based, while the second is expected accident frequency (or accident rate) based. A brief description of these two categories of methods is given as follows.

The first category of method is designed to identify blackspots based on a measure derived directly from the observed accident frequency or accident rate during a given time period. The accident rate is calculated as accidents per vehicle-kilometre for road segments or per entering vehicles for junction locations. Table 2.1 summarises the identification criteria of those proposed methods based on accident measures of observed accident frequency and accident rate reported in the literature.

Table 2.1 Summary of blackspot identification criteria of the methods based on observed accident frequency and accident rate

Literature	Criteria
Chung et al., 2011	Accident rate (accidents per unit distance) over **75th percentile**
Elvik, 2008	Accident frequency ranked in the top of **1% or 5%**
Loo, 2009	Accident rate (accidents per 100 meters) over **6, 8, or 10**
Stokes and Mutabazi, 1996	Accident rate over **critical rate**
Definition of blackspot in Hong Kong (Transport Department, 2001)	More than **9 accidents** or **6** accidents involved with **pedestrian casualties**

However, in the previous studies, the identification of blackspots based solely on observed accident frequency or accident rate has been found to be inefficient (Hauer, 1980; Maher and Mountain, 1988; Saccomanno et al., 2004). This is because accidents are rare random events that change among locations and over time. For instance, high collision frequency in one year at a location does not necessarily indicate high accident frequency in the next year at this location. The randomness of accident occurrence results from the effects of many factors, such as driver behavior, traffic conditions, road geometry and vehicle and environmental conditions (Miranda-Moreno, 2006). The observed accident frequency or accident rate, however, cannot reflect the effects of randomness of accident occurrence on blackspot identification. Identifying blackspots based on observed measures, therefore, would result in many false positives (claiming a site is unsafe when it is not) and false negatives (failing to claim a site is unsafe when it is).

Given the limitation of the methods based on observed accident frequency or accident rate, increasing attention has been paid to the development of the second category of method based on expected accident frequency or accident rate. These expected measures are estimated by using statistical models ranging from basic Poisson models to negative binomial models and zero-inflated negative binomial models. The observable and unobservable variations attributed to the abovementioned factors can be considered in the statistical models (Miranda-Moreno, 2006). The expected accident frequency or accident rate, therefore, is believed to more accurately reflect the expected risk level of a specific location in a given time period. The identification results based on expected accident frequency or accident rate may be better than those based on observed ones.

Although much attention has been paid to the development of identification methods based on accident frequency, few studies have considered the factor of accident severity in their proposed methods. Identification of blackspots based on accident frequency assumes that accidents with different severities would produce identical consequences. However, in practice, a collision resulting in a fatality may cause much more severe consequences than the one with slightly injured casualties. It can be assumed, therefore, that accident frequency cannot correctly reflect the consequences caused by accidents with different severities.

Some studies, such as Geurts et al. (2004), have attempted to consider the effects of accident severity on the identification of blackspots. In this published work, different weighted factors have been assigned to accidents with different severities to calculate priority value to rank blackspots. However, the weighted factors are subjectively determined by the decision makers, therefore, may not be adequately reveal the actual consequences caused by accidents with different severities. The weighted factors can be replaced by injury or accident costs, if injury or accident cost data is available, as either of them may be more accurately reveal accident consequences.

However, in practice, the injury or accident cost data is difficult to obtain as the traffic accident database is not linked to the hospital data in most cities.

2.2. Casualty Injury Modeling

In the literature, with the aim of improving road safety, many researchers have focused their attention on two types of studies. The first type focuses on the investigation of associations between certain factors, such as road geometrics, and accident frequency (Karlaftis and Golias, 2002; Lee et al., 2010; Sharkar et al., 1995; Wang et al., 2009). These studies aim to identify factors that have significant effects on accident frequency and therefore open the need for the generation of policies that can reduce the occurrence of accidents. The second type of study is the investigation of accident severity. The study concerning accident severity focuses not only on the prevention of accidents but also more specifically on the reduction of accident severity (Al-Ghamdi, 2002). In these studies, predictive models have been developed to examine and identify factors that affect accident severity (or injury severity) (Abdel-Aty, 2003; Barua and Tay, 2010; Nowakowska, 2010; Savolainen and Mannering, 2007; Sze and Wong, 2007; Tay and Rifaat, 2007; Yau, 2004; Yau et al., 2006).

Abdel-Aty (2003) adopted ordered probit models to estimate the association between certain contributory factors and driver injury severity levels at multiple locations. In the study of Abdel-Aty (2003), it was found that female and elderly drivers had a higher probability of being seriously injured. In addition, speeding and failing to wear a seat belt were found to increase the injury severity. Crashes at horizontal curves and those occurring in rural areas were also more likely to produce serious injuries. Crashes occurring during daytime, however, were found to be less injurious.

In the study of Savolainen and Mannering (2007), nested logit models and standard multinomial logit models were developed to predict the probability of motorcyclists' injury severity in single- and multi-vehicle crashes. It was found that the elderly are most likely to suffer fatalities and serious injuries during accidents. Additionally, this study concluded that the influence of alcohol, speeding, and collision with roadside objects, trees and poles resulted in an increase in the probability of a fatality. The use of helmets, however, reduced the severity of motorcyclists' injuries.

Yau (2004) conducted a population-based case-control study to examine factors affecting the severity of single-vehicle traffic accidents in Hong Kong. Three logistic regression models were developed to identify unique risk factors, associated with each of three types of vehicles (private vehicles, goods vehicles and motorcycles). In the study of Yau (2004), it was concluded that male drivers have an increased likelihood of death or serious injuries. Accidents involving aged vehicles have a higher risk of resulting in fatalities or serious injuries. Light rain was also found to be associated with a decreased risk of fatal or serious accidents, except in cases where the light rain falls on dry dusty road surfaces, causing aquaplaning.

In the study of Sze and Wong (2007), binary logistic regression models were developed to evaluate the injury risk to pedestrians and to identify factors contributing to fatalities and serious injuries. From the model results, it was concluded that the risk of fatalities and serious injuries was lower when crashes occurred at junction locations with traffic controls other than traffic signals. In addition, there is less likelihood of serious injury from crashes occurring on road sections with severe or moderate congestion or single lane or two lane roads. However, crashes that occur on road sections with a speed limit higher than 50 km/h or on dual or multiple carriageways have significantly higher injury risk.

Tay and Rifaat (2007) adopted an ordered-probit model to identify the risk factors that may affect the injury severity of accidents. It was also concluded that crashes which occur at night are found to be increasingly severe. The fatality risk of motorcycle related accidents and large-size vehicle (such as bus and truck) related accidents are significantly higher than accidents involving other types of vehicles.

Other studies (such as those of Kockelman and Kweon, 2002; Krull et al., 2000; Toshiyuki and Shankar, 2004) have similar findings. For instance the crashes of male drivers were less severe and alcohol use significantly increased the severity level of injuries. Krull et al. (2000) found that accidents which occurred on dry pavement were more severe than those occurring on slippery pavement. Similar findings were obtained from the study of Toshiyuki and Shankar (2004). They found that icy roadway surfaces and rain decrease the probability of more severe driver injuries.

2.3. Accident Effects on Traffic Conditions

2.3.1 Accident Detection

In the literature, many studies have been conducted to investigate accident effects on traffic conditions. The accident effects are usually measured as a reduction in traffic speed or an increase in occupancy at upstream, or an increase in travel times on the road segment where the accident occurred, etc. Most of these studies focused on automatic accident detection based on the degree of accident effects on traffic conditions. Different measures of accident effects have been adopted in various studies.

Discriminant analysis was adopted to classify accidents and non-accidents using various traffic data, such as travel time and occupancy (Sethi et al., 1995). The changes in travel time and occupancy are adopted to quantify the accident effects on traffic conditions in the study of Sethi et al. (1995). Two effective algorithms were proposed to detect accidents in heavy traffic flow conditions using traffic speed and occupancy to measure magnitude of accident effects (Mak and Fan, 2006). It was found that the pre-accident traffic flow conditions have a significant effect on the magnitude of accident effects. A threshold-based algorithm was proposed to detect accidents on a selected path in Hong Kong (Lam et al., 2008). In their study, the accident effects are measured as difference between estimated and predicted travel times calculated based on real-time automatic vehicle identification (AVI) data and off-line estimates. As an extension of the study of Lam et al. (2008), Ye et al. (2011) adopted both reduction in traffic speed and increase in travel times to represent accident effects, and to detect accidents based on the degree of accident effects.

All the studies mentioned above focused mainly on detecting the occurrence of accidents using accident effects as detection criteria. However, few studies have been conducted to specifically investigate how accidents affect traffic conditions. The degree of accident effects on traffic conditions would vary in accordance with accident severity, accident time and accident location. Further analysis of the specific accident effects and factors influencing accident effects would help to better understand their subsequent influences on traffic conditions and would contribute to an improved accident detection algorithm. In this study, traffic conditions are measured by traffic speed as this type of data can be obtained from the Hong Kong Journey Time Indication System (JTIS) which is described in details in Chapter 6. An extension of investigation of accident effects on traffic flow or occupancy is an interesting research topic for further study if these data are available.

2.3.2 Road Safety Before-After Studies

Before-After analysis method was firstly introduced to study road safety by Hauer

(1997). Since its introduction, many studies have been conducted to improve this method and to apply it for analysis of several road safety problems.

Wong et al. (2005) applied the Before-After analysis to study the effects of the changed speed limits on road safety for major roadways in Hong Kong. In the majority of the treatment sites, the accident counts was found to be worse after increasing speed limits, both as regards fatal, seriously injured and slightly injured accidents.

The implementation effectiveness of marking blue cycle crossings has been evaluated with the use of Before-After analysis method (Jensen, 2008). The area of conflict between motor vehicles and cyclists at the junction area is marked blue in order to draw more attention when crossing the junction area. In this published work, the number of accidents in 65 signalized junctions before and after marking blue cycle crossings was compared. It was concluded that the effects of implementing marking blue cycle crossings depend on the number of blue cycle crossings at the junction area.

In the study of Dommes et al. (2012), Before-After analysis method was adopted to assess the effectiveness of a training programme for older pedestrians after the implementation of the training programme. Dommes et al. (2012) also aimed at examining whether or to what extent age-related differences in street-crossing safety could be reduced after the training programme.

Luk et al. (2001) compared the upstream and downstream traffic speed before and after accidents occurring on the urban arterial roads. It was found that the occurrence of accidents would result in a reduction in upstream traffic speed, while an increase in downstream traffic speed. As an extension of the study of Luk et al. (2001), it is of interest to analyze the influences of factors on accident effects on traffic speed using

Before-After analysis.

2.4. Summary

In this chapter, the research background and works related to this study have been reviewed and discussed. The review deals with three research topics: 1. accident blackspot identification, 2. casualty injury modeling and 3. accident effects on traffic conditions.

In order to correctly identify traffic accident blackspots, many conducted studies have proposed numerous identification methods. In general, these methods can be grouped into two categories, 1. observed accident frequency (or accident rate) based and 2. expected accident frequency (or accident rate) based. However, the identification of accident blackspots based solely on accident frequency (or accident rate) has been found to have some limitations. For example, accident severity, an important accident feature, has not been considered in these existing identification methods. To overcome this limitation, some studies, such as Geurts et al. (2004), have proposed a method which assigns different weighted factors to accidents with different severities. This method can be easily extended if injury or accident cost data is available to replace the weighted factors.

A number of models to examine the factors that affect accident severity (or injury severity) have been identified in the literature review. Several factors, such as the sex of the casualty, age of the casualty, accident time, accident location, traffic control, speed limit, speeding, weather condition, whether wearing a seat belt or not, vehicle size and influence of alcohol, have been found to have significant influence on the accident or casualty injury severity.

Before-After analysis method has been applied in the analysis of several road safety problems since its introduction by Hauer (1997). These studies include studying the effects of the changed speed limits on road safety, evaluating the effectiveness of marking blue cycle crossings, assessing the effectiveness of a training programme for older pedestrians etc.

Based on previously related work recorded in this chapter, an effective accident blackspot identification method is proposed and given in Chapter 4. This method takes either injury costs or accident costs into consideration rather than having a sole reliance on accident frequency or accident rate. In Chapter 5, a binary regression model is proposed to compare the factors that affect the injury severity of different types of casualties: 1. driver casualties, 2. passenger casualties and 3. pedestrian casualties. The Before-After analysis method is applied to analyze accident effects on traffic speed in Chapter 6. Regression models are further proposed to quantify the influences of factors on the degree of accident effects.

3. Data Descriptions

The focus of this chapter is the preliminary statistical analysis of Hong Kong traffic accidents occurring during the years 2007, 2008 and 2009. Data for these years are extracted from the Hong Kong Traffic Accident Database System (TRADS). This set of data is later used for a blackspot identification case study presented in Chapter 4, to estimate the parameters of casualty injury severity models and accident costs model described in Chapter 5, and to analyze accident effects on traffic speed in Chapter 6. In the remainder of this chapter, details of TRADS are given, followed by descriptive statistics of accident data in the years 2007, 2008 and 2009.

3.1. Hong Kong Traffic Accident Database System (TRADS)

The Hong Kong Traffic Accident Database System (TRADS) is updated by the Hong Kong Police Force and the Transport Department (TD) of the Hong Kong Government. TRADS consists of three components: 1. crash environment profile, 2. casualty injury profile and 3. vehicle involvement profile. These three profiles are compiled according to specific accident reference numbers and inserted into TRADS (See sample data in Appendix).

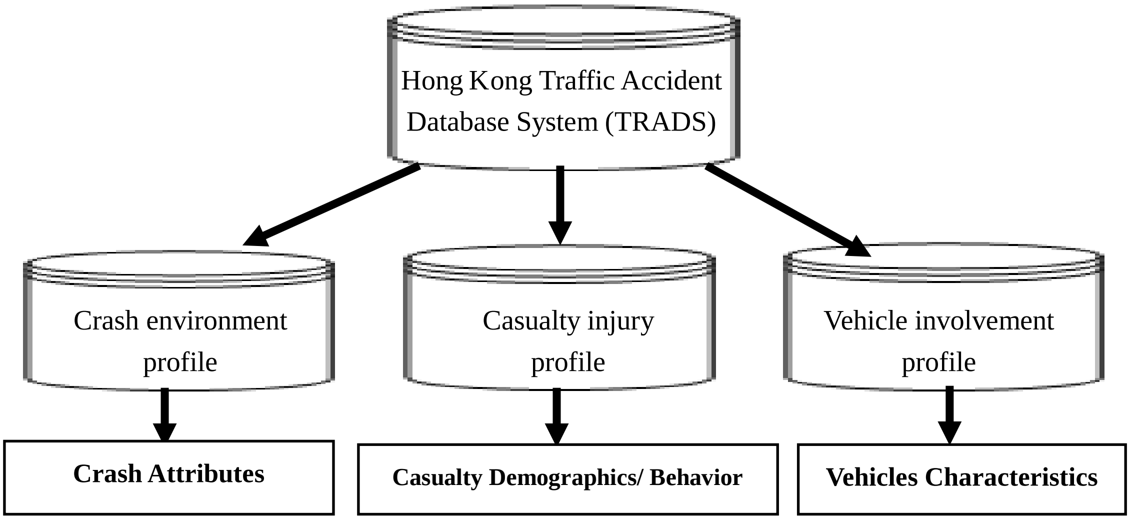


Figure 3.1 Hong Kong Traffic Accident Database System (TRADS)

The crash environment profile illustrates precisely the following accident features: accident severity, date and time of accident occurrence, the precise accident location, the number of vehicles and the number of casualties involved, weather condition, speed limit at the accident location, traffic condition, road surface condition, road type, junction type and the type of junction control.

The casualty injury profile provides records of the following: casualty injury severity (injuries are divided into three categories in TRADS: fatality, serious injury and slight injury), casualty age, casualty sex, casualty type (whether the casualty is driver, passenger or pedestrian), whether a seat belt or crash helmet was worn, injury location, vehicle seat position of the casualty and pedestrian location.

The vehicle involvement profile indicates details of all vehicles involved in the accidents. These details include vehicle class, vehicle age, the objects in collision with the vehicle, vehicle lighting condition, area of vehicle damaged, vehicle tyre condition, and other information regarding the involved vehicles.

Accidents are classified into three categories in TRADS: (a) fatal accident, (b) seriously injured accident and (c) slightly injured accident. The definition of these categories are as follows: (a) fatal accident is an accident in which one or more persons die within 30 days after the accident; (b) seriously injured accident describes a non-fatal accident in which one or more persons are detained in hospitals for more than twelve hours; (c) slightly injured accident describes an accident which involves personal injuries, other than the above.

3.2. Descriptive Statistics

In this sub-section, accident and casualty statistics for the years 2007, 2008 and 2009 are presented and discussed. Figures 3.2 and 3.3 show the number of accidents in terms of accident severity and location. Figures 3.4-3.6 depict the distribution of casualties by injury severity and casualty type (e.g. driver, passenger and pedestrian).

Accident Statistics

Figure 3.2 shows the number of accidents by severity for the years 2007, 2008 and 2009. For the year 2007, of the accidents recorded, 153 were fatal, 2376 involved seriously injured casualty and 12786 involved slightly injured casualty. Accidents involved slightly injured casualty were more than 5 times of the number of accidents involved fatality and seriously injured casualty. It is also observed that the number of all categories of accidents decreased continuously between the years 2007 to 2009, indicating an improvement in overall traffic safety conditions in Hong Kong during this period.

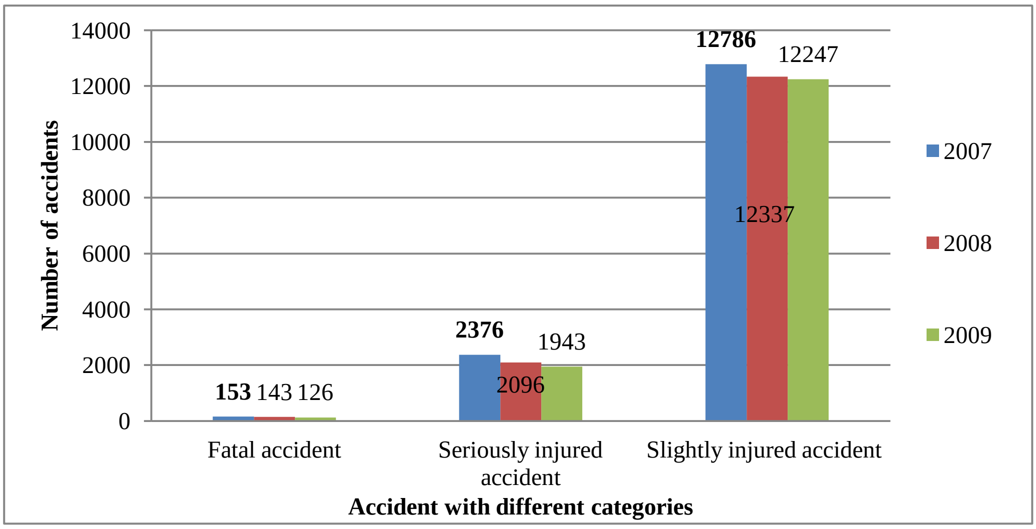


Figure 3.2 The number of accidents by accident category for the years 2007, 2008 and 2009

In Figure 3.3, it is noted that the number of accident at non-junction locations is twice than that at junction locations. One possible reason for this statistic is that road segment areas on road networks in Hong Kong have a greater coverage rate than junction areas and hence more accident records.

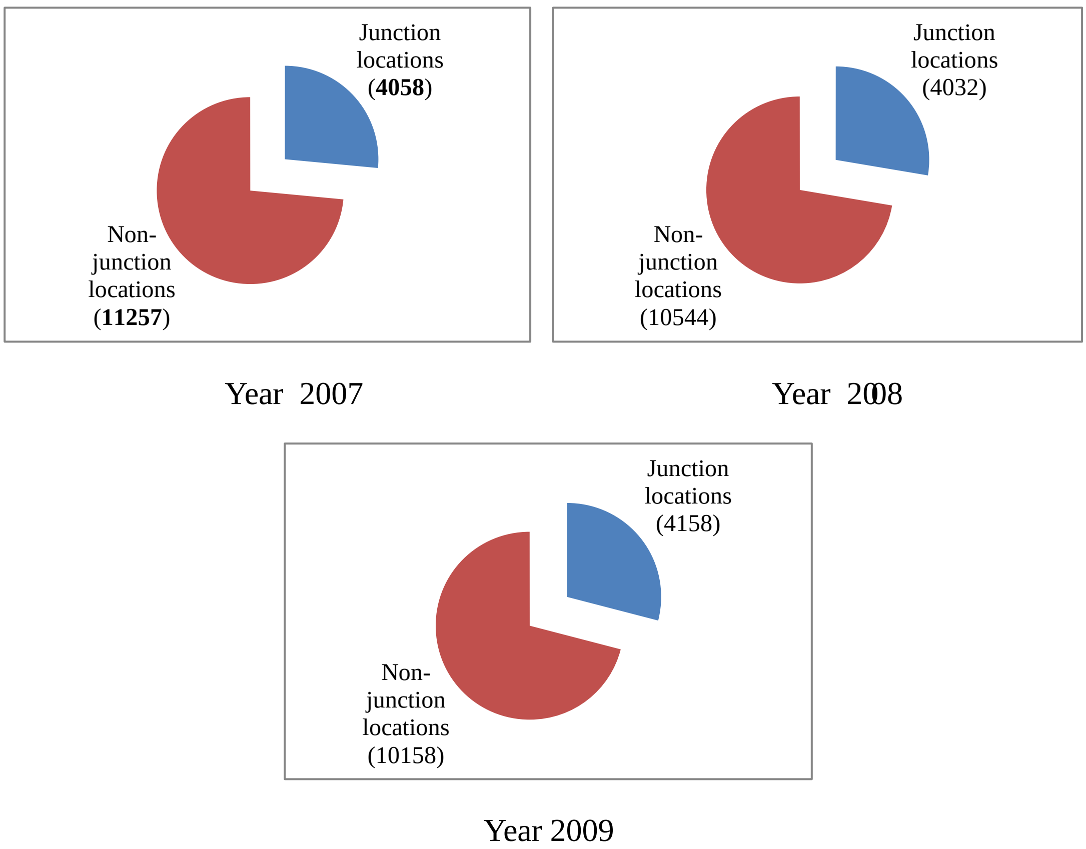


Figure 3.3 The number of accidents by location for the years 2007, 2008 and 2009

Casualty Statistics

Figure 3.4 gives the distribution of casualties with different degrees of injury severity for the years 2007, 2008 and 2009. For each of these three years, about 300 fatalities, 3000 seriously injured casualties and 15000 slightly injured casualties were recorded. Over these years, the trend, as regards seriously injured and slightly casualties, is seen to have reduced, while the trend in the number of fatalities is seen to have risen.

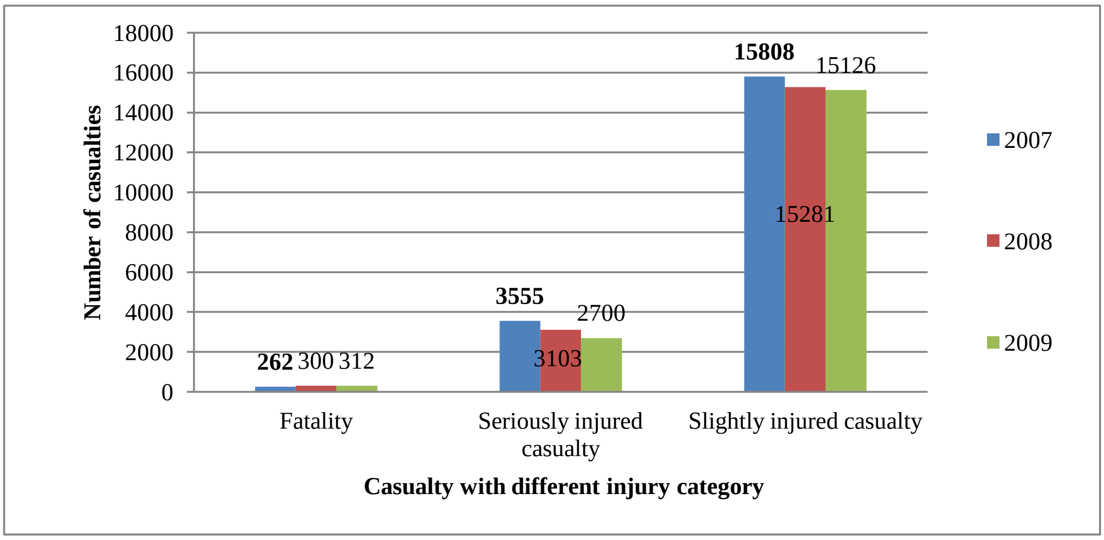


Figure 3.4 The number of casualties by injury category for the years 2007, 2008 and 2009

Figure 3.5 displays a general picture of casualty distributions by casualty type for the years 2007, 2008 and 2009. It can be seen that, in 2007, more than 19000 people were injured or killed as a result of accidents. Of these, 4078 were drivers, 6664 were passengers and 8883 were pedestrians. The number of driver casualties is greater than those of passengers and pedestrians.

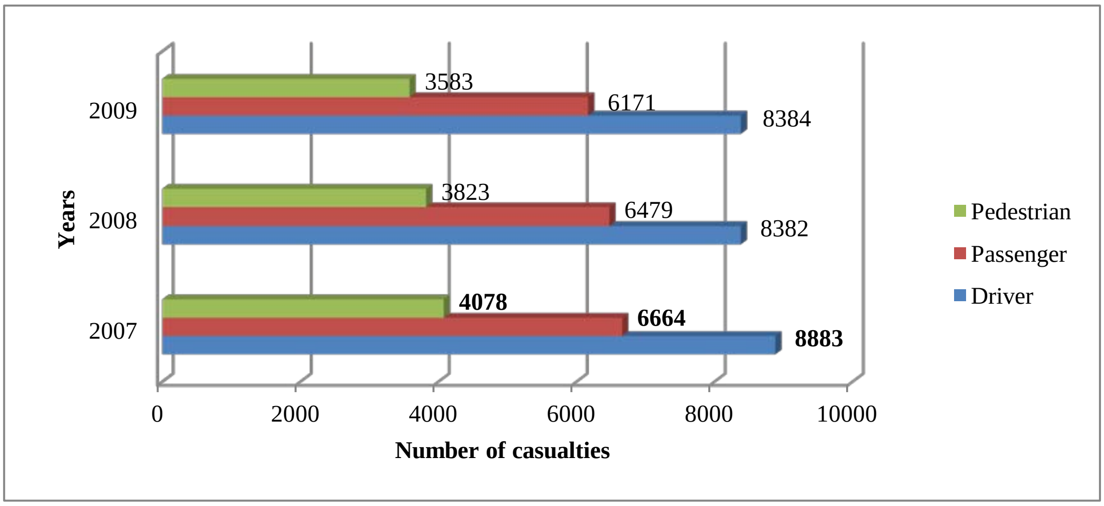


Figure 3.5 The number of casualties by casualty type for the years 2007, 2008 and 2009

Table 3.1 shows more details on the distribution of casualties by injury severity and casualty type. Among all of the fatalities for the years 2007, 2008 and 2009, more than half were pedestrians. However, drivers accounted for 22510 (45.9%) among the entire slightly injured casualty list. For the percentage of fatalities and seriously injured casualties, pedestrians had more than 20% (2.2%+20.8%), while driver and passenger had about 10% (0.5%+11.7% for driver and 0.4%+7.9% for passenger). From all these figures, it can be concluded that pedestrians are more likely to suffer severe injuries compared with vehicle occupants.

Table 3.1 Casualty by injury category and casualty type

		Injury category					
		Fatality		Seriously injured casualty		Slightly injured casualty	
		Count	Column Percentage	Count	Column Percentage	Count	Column Percentage
Casualty type	Driver	139	30.2%	3000	43.4%	22510	**45.9%**
	Passenger	72	15.6%	1522	22.0%	17720	36.1%
	Pedestrian	250	**54.2%**	2388	**34.6%**	8846	18.0%
		Count	Row Percentage	Count	Row Percentage	Count	Row Percentage
Casualty type	Driver	139	0.5%	3000	11.7%	22510	87.8%
	Passenger	72	0.4%	1522	7.9%	17720	91.7%
	Pedestrian	250	**2.2%**	2388	**20.8%**	8846	77.0%

3.3. Summary

The Hong Kong Traffic Accident Database System (TRADS) has been introduced and discussed in this chapter. The accident and casualty data for the years 2007, 2008 and 2009 have been extracted from TRADS and compared. The number of accidents by accident severity and location in different years have been given and discussed. The number of casualties described according to injury severity and casualty type has also been presented.

In Chapter 4, an accident blackspot identification method is proposed. To illustrate the proposed method, a case study using the extracted accident and casualty data for the years 2007, 2008 and 2009 for the whole Hong Kong territory is conducted and the results of the case study are also presented and discussed in Chapter 4.

4. Identifying Accident Blackspots with Consideration of Costs and Severity

4.1. Introduction

As indicated in Chapter 2, it was found that identification of accident blackspots, in practice, is based solely on accident frequency or accident rate. However, accident frequency or accident rate may not correctly reflect the total risk involved at each location. It is commonly believed that a collision resulting in a fatality may have greater economic costs than a casualty with slight injury, but the levels of accident severity have seldomly been considered in the identification of accident blackspots. Geurts et al. (2004) have attempted to consider accident severity using weighted factors assigned to accidents with different severities. Either injury costs or accident costs may be more accurately reveal accident consequences than weighted factors which are subjectively determined by the decision makers. However, as mentioned in Chapter 2, in practice, the injury or accident cost data is difficult to obtain as the traffic accident database is not linked to the hospital data in most cities.

In this chapter, an accident blackspot identification method to address the above situation is proposed. The proposed method is based on two alternative measures of accident consequences: 1. total accident costs or 2. total injury costs. The accident costs or injury costs data are obtained from the Third Comprehensive Transport Study (CTS-3) which is described in details in Section 4.2 of this chapter.

The total accident costs represent the total costs caused by all categories of accidents occurring at a location. Usually accidents are classified into different categories according to accident severity. As mentioned in Chapter 3, in Hong Kong, accidents are classified into three categories, fatal accident, seriously injured accident and slightly injured accident. It is proposed that the total accident costs are calculated as the weighted sum of costs of different categories of accidents. The merit of adopting total accident costs is that both accident severity and accident frequency are taken into consideration for identifying blackspots so as to reflect the consequences resulting from collisions at each location.

The effects of the number of injuries have not been considered when the total accident costs are adopted for identification of blackspots. The estimated cost of a specific category of accident is assumed to be a fixed value no matter the number of casualties involved in accidents. One accident, however, may result in several casualties with different categories of injuries. As indicated in Chapter 3, in Hong Kong, the injuries are classified into three categories, fatality, serious injury and slight injury. To overcome this problem, the total injury costs are proposed as another measure to reflect accident consequences. The total injury costs are quantified as the total consequences (in terms of costs) of all casualties with different categories of injuries at a location. The measure of total injury costs is believed to even more accurately reveal the consequences of accidents than the total accident costs.

The blackspot identification method proposed for junction and non-junction locations in this chapter is presented in Section 4.2. The junction locations in this chapter are defined as the road junctions and their surrounding road network with 70 meters, while the non-junction locations are defined as other road network except junction locations. Section 4.3 gives a brief summary of the data adopted for the case study. The results and findings of the case study are then discussed in Section 4.4. Finally, a summary of the key findings of Chapter 4 is given in Section 4.5.

4.2. Proposed Method for Blackspot Identification

In this chapter, it is proposed to rank blackspots using the measure of total accident consequences, instead of solely basing the ranking on observed accident frequency. The total accident consequences of a location are expressed in terms of total injury costs or total accident costs. One advantage of identifying blackspots based on the total injury costs or total accident costs is that it can provide a complete assessment of different categories of injuries or accidents occurring at each location. The adoption of total injury costs also accounts for the effects of the number of injuries in the identification of blackspots.

In order to calculate the total injury costs or total accident costs, the estimated injury cost per person or accident cost per accident should be obtained. The estimated injury cost per person or accident cost per accident is discussed in details in Section 4.2.1 of this chapter. Models are proposed to calculate the total injury costs or total accidents costs for all locations, including junction and non-junction locations. The details of these models are discussed in Sections 4.2.2 and 4.2.3 of this chapter.

4.2.1 Estimated Injury or Accident Costs

The estimated accident cost per accident in the Third Comprehensive Transport Study (CTS-3) is adopted in this chapter. As the estimated injury cost per person is not provided in CTS-3, a method is proposed to estimate the injury cost per person based on the estimated accident cost per accident.

In 1999, CTS-3 was conducted to develop a long-term transport plan for Hong Kong up to 2016 (Transport Department, 1999). In Appendix D of Evaluation of Policies and Projects in the final report of CTS-3, six types of evaluation were proposed to study the impacts of traffic policies and infrastructure projects. These six types are operational, economic, financial, environmental, accident and distributional evaluations. The estimated accident costs were applied to evaluate the impacts of traffic policies and infrastructure projects on accidents in terms of costs.

The accident cost per accident was firstly estimated by the Road Safety Division of the Transport Department in Hong Kong in 1981. The estimated accident cost per accident is calculated based on insurance cost of each accident record. In the CTS-3, the estimated accident cost was inflated to the dollar value in year 1997, and then forecasted to the dollar values in the years 2001, 2006, 2011 and 2016 in line with the growth in Gross Domestic Product (GDP)/head. In this chapter, the accident costs from the years 2007 to 2009 are estimated based on the cost value in year 1997 following the estimation approach proposed in the CTS-3.

Table 4.1 presents the estimated accident cost per accident (HK$7.8= US$1), for each category of accident (fatal, seriously injured and slightly injured accident) from the years 2007 to 2009. It is observed that the estimated accident cost per fatal accident is nearly 10 times and 100 times that of the respective costs per seriously injured accident and slightly injured accident.

Table 4.1 Estimated accident cost per accident for different categories of accidents in various years

Year	Estimated accident cost per accident (HK$×$10^3$)		
	Fatal accident	Seriously injured accident	Slightly injured accident
2007	4306	354	48
2008	4455	366	49
2009	4604	379	51

As mentioned above, a method is proposed to estimate the injury cost per person based on the estimated accident cost per accident, as the estimated injury cost per person is not provided in CTS-3. The proposed method to estimate the injury cost per person for different categories of injuries is described as follows (Shown as Equations (*4.1*-*4.3*)).

$$IC_{Sl,T} = \frac{AC_{Sl,T} N_{Sl,T}}{F^{Sl}_{Sl,T}} \tag{4.1}$$

$$IC_{Se,T} = \frac{AC_{Se,T} N_{Se,T} - F^{Se}_{Sl,T} IC_{Sl,T}}{F^{Se}_{Se,T}} \tag{4.2}$$

$$IC_{F,T} = \frac{AC_{F,T} N_{F,T} - F^{F}_{Se,T} IC_{Se,T} - F^{F}_{Sl,T} IC_{Sl,T}}{F^{F}_{F,T}} \tag{4.3}$$

where,

$AC_{F,T}$ is the estimated accident cost per fatal accident during time period T (for example, one year);

$AC_{Se,T}$ is the estimated accident cost per seriously injured accident during time period T;

$AC_{Sl,T}$ is the estimated accident cost per slightly injured accident during time period T;

$N_{F,T}$ is the number of all fatal accidents in Hong Kong during time period T;

$N_{Se,T}$ is the number of all seriously injured accidents in Hong Kong during time period T;

$N_{Sl,T}$ is the number of all slightly injured accidents in Hong Kong during time period T;

$IC_{F,T}$ is the estimated injury cost per fatality during time period T;

$IC_{Se,T}$ is the estimated injury cost per serious injury during time period T;

$IC_{Sl,T}$ is the estimated injury cost per slight injury during time period T;

$F^{F}_{F,T}$ is the number of fatalities in all fatal accidents in Hong Kong during time period T;

$F^{F}_{Se,T}$ is the number of serious injuries in all fatal accidents in Hong Kong during time period T;

$F^{F}_{Sl,T}$ is the number of slight injuries in all fatal accidents in Hong Kong during time period T;

$F^{Se}_{Se,T}$ is the number of serious injuries in all seriously injured accidents in Hong Kong during time period T;

$F^{Se}_{Sl,T}$ is the number of slight injuries in all seriously injured accidents in Hong Kong during time period T;

$F^{Sl}_{Sl,T}$ is the number of slight injuries in all slightly injured accidents in Hong Kong during time period T.

For example, in 2007, one fatal accident, two seriously injured accidents and three slightly injured accidents occurred. There were one fatality, one serious injury and one slight injury in the fatal accident, two serious injuries and one slight injury in those two seriously injured accidents and three slight injuries in those three slightly injured accidents. The estimated accident costs per fatal, seriously injured and slightly injured accident are HK$1000, HK$100 and HK$10, respectively. The estimated injury costs per fatality, serious injury and slight injury are HK$10 (10×3 ÷3=HK$10), HK$95 ((100×2－10×1) ÷2=HK$95) and HK$895 (1000×1－95 ×1－10×1=HK$895), respectively.

Based on the estimated accident cost and injury information for each accident, the injury cost per person for each category of injury (fatality, serious injury and slight injury) from the years 2007 to 2009 are estimated and shown in Table 4.2. Similarly, it was found that the estimated injury cost per fatality is nearly 10 times and 100 times that of the respective injury cost per serious injury and slight injury.

Table 4.2 Estimated injury cost per person for different categories of injuries in various years

Year	**Estimated injury cost per person (HK$×10³)**		
	Fatality	**Serious injuries**	**Slight injuries**
2007	4060	319	39
2008	3803	330	40
2009	3980	349	41

In practice, however, the actual costs of each injury or accident may be varied due to some factors, such as the number of casualties, casualty characteristics (e.g., age, gender). Using either the estimated injury or accident costs instead of actual costs, therefore, may to some extent reduce the accuracy of blackspot identification results. Two approaches can be used to address this problem and they are discussed as follows.

The first approach is to make use of different probability distribution functions (PDF) for injury or accident cost variation. The injury or accident costs can then be calculated as the product of estimated injury or accident costs and their standard deviation (shown as Equation (*4.4*)).

$$C_{S,T} = \overline{C}_{S,T} + z\sigma_{S,T} \tag{4.4}$$

where,

$C_{S,T}$ is the cost of injury or accident with specific severity S (F represents fatality or fatal accident, Se represents serious injury or seriously injured accident and Sl represents slight injury or slightly injured accident) during time period T;

$\overline{C}_{S,T}$ is the estimated cost of injury or accident with specific severity *S* during time period *T*;

$\sigma_{S,T}$ is the standard deviation of the cost of injury or accident with specific type of severity *S* during time period *T*.

As the standard deviation of injury or accident costs has not been estimated and provided in CTS-3, it is assumed that the standard deviation of injury or accident costs is a product of estimated injury or accident costs and a parameter $\beta_{S,T}$ in this chapter. Equation (*4.4*) can be expressed as Equation (*4.5*). However, this assumption has a limitation that the standard deviation is a stochastic term while the term of the estimated accident costs times a parameter is deterministic. However, this thesis focuses mainly on the application of injury or accident costs to identify accident blackspots for demonstration rather than on the estimation of standard deviation of accident costs. The estimation of standard deviation of accident costs using regression models is an interesting extension of the study when the Hong Kong Road Casualty Information System (RoCIS) is available and linked to TRADS in the future. In RoCIS, the traffic accident data is linked to the hospital data and some data such as human losses, medical costs, and police and administrative costs can be obtained (Loo and Tsui, 2007). The use of a more reliable estimate on standard deviation of accident costs would improve the accuracy of blackspot identification results.

$$C_{S,T} = \overline{C}_{S,T} + z\sigma_{S,T} = \overline{C}_{S,T} + z\beta_{S,T}\overline{C}_{S,T} = \overline{C}_{S,T}(1 + z\beta_{S,T}) = \theta_{S,T}\overline{C}_{S,T} \qquad (4.5)$$

where,

$\theta_{S,T}$ is the weighted factors for injury or accident with specific type of severity *S* (*F* represents fatality or fatal accident, *Se* represents serious injury or seriously injured accident and *Sl* represents slight injury or slightly injured accident) during time period *T*.

The second approach is to develop a model to estimate the injury or accident costs. As indicated above, in CTS-3, the accident cost per accident is estimated based on insurance cost of each accident. However, insurance cost is usually not the exact cost of the accident, and lower than the exact accident cost (under estimation). It may be more accurately to estimate injury or accidents using a regression model if the traffic accident data can be linked to the hospital data.

It has been concluded that injury or accident costs can be divided into human losses, medical costs, and police and administrative costs in some studies (Evans, 2009). It has also been found that the injury or accident costs are correlated with several factors, such as the number of casualties, injury severity, body part injured, blood alcohol level, victim's age, involved vehicle type, collision type and speed limit (Miller et al., 1998; Zaloshnja and Miller, 2004; Zaloshnja et al., 2006). Therefore, the injury or accident costs can be formulated as a model with the independent variables of certain factors (Equation (*4.6*)).

$$C_{S,T} = f(\boldsymbol{\beta}\mathbf{X}_{T}) \tag{4.6}$$

where,

$C_{S,T}$ is the cost of injury or accident with specific severity *S* (*F* represents fatality or fatal accident, *Se* represents serious injury or seriously injured accident and *Sl* represents slight injury or slightly injured accident) during time period *T*;

$\mathbf{X_T}$ is the vector of factors to be considered;

$\boldsymbol{\beta}$ is the vector of corresponding coefficients to be calibrated.

This chapter, however, focuses mainly on the application of injury or accident costs to identify accident blackspots for demonstration rather than on the estimation of injury or accident costs. To calibrate the injury or accident cost estimation model, some data such as human losses, medical costs, and police and administrative costs should be obtained from the Hong Kong Road Casualty Information System (RoCIS) in which the traffic accident data is linked to the hospital data (Loo and Tsui, 2007). For this study, the first approach to calculate the actual injury or accident costs is adopted but it can be easily extended for updating the injury or accident costs if the RoCIS data is available.

4.2.2 Modeling Total Accident Costs for Junction and Non-junction Locations

It is proposed that the total accident costs are calculated as the weighted sum of different categories of accidents. The weighted factors are the estimated accident costs for different categories of accidents. The proposed models to calculate total accident costs for junction and non-junction locations are presented as follows.

Junction locations

For junction locations, the total accident costs at junction location d during time period T (for example, one year) are calculated as $TAC_T^J(d)$.

$$TAC_T^J(d) = AC_{F,T}N_{F,T}^J(d) + AC_{Se,T}N_{Se,T}^J(d) + AC_{Sl,T}N_{Sl,T}^J(d) \qquad (4.7)$$

where,

$AC_{F,T}$ is the estimated accident cost per fatal accident during time period T;

$AC_{Se,T}$ is the estimated accident cost per seriously injured accident during time period T;

$AC_{Sl,T}$ is the estimated accident cost per slightly injured accident during time period T;

$N^{J}_{F,T}(d)$ is the number of fatal accidents at junction location d during time period T;

$N^{J}_{Se,T}(d)$ is the number of seriously injured accidents at junction location d during time period T;

$N^{J}_{Sl,T}(d)$ is the number of slightly injured accidents at junction location d during time period T.

For example, in 2007, one fatal accident, two seriously injured accidents and three slightly injured accidents occurred at junction A. The estimated accident costs per fatal, seriously injured and slightly injured accident are HK$1000, HK$100 and HK$10, respectively. The total accident costs at junction A in 2010, therefore, are HK$1230 (1000×1+100×2+10×3= HK$1230).

Non-junction locations

For non-junction locations, the total accident costs (per unit distance) at road segment d during time period T (for example, one year) are calculated as $TAC^{NJ}_{T}(d)$.

$$TAC^{NJ}_{T}(d) = \frac{AC_{F,T}N^{NJ}_{F,T}(d) + AC_{Se,T}N^{NJ}_{Se,T}(d) + AC_{Sl,T}N^{NJ}_{Sl,T}(d)}{L(d)/L_U} \tag{4.8}$$

where,

$AC_{F,T}$ is the estimated accident cost per fatal accident during time period T;

$AC_{Se,T}$ is the estimated accident cost per seriously injured accident during time period T;

$AC_{Sl,T}$ is the estimated accident cost per slightly injured accident during time period T;

$N_{F,T}^{NJ}(d)$ is the number of fatal accidents at non-junction location d during time period T;

$N_{Se,T}^{NJ}(d)$ is the number of seriously injured accidents at non-junction location d during time period T;

$N_{Sl,T}^{NJ}(d)$ is the number of slightly injured accidents at non-junction location d during time period T;

$L(d)$ is the length (in meters) of non-junction location d;

L_U is the unit distance adopted (in meters) (In this chapter, the unit distance is adopted as 100 meters, following the study of Loo (2009)).

The same example discussed above is referred to again. However, those six accidents occurred at non-junction location B with the length of one kilometer but at junction A. The total accident costs (per 100 meters) at non-junction location B in 2007, therefore, are HK$123 ((1000×1+100×2+10×3)/(1000/100)= HK$123).

4.2.3 Modeling Total Injury Costs for Junction and Non-junction Locations

In this chapter, the total injury costs are adopted as the weighted sum of different categories of injuries. The weighted factors are the estimated injury costs for different categories of injuries. The proposed models to calculate total injury costs for junction and non-junction locations are illustrated.

Junction locations

For junction locations, the total injury costs at junction location *d* during time period *T* (for example, one year) are calculated as $TIC_T^J(d)$.

$$TIC_T^J(d) = IC_{F,T}F_{F,T}^J(d) + IC_{Se,T}F_{Se,T}^J(d) + IC_{Sl,T}F_{Sl,T}^J(d) \quad (4.9)$$

where,

$IC_{F,T}$ is the estimated injury cost per fatality during time period *T*;

$IC_{Se,T}$ is the estimated injury cost per serious injury during time period *T*;

$IC_{Sl,T}$ is the estimated injury cost per slight injury during time period *T*;

$F_{F,T}^J(d)$ is the number of fatalities at junction location *d* during time period *T*;

$F_{Se,T}^J(d)$ is the number of serious injuries at junction location *d* during time period *T*;

$F_{Sl,T}^J(d)$ is the number of slight injuries at junction location *d* during time period *T*.

The example given above is used again. Six accidents occurred at junction A in 2007. There were one fatality and one serious injury in the fatal accident, two serious injuries and one slight injury in those two seriously injured accidents and three slight injuries in those three slightly injured accidents. The estimated injury costs per fatality, serious injury and slight injury are HK$1000, HK$100 and HK$10, respectively. The total injury costs at junction A in 2010, therefore, are HK$1340 (1000×1+100×3+10×4= HK$1340).

Non-junction locations

For non-junction locations, the total injury costs (per unit distance) at road segment d during time period T (for example, one year) are calculated as $TIC_T^{NJ}(d)$.

$$TIC_T^{NJ}(d) = \frac{IC_{F,T}F_{F,T}^{NJ}(d) + IC_{Se,T}F_{Se,T}^{NJ}(d) + IC_{Sl,T}F_{Sl,T}^{NJ}(d)}{L(d)/L_U} \qquad (4.10)$$

where,

$IC_{F,T}$ is the estimated injury cost per fatality during time period T;

$IC_{Se,T}$ is the estimated injury cost per serious injury during time period T;

$IC_{Sl,T}$ is the estimated injury cost per slight injury during time period T;

$F_{F,T}^{NJ}(d)$ is the number of fatalities at non-junction location d during time period T;

$F_{Se,T}^{NJ}(d)$ is the number of serious injuries at non-junction location d during time period T;

$F_{Sl,T}^{NJ}(d)$ is the number of slight injuries at non-junction location d during time period T;

$L(d)$ is the length (in meters) of non-junction location d;

L_U is the unit distance adopted (in meters).

The example given above is adopted again. The total injury costs (per 100 meters) at non-junction location B in 2007, therefore, are HK$134 ((1000×1+100×3+10×4)/(1000/100)= HK$134).

The number of injuries or accidents of each category of injury or accident should also be obtained to calculate the total injury costs or accident costs of each location (Equations (*4.7-4.10*)). The number of injuries can be the observed number of injuries or the estimated excepted number of injuries based on multinomial model and logit model (Miranda-Moreno et al., 2009). Similarly, the accident frequency can be the observed accident frequency or the estimated excepted accident frequency using a specific model, such as Poisson regression and negative binomial regression models (Miaou, 1994; Washington et al., 2003; Wong et al., 2007). In this chapter, the observed number of injuries or accidents is adopted rather than estimated excepted ones. In the future, an extension of this study can be to adopt the expected number of injuries or accidents instead of observed ones.

4.3. Data Sources

As discussed in Chapter 3, TRADS records relevant information about traffic accidents occurring in the Hong Kong territory. In this chapter, Hong Kong traffic accident data from the years 2007 to 2009 are extracted from TRADS for analyses. There are 15315, 14576, and 14316 accidents in the years 2007, 2008 and 2009, respectively. Among these accidents, about 4000 accidents in each year occurred at junction locations. The approximate remaining 10000 accidents in each year happened at non-junction locations (road segments). In Hong Kong, during these three years, there are a total of 4406 junction locations and 2145 non-junction locations where accidents have occurred.

4.4. Results of Analyses and Discussions

The results of the analyses are shown and discussed in this section. Firstly, the results of blackspots identified using different accident measures, namely accident frequency, total injury costs or accident costs, are compared. In the next sub-section, comparisons of blackspots identified by TD in Hong Kong and those produced by the proposed method, based on total injury costs or accident costs, are given. Finally, the results of sensitivity analyses by changing the injury costs or accident costs are discussed.

4.4.1 Comparisons of Blackspots Identified by Different Accident Measures

In this chapter, locations with rankings in the top 1% by aforementioned accident measures are identified as blackspots. 45 junction blackspots and 22 non-junction blackspots were obtained for each of these three accident measures. To quantify the effects of changing different accident measures to identify blackspots, the percentage deviation value is adopted. The measure of percentage deviation value has been previously used to conduct similar analyses (Geurts et al., 2004). A comparison of the rankings between two datasets can be made using this value. The definition of percentage deviation value is described as Equation (*4.11*).

$$p_d = (1 - \frac{G}{T}) \times 100\% \qquad (4.11)$$

where,

G is the number of common blackspots in the two comparison datasets;

T is the total number of blackspots in each dataset.

Table 4.3 presents the comparison results using different accident measures for identifying junction blackspots. More specifically, Table 4.3 shows the percentage deviation values for junction blackspots using different accident measures in the years 2007, 2008 and 2009. From Table 4.3, it is observed that more than about 90% of junction blackspots identified based on accident frequency deviate from those identified based on total injury costs (TIC) or total accident costs (TAC). Table 4.4 presents the spearman’s rank correlation between junction blackspots using different accident measures. All results are statistically significant at the 1% level. It can be seen that the correlation between blackspots identified based on TIC and TAC is strong (0.71), while the correlations between blackspots identified based on N and TAC or N and TIC are weak (0.32 or 0.34).

These results imply that most junction blackspots identified based on total injury costs or accident costs cannot be identified based on accident frequency. In other words, most junction locations with high injury or accident costs cannot be found based on accident frequency only. Using total injury or accident costs, therefore, can locate blackspots with higher injury or accident costs than using accident frequency only. The identified junction blackspots based on total accident costs are almost the same as those based on total injury costs (the percentage deviation values are 0%, except for 2.2% in year 2009).

Table 4.3 Percentage deviation values for junction blackspots using different accident measures

	Year 2007			**Year 2008**			**Year 2009**		
	N	**TAC**	**TIC**	**N**	**TAC**	**TIC**	**N**	**TAC**	**TIC**
N	0.0%			0.0%			0.0%		
TAC	**95.6%**	0.0%		91.1%	0.0%		86.7%	0.0%	
TIC	**95.6%**	0.0%	0.0%	91.1%	0.0%	0.0%	88.9%	**2.2%**	0.0%

Note: N stands for the number of accident (i.e. accident frequency); TAC stands for the total accident costs; TIC stands for the total injury costs.

Table 4.4 Spearman's rank correlation between junction blackspots using different accident measures

	Year 2007			**Year 2008**			**Year 2009**		
	N	**AC**	**IC**	**N**	**AC**	**IC**	**N**	**AC**	**IC**
N	1.00			1.00			1.00		
AC	0.32**	1.00		030**	1.00		0.31**	1.00	
IC	0.34**	0.71**	1.00	0.29**	0.70**	1.00	0.32**	0.67**	1.00

**Statistically significant at the 1% level

In Table 4.5, the junction blackspot distributions for the years 2007, 2008 and 2009 present further elaboration of the results based on different accident measures. From Table 4.5, it can be observed that all the blackspots identified using accident frequency have more than 7 accidents occurring in 2007. Nearly all the blackspots identified based on total injury or accident costs, however, have had less than 7 accidents. Despite the small number of accidents, these locations are identified as blackspots mainly because fatal accidents or fatalities occurred at these locations in 2007.

Table 4.5 Number of junction blackspots using different accident measures in the years 2007, 2008 and 2009

Accident frequency	**Number of blackspots**								
	2007			2008			2009		
	N	TAC	TIC	N	TAC	TIC	N	TAC	TIC
1	0	**13**	**13**	0	24	24	0	15	15
2	0	**14**	**14**	0	8	8	0	12	12
3	0	**9**	**9**	0	5	5	0	2	2
4	0	**5**	**5**	0	2	2	0	4	4
5	0	**2**	**2**	0	1	1	0	3	3
6	0	0	0	7	1	1	0	3	3
7	**18**	**2**	**2**	13	2	2	12	0	1
8	**11**	0	0	6	1	1	9	2	1
9	**8**	0	0	5	1	1	6	1	1
10	**5**	0	0	4	0	0	7	2	2
More than 10	**3**	0	0	10	0	0	11	0	0

The percentage deviation values and the spearman's rank correlations for non-junction blackspots using different accident measures in the years 2007, 2008 and 2009 are shown in Tables 4.6 and 4.7, respectively. Similarly, the deviations between the resultant non-junction blackspots identified based on accident frequency and those based on total injury or accident costs are significant. The percentage deviation values range from 63.6% to 81.8% over three years. The correlations between blackspots identified based on N and TAC or N and TIC are about 0.3 statistically significant at the 1% level. These results also indicate that adopting injury or accident costs can identify non-junction blackspots with higher injury or accident costs than using accident frequency only.

Table 4.6 Percentage deviation values for non-junction blackspots using different accident measures

	Year 2007			Year 2008			Year 2009		
	N	**TAC**	**TIC**	**N**	**TAC**	**TIC**	**N**	**TAC**	**TIC**
N	0.0%			0.0%			0.0%		
TAC	72.7%	0.0%		68.2%	0.0%		77.3%	0.0%	
TIC	72.7%	0.0%	0.0%	**63.6%**	**9.1%**	0.0%	**81.8%**	**9.1%**	0.0%

Note: N stands for the number of accident per 100 meters; TAC stands for the total accident costs 100 meters; TIC stands for the total injury costs 100 meters.

Table 4.7 Spearman's rank correlation between non-junction blackspots using different accident measures

	Year 2007			Year 2008			Year 2009		
	N	**AC**	**IC**	**N**	**AC**	**IC**	**N**	**AC**	**IC**
N	1.00			1.00			1.00		
AC	0.32**	1.00		0.29**	1.00		0.31**	1.00	
IC	0.31**	0.68**	1.00	0.31**	0.70**	1.00	0.31**	0.69**	1.00

**Statistically significant at the 1% level

4.4.2 Comparison Analyses of Blackspots Identified by the Transport Department and Total Injury/Accident Costs

In this sub-section, the blackspots identified by TD are compared against those based on total injury costs or accident costs. As discussed in Table 2.1 in Chapter 2, junction locations where more than 9 injury accidents or 6 accidents involved with pedestrian injuries over the past year are defined as blackspots by the TD in Hong Kong. There are 85, 62 and 126 junction locations identified as blackspots in the years 2007, 2008 and 2009, respectively.

Figure 4.1 shows the numbers and percentages of common blackspots identified based on TAC, TIC and TD's definition in the years 2007, 2008 and 2009. It was found that only about 20% of blackspots are simultaneously identified based on the three methods in the years 2007 and 2008, but 42.8% in year 2009. The results from Figure 4.1 also indicate that more than half of blackspots identified based on TD's definition deviate from those identified based on total injury or accident costs. For example, in the year 2007, 77.6% (64.7%+12.9%) and 75.3% (64.7%+10.6%) of blackspots identified based on total injury or accident costs cannot be identified based on TD's definition. In other words, a large number of junction locations with high total injury or accident costs cannot be identified as blackspots according to the TD's definition. This indicates that adopting total injury or accident costs can identify blackspots with higher injury or accidents costs than using TD's definition. This finding is further justified by later analysis.

Another finding which can be drawn from the results shown in Figure 4.1 is that a large number of junction locations are simultaneously identified as blackspots based on total injury or accident costs. The percentages of blackspots in common, identified based on total injury or accident costs, are 84.7% (64.7%+20%), 88.7% (69.3%+19.4%) and 84.9% (42.8%+42.1%) in the years 2007, 2008 and 2009, respectively. This shows that majority of blackspots identified based on total injury or accident costs are identical.

In order to further justify the finding that using total injury costs or accident costs can identify blackspots with higher injury or accidents costs than using TD's definition, another analysis was conducted. The reduction of total injury costs per reduction of one junction blackspot was calculated for the blackspots identified based on TD's definition and total injury or accident costs. Table 4.8 shows the results in the years 2007, 2008 and 2009. In Table 4.8, the percentages in the parentheses are calculated by dividing the injury cost values by the total injury costs of all junction locations in that year.

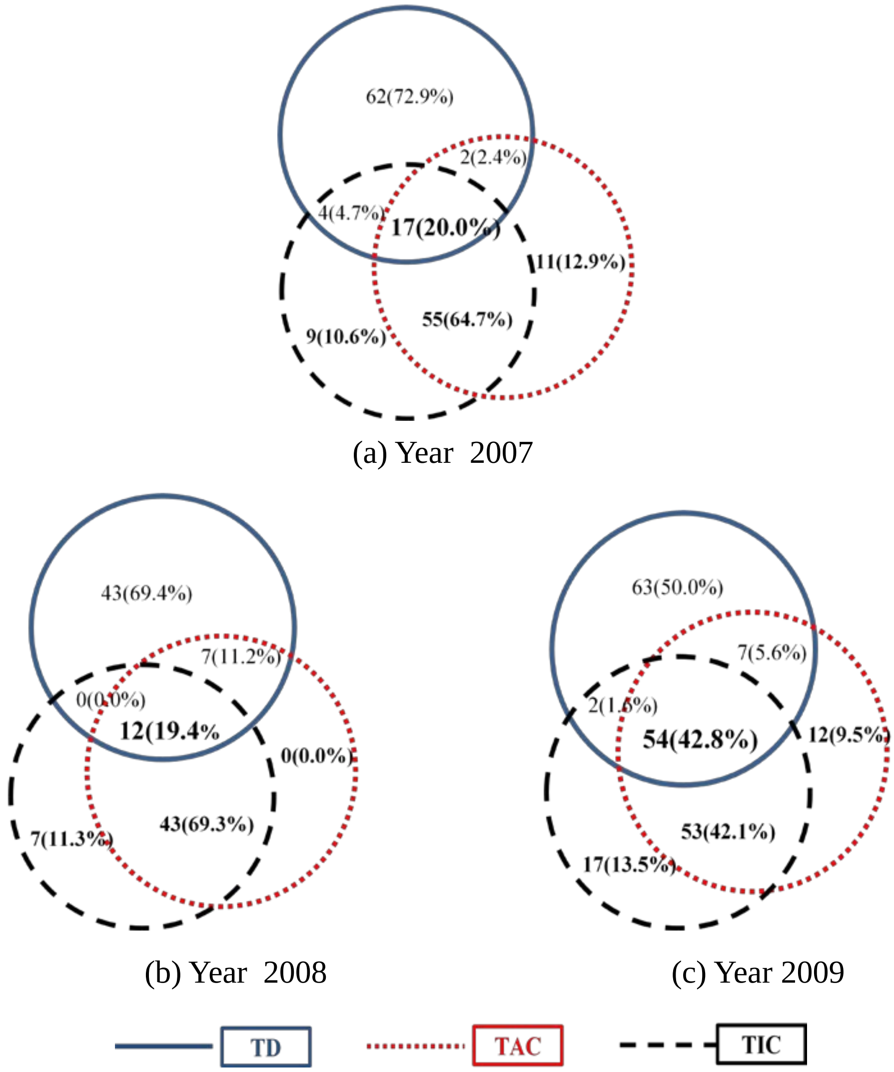


Figure 4.1 Numbers and percentages of common accident blackspots identified based on total injury or accident costs and by the TD

Table 4.8 Reduction of total injury costs per reduction of one junction blackspot (HK$×10³)

Total injury costs	Year 2007			Year 2008			Year 2009		
	TD	TAC	TIC	TD	TAC	TIC	TD	TAC	TIC
Mean	679 (0.12%)	**2838** **(0.52%)**	**2897** **(0.53%)**	968 (0.18%)	**3337** **(0.62%)**	**3406** **(0.64%)**	832 (0.16%)	**2056** **(0.38%)**	**2095** **(0.39%)**
1st ranked blackspot	1139 (0.21%)	8554 (1.56%)	8554 (1.56%)	1326 (0.25%)	4464 (0.84%)	4532 (0.85%)	949 (0.18%)	4803 (0.90%)	16493 (3.08%)

Table 4.8 shows that the mean values of the total injury cost reduction per reduction of one blackspot. It was found that the mean values for blackspots identified based on total injury or accident costs are dramatically larger than those for blackspots identified based on TD's definition. The values in the row of "1st ranked blackspot" represent the reduction of total injury costs if reducing the 1st ranked blackspot. It is also observed that these values for blackspots identified based on total injury or accident costs are much larger than those for TD's blackspots. Figure 4.2 shows the total injury costs for each blackspot identified based on TD's definition and total injury or accident costs in the year 2007 as an example.

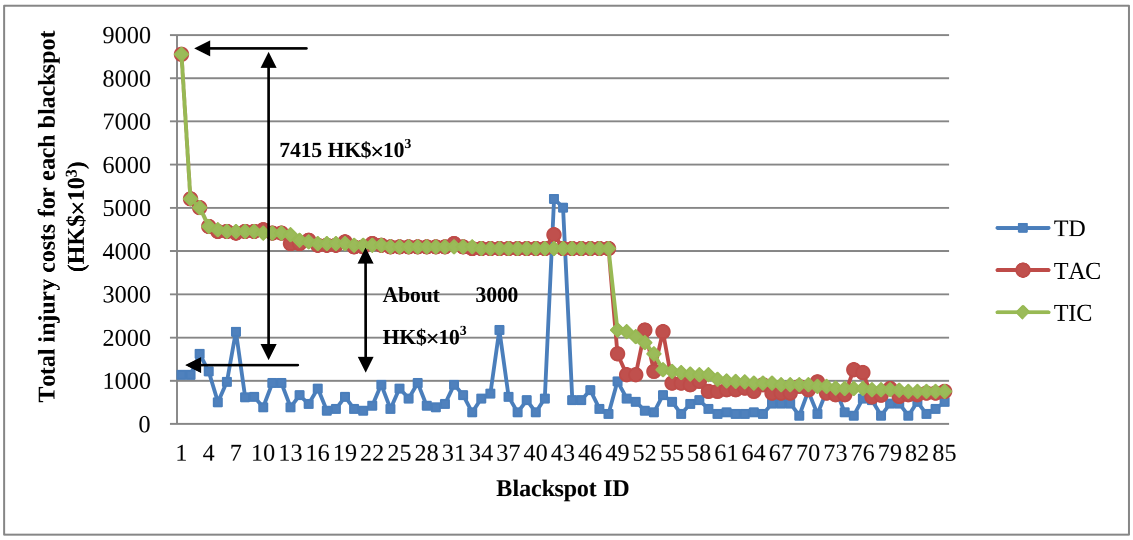


Figure 4.2 Total injury costs for each junction blackspot in the year 2007

From Figure 4.2, it can be seen that total injury costs of nearly all of blackspots identified based on total injury or accident costs are larger than those of blackspots identified by TD. All these results further reveal that the application of total injury or accident costs can rank blackspots with higher injury or accident costs than the use of the TD's definition.

4.4.3 Sensitivity Analyses Using Different Combinations of Injury/Accident Costs

From the description of the method given in Section 4.2, it was assumed that the actual injury or accident costs could be weighted values of the estimated injury or accident costs. The choice of using different weighted factor combinations ($\theta_F\_\theta_{Se}\_\theta_{SI}$) to calculate the actual injury or accident costs with different severities will greatly influence the ranking results of most dangerous blackspots. Therefore, in this chapter, a sensitivity analysis has been carried out to evaluate the effects of changing weighted factor combinations on the ranking and identification of blackspots. Table 4.9 presents three examples of weighted factor combinations adopted to calculate actual injury or accident costs. They are 1_1_1, 0.5_1_2 and 0.2_1_5.

Table 4.9 The actual accident costs using different weighted factor combinations in the year 2007

Weighted factor combination	**Actual accident costs (HK$\times 10^3$)**		
($\theta_F\_\theta_{Se}\_\theta_{SI}$)	**Fatal accident**	**Seriously injured accident**	**Slightly injured accident**
1_1_1	4306	354	48
0.5_1_2	2153	354	96
0.2_1_5	861	354	240

The actual accident costs calculated based on three combinations of weighted factors in the year 2007 are taken as examples to illustrate the physical meaning of these three weighted factor combination examples. For the weighted factor combination 1_1_1, it is assumed that the consequences in terms of costs caused by fatal accident are much larger than those caused by seriously and slightly injured accidents. The weighted factor combination 0.5_1_2 uses a more moderate approach to stress accident consequences of fatal accidents. Using the last weighted factor combination 0.2_1_5, it is proposed that costs caused by fatal accident are slightly larger than those caused by seriously and slightly injured accidents.

Table 4.10 presents the sensitivity results when changing the combinations of weighted factors to calculate the actual accident costs in the years 2007, 2008 and 2009. It is observed that the percentage deviation values between weighted factor combinations 1_1_1 and 0.5_1_2 are small for junction locations (from 2.2% to 4.4%). This implies that the resultant blackspots identified based on these two weighted factor combinations are almost the same. However, the percentage deviation values between weighted factor combinations 0.2_1_5 and 1_1_1, 0.5_1_2 are much larger. This indicates that more than 80% of the junction blackspots identified based on weighted factor combination 0.2_1_5 do not appear in the dataset of blackspots identified based on weighted factor combinations 1_1_1 and 0.5_1_2. The effects of changing weighted factor combinations from 1_1_1 and 0.5_1_2 to 0.2_1_5 on the identification results of junction blackspots are pronounced. For the non-junction locations, the effects of changing weighted factor combinations from 1_1_1 and 0.5_1_2 to 0.2_1_5 are relatively smaller. The percentage deviation values range from 45.5% to 72.7%.

Table 4.10 Percentage deviation values for blackspots using different weighted factor combinations (based on accident costs)

Weighted factor combination ($\theta_F - \theta_{Se} - \theta_{Sl}$)	**Year 2007**			**Year 2008**			**Year 2009**		
Junction locations									
	1_1_1	**0.5_1_2**	**0.2_1_5**	**1_1_1**	**0.5_1_2**	**0.2_1_5**	**1_1_1**	**0.5_1_2**	**0.2_1_5**
1_1_1	0.0%			0.0%			0.0%		
0.5_1_2	**2.2%**	0.0%		**4.4%**	0.0%		**2.2%**	0.0%	
0.2_1_5	86.7%	84.4%	0.0%	82.2%	77.8%	0.0%	71.1%	71.1%	0.0%
Non-junction locations									
	1_1_1	**0.5_1_2**	**0.2_1_5**	**1_1_1**	**0.5_1_2**	**0.2_1_5**	**1_1_1**	**0.5_1_2**	**0.2_1_5**
1_1_1	0.0%			0.0%			0.0%		
0.5_1_2	36.4%	0.0%		22.7%	0.0%		22.7%	0.0%	
0.2_1_5	**72.7%**	**45.5%**	0.0%	**68.2%**	**45.5%**	0.0%	**68.2%**	**45.5%**	0.0%

Table 4.11 shows the percentage deviation values for blackspots identified based on injury costs calculated using different weighted factor combinations for the years 2007, 2008 and 2009. The results are quite similar to those presented in Table 4.10 for blackspots identified based on accident costs. It is also seen that large percentage deviation values exit between the resultant blackspots identified based on the weighted factor combinations 0.2_1_5 and 1_1_1, 0.5_1_2, from 71.1% to 91.1% for junction locations and from 36.4% to 72.7% for non-junction locations. This also means that most blackspots identified, based on the weighted factor combination 0.2_1_5, do not belong to the blackspot dataset ranked based on weighted factor combinations 1_1_1 and 0.5_1_2.

Table 4.11 Percentage deviation values for blackspots using different weighted factor combinations (based on injury costs)

Weighted factor combination ($\theta_F - \theta_{Se} - \theta_{Sl}$)	**Year 2007**			**Year 2008**			**Year 2009**		
Junction locations									
	1_1_1	**0.5_1_2**	**0.2_1_5**	**1_1_1**	**0.5_1_2**	**0.2_1_5**	**1_1_1**	**0.5_1_2**	**0.2_1_5**
1_1_1	0.0%			0.0%			0.0%		
0.5_1_2	11.1%	0.0%		20.0%	0.0%		2.2%	0.0%	
0.2_1_5	**91.1%**	80.0%	0.0%	91.1%	**71.1%**	0.0%	75.6%	73.3%	0.0%
Non-junction locations									
	1_1_1	**0.5_1_2**	**0.2_1_5**	**1_1_1**	**0.5_1_2**	**0.2_1_5**	**1_1_1**	**0.5_1_2**	**0.2_1_5**
1_1_1	0.0%			0.0%			0.0%		
0.5_1_2	31.8%	0.0%		18.2%	0.0%		31.8%	0.0%	
0.2_1_5	**72.7%**	40.9%	0.0%	63.6%	45.5%	0.0%	68.2%	**36.4%**	0.0%

4.5. Summary

In this chapter, a method has been proposed for traffic accident blackspot identification. The merit of this proposed method is in the adoption of accident consequences instead of accident frequency for blackspot ranking. The measure of accident consequences is alternatively represented as total injury costs or total accident costs. Using total accident costs allows the integration of accident frequency and accident severity into the process of blackspot identification. Total injury costs take both the effects of the number of injuries and their severities into consideration when identifying accident blackspots.

It was found that using the measures of total injury or accident costs can locate blackspots with higher injury or accident costs than adopting the measure of accident frequency only. The proposed identification method is also proved to be easier to identify blackspots with higher injury or accident costs than using the blackspot definition proposed by the Hong Kong Transport Department. Sensitivity analyses have been carried out to analyze the effects of changing weighted factor combinations when calculating total injury or accident costs. The results reveal that the choice of different costs for injuries or accidents has significant effects on the identification of blackspots.

Injury severity and injury costs of casualties are further investigated in Chapter 5 as both of them are importance factors for identifying accident blackspots. Models are built to quantify the effects of contributory factors on injury severity of pedestrian, driver and passenger casualties and crash injuring costs. The modeling results are presented and discussed in Chapter 5.

5. Modeling Casualty Injury Severity and Accident Costs

5.1. Introduction

In Chapter 4, it is found that identifying accident blackspots based on injury costs can identify locations with higher accident costs than that using accident frequency. Costs is adopted to reflect the accident or injury severity in Chapter 4, as definitely fatalities and serious injuries cause more social and economic lost than slight injuries. It is of interest, therefore, to further analyze injury severity of casualties and accident costs, and to find out useful empirical findings. These empirical findings may contribute to the proposal of effective road safety treatments which aim at reducing injury severity and accident costs.

Casualty distributions by injury severity and casualty type for the years 2007, 2008 and 2009 are presented in Chapter 3. It was found that driver casualties account for more than 40% of total casualties, however, among all the fatalities, more than 50% are pedestrians. Of all driver and passenger casualties, the fatality and seriously injured casualty rates are about 10%, while the rate for pedestrian casualties is more than 20%. These characteristics of casualty distribution between driver, passenger and pedestrian appear to be attributed to a combination of causal circumstances in the region of the accident. For instance, pedestrians in the path of vehicular collisions have no protection from possible impacts and are therefore likely to be severely injured. The driver in the well protected cab of a goods vehicle is likely to be less badly injured.

As summarized in Chapter 2, numerous studies have been conducted to model the associations between contributory factors and casualty severity suffered by such as drivers and pedestrians. Abdel-Aty (2003) specifically, adopted the ordered probit models to estimate the effects of contributory factors related to driver casualty injury severity levels at multiple highway locations. Savolainen and Mannering (2007) developed nested logit models and standard multinomial logit models to predict the probability of motorcyclists' injury severity in single- and multi-vehicle crashes. In the study of Sze and Wong (2007), binary logistic regression models were developed to evaluate the injury risk to pedestrians and to identify factors contributing to fatalities and serious injuries.

Although many studies have separately investigated the effects of contributory factors on driver and pedestrian casualty injury severity, few results can be directly applied to explain the characteristics of distribution of driver, passenger and pedestrian casualties mentioned above. This chapter aims to extend knowledge reported in the literature by comparing contributory factors that affect the injury severity of driver, passenger and pedestrian casualties. A binary logistic regression model is used to model and quantify the effects of these contributory factors.

The other contribution of this chapter is to propose a regression model to quantify the associations between accident costs and contributory factors. The main advantage of this regression model is that it assigns a dollar value to the safety impact of each contributory factor. Understanding the effect of factors in money term can help Transport Department to carry out the cost-benefit analysis for improvements of road safety programmes.

Section 5.2 presents a brief summary of the data used in this chapter. The models adopted to quantify the associations between casualty injury severity and accident costs and contributory factors are described in Section 5.3. The results of three injury severity models (one each for driver, passenger and pedestrian casualties) and one accident costs model are discussed in Section 5.4. Section 5.5 gives a brief summary of the key findings in this chapter.

5.2. Data Description

In this chapter, the associations between contributory factors and the injury severity of different casualties and accident costs are estimated. The accident and casualty data used to estimate the associations is extracted from the Hong Kong Traffic Accident Database System (TRADS). Details of TRADS are given in Chapter 3.

Table 5.1 presents a summary of the 56447 traffic accident casualties from the years 2007 to 2009, which are used in this chapter. Among these 56447 casualties, more than 13% (7371/56447) are killed or seriously injured. Nearly half of the total casualties are drivers (25649/56447), while the remainders are passengers and pedestrians.

Table 5.1 Summary of casualties from the years 2007 to 2009

Degree of Injury Severity	**Casualty Type**			**Total**
	Driver	Passenger	Pedestrian	
Slightly Injured	22510	17720	8846	49076
Fatal/Seriously Injured	3139	1594	2638	7371
Total	25649	19314	11484	**56447**

With the use of these casualty data, regression models to estimate the associations between injury severity of driver, passenger, and pedestrian casualties and accident cost, and all possible contributory factors are proposed. For the injury severity models, the dependent variable is the injury severity of the casualty, while for the accident costs model, the dependent variable is the total injury costs (TIC) of the accident referred to Chapter 4. Independent variables are some contributory factors which are likely to affect the outcome of injury severity and accident costs.

From the literature review presented in Chapter 2, it is observed in many papers that the severity of a crash (or a casualty) have been found to be influenced by various contributory factors related to casualty demographic characteristics, the environment, vehicle and road characteristics. In line with the factor classification logic of the previous research work on this topic, all possible contributory factors considered in this Chapter are divided into casualty related factors, environmental factors, site factors and vehicle factor. The details of these contributory factors are described as follows. The distributions of casualties and accidents by these factors are given in Table 5.2 and Table 5.3.

Casualty related factors

The casualty related factors considered in this chapter include: 1. age, as classified into three groups, under 15, 15-65, and above 65, as suggested by Sze and Wong (2007); 2. sex, i.e., male and female; 3. whether a seat belt or crash helmet was worn, this factor is divided into three groups: yes, no, and pedestrian; 4. seat occupied, which is then classified into five groups including driver, front nearside, rear, standing, others and pedestrian; 5. pedestrian location, which is divided into four groups: on the crossing, within 15m of the crossing, others and driver/passenger.

Environmental factors

Three environmental factors are considered as follows: 1. the day of week on which the accident occurred which is then classified into two groups, Monday-Friday and Weekend; 2. the time the accident occurred, which is divided into four groups in relation to working hours: 7:00am-9:59am, 10:00am-3:59pm, 4:00pm-6:59pm and 7:00pm-6:59am, as suggested by Sze and Wong (2007); 3. level of rain: no rain, light rain and heavy rain. It should be mentioned that the rain condition when an accident occurs is obtained from TRADS. The classification is based on the observation and subjective judgment of traffic police handling the accident. The estimated effects of rain conditions on injury severity may be biased owing to this subjectivity.

Site factors

Five types of site factors are considered. They include: 1. speed limit, which is classified into the following three groups: below 50 km/h, 50 km/h and above 50 km/h; 2. traffic congestion condition, which is classified into three levels: severe, moderate and none; 3. road surface condition, i.e., whether the conditions are wet or dry; 4. junction control type, which is divided into the following four groups: traffic signals, other control types, no control, and non-junction; 5. road type, which is classified into three categories: one way carriageway, two-way carriageway and multi-/dual carriageway. Similar to the rain condition details, the traffic congestion conditions are also provided by TRADS and are also determined subjectively based on the observation of the traffic police. The estimated results regarding this factor, therefore, may be likewise biased.

Vehicle factor

The vehicle factor provides information regarding the type of vehicle involved in the investigated accident. Vehicles are classified into seven types: motorcycles, private cars, goods vehicles, vans and light buses, buses, taxis and others.

One limitation of this study is that the factor of traffic exposure has not been considered in the regression models of this study. It is because that TRADS is not directly linked to Hong Kong Annual Traffic Census (ATC) in which the traffic data are collected regularly at pre-determined locations of major roads. In the literature, the traffic exposure factor was found to be an important factor when analyzing the occurrence of accidents and accident severity modeling (Karlaftis and Golias, 2002;

Wang et al., 2009). Therefore, analyzing the effects of traffic exposure on injury severity or accident costs can be an interesting extension of this study. The traffic exposure at road segments or junction locations with accident occurred would be obtained if ATC in Hong Kong can be linked to TRADS using the name of roads as the reference index. Although about 87% of trafficable roads in Hong Kong are covered by the ATC stations, less than one-third of these stations are with permanent detectors for capturing the traffic flow data over the whole year. Therefore, accidents occurred on roads that are not covered by ATC stations with permanent detectors should be excluded from the regression models when modeling the effects of traffic exposure on injury severity and accident costs.

5.3. Regression Model

5.3.1 Binary Logistic Regression Model

Binary Logistic Regression

Associations between driver, passenger, and pedestrian casualty injury severity and contributory factors are measured in this chapter. The dependent variable, injury severity, is divided into two types, 1. fatal and seriously injured, and 2. slightly injured. This binary nature of the dependent variable enables the application of binary logistic regression. In the binary logistic regression the association between the dependent variable (injury severity) and independent variables (contributory factors) can be formulated by Equation (*5.1*) (Hosmer and Lemeshow, 1989):

Table 5.2 Summary of casualties by factors considered

Category	Factor	Attribute	Count	Proportion
Casualty related factor	Age (years)	Under 15	4357	7.7%
		Above 65	4916	8.7%
		15-65	47174	83.6%
	Sex	Male	36004	63.8%
		Female	20443	36.2%
	Seat Belt or Crash Helmet Worn	Yes	36432	64.5%
		No	8531	15.1%
		Pedestrian	11484	20.3%
	Seat Occupied	Driver	25649	45.4%
		Front Nearside	3620	6.4%
		Rear	11799	20.9%
		Standing	3031	5.4%
		Others	864	1.5%
		Pedestrian	11484	20.3%
	Pedestrian Location	On the Crossing	1754	3.1%
		Within 15m of Crossing	838	1.5%
		Others	8892	15.8%
		Driver/Passenger	44963	79.7%
Environmental factor	Day of Week	Monday-Friday	39905	70.7%
		Weekend	16542	29.3%
	Time of Day	7:00am-9:59am	8945	15.8%
		10:00am-3:59pm	18087	32.0%
		4:00pm-6:59pm	10928	19.4%
		7:00pm-6:59am	18484	32.7%
	Rain Condition	Light Rain	6733	11.9%
		Heavy Rain	1268	2.2%
		No Rain	48441	85.8%
Site factor	Speed Limit	Below 50km/h	1285	2.3%
		50km/h	48141	85.3%
		Above 50km/h	7015	12.4%
	Traffic Congestion	Severe Congestion	6827	12.1%
		Moderate Congestion	7523	13.3%
		None	42097	74.6%
	Road Surface Condition	Wet	9455	16.8%
		Dry	46992	83.2%
	Junction Control Type	Traffic Signal	8896	15.8%
		Other Control Types	2888	5.1%
		No Control	3507	6.2%
		Non-junction	41156	72.9%
	Road Type	One-way Carriageway	22575	40.0%
		Two-way Carriageway	19139	33.9%
		Multi-/dual Carriageway	14733	26.1%
Vehicle factor	Vehicle Class	Motorcycle	8509	15.1%
		Private Car	13671	24.2%
		Goods Vehicle	1229	2.2%
		Van and Light Bus	10454	18.5%
		Bus	6177	10.9%
		Taxi	9612	17.0%
		Others	6795	12.0%
Year	Year	2007	19625	34.8%
		2008	18684	33.1%
		2009	18138	32.1%

Table 5.3 Summary of accidents by factors considered

Category	Factor	Attribute	Count	Proportion
Environmental factor	Day of Week	Monday-Friday	31210	70.6%
		Weekend	12997	29.4%
	Time of Day	7:00am-9:59am	6810	15.4%
		10:00am-3:59pm	14760	33.4%
		4:00pm-6:59pm	8887	20.1%
		7:00pm-6:59am	13750	31.1%
	Rain Conditions	Light raining	4919	11.1%
		Heavy raining	880	2.0%
		Not raining	38408	86.9%
Site factor	Speed Limit	Below 50km/h	1145	2.6%
		50km/h	38751	87.7%
		Above 50km/h	4311	9.8%
	Traffic Congestion	Severe Congestion	5409	12.2%
		Moderate Congestion	5852	13.2%
		None	32946	74.5%
	Road Surface Condition	Wet	6910	15.6%
		Dry	37297	84.4%
	Junction Control Type	Traffic signal	6954	15.7%
		Other control types	2283	5.2%
		No control	3011	6.8%
		Non-junction	31959	72.3%
	Road Type	One way carriageway	18369	41.6%
		Two-way carriageway	15234	34.5%
		Multi-/dual carriageway	10604	24.0%
Year	Year	2007	15315	34.6%
		2008	14576	33.0%
		2009	14316	32.4%

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots \beta_m x_m \qquad (5.1)$$

where,

$g(x)$ is the latent dependent variable,

x_j is the jth independent variable which is the contributory factor,

β_j is the corresponding coefficient of the variable (for j = 1, 2, 3, …., m),

m is the number of independent variables or contributory factors.

The association between $g(x)$ and the conditional probability of the outcome is modeled by the following equations:

$$\pi(x) = \frac{\exp(g(x))}{1+\exp(g(x))} \qquad (5.2)$$

or

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) \qquad (5.3)$$

where,

$\pi(x)$ is the conditional probability of the outcome.

Therefore, the conditional probability of the outcome $\pi(x)$ can be further expressed in Equation (*5.4*):

$$\pi(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_j x_j+\cdots\beta_m x_m)}} \qquad (5.4)$$

for those pairs (x_i, y_i),

The probability of the outcome is $\pi(x_i)$ for those pairs with $y_i = 1$,

The probability of the outcome is $1-\pi(x_i)$ for those pairs with $y_i = 0$,

where,

i is the index of the observation,

y_i is the *i*th observed outcome.

In order to estimate the coefficients β_j for each factor, the maximum likelihood method is employed. The likelihood function is calculated as a product of $\pi(x_i)$ and $1-\pi(x_i)$.

$$l(\beta)=\prod_{i=1}^{n}\pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i} \tag{5.5}$$

where,

n is the number of observations.

It is mathematically easier to work with the log of Equation (5.5). The log likelihood expression is given as:

$$LL(\beta)=\ln(l(\beta))=\sum_{i=1}^{n}[y_i\ln(\pi(x_i))+(1-y_i)\ln(1-\pi(x_i))] \tag{5.6}$$

The coefficients can be estimated by maximizing the log likelihood function with respect to β and setting the resultant expressions equal to zero.

$$\sum_{i=1}^{n}(y_i-\pi(x_i))=0 \tag{5.7}$$

$$\sum_{i=1}^{n}x_i(y_i-\pi(x_i))=0 \tag{5.8}$$

With the estimated β s, the influence of contributory factor j on the injury outcome can be calculated as the odds ratio:

$$\text{Odds ratio}=\exp(\beta_j) \tag{5.9}$$

For the odds ratio of a particular contributory factor, with a value larger than 1, the injury level tends to be higher, whereas, an odds ratio smaller than 1 indicates that this contributory factor leads to lower injury risk.

Hosmer-Lemeshow Statistic

To verify the goodness-of-fit of the logistic regression model, the Hosmer-Lemeshow statistic is applied. This statistic is applicable for any number of contributory factors, regardless of their nature. Both categorical and continuous variables can be calibrated in this test. Hosmer and Lemeshow (1980) proposed a Pearson chi-square statistic for logistic regression, based on a grouping of estimated probabilities. Suppose that the model consists of *n* observations, divided into *g* groups. Different groups consist of *n/g* observations with different estimated probabilities. For example, Group 1 contains *n/g* observations with the lowest estimated probabilities, followed by Group 2 containing *n/g* observations with the subsequent lowest predicted probabilities. The remaining groups follow a similar sequence. After the creation of these groups, the Hosmer-Lemeshow statistic is calculated based on the Pearson chi-square statistic with the use of observed and expected frequencies. The number of observed responses of interest for the *k*th group is then, calculated as:

$$o_k = \sum_{i=1}^{n_k} y_i \qquad (5.10)$$

where,

n_k is the number of observations in the *k*th group,

y_i is the response of the *i*th observation.

The average predicted probability for the *k*th group is computed by the following equation:

$$\bar{\pi}_k = \sum_{i=1}^{n_k} \frac{\hat{\pi}_i}{n_k} \qquad (5.11)$$

where,

$\hat{\pi}_i$ is the predicted probability for the *i*th observation.

Finally, the Hosmer-Lemeshow statistic is calculated by the following equation:

$$H_L = \hat{C} = \sum_{k=1}^{g} \frac{(o_k - n_k \bar{\pi}_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)} \qquad (5.12)$$

with a (g-2) degree of freedom.

When the Pearson chi-square statistic is significant, the observed and predicted probabilities deviate. Therefore, in order to gain a model with a satisfactory goodness-of-fit, the p value should be larger than 0.05 at 5% significant level, which means that the observed and predicted probabilities match well.

5.3.2 Linear Regression Model

In this chapter, linear regression model is used to model the associations between accident costs and contributory factors. In the linear regression model, the dependent variable is the accident costs which is calculated as the total injury costs (TIC) of all casualties of each accident. Details of accident costs can be referred to Chapter 4. The independent variables are the contributory factors. In the linear regression model, the associations between the accident costs and contributory factors can be formulated by Equation (*5.13*):

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots \beta_m x_m + \varepsilon \qquad (5.13)$$

where,

Y is the dependent variable which is the accident cost,

x_j is the jth independent variable which is the contributory factor,

β_j is the corresponding coefficient of the variable (for j = 1, 2, 3, …., m),

m is the number of independent variables or contributory factors,

ε is the disturbance term.

In order to estimate the coefficients β_j for each contributory factor, the least squares estimation method is employed. This estimation method is often referred to as “ordinary least squares” or OLS.

5.4. Estimated Results of the Regression Models and Discussions

5.4.1 Binary Logistic Regression Models Results

The associations between driver, passenger, and pedestrian casualty injury severity and casualty related factors, environmental factors, site factors and vehicle factor were estimated. The estimated results of the associations in terms of the odds ratio are presented as follows. The implications of estimated odds ratios of different contributory factors are discussed and compared.

Estimated results for driver casualties

Before estimate the coefficients of each contributory factor of regression models, a Chi-Square test has been conducted to statistically test the correlation between injury severity of driver casualty and each factor. Table 5.4 shows the Chi-Square test results. From Table 5.4, it can been seen that the contributory factors of age, seat belt or crash helmet worn, day of week, time of day, rain conditions, speed limit, traffic congestion, junction control type, road type and class of vehicle driver casualty drives are statistically correlated with the injury severity of driver casualty.

Table 5.4 Results of Chi-Square test for the correlation between injury severity of driver casualty and each contributory factor

Factor	X^2	degree of freedom	p value
Age (years)	45.62*	2	.000
Sex	1.753	1	.186
Seat Belt or Crash Helmet Worn	22.247*	1	.000
Day of Week	5.003*	1	.025
Time of Day	69.582*	3	.000
Rain Conditions	14.96*	2	.001
Speed Limit	49.437*	2	.000
Traffic Congestion	13.77*	2	.001
Road Surface Condition	2.431	1	.119
Junction Control Type	123.078*	3	.000
Road Type	19.271*	2	.000
Class of vehicle driver casualty drivers	235.824*	6	.000
Year	12.956*	2	.002

*Statistically significant at the 5% level

**Statistically significant at the 10% level

Table 5.5 shows the results of the odds ratio estimation for driver casualties. As mentioned in the methodology section of this chapter (Section 5.3), the odds ratio can be a measurement reflecting associations between contributory factors and injury severity. An odds ratio that is larger than 1 indicates that the concerned attribute (of the particular contributory factor) tends to generate injuries with higher severity, and vice versa.

This model has a high goodness-of-fit (with a Hosemer-Lemeshow statistic of 3.192). The contributory factors of age, sex, seat belt or crash helmet worn, day of week, time of day, rain conditions, speed limit, junction control type, road type, class of vehicle casualty drives and year of accident occurred were found to significantly influence the injury severity measurement of driver casualties at the 5% or 10% level.

The following attributes (of contributory factors) lead to a significant lower degree of injury severity of driver casualty: wearing a seat belt or crash helmet (0.826); Monday-Friday (0.923); 7:00 am-9:59 am (0.719), 10:00 am-3:59 pm (0.661), and 4:00 pm-6:59 pm (0.632); light rain (0.784); with a speed limit below 50 km/h (0.588) and of 50 km/h (0.682); traffic signal (0.563), other control types (0.679), no control (0.573); and one-way carriageway (0.847). In contrast, the following attributes lead to a significantly higher degree of severity of pedestrian injuries at junction locations: above 65 years old (2.172); male (1.057); motorcycle (2.146), goods vehicle (1.793), van and light bus (1.389), and others (1.707); and year 2007 (1.173).

With regard to casualty related factors, the fatal and serious injury risk to older drivers (Age above 65) was more than twice the rate for younger adults (the odds ratio is 2.172). This result is expected as it is more likely that the older driver has weaker driving ability due to a poorer physical condition and slower reactions. This result also implies that some policies for older drivers should be proposed to reduce the occurrence of fatal accidents, for example, request older drivers to attend driving training programme or to submit health examination report to prove their ability for driving more frequently (similar findings were reported in the studies of Owsley et al. (2004) and Kostyniuk and Shope (2003)). Male drivers in collisions were found to have a slightly higher probability of death or serious injury than female drivers in a similar situation. This is because male drivers are more aggressive and tend to drive faster than female drivers. Wearing a seat belt or crash helmet is found to reduce driver fatalities during crashes (with an odds ratio of 0.826).

From Table 5.5, it can be seen that driver casualties from crashes occurring during the daytime have a lower degree of injury severity (a similar conclusion was also drawn by the studies of Abdel-Aty, 2003 and Tay and Rifaat, 2007). The probabilities of drivers being killed or seriously injured are greater when driving at night, owing to poorer visual conditions, and a tendency for drivers to be careless about normal road safety rules such as speed limits or junction signals.

Table 5.5 Estimated results for driver casualties

Factor	Attribute	Control	p value	Odd ratio
Age (years)	Under 15	15-65	.000	.538*
	Above 65		.000	**2.172***
Sex	Male	Female	.379	1.057
Seat Belt or Crash Helmet Worn	Yes	No	.008	**.826***
Day of Week	Monday-Friday	Weekend	.056	.923**
Time of Day	7:00am-9:59am	7:00pm-6:59am	.000	.719*
	10:00am-3:59pm		.000	.661*
	4:00pm-6:59pm		.000	.632*
Rain Conditions	Light Rain	No Rain	.013	.784*
	Heavy Rain		.943	.990
Speed Limit	Below 50km/h	Above 50km/h	.000	.588*
	50km/h		.000	.682*
Traffic Congestion	Severe Congestion	None	.106	.901
	Moderate Congestion		.130	.909
Road Surface Condition	Wet	Dry	.872	1.014
Junction Control Type	Traffic Signal	Non-junction	.000	.563*
	Other Control Types		.000	.679*
	No Control		.000	.573*
Road Type	One-way Carriageway	Multi-/dual Carriageway	.001	.847*
	Two-way Carriageway		.814	1.013
Class of Vehicle Driver Casualty Drives	Motorcycle	Private Car	.000	**2.146***
	Goods Vehicle		.000	1.793*
	Van and Light Bus		.000	1.389*
	Bus		.424	.861
	Taxi		.928	1.007
	Others		.000	1.707*
Year	2007	2009	.001	1.173*
	2008		.230	1.060
Constant			.000	.249*
Number of Observations				25649
Hosmer-Lemeshow Statistic			.922	3.192

*Statistically significant at the 5% level

**Statistically significant at the 10% level

Drivers in accidents in light rain conditions are less at risk of suffering death or serious injuries than those in accidents in dry conditions. This result is unexpected as the crash risk is usually higher under rain condition because the visual condition is poorer and the road pavement is more slippery. Two reasons are given to explain this unexpected result: 1. as mentioned above, the classification of the strength of the wet conditions is subjectively determined. Observation may be biased which would have some bearing on the estimation results; 2. the driver may tend to be more attentive under light rain conditions and to drive more slowly. This may reduce the possibility and consequences of crashes.

With regard to site factors, accidents occurring at locations with a higher speed limit result in significantly more severe driver casualties. This is expected, as the average vehicular speed is much higher under high speed limit circumstance. This circumstance, in turn, determines the collision speed and crash consequence. Accidents occurring at junction locations and one-way carriageways are found to result in driver casualties with a lower degree of injury severity, whereas accidents occurring at non-junction locations and multi-/dual carriageways result in higher casualty severity. Again, this is probably because the average vehicular speed at junction locations and one-way carriageways is slower. For crashes occurring at junction locations with traffic signal controls other than other control types or no control, the risk of fatality and/or seriously injured casualty is lower. This is probably because of poor traffic aids and traffic violations at junction locations with other control types or no control at all.

Regarding vehicle factor, motorcycles, private cars, vans and light buses, large buses, and taxis are found to significantly affect the injury severity of driver casualties. From Table 5.5, it is observed that motorcyclists are subjected to greater danger than those driving other types of vehicles (with the largest odds ratio of 2.146). This is partly because, in Hong Kong, motorcyclists tend, on average, to travel faster and consequently frequently run risks by overtaking other vehicles. In addition, motorcycles are less stable, in that they suffer from the suction generated by passing large-size vehicles. Hence, motorcyclists, when driving at fast speeds and unprepared for this destabilizing effect and also being less protected than other drivers, are at greater risk of severe casualty than drivers of other types of vehicles.

Regarding the "year" factor, it is noted in Table 5.5 that as the year passes, severe driver casualty risk tends to decrease. This is possibly due to the road safety policies and/or remedial measures implemented by TD in Hong Kong. One of these policies and/or remedial measures is MASS action (Multiple Application of Standard Solutions to particular accident sites). The general solutions currently in use are the application of anti-skid treatment and street lighting improvements for skidding on wet surfaces and accidents occurring at night time respectively. The other policy is the accident black site investigation programme. There are around 150 sites are investigated each year and some small-scale and low-cost schemes are proposed to improve the road safety conditions at individual locations.

Estimated results for passenger casualties

Table 5.6 presents the results of Chi-Square test for the correlation between injury severity of passenger casualty and each contributory factor. It can be observed that the injury severity is statistically influenced by the contributory factors of age, seat belt or crash helmet worn, seat occupied, time of day, rain conditions, speed limit, traffic congestion, road surface condition, junction control type, road type, class of vehicle passenger casualty takes and year when accident occurred.

Table 5.6 Results of Chi-Square test for the correlation between injury severity of passenger casualty and each contributory factor

Factor	X^2	degree of freedom	p value
Age (years)	220.437*	2	.000
Sex	0.111	1	.739
Seat Belt or Crash Helmet Worn	41.056*	1	.000
Seat Occupied	32.381*	3	.000
Day of Week	0.004	1	.953
Time of Day	24.22*	3	.000
Rain Conditions	5.271**	2	.072
Speed Limit	7.377*	2	.025
Traffic Congestion	15.078*	2	.001
Road Surface Condition	1.318*	1	.251
Junction Control Type	42.772*	3	.000
Road Type	1.758*	2	.415
Class of Vehicle Passenger Casualty takes	38.89*	6	.000
Year	4.829**	2	.089

Passenger casualty estimation results are given in Table 5.7. It was found that the contributory factors of age, seat belt or crash helmet worn, time of day, rain condition, speed limit, traffic congestion, road surface condition, junction control type, class of vehicle passenger casualty takes and year are all significant at the 5% or 10% level. This model also, has a high goodness-of-fit (with a Hosemer-Lemeshow statistic of 10.999).

From Table 5.7, several estimation results for passenger casualties are found to be similar to those for driver casualties. Older passengers (Age above 65) are at a much higher risk of being killed or seriously injured than the younger adults. Passengers who wear seat belts or crash helmets are found to have a lower probability of fatality and seriously injured casualty than those who do not wear seat belts or crash helmets.

Table 5.7 Estimated results for passenger casualties

Factor	Attribute	Control	p value	Odd ratio
Age (years)	Under 15	15-65	.000	.298*
	Above 65		.000	2.102*
Sex	Male	Female	.217	1.071
Seat Belt or Crash Helmet Worn	Yes	No	.000	.683*
Seat Occupied	Front Nearside	Others	.592	1.084
	Rear		.780	1.039
	Standing		.102	1.277
Weekday	Monday-Friday	Weekend	.845	1.012
Time	7:00am-9:59am	7:00pm-6:59am	.192	.902
	10:00am-3:59pm		.000	.726*
	4:00pm-6:59pm		.000	.729*
Rain Conditions	Light Raining	Not Raining	.009	**.692***
	Heavy Raining		.905	1.023
Speed Limit	Below 50km/h	Above 50km/h	.349	.762
	50km/h		.049	.861*
Traffic Congestion	Severe Congestion	None	.009	**.797***
	Moderate Congestion		.027	**.834***
Road Surface Condition	Wet	Dry	.009	**1.393***
Junction Control Type	Traffic Signal	Non-junction	.000	.668*
	Other Control Types		.003	.645*
	No Control		.003	.667*
Road Type	One-way Carriageway	Multi-/dual carriageway	.702	.975
	Two-way carriageway		.180	1.101
Class of Vehicle Passenger Casualty takes	Motorcycle	Bus	.000	1.880*
	Private Car		.035	1.236*
	Goods Vehicle		.007	1.785*
	Van and Light Bus		.644	.963
	Taxi		.303	.895
	Others		.156	.838
Year	2007	2009	.549	1.041
	2008		.014	1.175*
Constant			.000	.134*
Number of Observations				19314
Hosmer-Lemeshow Statistic			.202	10.999

*Statistically significant at the 5% level

**Statistically significant at the 10% level

Accidents that occur at night are found to have a higher probability of resulting in more severe passenger casualties. This is probably an effect of poorer visual conditions, higher vehicular speed and the poorer driving attention often associated with nighttime driving. Passengers are at a lower risk of fatality and serious injury if involved in accidents in light rain conditions compared to those involved in accidents in dry conditions (with an odds ratio of 0.692).

The speed limit contributory factor is also found to significantly affect the injury severity degree of passenger casualties. Passengers in accidents occurring at locations with a higher speed limit are more at risk of being killed or seriously injured. A significant association between severe passenger casualties and the traffic congestion level is found (with odds ratios of 0.797 and 0.834 for severe congestion and moderate congestion). The implication is that the traffic congestion level is negatively associated with fatal and seriously injured passenger casualty. This is because the average vehicular speed is much lower under traffic congestion condition.

Crashes that occur on roads with wet road surfaces have a higher injury risk (with an odds ratio of 1.393) than those occurring on dry roads. This is probably due to more slippery surfaces under wet conditions and a related increase in vehicle braking distances and hence skids. It is noted that vehicular speed at junction locations is slower, possibly resulting in less fatalities or seriously injured casualties.

Motorcycles, private cars and goods vehicles are found to significantly affect the severity of the injury suffered by passenger casualties. Similar to the estimation results of the driver casualty model, motorcycles, more than any other vehicle, present the greatest danger to passengers. Passengers in private cars and goods vehicles are found to be at more risk than those on buses, although this is not always the case as a bus has larger passenger capacity, therefore a larger number of potential victims. However, because of the good training and experience of bus drivers, the probability of crashing and causing severe casualties is lower than private cars. From this point of view, a bus may provide safer travel for passengers than other vehicles such as motorcycles, private cars and goods vehicles.

Estimated results for pedestrian casualties

The results of Chi-Square test for the correlation between injury severity of passenger casualty and each contributory factor are shown in Table 5.8. It can be found that the contributory factors of age, sex, pedestrian location, time of day, speed limit, traffic congestion, road surface condition, junction control type, road type and class of vehicle collided with pedestrians are statistically correlated with the injury severity of pedestrian casualty.

Table 5.8 Results of Chi-Square test for the correlation between injury severity of passenger casualty and each contributory factor

Factor	X^2	degree of freedom	p value
Age (years)	331.734*	2	.000
Sex	13.155*	1	.000
Pedestrian Location	59.259*	2	.000
Day of Week	2.015	1	.156
Time of Day	59.826*	3	.000
Rain Conditions	2.318	2	.314
Speed Limit	35.295*	2	.000
Traffic Congestion	13.607*	2	.001
Road Surface Condition	8.4*	1	.004
Junction Control Type	26.781*	3	.000
Road Type	86.917*	2	.000
Class of Vehicle Collided with Pedestrians	193.264*	6	.000
Year	3.352	2	.187

Table 5.9 presents the estimated associations between the injury severity of experienced by pedestrian casualties and the contributory factors of age, sex, pedestrian location, time of day, rain condition, speed limit, traffic congestion, road

surface condition, junction control type, road type, class of vehicle collided with pedestrians and year. These contributory factors significantly affect the severity of injury suffered by pedestrian causalities. This model is well fitted and shows agreement with a Hosemer-Lemeshow statistic of 11.533.

Female pedestrians and older pedestrians have a greater risk of being killed or seriously injured when involved in traffic accidents. This is because both female and older pedestrians usually have a weaker physical frame. Pedestrians on a crossing and within 15m of a crossing are more likely to have more severe injuries than those walking along a footpath. Pedestrians when walking along a footpath are free of direct interaction with vehicles and are consequently less likely to be subjected to driver error, as is the case of those on, or close to a crossing (Sze and Wong, 2007), unless, of course the vehicle mounts the footpath.

Accidents that occur at night and during light rain conditions are found to result in less pedestrian fatalities and serious injuries. However, pedestrians who are involved in accidents on the roads with higher speed limit, non-congested traffic conditions and wet road surfaces are at greater risk of serious injury and fatality. A higher proportion of fatal and seriously injured accidents occur on the multi-/dual carriageways, where vehicular speed is higher than average.

When investigating the vehicle factor, it was found that crashes involving goods vehicles, vans and light buses, and large buses have a significant increase in pedestrian injury severity (This finding is in agreement with that of Tay and Rifaat (2007)). As indicated above, the larger the vehicle size, the greater the area of impact and the greater the severity of pedestrian casualty. The results of vehicle collisions as regards pedestrian and driver/passenger casualties are different due to different crash impacts on all casualties involved. Drivers/passengers, inside large-size vehicles involved in collisions appear to be better protected, while pedestrians, in the same collision, are more at risk due to their smaller ability to avoid the larger impact area.

Table 5.9 Estimated results for pedestrian casualties

Factor	Attribute	Control	p value	Odd ratio
Age (years)	Under 15	15-65	.339	.929
	Above 65		.000	2.583*
Sex	Male	Female	.007	.881*
Pedestrian Location	On the Crossing	Others	.000	1.459*
	Within 15m of Crossing		.000	1.460*
Day of Week	Monday-Friday	Weekend	.829	1.011
Time of Day	7:00am-9:59am	7:00pm-6:59am	.001	.785*
	10:00am-3:59pm		.000	.598*
	4:00pm-6:59pm		.000	.642*
Rain Conditions	Light Rain	No Rain	.024	.721*
	Heavy Rain		.104	.676
Speed Limit	Below 50km/h	Above 50km/h	.000	.325*
	50km/h		.000	.317*
Traffic Congestion	Severe Congestion	None	.002	.800*
	Moderate Congestion		.000	.726*
Road Surface Condition	Wet	Dry	.009	1.419*
Junction Control Type	Traffic Signal	Non-junction	.287	.931
	Other Control Types		.545	.938
	No Control		.000	.723*
Road Type	One-way Carriageway	Multi-/dual Carriageway	.000	.649*
	Two-way Carriageway		.848	.987
Class of Vehicle Collided with Pedestrians	Motorcycle	Taxi	.065	1.264**
	Private Car		.437	.950
	Goods Vehicle		.000	2.080*
	Van and Light Bus		.000	1.557*
	Bus		.000	2.156*
	Others		.001	.712*
Year	2007	2009	.068	1.110**
	2008		.901	1.007
Constant			.622	1.131
Number of Observations				11481
Hosmer-Lemeshow Statistic			.173	11.533

*Statistically significant at the 5% level

**Statistically significant at the 10% level

5.4.2 Linear Regression Model Results

The associations between accident costs and contributory factors were estimated. The estimated results of the associations in terms of the coefficients of each contributory factor are presented as follows. The implications of estimated coefficients of different contributory factors are discussed and compared.

The contributory factors of time of day, speed limit, traffic congestion, junction control type, road type, number of casualties were found to significantly influence the accident costs at the 5% or 10% level.

The following attributes (of contributory factors) lead to a significant lower accident costs: 4:00 pm-6:59 pm (-10723.136); with a speed limit below 50 km/h (-25307.956); severe congestion (-20610.677), moderate congestion (-28907.017); traffic signal (-22542.144), other control types (-12693.377); and one-way carriageway (-17786.107), multi-/dual carriageway (-28805.978). In contrast, the following attributes lead to a significantly higher accident costs: 7:00 pm-6:59 am (6953.925); with a speed limit above 50 km/h (7226.675); and no. of casualty (192690.638).

From Table 5.10, it can be seen that accidents occurring during the nighttime have a higher crash injury costs (HK$ 6953.925 greater than 10:00am-3:59pm). The probabilities of accidents occurring and casualties being killed or seriously injured are greater at night, owing to poorer visual conditions, and there is a higher tendency for drivers to be careless in control at speed limits or junction signals.

Table 5.10 Estimated results of linear regression on accident costs

Factor	Attribute	Control	Coef.(HK$)	t
Day of Week	Weekend	Monday-Friday	-1393.048	-.184
Time of Day	7:00am-9:59am	10:00am-3:59pm	20257.991	0.811
	4:00pm-6:59pm		-10723.136	-1.808**
	7:00pm-6:59am		6953.925	2.111**
Rain Conditions	Light Raining	Not Raining	-25295.253	-1.308
	Heavy Raining		-37028.814	-1.224
Speed Limit	Below 50km/h	50km/h	-25307.956	-2.070*
	Above 50km/h		7226.675	2.329*
Traffic Congestion	Severe Congestion	None	-20610.677	-1.899**
	Moderate Congestion		-28907.017	-2.715*
Road Surface Condition	Wet	Dry	13457.840	.763
Junction Control Type	No Control	Non-junction	3063.206	.219
	Traffic Signal		-22542.144	-1.851**
	Other Control Types		-12693.377	-1.954**
Road Type	One-way Carriageway	Multi-/dual Carriageway	-17786.107	-2.202*
	Two-way Carriageway		-28805.978	-2.941*
Year	2008	2007	-185.009	-.026
	2009		24744.633	1.627
Number of Casualty			192690.638	63.699*
Constant			-88855.876	-9.027*

*Statistically significant at the 5% level

**Statistically significant at the 10% level

The speed limit contributory factor is also found to significantly affect the degree of accident costs. Accidents occurring at locations with a higher speed limit (Above 50 km/h) result in significantly much higher accident costs (HK$ 7226.675 greater than

50 km/h). This is expected, as the average vehicular speed is much higher under high speed limit circumstance. Therefore, the probabilities of accidents, especially for those accidents causing fatalities and seriously injured casualties, are much greater.

Accidents occurring at locations with congestion, junction locations, one-way carriageways and two-way carriageways are found to result in less crash injury costs, whereas accidents occurring at locations without congestion, non-junction locations and multi-/dual carriageways result in greater accident costs. This is probably because the average vehicular speed at congestion locations, junction locations, one-way carriageways and two-way carriageways is slower. Therefore, the probabilities of accidents with more casualties and seriously injured casualties are much lower.

The factor of the number of casualties is also found to have a positive effect on the accident costs. The increase of one casualty would result in HK$ 192690.638 more accident costs.

5.5. Summary

The associations between casualty severity and accident costs, and casualty related factors, environmental factors, site factors and vehicle factor have been estimated in this chapter. Three binary regression models have been developed for driver, passenger and pedestrian casualties and one linear regression model has been developed for accident costs based on the traffic accident database from 2007 to 2009 inclusive for Hong Kong.

From the estimation results of injury severity models, several contributory factors were found to significantly affect the injury severity of casualties. The contributory factors that led to a noticeably higher injury severity of all casualties were aged over 65, involved in daytime crashes, and involved in crashes on a road segment with a speed limit above 50km/h and on multi-/dual carriageways (a similar conclusion was also drawn by the studies of Abdel-Aty, 2003, Tay and Rifaat, 2007 and Sez and Wong, 2007).

Male drivers were at more risk of being killed and seriously injured due to their aggressive driving behavior. This conclusion is similar to that drawn from Yau (2004), while different from that drawn from Abdel-Aty (2003) in which female drivers are found to have a higher probability of being seriously injured. The former one (Yau, 2004) concerned with traffic accidents in Hong Kong, while the latter one (Abdel-Aty, 2003) focused on Central Florida area in America. The difference between findings may attribute to different driving behavior in these two areas.

Wearing seat belt or crash helmet helped to protect drivers and passengers from accidents with less serious injuries and fatalities. Drivers of motorcycles and their passengers were found to be at the most significant risk regarding injury severity. This is mainly due to many motorcyclists' risk-taking behavior (such as higher driving speed and overtaking other vehicles), lack of physical protection other than "leathers" and less stability when driving (This finding is in agreement with that of Yau (2004)).

Female pedestrians and those who are on the crossing or within 15m of a crossing underwent a noticeably higher proportion of fatalities and seriously injured casualties. Accidents involving large-size vehicles have a higher probability of causing severe pedestrian casualties. This is because of the larger impact area and collision force inflicted by large-size vehicles, both of which can result in more severe damage to a pedestrian.

From the estimation results of accident costs regression model, it was found that accidents occurring during the nighttime and at locations with higher speed limit have a greater crash injuring costs. However, accidents occurring at locations with congestion, junction locations, one-way carriageways and two-way carriageways are found to result in less crash injuring costs. The factor of the number of casualties was also found to have a positive effect on the accident costs.

In this chapter, several speed related factors, such as speed limit and traffic congestion condition were found to significantly affect the injury severity of casualties and accident costs. It is of interest to further investigate the relationships between the occurrence of accidents and traffic speed, such as how and to what degree accidents affect traffic speed. Further information on this issue is presented and discussed in Chapter 6.

6. Investigation of Accident Effects on Traffic Speed: Before-After Analysis

6.1. Introduction

In Chapter 5, several speed related factors, such as speed limit and traffic congestion have been found to significantly affect accident severity. The focus of this chapter is the effects of accidents on traffic speed as indicated in Chapter 1. It is tried to identify factors influencing the effects of accidents on traffic speed and to analyze their influences quantitatively in this chapter.

Accident effects are used in many studies as criteria for accident detection as indicated in Chapter 2. However, few studies have been carried out to specifically analyze how accidents affect traffic speed and precisely what factors influence the accident effects. Factors such as accident severity and the time of day during which the accident happens are likely to influence the accident effects on traffic speed to different degrees.

The literature review presented in Chapter 2 reveals that a Before-After analysis method was introduced to study road safety by Hauer (1997). As stated in Chapter 2, since its introduction, this method has been applied in the analysis of several road safety problems, including the study of the effects of changed speed limits on road safety, the evaluation of the effectiveness of marking blue cycle crossings, and the assessment of the effectiveness of a training programme for older pedestrians (Wong et al., 2005; Jensen, 2008; Dommes et al., 2012).

In this chapter, the Before-After analysis method is applied in the investigation of accident effects on traffic speed. Factors that would affect the degree of accident effects on traffic speed are identified. Regression models are adopted to quantify the influences of these contributory factors on accident effects.

A brief description of the data used in this chapter is presented in Section 6.2, followed by a presentation in Section 6.3, of the Before-After analysis as applied to analyze accident effects on traffic speed. The results and findings of the case study are discussed in Section 6.3. A summary of the key findings is then given in Section 6.4.

6.2. Data Description

The speed data used in the Before-After analysis are extracted from the Hong Kong Journey Time Indication System (JTIS) for the urban area, while the relevant accident data are extracted from the Hong Kong Traffic Accident Database System (TRADS). These two systems and the extracted speed and accident data are described below.

Speed data

JTIS was firstly commissioned on the major roads of Hong Kong Island by the Transport Department (TD) of the Hong Kong Government in 2003. The aim of JTIS is to provide road users with real-time cross-harbour traffic information. In 2008, JTIS was re-developed by means of a novel algorithm for general path journey time estimation. Path journey times are estimated based on the real-time and the offline travel time data in the algorithm. The new JTIS was launched in mid-2009. In mid-2010, TD commissioned new journey time indicators to be installed in Kowloon urban area and Eastern District on Hong Kong Island (Tam and Lam, 2011).

Figure 6.1 shows the location of the whole JTIS network in the territory of Hong Kong, including the JTIS network on Hong Kong Island and the JTIS network in Kowloon. The speed limit of the whole JTIS network ranges from 50km/h to 80km/h. Approximately half of the roads in this network are designed to have a speed limit of 70km/h.

The JTIS network has a total of 221 links, 95 links on Hong Kong Island and 126 links in Kowloon. Travel time estimates of each link, together with speed estimates (obtained by dividing link length by estimated travel time), are provided every two minutes by the JTIS (see website: http://www.td.gov.hk/en/transport_in_hong_kong/its/its_achievements/journey_time_indication_system_/index.html). Speed data before and after the studied accidents are extracted from this system for analysis in this chapter.

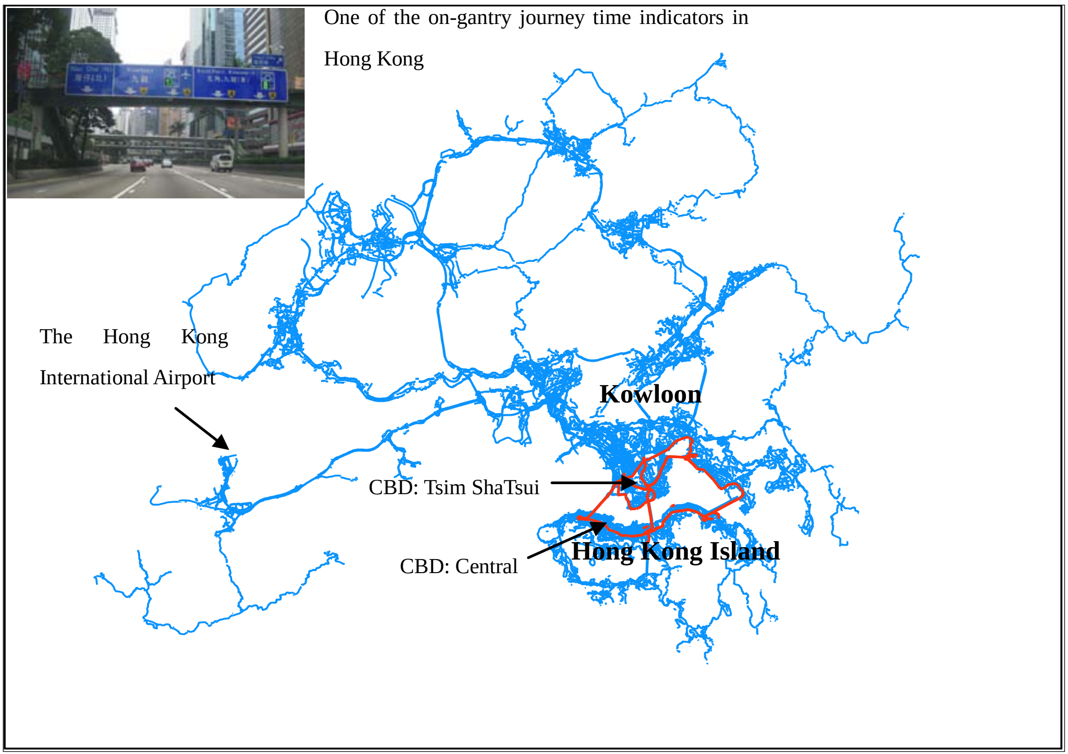


Figure 6.1 Location of JTIS network in Hong Kong

Accident data

Accidents which have occurred since 2009 (the launch of JITS) are considered in this chapter. Data relating to these accidents are extracted from TRADS. All these accidents are matched to the Hong Kong road network and JTIS network using ArcGIS tools, and according to the coordinate information of each accident provided by TRADS. To exactly match those accidents on the JTIS network, other information is also taken into consideration. This information includes vehicle direction and precise accident location. A total of 313 accidents were found to occur on the JTIS network during the study period: September 2009 to December 2010.

Table 6.1 presents the accident distribution in terms of accident severity and location during the study period. A total of 313 accidents took place during the study period, one of which was fatal and 20 involved serious injury. 139 accidents occurred on Hong Kong Island and 174 accidents occurred in Kowloon.

Table 6.1 Number of accidents during the study period

Accident Severity	**Hong Kong Island**		**Kowloon**	**Total**
	2009.9-2009.12	2010.01-2010.12	2010.06-2010.12	
Fatal	1	0	0	1
Seriously Injured	1	8	11	20
Slightly Injured	18	111	163	292
Total	20	119	174	313

Figure 6.2 shows all accident records on the JTIS network during the study period. All accidents involving fatalities and serious injuries recorded on the JTIS network during the study period are presented in Figure 6.3.

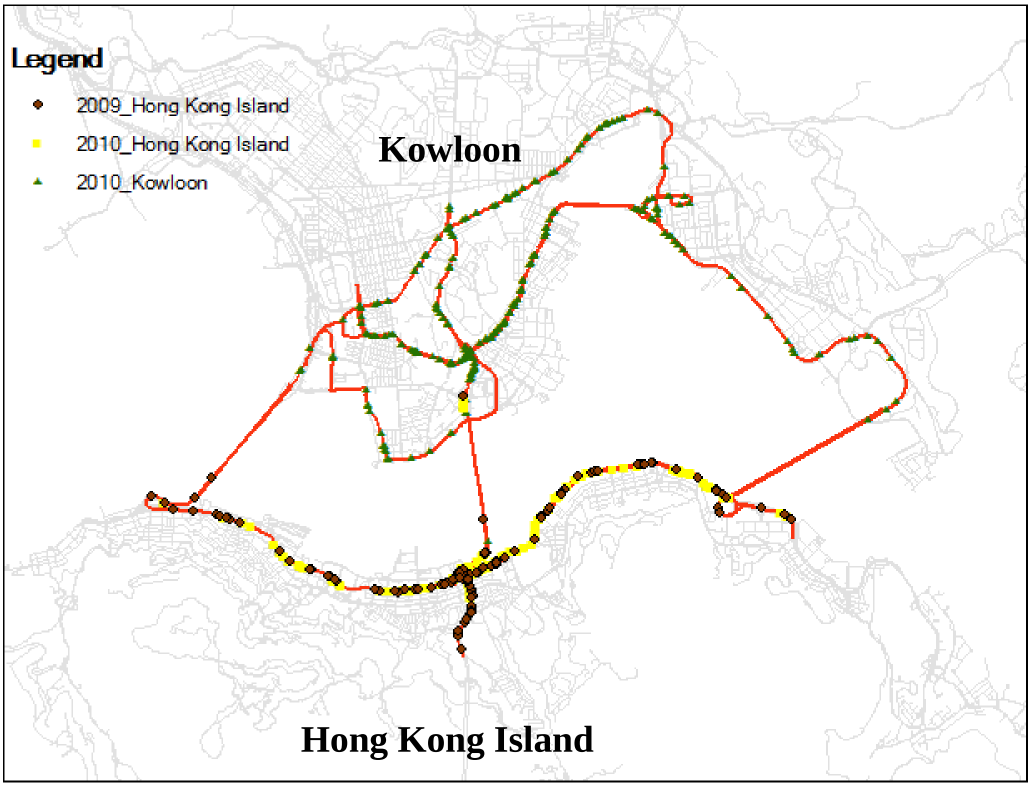


Figure 6.2 All accidents on the JTIS network

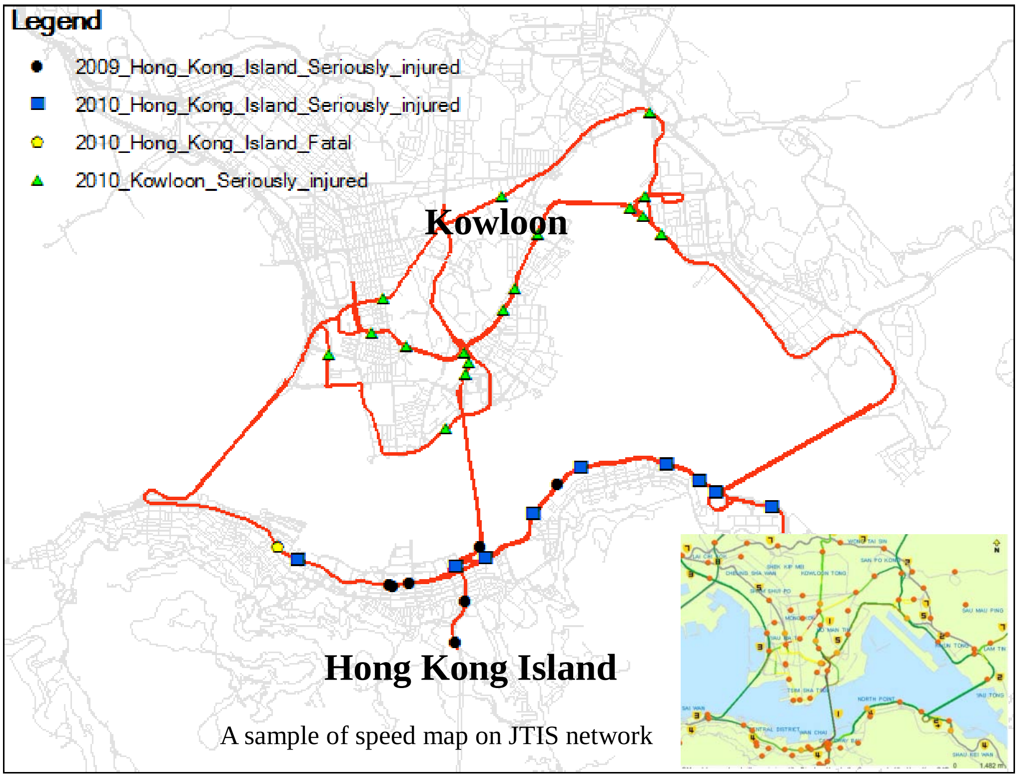


Figure 6.3 Fatal and seriously injured accidents on the JTIS network

6.3. Accident Effects on Traffic Speed Using the Before-After Analysis

In this section, the "Comparison Group" Before-After analysis method is firstly described. This method is adopted to analyze accident effects on traffic speed in this chapter. The results of the Before-After analysis on accident effects are then presented and discussed.

6.3.1 Before-After Analysis

When a road accident occurs, traffic speed decreases due to a reduction in roadway capacity. The effects of the accident on the ensuing road speed can be quantified as reduction in traffic speed. To illustrate, let A be the road speed after the occurrence of the accident and B be the road speed in what is estimated to be the 'after' time period had the accident not occurred. The main idea of Before-After analysis on accident effects is to compare A with B and to measure the speed reduction from B to A.

A. the road speed in the 'after' period when the accident happened	**With**	B. What would have been the road speed in the 'after' period had the accident not happened

The Before-After analysis method proposed by Hauer (1997), requires two tasks to be conducted to enable the accident effects on traffic speed to be measured:

Task 1: **Predict** what would have been the road speed of a specific area in the 'after' period had the accident not happened.

Task 2: **Estimate** the road speed of the area where the accident did happen in the ‘after’ period.

Let π be the **predicted** road speed of the area in the ‘after’ period had the accident not happened.

Let λ be the **estimated** road speed of the area where the accident did happen in the ‘after’ period.

The accident effects on traffic speed can be measured in two ways:

$\delta = \lambda - \pi$: the difference between “the road speed of the area where the accident did happen” and “what would have been the road speed of the same area had the accident not happened” in the ‘after’ period, or

$\theta = \lambda / \pi$: the ratio of “the road speed of the area where the accident did happen” to “what would have been the road speed of the same area had the accident not happened” in the ‘after’ period.

Note the following example. The time series of speed before and after the occurrence of an accident in Hong Kong is taken as an illustration (shown in Figure 6.4). In this chapter, it is assumed that the estimated speed λ equals the observed speed in the ‘after’ period when the accident did happen. In fact, the observed speeds estimated by JTIS have been validated with observation data satisfactorily. It was reported that the estimation errors are less than 20% with a probability of 95%.

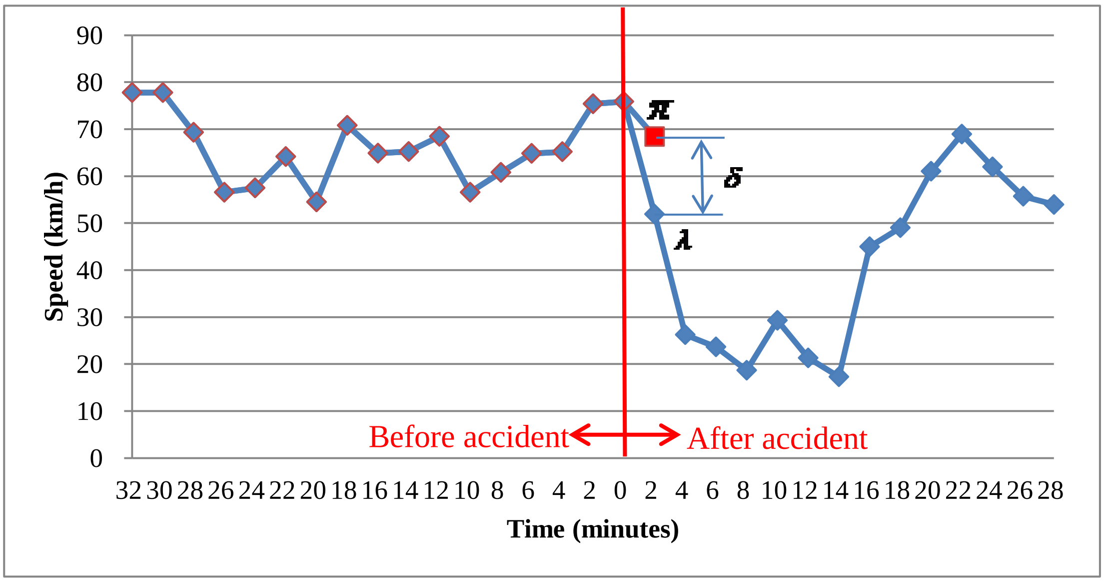


Figure 6.4 An example of time series of speed before and after accident

Many methods are available to predict what would have been the road speed in the ‘after’ period had the accident not happened (Hauer, 1997). One of these prediction methods is “Averaging”. The speed π represented as a square in Figure 6.4 is predicted with the use of the “Averaging” method. The speed π is calculated as the average value of the last ten minutes of speed. The speed difference δ is negative in this case and represents a reduction in traffic speed after the accident happened.

As stated above, this chapter focuses mainly on an analysis of accident effects using Before-After analysis rather than proposing new prediction method for speed π. To achieve the prediction of speed π, one existing and widely adopted prediction method is applied. This method is usually named as the “Comparison Group” method. Details of the Before-After analysis based on the “Comparison Group” method are discussed below.

“Comparison Group” Before-After analysis:

In this chapter, accident cases are defined as an accident group, while a comparison group is defined as a group of non-accident cases in which the situations are similar to individual accident cases. The comparison group involves non-accident cases occurring on the same weekday, at the same time and at the same location where the accident cases had originally taken place. For example, if an accident occurred at 7:30am on Thursday on road A, the corresponding non-accident case in the comparison group would be 7:30am on certain Thursday (for example, one week before or after) on road A but without the presence of an accident.

The use of the Comparison Group method is based on a hypothesis. The hypothesis is that the change in traffic speed of a non-accident case from 'before' to 'after' periods at the study site is indicative of how the speed of the corresponding accident case would have changed if the accident has not been happened. The hypothesis is based on two assumptions: (1). all factors at the study site other than the occurrence of accident affecting the speed have changed in the same manner from 'before' to 'after' periods for both the accident group and comparison group; and (2). the change in the "sundry" factors affects the speed of the accident group and comparison group in the same way (following the logic of Comparison Group method proposed by Hauer (1997)).

To be precise, let

μ be the **estimated** road speed in the comparison group in the 'before' period;

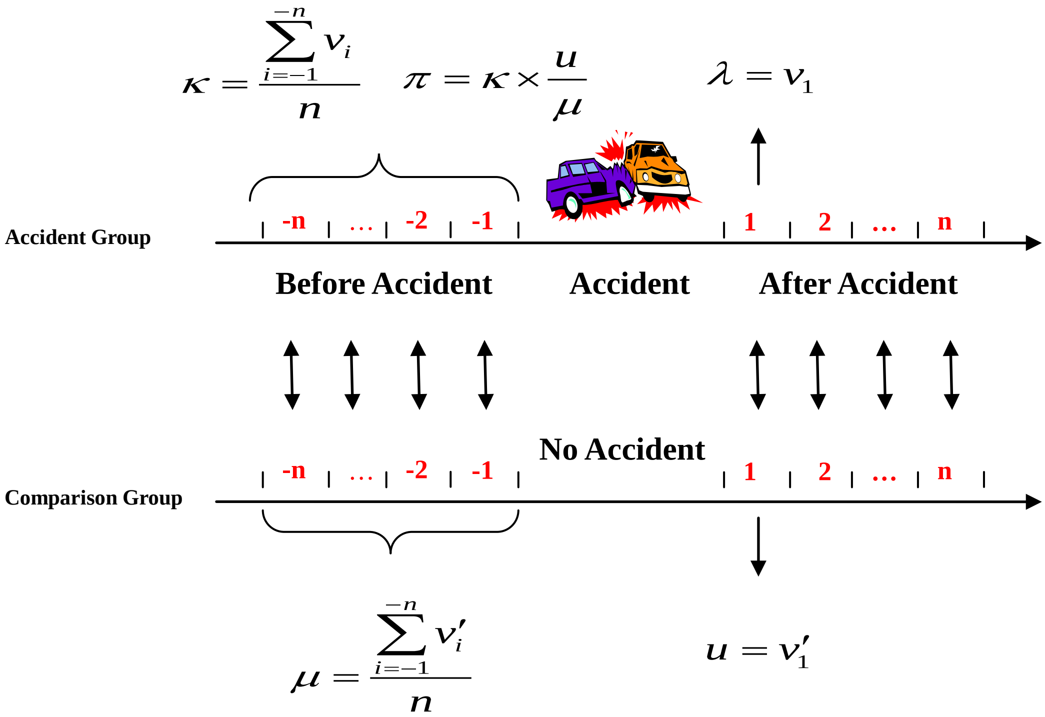
u be the **estimated** road speed in the comparison group in the 'after' period;

κ be the **estimated** road speed in the accident group in the 'before' period.

The Comparison Group method is based on the hypothesis that the ratio of estimated ‘before’ to ‘after’ road speed would be the same in the accident group and comparison group if the accident had not occurred,:

$$r_C = r_A = \kappa / \pi = \mu / u \tag{6.1}$$

Therefore, the design and definition of “Comparison Group” Before-After analysis is proposed and presented in Figure 6.5.



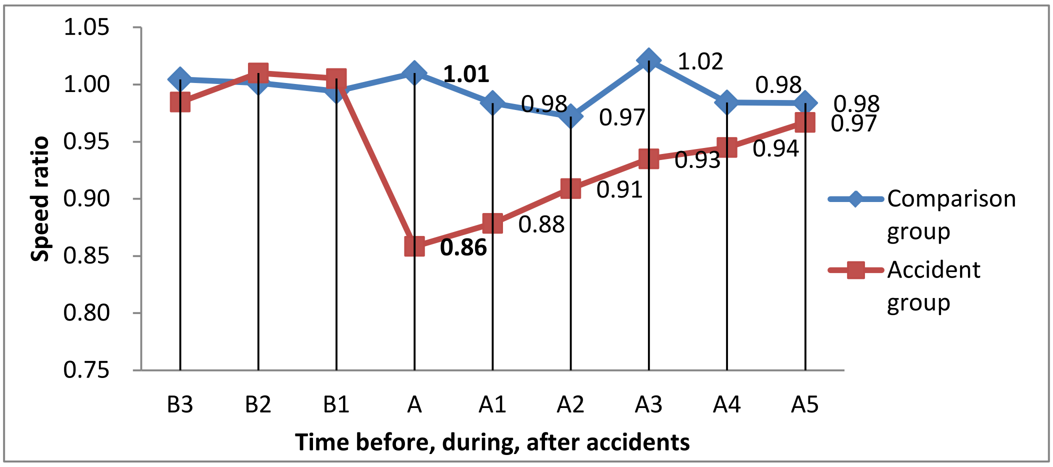
Notes: -n is the n^th time interval before accident (one time interval is two minutes);

n is the n^th time interval after accident.

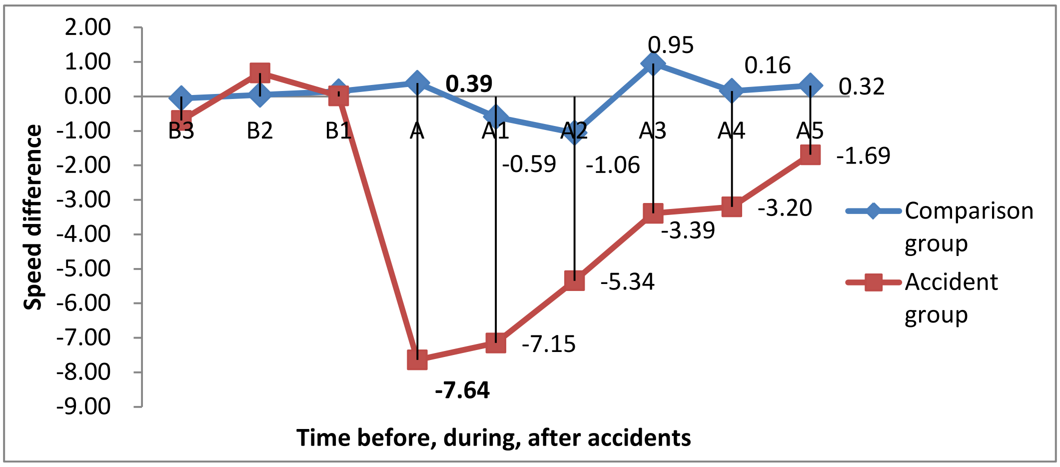
Figure 6.5 The design and definition of “Comparison Group” Before-After analysis

6.3.2 Results of Before-After Analysis on Accident Effects

In this section, the results of the Before-After analysis on accident effects are presented and discussed. To begin with, results are shown to illustrate how and to what degree accidents impact traffic speed. Figure 6.6 presents the average speed ratio (θ) and difference (δ) for the accident group and comparison group by the relevant time intervals.



(a) Speed ratio



(b) Speed difference

Notes: B1-B3: the first to the third time interval before accidents

A: The time period when accidents occurred

A1-A3: the first to the third time interval after accidents

Figure 6.6 Average speed ratio and difference for the accident and comparison group by time interval

From Figure 6.6 (a), it can be observed that both accident and comparison group curves change over various time intervals. In the accident group curve, point A, which represents the time period when accidents occurred, has a minimum speed ratio value of 0.86. This value is far lower than those experienced during the time period before accidents and the value of point A in the curve of comparison group (1.01). The accident occurrence results in blocking one or more lanes and therefore a reduction in road capacity. The sudden reduction in road capacity may decrease the number of vehicles that can pass and the vehicular speed when passing (This finding is in agreement with that of Luk et al. (2001)). However, after accidents, the speed ratio value may gradually increase as the traffic police may take action to clear all debris and so enable traffic speed back to normal condition.

From the above, it is concluded that the occurrence of accidents can cause significant adverse impacts on traffic speed. The accident effects are quantified as changes in speed ratio or speed difference before and after accidents. Further investigation is necessary to find how and to what degree specific factors influence accident effects on traffic speed. The corresponding results are presented and discussed below.

It was found that accident effects on traffic speed were influenced by the following factors: accident severity, accident time and accident location. Fatal and seriously injured accidents were found to have larger adverse effects on traffic speed than slightly injured accidents. Under heavy traffic flow conditions, such as peak-hour periods, accidents were found to cause greater reduction in traffic speed. The length of the link (the road segments on the JTIS network which is introduced in Section 6.2) on which accidents occurred was also found to be associated with the degree of speed decrease.

Figures 6.7 and 6.8 present the respective cumulative percentage speed ratio (θ) and speed difference (δ) curves for fatal and seriously injured accidents and slightly injured accidents.

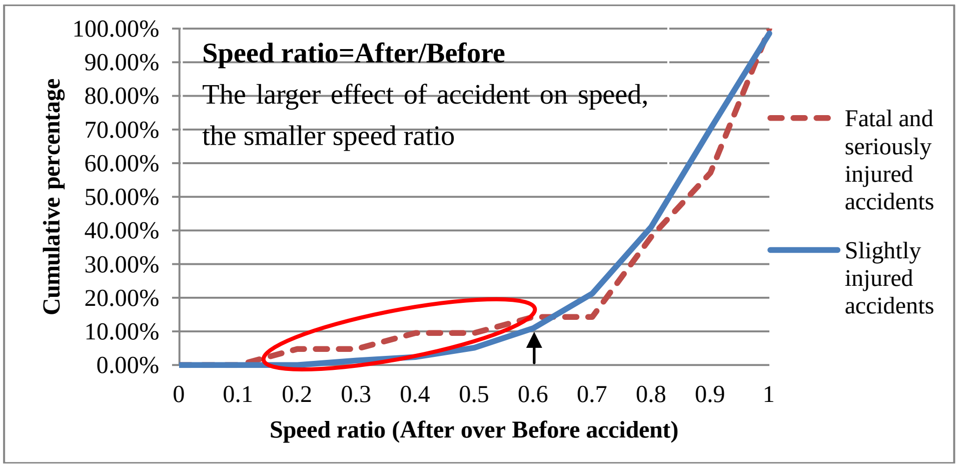


Figure 6.7 Cumulative percentage curve of speed ratio

Figure 6.7 shows the comparison of cumulative percentage speed ratio curves between fatal and seriously injured accidents and slightly injured accidents. From Figure 6.7, it is observed that the cumulative percentage of speed ratio smaller than 0.6 for fatal and seriously injured accidents is larger than that for slightly injured accidents, 15% for the former and 10% for the latter. This result implies that the occurrences of fatal and seriously injured accidents are more likely to cause larger reduction in speed.

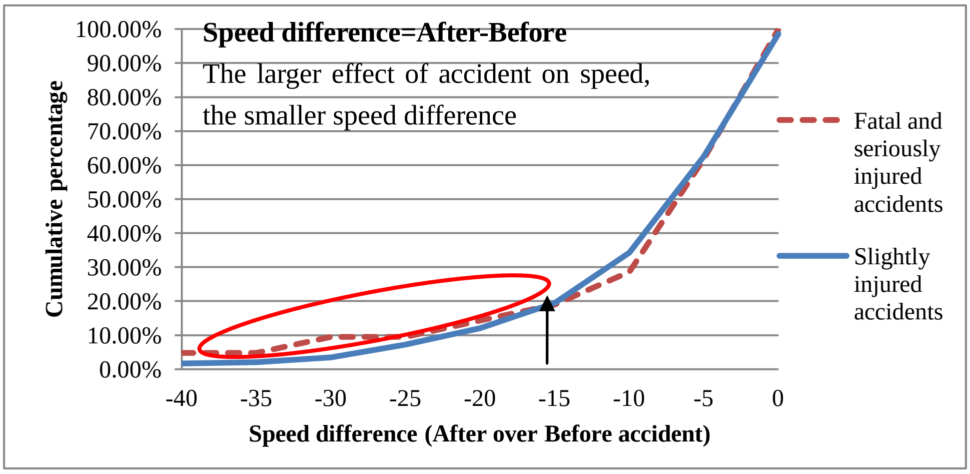
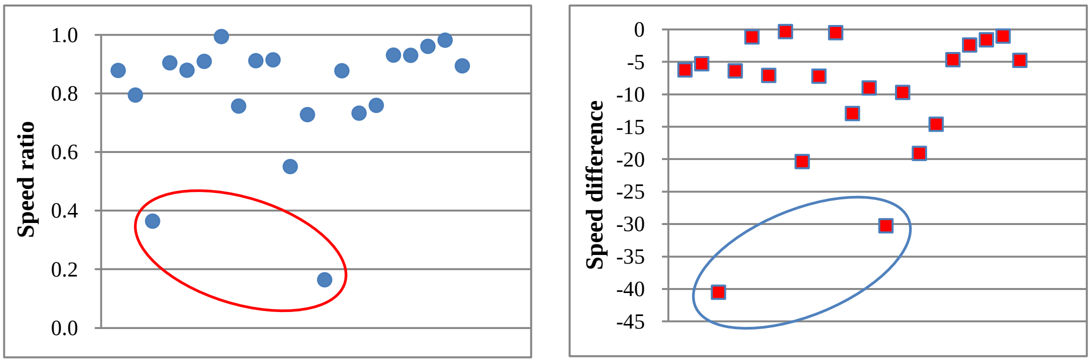


Figure 6.8 Cumulative percentage curve of speed difference

Figure 6.8 compares the cumulative percentage speed difference curves between fatal and seriously injured accidents and slightly injured accidents. It is also noted that the cumulative percentage of small speed difference (such as, smaller than -15 km/h) for fatal and seriously injured accidents is larger. This result also indicates that the occurrence of fatal and seriously injured accidents would cause more significant adverse impacts on traffic speed.

From Figures 6.7 and 6.8, it can be concluded that fatal and seriously injured accidents would cause a greater reduction in traffic speed. However, specific effect of each accident on traffic speed cannot be observed from Figures 6.7 and 6.8. To gain a clear picture of the specific effect of each accident on traffic speed, all fatal and seriously accidents are given as examples. Figure 6.9 shows the values of speed ratio and speed difference for all fatal and seriously injured accidents.

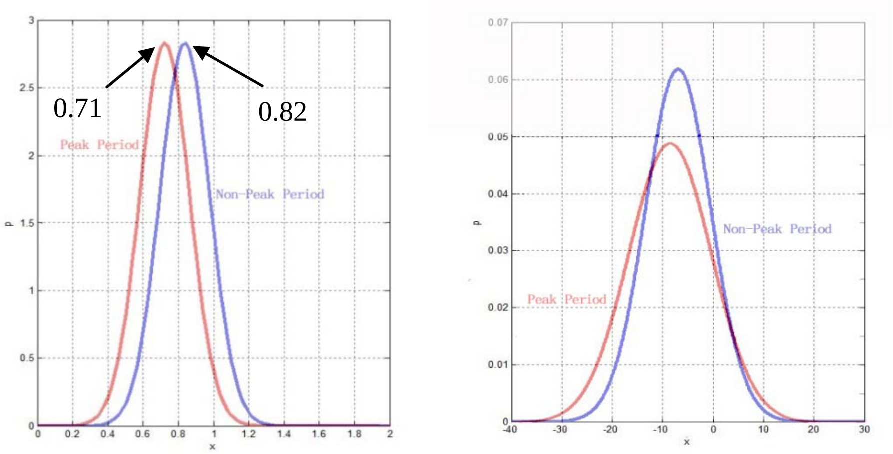


(a) Speed ratio (b) Speed difference

Figure 6.9 Distribution of speed ratio and speed difference for all fatal and seriously injured accidents

From Figure 6.9, it is observed that two speed ratio values are lower than 0.4 with one even lower than 0.2. This implies that these two accidents have caused more than a 60% speed reduction. Specifically, one accident has caused about 30km/h decrease in speed and the other has caused more than 40km/h reduction in speed (refer to Figure 6.9(b)). Two reasons can explain why these two accidents have had such significant effects on traffic speed. Firstly, both accidents are categorized as fatal and seriously injured accidents as fatalities and serious injuries have been caused. The second reason is that both accidents happened during peak-hour periods (at about 18:00). During peak-hour periods, the traffic flow is much heavier than that during non-peak-hour periods. Under heavy traffic flow condition, accident occurrences would be more likely to cause long traffic queues and such queues would take longer to dissipate owing to the heavy traffic rate.

For a more general finding regarding the association of accident effects and the time of the day during which the accident took place, further analysis was carried out. Figure 6.10 shows fitted distribution curves of speed ratio and speed difference during peak-hour periods and non-peak-hour periods (peak-hour periods are taken as 7:00am-10:00am and 17:00pm-19:00pm, while non-peak-hour periods are taken as 10:00am-17:00pm and 19:00pm-7:00am). It was found that accidents occurring during peak-hour periods resulted in a larger reduction in speed than when accidents occurred during non-peak-hour periods. From Figure 6.10, it can be seen that the mean of speed ratio curve for peak-hour periods is about 0.71 which is significantly smaller than that (approximately 0.82) for non-peak-hour periods. A similar finding, that the pre-accident traffic flow conditions have an effect on the magnitude of accident effects on traffic speed statistically, was also drawn by Mak and Fan (2006).



(a) Speed ratio (b) Speed difference

Figure 6.10 Fitted distribution curves of speed ratio and speed difference for all fatal and seriously injured accidents

The analysis of the associations between length of link on which accidents occurred and also accident effects is given below. All accidents are divided into different groups. The grouping is based on the degree of accident effects on traffic speed. For instance, accidents with a speed ratio larger than 0.6 are aggregated into one group. Tables 6.2 and 6.3 present the average length of links on which any specific group of accidents occurred.

It was found that the average link length is smaller for groups of accidents causing larger effects on traffic speed. For example, the link length for the accident group with a speed ratio larger than 0.6 is 465m and the length for the accident group with a speed ratio less than or equal to 0.6 is 321m (shown in Table 6.2). The length for the accident group with a speed difference larger than -20 is 460m and the length for the accident group with a speed difference less than or equal to -20 is 367m (shown in Table 6.3).

Table 6.2 Average length of links on which accident occurred for different speed ratio

	All				**Fatal and seriously injured**			
	Speed difference (km/h)				Speed difference (km/h)			
	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
Larger than	522	507	471	465	491	455	452	452
Less than or equal to	**417**	**365**	**366**	**321**	**348**	**336**	**154**	**154**

Table 6.3 Average length of links on which accident occurred for different speed difference

	All				**Fatal and seriously injured**			
	Speed difference (km/h)				Speed difference (km/h)			
	-5	-10	-15	-20	-5	-10	-15	-20
Larger than	422	460	457	460	439	451	441	428
Less than or equal to	465	**429**	**417**	**367**	**391**	**305**	**278**	**298**

In order to statistically quantify the influences of the following contributory factors: 1. accident severity, 2. accident time and 3. link length on accident effects, a linear regression model is adopted. The dependent variable is speed ratio or speed difference, while the independent variables are the contributory factors of accident severity, accident time and link length. In Table 6.4, the details of these three contributory factors are shown. Table 6.5 presents the regression results when speed ratio is the dependent variable, while Table 6.6 gives the regression results when adopting speed difference as the dependent variable.

Table 6.4 Summary of contributory factors of regression models

Contributory factors	**Attribute**		**Proportion**
Accident severity	Fatal and seriously injured	1	6.7%
	Slightly injured	0	93.3%
Accident time	Peak-time period	1	30.9%
	Non-peak-time period	0	69.1%
Link length	Continuous variable		-

Table 6.5 Regression results when speed ratio is the dependent variable

	Beta	t	Sig.
(Constant)		39.449	.000
Accident severity	**-.038**	-.684	.494
Accident time	**-.068**	-1.228	.220
Link length	.194	3.491	.001

Table 6.6 Regression results when speed difference is the dependent variable

	Beta	t	Sig.
(Constant)		18.414	.000
Accident severity	**-.028**	.151	.880
Accident time	**-.038**	.026	.979
Link length	-.011	-.006	.995

From Table 6.5, it is observed that the coefficients of severity, time and link length in the speed ratio regression model are -0.038, -0.068 and 0.194, respectively. From Table 6.6, the coefficients of severity, time and link length in the speed difference regression model are observed as -0.028, -0.038 and -0.011, respectively. The negative coefficient of severity factor implies that fatal and seriously injured accidents would cause larger decrease in speed. Accidents occurring during peak-hour periods generate more significant adverse effects on traffic speed (with a negative coefficient of time factor). However, the estimated results are not statistically significant at the 10% level as the significant values of the factors of accident severity and accident time are larger 0.1. This is probably due to inadequate sample data as only 313 accidents occurred in the JTIS area during the study period (September 2009 to December 2010). In the future, the proposed model can be re-calibrated with a higher significant level if more empirical data is available.

6.4. Summary

In Chapter 6, the Before-After analysis method has been applied to investigate the accident effects on traffic speed. A total of 313 accidents and speed data before and after these accidents were used for a case study. Three factors: 1. accident severity, 2. accident time and 3. link length, were found to affect the accident effects on traffic speed. The degree of the influence of these contributory factors on accident effects was further investigated quantitatively. Regression models were adopted to statistically quantify the effects of these three contributory factors on the accident effects.

The results presented in this chapter indicated that the occurrence of accidents have significant adverse effects on traffic speed (a similar conclusion was also drawn by the study of Luk et al. (2001)). It was also noted that accident severity and accident time significantly influenced the accident effects on traffic speed (specifically the degree of reduction in traffic speed). More severe accidents, such as fatal and seriously injured accidents, were found to cause more adverse effects on traffic speed than those of lower severity, such as slightly injured accidents. Accident time was also found to have a significant impact on accident effects. During peak-hour periods when the traffic flow is heavier, the occurrence of accidents was found to result in larger scale of reduction in traffic speed. This finding is in agreement with that of Mak and Fan (2006). In the study of Mak and Fan (2006), it was found that the pre-accident traffic flow conditions statistically affect the magnitude of accident effects on traffic speed.

7. Conclusions and Further Studies

7.1. Summary and Conclusions

In this study, statistical analyses on road safety in Hong Kong have been conducted. The subjects of these analyses are accident blackspot identification, casualty injury severity modeling and accident effects on traffic speed. The aim of this study is to provide some insightful findings of traffic accidents in Hong Kong based on the statistical analysis of them. Some possible road safety improvement policies may be proposed based on these insightful findings in order to reduce accident frequency, severity and accident effects on traffic speed.

A method is proposed to identify traffic accident blackspots based on injury costs or accident costs rather than accident frequency or accident rate. Identifying accident blackspots based solely on accident frequency or accident rate has been found, in practice, to be inefficient as neither can correctly reveal the extent of accident consequences. Injury or accident costs is believed to be a more correct reflection of accident consequences as both injury (or accident) severity and costs in addition to number of injuries (or accidents) are considered.

A binary regression model is adopted to quantify and compare the associations between injury severity of driver, passenger and pedestrian casualties in traffic accidents and contributory factors. A linear regression model is applied for modeling the effects of contributory factors on accident costs. The contributory factors are categorized as environmental factors, site factors and vehicle factor. The coefficients of the proposed regression models are estimated using traffic accidents in Hong Kong from the years 2007 to 2009 inclusively.

Before-After analysis is applied to analyze accident effects on traffic speed. Most existing studies on accident effects focus on the detection of accidents on a basis of their effects on traffic conditions. Few studies, however, have been conducted to specifically investigate what contributory factors and how they affect accident effects. In this study, the influences of three contributory factors, namely accident severity, accident time and accident location, on accident effects are statistically analyzed. Regression models are adopted to quantify the influences of these contributory factors on accident effects.

The main conclusions drawn from this study are divided into the following four aspects.

(1) The proposed blackspot identification method is found to identify accident blackspots with higher injury or accident costs than that using accident frequency or based on Hong Kong Transport Department's blackspot definition only. One limitation in using accident frequency to identify blackspots is that it may be not able to identify locations or potential blackspots which have few accidents but the fatal accident rate is high. The proposed method can solve this problem and identify this type of blackspot. The drawback of the proposed method, however, is that it may fail to identify blackspots with a large number of slightly injured accidents since the total accident costs of these locations are not significant high enough. Sensitivity analyses were carried out to analyze the effects of changing weighted factor combinations in the proposed method. The results reveal that the choice of different costs for injuries or accidents has a significant influence on the identification of blackspots.

(2) As mentioned above, binary regression models are adopted to quantify the association between injury severity and contributory factors. The effects of contributory factors on injury severity are found to be varied in accordance with

different categories of casualties. Some contributory factors are found to have identical effects on all categories of casualties. These contributory factors include the age of the casualty, accident time, speed limit and road conditions. It is found that the attributes of aged over 65, involved in daytime crashes, and involved in crashes on a road segment with a speed limit above 50km/h and on multi-/dual carriageways lead to a noticeably higher injury severity of all casualties. However, the effects of certain contributory factors, such as vehicle size and the sex of the casualty, were found to vary in accordance with casualty category. Accidents involving large-size vehicles, such as buses, are found to have a higher probability of causing severe pedestrian casualties, but result in lower risk to drivers and passengers. Male drivers are at greater risk of being killed and seriously injured due to their aggressive driving behavior, while female pedestrians underwent a noticeably higher proportion of fatalities and seriously injured casualties.

(3) From the estimation results of the proposed linear regression model for accident costs, it was found that accidents occurring during the nighttime and at locations with higher speed limit have a greater crash injury costs. However, accidents occurring at locations with congestion, junction locations, one-way carriageways and two-way carriageways are found to result in less crash injury costs. The factor of number of casualties was also found to have a positive effect on the accident costs.

(4) Factors of accident severity and accident time significantly influence the accident effects on traffic speed. Fatal and seriously injured accidents are found to cause more severe effects on traffic speed than those accidents of lower severity, such as slightly injured accidents. It is also found that the accident time of day have a significant impact on the degree of reduction in traffic speed after accidents. During peak-hour periods when the traffic flow is heavier, the

occurrence of accident causes a larger scale of reduction in traffic speed.

7.2. Recommendations for Future Studies

Although the study presented in this thesis covers much statistical analyses on road safety in Hong Kong, there remain many interesting questions and important issues for which answers have not been provided. Some of these issues are outlined below for future studies:

(1) In Chapter 4, the accident costs extracted from CTS-3 may be underestimated as they are calculated only based on insurance cost. One of valuable extension of this study is to propose model to estimate the injury or accident costs and their standard deviation with the use of cost related data from the Hong Kong Road Casualty Information System (RoCIS), and to improve the accuracy of the proposed blackspot identification method.

(2) Ordered logit models, ordered probit models and nested logit models can be the alternative models adopted to estimate the associations between contributory factors and injury severity (Abdel-Aty, 2003; Savolainen and Mannering, 2007). The estimation results from these models can be compared with those estimated in Chapter 5 with the use of binary regression models.

(3) It is of interest to further analyze the effects of accident on traffic flow, occupancy or travel times instead of traffic speed with the use of the Network Fundamental Diagrams (NFD). The analysis results can contribute to the development of an efficient automatic accident detection method.

(4) In Chapter 6, a linear regression model is applied to quantify the associations

between certain contributory factors and accident effects. Linear regression models, however, may not be the best method to statistically analyze the influences of these contributory factors. Therefore, it may be an interesting extension to evaluate various advanced methods to estimate the influences of these contributory factors on accident effects.

(5) In Chapter 6, the estimated results of the proposed linear regression models are not statistically significant at the 10% level. This is probably due to inadequate sample data in the database adopted for analysis. In the future, the proposed model can be estimated with a higher significant level if more empirical data is available.

Appendix-Sample Data Form of Hong Kong Traffic Accident Database (TRADS)

Table A1 Data items in the crash environment profile of TRADS

Item	Name	Variable	Type	Values
1	REF	Reference No.	Char	
2	SEVERITY	Severity	Num	1=Fatal, 2= Serious, 3= Slight
3	ACC_DATE	Date of accident	Num	DD/MM/YY
4	ACC_TIME	Time	Num	HH/MM
5	WEEK_DAY	Day of week	Num	1= Mon, 2= Tue, 3= Wed, 4= Thu, 5= Fri, 6= Sat, 7= Sun
6	ST_NM	Street Name	Num	Coded by number
7	GRID_E	Easting Grid	Num	
8	GRID_N	Northing Grid	Num	
9	NO_VEH	Number of vehicles	Num	
10	NO_CSU	Number of casualties	Num	
11	WEATHER	Weather	Num	1= Clear, 2= Dull, 3= Fog/mist, 4= Strong Wind, 9= Not known
12	RAIN	Rain	Num	1= Not raining, 2= Light rain, 3= Heavy rain, 9= Not known
13	NAT_LGT	Natural Light	Num	1= Daylight, 2= Dawn/ Dusk, 3= Dark, 9= Not known
14	ST_LGT	Street Lighting	Num	1= Good, 2= Poor, 3= Obscured, 4= Not lit, 5= None, 6= Daylight, 9= Not known
15	SPEED_LMT	Speed Limit	Num	
16	TRAFF_CON	Traffic Congestion	Num	1= Severe, 2= Moderate, 3= None, 9= Not known
17	RD_SURFAC	Road Surface	Num	1= Wet, 2= Dry, 9= Not known
18	XING_LMT	On a crossing controlled by	Num	1= Zebra, 2= Traffic signal, 3= Police, 4= Crossing patrol, 5= Cautionary Crossing, 8= None
19	XING_15M	Within 15m of crossing controlled	Num	1= Zebra, 2= Traffic signal, 3= Police, 4= Crossing patrol, 5= Cautionary Crossing, 6= Footbridge/ subway, 8= None
20	JCN_CTRL	Junction control	Num	1= No, 2= Stop, 3= Give way, 4= Traffic signal, 5= Police, 6= Not junction
21	JCN_TYPE	Junction type	Num	1= Roundabout, 2= T-junction, 3= Staggered, 4= Y-junction, 5= Slip road, 6= Cross-roads, 7= Multiple, 8= Private access, 9= Other,
22	RD_TYPE	Road type	Num	1= One way, 2= Two way, 3= Dual Carriageway, 4= More than 2 carriageway
23	RD_CLASS	Road Classification	Num	1= Primary Distributor, 2= Private Road, 3= Other

Table A2 Sample accident data extracted from the crash environment profile of TRADS in the year 2010

ref	severity	acc_date	acc_time	week_day	st_nm	grid_E	grid_N	no_veh	no_csu	weather	rain	nat_lgt	st_lgt	speed_lmt	traff_cong	rd_surface	xing_lmt	xing_15m	jcn_ctrl	jcn_type	rd_type	rd_class
1	3	2010/1/1	0010	5	1864	12491	11186	1	1	1	1	3	1	50	2	2	8	8	6	10	2	4
2	3	2010/1/1	0019	5	1415	35115	22266	2	1	1	1	2	1	50	1	2	8	8	1	2	3	4
3	3	2010/1/1	0122	5	3223	15617	29093	2	1	1	1	3	1	50	3	2	8	8	6	10	2	3
4	2	2010/1/1	0129	5	1638	35284	20957	1	1	1	1	3	1	50	1	2	2	2	4	2	1	4
5	3	2010/1/1	0148	5	1420	35667	19104	2	1	1	1	2	1	50	1	2	2	2	4	2	3	4
6	3	2010/1/1	0153	5	317	36014	15317	1	1	1	1	3	1	50	3	2	8	8	6	2	2	4
7	2	2010/1/1	0214	5	1160	35529	19495	2	1	1	1	3	1	50	3	2	8	2	3	6	1	4
8	3	2010/1/1	0245	5	1002	35936	17783	2	1	1	1	1	4	50	2	2	2	2	4	2	3	4
9	3	2010/1/1	0415	5	914	42846	14028	1	1	2	2	3	1	50	3	1	8	8	6	10	2	4
10	3	2010/1/1	0615	5	1420	35464	20215	1	1	1	1	3	1	50	1	2	8	8	1	6	1	4
11	3	2010/1/1	0645	5	1407	35951	17576	1	1	1	1	1	1	50	1	2	2	2	4	2	1	4
12	3	2010/1/1	0735	5	2112	31227	24257	1	1	1	1	1	6	50	3	2	8	8	6	10	1	4
13	3	2010/1/1	0845	5	398	37206	15457	2	1	1	1	1	1	50	3	2	2	2	4	7	1	4
14	3	2010/1/1	1015	5	3300	39431	19998	2	1	1	1	1	4	70	3	2	8	8	6	10	3	4
15	3	2010/1/1	1030	5	2230	35338	35138	2	1	1	1	1	6	50	3	2	8	8	6	10	2	4
16	3	2010/1/1	1100	5	581	35301	15427	1	1	1	1	1	1	50	1	2	2	2	6	10	2	4
17	3	2010/1/1	1150	5	3760	26962	23568	3	2	1	1	1	6	80	3	2	8	8	6	10	4	4
18	3	2010/1/1	1244	5	3582	15954	31211	2	3	1	1	1	6	50	1	2	8	8	6	10	2	4
19	3	2010/1/1	1305	5	1492	39083	23047	1	1	1	1	1	4	50	3	2	2	2	4	2	3	4
20	3	2010/1/1	1305	5	1880	16412	26479	1	1	1	1	1	4	50	3	2	8	8	6	10	2	3

Table A3 Data items in the casualty injury profile of TRADS

Item	Name	Variable	Type	Values
1	REF	Reference No.	Char	
2	CAS_NO	Casualty Number	Num	
3	CAS_AGE	Casualty Age	Num	
4	CAS_SEX	Casualty Sex	Num	1= M, 2= F, 9= Not known
5	INJURY	Degree of injury	Num	1= Fatal, 2= Serious, 3= Slight
6	ROLE	Role of casualty	Num	1= Driver, 2= Passenger, 3= Pedestrian
7	SB_WORN	Seat belt or crash helmet worn	Num	1= Yes, 2= No, 9= Not known
8	IN_VEH_NO	In vehicle number	Num	
9	SEAT	Seat occupied	Num	1= Rear, 2= Front nearside, 3= Driver, 4= Standing in lower deck, 5= G/V Compartment (fixed), 6= G/V Compartment (w/o fixed), 8= Standing in upper deck, 9= Not known
10	PED_LOCATN	Pedestrian Location	Num	1= Footpath/ verge, 2= Refuge/ Central strip, 3= On controlled crossing, 4= Within 15M of controlled crossing, 5= Carriageway, 8= Other, 9= Not known
11	PED_ACTION	Pedestrian Action	Num	1= Walking (back), 2= Walking (face), 3= Standing, 4= Boarding, 5= Alighting, 6= Falling or jumping from, 7= Working at vehicle, 8= Other working, 9= Playing, 10= Crossing from nearside, 11= Crossing from offside, 99= Not known
12	PED_CIRCUM	Special Circumstances	Num	1= Footpath overcrowded, 2= Footpath obstructed, 3= One side no footpath, 4= Two side no footpath, 5= Ran onto road, 6= Climbed over barrier, 9= None
13	DIRECTN_FR	Direction from	Num	1=North East, 2=East, 3=South East, 4=South, 5=South West, 6=West, 7=North West, 8=North, 9=Not known, 99=Invalid input
14	DIRECTN_TO	Direction to	Num	1=North East, 2=East, 3=South East, 4=South, 5=South West, 6=West, 7=North West, 8=North, 9=Not known, 99=Invalid input

Table A4 Sample casualty data extracted from the casualty injury profile of TRADS in the year 2010

ref	cas_no	cas_age	cas_sex	injury	role	sb_worn	in_veh_no	seat	ped_locatn	ped_action	ped_circum	directn_fr	directn_to
1	1	53	1	3	1	2	1	3	0	0	0	0	0
2	1	34	1	3	1	1	2	3	0	0	0	0	0
3	1	40	1	3	1	1	1	3	0	0	0	0	0
4	1	26	2	2	3	0	0	0	4	11	4	6	2
5	1	47	1	3	1	1	1	3	0	0	0	0	0
6	1	32	2	3	3	0	0	0	1	2	5	4	8
7	1	34	1	2	1	1	2	3	0	0	0	0	0
8	1	57	1	3	1	1	2	3	0	0	0	0	0
9	1	26	1	3	1	1	1	3	0	0	0	0	0
10	1	48	1	3	2	9	1	1	0	0	0	0	0
11	1	33	1	3	3	0	0	0	8	3	4	2	6
12	1	55	1	3	2	2	1	1	0	0	0	0	0
13	1	51	2	3	3	0	0	0	5	2	5	2	6
14	1	45	1	3	1	1	1	3	0	0	0	0	0
15	1	55	1	3	1	1	1	3	0	0	0	0	0
16	1	51	2	3	2	1	1	4	0	0	0	0	0
17	1	61	1	3	1	1	3	3	0	0	0	0	0
17	2	56	2	3	2	1	3	2	0	0	0	0	0
18	1	40	1	3	1	2	1	3	0	0	0	0	0
18	2	10	1	3	2	2	1	1	0	0	0	0	0
18	3	4	1	3	2	2	1	1	0	0	0	0	0
19	1	49	1	3	2	2	1	1	0	0	0	0	0
20	1	57	1	3	1	1	1	3	0	0	0	0	0

Table A5 Data items in the vehicle involvement profile of TRADS

Item	Name	Variable	Type	Values
1	REF	Reference No.	Char	
2	VEH_NO	Vehicle Number	Num	
3	DRIVER_AGE	Age of driver	Num	
4	DRIVER_SEX	Sex of driver	Num	1= M, 2= F, 9= Not known
5	VEH_AGE	Vehicle age	Num	
6	VALID_LIC	Valid License	Num	1= Yes, 2= No, 8= N/A, 9= Not known
7	VALID_INS	Valid Insurance	Num	1= Yes, 2= No, 8= N/A, 9= Not known
8	VEH_CLASS	Vehicle Class	Num	Coded by number (1=Motorcycle, 2=Private car, 3=Private light bus, etc.)
9	COLLIDE	Vehicle Collision With	Num	1= Vehicle, 2= Pedestrian, 3= Animal, 4= Object on c'way, 5= Traffic sign post, 6= Lamp/ teleph post, 7= Road sign, 8= Tree, 9= Wall/ bridge prpt, 10= Utility co equip, 11= Bollard, 12= Fire hydrant, 13= Pedestrn barrier, 14= Crash barrier, 15= Road works, 16= Hoarding/ walkway, 17= Hawker stall, 18= Other, 19= None, 99= Not known
10	VEH_LGT	Vehicle light	Num	1= None, 2= Parking lights, 3= Headlights dipped, 4= Headlight main beam, 8= N/A, 9= Not known
11	DIR_FROM	Direction from	Num	1=North East, 2=East, 3=South East, 4=South, 5=South West, 6=West, 7=North West, 8=North, 9=Not known, 99=Invalid input
12	DIR_TO	Direction to	Num	1=North East, 2=East, 3=South East, 4=South, 5=South West, 6=West, 7=North West, 8=North, 9=Not known, 99=Invalid input
13	DAMAGE_SEV	Damage Severity	Num	1= No, 2= Slight, 3= Severe, 9= Not known
14	GV_LOAD	Goods vehicle loading	Num	1= None, 2= Secure, 3= Insecure
15	OVERLOAD	Vehicle overloaded	Num	1= Yes, 2= No

Table A6 Sample vehicle data extracted from the vehicle involvement profile of TRADS in the year 2010

serial	veh_no	driver_age	driver_sex	veh_age	valid_lic	valid_ins	veh_class	collide	veh_lgt	dir_from	dir_to	damage_sev	gv_load	overload
1	1	53	1	0	8	8	82	19	4	4	8	0	0	0
2	2	34	1	12	1	1	2	1	3	2	6	0	0	0
2	1	57	1	9	1	1	72	1	3	2	6	0	0	0
3	2	19	1	5	1	1	2	1	3	6	2	0	0	0
3	1	40	1	9	1	1	72	1	3	6	2	0	0	0
4	1	43	1	9	1	1	72	2	4	8	4	0	0	0
5	2	41	1	10	1	1	62	1	3	4	8	0	0	0
5	1	47	1	9	1	1	72	1	1	4	8	0	0	0
6	1	55	1	7	1	1	72	2	9	2	6	0	0	0
7	2	34	1	16	1	1	1	1	4	6	2	0	0	0
7	1	41	1	9	1	1	72	1	3	4	8	0	0	0
8	1	55	1	9	1	1	75	1	1	4	6	0	0	0
8	2	57	1	9	1	1	75	1	1	4	6	0	0	0
9	1	26	1	8	1	1	1	19	9	2	6	0	0	0
10	1	999	9	0	9	9	72	19	9	4	8	0	0	0
11	1	50	1	9	1	1	72	2	4	2	6	0	0	0
12	1	59	2	19	1	1	97	19	8	4	8	0	0	0
13	1	64	1	9	1	1	72	18	1	6	8	0	0	0
13	2	51	2	0	8	8	84	1	8	2	6	0	0	0
14	2	67	1	4	1	1	12	1	1	6	2	0	1	2
14	1	45	1	8	1	1	63	1	1	6	2	0	0	0
15	2	29	1	13	1	1	2	1	1	8	4	0	0	0
15	1	55	1	5	1	1	73	1	1	4	8	0	0	0
16	1	54	1	15	1	1	67	19	1	2	6	0	0	0
17	3	61	1	7	1	1	2	1	8	2	6	0	0	0
17	1	56	1	15	1	1	7	1	8	2	6	0	2	2
17	2	57	1	5	1	1	7	1	8	2	6	0	1	2
18	1	40	1	16	1	1	2	1	8	4	8	0	0	0
18	2	56	1	2	1	1	2	1	8	4	8	0	0	0
19	1	46	1	6	1	1	52	19	1	2	8	0	0	0

References

Abdel-Aty, M., 2003, Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research*, Vol. 34, Issue 5, pp. 597-603.

Al-Ghamdi, A.S., 2002, Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis and Prevention*, Vol. 34, Issue 6, pp. 729-741.

Barua, U. and Tay, R., 2010, Severity of urban transit bus crashes in Bangladesh. *Journal of Advanced Transportation* Vol. 44, Issue 1, 34-41.

Chung, K., Jang, K., Madanat, S. and Washington, S., 2011, Proactive detection of high collision concentration locations on highways. *Proceeding of 19th International Symposium on Transportation and Traffic Theory*, 650-661.

Dommes, A., Cavallo, V., Vienne, F. and Aillerie, I., 2012, Age-related differences in street-crossing safety before and after training of older pedestrians. *Accident Analysis and Prevention*, Vol. 44, Issue 1, 43-47.

Elvik, R., 2008, A comparative analysis for techniques for identifying hazardous road locations. *Transportation Research Record* 2083, 72-75.

Evans, A.W., 2009, Economic Appraisal of Road Safety Measures in Great Britain. *International Journal of Sustainable Transportation* Vol. 3, Issue 3, 160-177.

Geurts, K., Wets, G., Brijs, T. and Vanhoof, K., 2004, Identification and ranking of black spots: sensitivity analysis. *Transportation Research Record* 1897, 34-42.

Hauer, E., 1980, Bias-by-selection: overestimation of the effectiveness of safety countermeasures caused by the process of selection for treatment. *Accident Analysis and Prevention*, Vol. 12, Issue 2, 113-117.

Hauer, E., 1996, Identification of sites with promise. *Transportation Research Record* 1542, 54-60.

Hauer, E., 1997, *Observational before-after studies in road safety*. Pergamon, Oxford.

Hosmer, D.W. and Lemeshow, S., 1980, Goodness-of-fit test for the multiple logistic regression model. *Communication in Statistics-Theory and Methods*, Vol. 9, Issue 10, pp. 1043-1069.

Hosmer, D.W. and Lemeshow, S., 1989, *Applied Logistic Regression*. Wiley-Interscience, New York, USA.

Jesen, S.U., 2008, Safety effects of blue cycle crossings: A before-after study. *Accident Analysis and Prevention*, Vol. 40, Issue 2, pp. 742-750.

Karlaftis, M.G. and Golias, I., 2002, Effects of road geometry and traffic volumes on rural roadway accident rates. *Accident Analysis and Prevention*, Vol. 34, Issue 3, pp. 357-365.

Kockelman, K.M. and Kweon, Y.J., 2002, Driver injury severity: an application of ordered probit models. *Accident Analysis and Prevention*, Vol. 34, Issue 3, pp. 313-321.

Kostyniuk, L.P. and Shope, J.T., 2003, Driving and alternatives: Older drivers in Michigan. *Journal of Safety Research*, Vol. 34, Issue 4, pp. 407-414.

Krull, K., Khattak, A. and Council, F., 2000, Injury effects of rollovers and events sequence in single-vehicle crashes. *In Transportation Research Record: Journal of the Transportation Research Board*, No. 1717, Transportation Research Board of the National Academies, Washington, D.C., pp. 46-54.

Lam, W.H.K., Tam, M.L., Sumalee, A., Li, C.L., Chen, W., Kwok, S.K., Li, Z.L. and Ngai, W.T., 2008, Incident detection based on short-term travel time forecasting. *Proceedings of 13th International Conference of the Hong Kong Society for Transportation Studies*, pp. 83-92.

Lawrence, J., 1995, The Economic Cost of Motor Vehicle Crashes, *1994 NHTSA Technical Report*. Washington, DC: U.S. Department of Transportation.

Lee, S., Choi, J. and Kin, S.W., 2010, Bayesian approach with the power prior for road safety analysis. *Transportmetrica*, Vol. 6, Issue 1, pp. 39-51.

Loo, B.P.Y., 2009, The identification of hazardous road locations: a comparison of the blacksite and hot zone methodologies in Hong Kong. *International Journal of Sustainable Transportation* Vol. 3, Issue 3, 187-202.

Loo, B.P.Y. and Tsui, K.L., 2007, Factors affecting the likelihood of reporting road crashes resulting in medical treatment to the police. *Injury Prevention*, Vol. 13, Issue 3, pp. 186-189.

Luk, J.Y.K., Chung, E.C.S. and Sin, F.Y.C., 2001, Characterization of incidents on an urban arterial road. *Journal of Advanced Transportation* Vol. 35, Issue 1, pp. 67-92.

Maher, M.J. and Mountain, L.J., 1988, The identification of accident blackspots: a comparison of current methods. *Accident Analysis and Prevention*, Vol. 20, Issue 2, 143-151.

Mak, C.L. and Fan, H.S.L., 2006, Heavy flow-based incident detection algorithm using information from two adjacent detector stations. *Journal of Intelligent Transportation Systems*, Vol. 10, Issue 1, pp. 23-31.

Miaou, S., 1994, The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis and Prevention*, Vol. 28, Issue 6, 695-707.

Miller, T.R., Lestina, D.C. and Spicer, R.S., 1998, Highway crash costs in the United States by driver age, blood alcohol level, victim age, and restraint use. *Accident Analysis and Prevention*, Vol. 30, Issue 2, 137-150.

Miranda-Moreno, L. F., 2006, Statistical Models and Methods for Identifying Hazardous Locations for Safety Improvements. PhD thesis. University of Waterloo, Canada.

Miranda-Moreno, L.F., Fu, L., Ukkusuri, S. and Lord, D., 2009, How to incorporate accident severity and vehicle occupancy into the hot spot identification process? *Transportation Research Record* 2102, 53-60.

Montella, A., 2010, A comparative analysis of hotspot identification methods. *Accident Analysis and Prevention*, Vol. 42, Issue 2, 571-581.

Nowakowska, M., 2010, Logistic models in crash severity classification based on road characteristics. In Transportation Research Record: Journal of the Transportation Research Board, No. 2148, *Transportation Research Board of the National Academies*, Washington, D.C., pp. 16-26.

Owsley, C., McGwin, G., Phillips, J.M., McNeal, S.F. and Stalvey, B.T., 2004, Impact of an educational program on the safety of high-risk, visually impaired, older drivers. *American Journal of Preventive Medicine*, Vol. 26, Issue 3, pp. 222-229.

Saccomanno, F.F., Fu, L. and Miranda-Moreno, L.F., 2004, Risk-Based Model for Identifying Highway–Rail Grade Crossing Blackspots. *Transportation Research Record* 1862, 127-135.

Savolainen, P. and Mannering, F., 2007, Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. *Accident Analysis and Prevention*, Vol. 39, Issue 5, pp. 955-963.

Sethi, V., Bhandari, N., Koppelman, F.S. and Schofer, J.L., 1995, Arterial incident detection using fixed detector and probe vehicle data. *Transportation Research Part C*, Vol. 3, Issue. 2, pp. 99-112.

Sharkar, V., Mannering, F. and Barfield, W., 1995, Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis and Prevention*, Vol. 27, Issue 3, pp. 371-389.

Stokes, R.W. and Mutabazi, M.I., 1996, Rate-quality control method of identifying hazardous road locations. *Transportation Research Record* 1542, 44-48.

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

Tam, M.L. and Lam, W.H.K., 2011, Validation of Instantaneous Journey Time Estimates: A Journey Time Indication System in Hong Kong. *Proceedings of the Eastern Asia Society for Transportation Studies*, Vol.8.

Tay, R. and Rifaat, S.M., 2007, Factors contributing to the severity of intersection crashes. *Journal of Advanced Transportation*, Vol. 41, Issue. 3, pp. 245-265.

Toshiyuki, Y. and Shankar, V., 2004, Bivariate ordered-response probit model of driver's and passenger's injury severities in collision with fixed objects. *Accident Analysis and Prevention*, Vol. 36, Issue 5, pp. 869-876.

Transport Department, 1999, The Third Comprehensive Transport Study : Final Report.

Transport Department, 2001, Transport Planning and Design Manual. Vol. 5, Government Printer, Hong Kong SAR.

Washington, S.P., Karlaftis, M.G. and Mannering, F.L., 2003, Statistical and econometric methods for transportation data analysis. CRC Press, New York, United States.

Wang, C., Quddus, M.A. and Ison, S.G., 2009, Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. *Accident Analysis and Prevention*, Vol. 41, Issue 4, pp. 798-808.

Wong, S.C., Sze, N.N., Lo, H.K., Hung, W.T. and Loo, B.P.Y., 2005, Would relaxing speed limits aggravate safety? A case study of Hong Kong. *Accident Analysis and Prevention*, Vol. 37, Issue 2, pp. 377-388.

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

Yau, K.K.W., 2004, Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. *Accident Analysis and Prevention*, Vol. 36, Issue 3, pp. 333-340.

Yau, K.K.W., Lo, H.P. and Fung, S.H.H., 2006, Multiple-vehicle traffic accidents in Hong Kong. *Accident Analysis and Prevention*, Vol. 38, Issue 6, pp. 1157-1161.

Ye, D.J., Lam, W.H.K. and Tam, M.L., 2011, Automatic incident detection for urban roads in Hong Kong. *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 9, pp. 1897-1912.

Zaloshnja, E. and Miller, T.R., 2004, Costs of large truck-involved crashes in the United States. *Accident Analysis and Prevention*, Vol. 36, Issue 5, 801-808.

Zaloshnja, E., Miller, T.R., Council, F. and Persaud, B., 2006, Crash costs in the United States by crash geometry. *Accident Analysis and Prevention*, Vol. 38, Issue 4, 644-651.