An Empirical Investigation of Harmfulness, Pattern and Influential Factor Associated with Fatigue-related Crash

by

LI Yanyan

Department of Civil Engineering

Nagoya University

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Engineering

Nagoya, Japan
September 2018

ABSTRACT

Road traffic deaths and injuries have already become a major cause of concern and aroused great attention all around the world. According to Global Status Report on Road Safety 2015 by World Health Organization, more than 1.2 million people die each year, with millions more sustaining serious injuries and living with long-term adverse health consequences. In low- and middle-income countries, traffic injuries have become one of the leading causes of death and cost approximately 3% of their GDP as a result of traffic crashes. Fatigue driving was identified as one of the four most risky driving-related behaviors, especially in fatal traffic crashes and represented a significant social and economic cost to the community. Despite extensive body of research addressing the harmfulness of driver fatigue on road safety, it has not attracted enough attention. Drivers are less concerned about driver fatigued than other traffic safety issues. Besides drivers, public are also not fully aware of the potential risk of driver fatigue because it is difficult to evaluate its effect accurately.

The focus of this dissertation is to examine possible reasons for disregarding the harmfulness of fatigue-related crash, and identify factors contributing to the occurrence of fatigue-related crash as well as severe outcome in the crash. The first problem addressed in the dissertation is the misclassification problem of fatigue-related crash. Reliable and accurate records are essential for assessing the scope of fatigue-related crash problems, monitoring, and evaluating the effectiveness

of intervention measures. An analysis framework is developed to identify factors that cause police officers misclassify fatigue-related crashes and examine the interactive effects of those factors. It can be inferred that the stereotype of certain groups of drivers, crash types, and roadway conditions affects police officers' judgment on fatigue-related crashes.

Another possible reason for impeding understanding the harmfulness of fatigue-related crash is examined. Fatigue driving and injury severity in the crash may share some common influential factors. Ignoring the impact of these common factors will lead to endogeneity problem and result in biased parameter estimation. Therefore, a bivariate endogenous binary-ordered probit model is developed to examine the relationship between fatigue driving and injury severity considering endogeneity. Regarding the potential systematic differences between commercial and non-commercial vehicle drivers, the difference of influential factors between commercial and non-commercial vehicle drivers is also discussed. The results show that the influence of fatigue driving on injury severity is significantly underestimated if ignoring the endogeneity.

Lastly, the dissertation investigates the fatigue-related crash from macro-level. A spatial filtering technique is applied to capture the unobserved spatial correlation of fatigue-related crash frequency. With the filtered spatial components, a semi-parametric Poisson model can be developed to explore the impacts of both road and macroscopic variables on the occurrence of fatigue-related crashes. Also, the effects of omitted spatial components and macroscopic variables can also be calculated. The calculation results indicate that the filtered spatial components and macroscopic variables explain more than half of the unobserved variation in the error term.

ACKNOWLEDGEMENTS

I would like to express my appreciation and gratitude to my supervisor Professor Toshiyuki Yamamoto. His support and brilliant guidance contributed greatly to my research. Also, his spirits of seriousness and enthusiasm of research inspire me in this three years. I also would like to thank the dissertation committee members, Professor Nakamura, Professor Miwa and Professor Fujita for valuable advices for my dissertation. I want to express my appreciation to Professor Morikawa and Dr. Sato for providing comments for my research as well as all members in Morikawa, Yamamoto and Miwa Lab. I grateful acknowledge the help of Professor Guangnan Zhang and my friend He Tian. They helped me a lot in data collection.

Many thanks to my friends, Dr. Ran Song, Dr. Lei Gong, Dr. Mothafer Ghasak I.M. A, Dr. Weiliang Zeng, Nguyễn Thị Ngọc, Sirion Pita, Mingyang Hao, Zhiguang Liu, Lanhang Ye, Ruotian Tang, Jiahang He, for sharing great idea and discussing my research with me. I would also like to thanks Yang Peng, Jing Feng, Dr. Qiulu Han, Yizhou Yang, Xingyi Huang, Yefang Zhou, Xun Zheng, for encouraging me when I have difficulties in life and research work. Special thanks to Tun Ling, Gou Ge, Hate Ling, Nen Zai and Tu Yang for company in these three years.

Last but not the least, I would like to thank my families and my boyfriend. Without them, I couldn’t accomplish the work I have done.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
ATC	Australian Transport Council
ATSB	The Australian Transport Safety Bureau
BP	Basic Poisson Model
CCMTA	Canadian Council of Motor Transport Administrator
DaCoTA	Road Safety Data Collection, Transfer and Analysis
F-F	Fatigue to Fatigue
F-N	Fatigue to Non-fatigue
GDP	Gross Domestic Product
GIS	Geographic Information System
GTAD	The Guangdong Traffic Accident Dataset
GWR	Geographically Weighted Regression
LHS	Left-hand-side
LR	Log-Likelihood
MI	Moran’s I statistic
NCSDR	National Center on Sleep Disorders Research
N-F	Non-fatigue to Fatigue
NHTSA	National Highway Traffic Safety Administration
OR	Odds Ratio
PCNM	The Principal Coordinate Analysis of Neighbor Matrices
PCoA	The Principal Coordinate Analysis
PRC	People's Republic of China
PRD	The Pearl River Delta

PRV	The Proportion of Reduction in Variance
QSFP	Queen Spatial Filtering Poisson Model
RHS	Right-hand-side
RSFP	Rook Spatial Filtering Poisson Model
UK	The United Kingdom
VIF	Variance Inflation Factor
WHO	World Health Organization

CHAPTER 1

Introduction

1.1 Background

Road traffic deaths and injuries have already become a major cause of concern and aroused great attention all around the world. More than 1.2 million people die each year, with millions more sustaining serious injuries and living with long-term adverse health consequences (World Health Organization, WHO, 2015). Particularly, in low- and middle-income countries, traffic injuries have become one of the leading causes of death and cost approximately 3% of their GDP as a result of traffic crashes (WHO, 2015).

Over the past few years, the concern over driver fatigue has risen. Fatigue driving was identified as one of the four most risky driving-related behaviors, especially in fatal traffic crashes (Fernandes et al., 2010) and represented a significant social and economic cost to the community. Approximately 20% of fatal crashes in Canada involved driver fatigue, eliminating the influence of alcohol, speeding and unsafe passing (Canadian Council of Motor Transport Administrator, CCMTA, 2010). In Australia, 20 - 30% of all fatal traffic crashes were found due to fatigue driving (Australian Transport Council, ATC, 2011). However, this situation could be worse in developing countries since those countries include most of traffic crashes worldwide (WHO, 2015). A questionnaire-based research among commercial bus drivers in Malaysia found that the prevalence of fatigue among commercial bus drivers was 37.7% (Fadhli et al., 2008). Statistics from China also showed that 1,271 (0.83% of total number of crashes due to any cause) crashes were caused by fatigue driving in 2013, with 677 (1.16% of total number of people killed in the crashes due to any cause) people killed, 1,600 (0.75% of total number of people injured in crashes due to any cause) people injured, and over RMB 37 million in property losses (Traffic Management Bureau, Ministry of Public Security, PRC, 2013). Although China seems to have lower fatal fatigue-related crash rate than Canada and Australia, the reason for this contrast may be related to their criterion for calculating the "crash rate". The fatigue crash rate in the statistics of Canada and Australia is calculated using the number of crash which "fatigue is one of the contributing factors". However, the fatigue-related crash rate for China is calculated by the number of crashes which "fatigue is the major cause of crash". In this case, China is applying a much narrower concept in calculating fatigue-related crash rate than Canada and Australia. Applying the similar criteria, UK estimated the fatigue-related crash rate should be around 2% of all crashes in 2015 (Department for Transport, UK, 2016), which the fatigue-related rate is much closer to China. Although the reported fatigue-related crash rate of China is not so high, we can still speculate that the crash rate for "fatigue is one of the contributors of crash" would be much higher.

Despite extensive body of research addressing the harmfulness of driver fatigue on road safety, it has not aroused enough attention. Drivers were less concerned about fatigued driving than other traffic safety issues (Vanlaar et al., 2008). Studies from different countries showed that many people still drove when they felt fatigue (Beirness et al., 2005; Nordbakke & Sagberg, 2007; Tefft, 2010). Besides drivers, public are also not fully aware of the potential risk of driver fatigue due to the inaccurate evaluation of its harmfulness.

1.2 Problem Statement

Although fatigue-related crashes represent a significant social and economic cost to the community, their influential mechanism, contributing risk factors and countermeasure are not fully understood. Generally speaking, three major barriers hinder the investigation of fatigue-related crashes. First and foremost, the absence of consistent definition for driver fatigue makes it hard to evaluate. One of the obstacles is that is hard to quantify the relationship between fatigue and working strength since both high-demand and low-demand road condition could induce driver fatigue (Oron-Gilad et al., 2008; Zhao & Rong, 2013). For example, poor road condition (Arnold et al., 1997), complex traffic conditions, and road environments (Pilcher & Huffcutt, 1996) required more attention and could easily induce physical and mental fatigue. However, fatigue can also be induced in simple and monotonous condition, which has been confirmed by simulated driving studies (Desmond & Hancock, 2001; Thiffault & Bergeron, 2003). In addition, fatigue is not a strictly linear progress, but a gradual and cumulative process, and the syndrome of fatigue can be resolved after a period of rest. More importantly, individual differences hinder the objective quantification of fatigue degree in a crash (Haworth et al., 1998; Stutts et al., 1999; Rajaratnam & Arendt, 2001; Karrer et al., 2004; Philip et al., 2005). In this respect, it is still difficult to develop a general definition for driver fatigue like other violation behaviors.

The second barrier is related to the data quality. Data quality is essential for accurately analyzing the influence and contributing factors of fatigue-related crashes, especially in statistically analysis. However, as mentioned above, fatigue-related crash data is hard to collect due to the absence of universal accepted definition. The sources of fatigue-related crash data can be can be classified into three categories: (1) self-report data collected by surveys, (2) experiment data obtained from lab experiment, and (3) police records. Self-report data are collected by asking drivers about their experience of fatigue and their involvement in automobile crashes (National Sleep Foundation, 2008; Pennay, 2008). The disadvantage of self-report data is that it can be affected by social desirability and recall bias (Neugebauer & Ng, 1990; Wåhlberg et al., 2010). Moreover, the narrow definition of fatigue in survey can also cause underestimation in the impact of fatigue in a crash (Armstrong et al., 2013). On the other hand, the data collected from lab experiment is believed to be much accurate for the reason that it is automatically recorded and saved by equipment. In this way, some objective bias of measuring fatigue can be avoided. It should be noted that experiment data is also criticized for whether it can reveal the reality since people's driving behavior under experimental setting may differ from the real driving situation.

Moreover, individual characteristics have great influence on the fatigue level given the same condition. As a result, the results obtained from a small group of people may not be able to generalize to general public.

The police records are perhaps the most often adopted data in fatigue-related crash analysis. This kind of data is recorded by police officers and can be easily combined to other data (e.g. medical records, geographic data, etc.). Nevertheless, it is criticized by some researchers that police officers may not has adequate knowledge and information to judge whether a crash is cause by driver fatigue (Armstrong et al., 2013). In practice, to assist in identifying fatigue in crash, proxy measurements are developed aiming to improve reporting accuracy of fatigue-related crashes (Filtness et al., 2015). Although these surrogate definitions are also based on experience or scientific research, they are criticized for being too specific (Crummy et al., 2008; Armstrong et al., 2013) and may provide misleading instructions for police officers. Additionally, police officers also tend to assign the reason of a crash to current interest when there exist more than one possible causes for the crash (Ogden & Moskowitz, 2004).

As one of the fundamental methodologies for analysis traffic crashes on road, statistical model provides useful tools to investigate the risk factors contributing to fatigue-related crash (e.g. Pack et al., 1995; Zhang et al., 2016), and the relationship between fatigue driving behavior and severity outcome in the crash (e.g. Summala & Mikkola, 1994). Some important ongoing methodological issues should be considered when developing statistical models so as to achieve better evaluation.

Those issues are discussed in a review by Mannering and Bhat (2014): parsimonious vs. fully specified models; unobserved heterogeneity; risk compensation; choice of methodological approach; under-reporting of crashes with less severe injuries; selectivity-bias/endogeneity; spatial and temporal correlations.

1.3 Objectives

The primary objective of this dissertation is to develop statistical models to examine the harmfulness, pattern, and influential factor associated with fatigue-related crash from different prospects. The following are the specific objectives.

The first purpose is to examine possible reasons for neglecting the harmfulness of fatigue-related crash. Due to lacking of proper criteria, the identification of fatigue-related crashes by police officers largely depends on inferential evidence and their own experience. As a result, many fatigue-related crashes are misclassified and the harmfulness of fatigue on road safety is misestimated. Moreover, problematic statistical models can also induce inaccurate evaluation of the impact of driver fatigue. Therefore, in this dissertation, endogeneity will be addressed when analyzing the impact of fatigue on its consequently injury severity in fatigue-related crashes

The second purpose is to identify factors contributing to the occurrence of fatigue-related crash as well as severe outcome in a crash. The micro-level analysis focuses on identifying factors affecting both driver's fatigue propensity and driver's injury severity in a crash, including observed and unobserved factors. This analysis is based on individual data. On the other hand, this study also investigates factors associated with crash frequency using county-level fatigue-related crash data. Macro-level data is often found to be spatial dependent that the spatial correlation of fatigue-related crash occurrence should be discussed. Moreover, the contribution of macroscopic variables will also be investigated.

Based on the discussion, suggestions for better preventing fatigue-related crash and its consequently injury will be proposed.

1.4 Structure

The structure of the dissertation is shown in the following Figure 1.1. This dissertation contains seven chapters. Chapter 1 introduces the background and objectives of this research. Chapter 2 summarizes the researches of driver fatigue and modelling issues relevant to traffic crashes, especially to fatigue-related crash. The description of data applied in the dissertation is shown in Chapter 3. In Chapter 4, a joint model framework is introduced to analyze factors contributing to the misclassification of fatigue-related crashes in police reports. Association rule data mining technique is employed to identify the potential interactions of factors, and logistic regression models are applied to analyze factors that hinder police officers' identification of fatigue-related crashes. In Chapter 5, an empirical analysis is conducted to examine the relationship between fatigue driving propensity and fatal injury severity by comparing bivariate and univariate endogenous binary-ordered probit model. In Chapter 6, the spatial correlation of fatigue-related crash is captured by spatial filtering technique. In addition, the influence of spatial correlation and macroscopic variables on fatigue-related crash frequency is calculated. Finally, in Chapter 7, conclusions and recommendations for future studies are given.

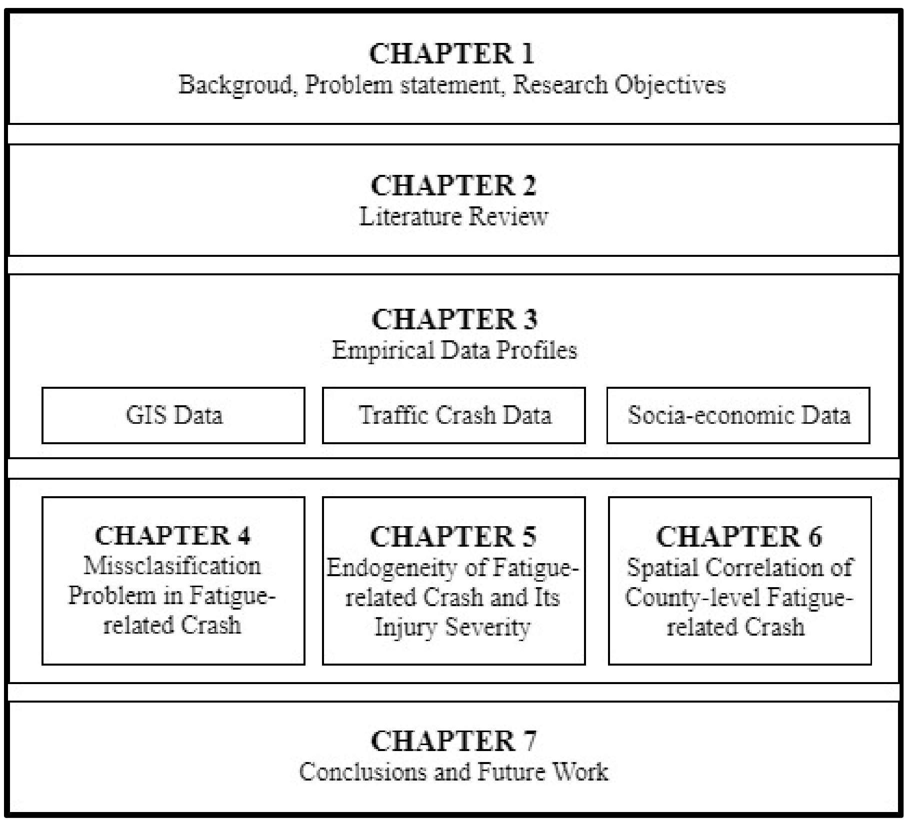

Figure 1.1 Structure of the dissertation

CHAPTER 2

Literature Review

The review of literature is divided into three main sections: First, the previous studies on fatigue-related crashes are summarized. Second, studies on misclassification of traffic crash will be discussed in detail. Particularly, the researches based on fatigue-related crash will be addressed. Finally, a review of endogeneity and spatial correlation will be presented.

2.1 Driver fatigue and fatigue-related crash

Although the phenomenon of driver fatigue has been heavily researched, there is not a widely accepted definition for it. Fatigue is a physiological condition that can occur long before you fall asleep at the wheel, and impairs the reaction time, vigilance and judgement on traffic condition (The Ministry of Transport, New Zealand, 2017). There are three main determinants of fatigue identified by previous researches: lacking of sleep, time of the day, and time on the task (Dobbie, 2002).

Lacking of sleep is the most well-known cause of driver fatigue that would reduce the alertness and performance of drivers (Dinges, 1995; Hartley & Arnold, 1995; Rosekind, 1999). Different individuals require different length of sleep. Moreover, the impact of sleep loss can be accumulated and manifests themselves during inappropriate or dangerous situations (e.g. falling asleep during driving) (Dinges, 1995; Feyer, 2001). On the other hand, time of the day is closely related to fatigue because it affects the circadian rhythm or body clock of human beings. Crash facts also confirm the influence of time of the day on the occurrence of crashes (Mitler et al., 1988). Several studies have claimed that there are two peaks for when the level of sleepiness is high: night to early morning and in the afternoon (Pack et al., 1995; Haworth, 1998; Hartley, 2000; Horne & Reyner, 2001; Philip et al., 2005). Additionally, driving task-based research shown that with the increase of time spending on a task, the level of fatigue also increases. Also, the time spent in other tasks, such as working and studying, can also increase fatigue and affect subsequent driving behavior (The Ministry of Transport, New Zealand, 2017).

Unlike drunk driving crashes, no blood or breath test can be applied to quantify driver's fatigue level at the crash scene (Pack et al., 1995; Connor et al., 2001a; DaCoTA, 2012). As a result, there is currently no standard methodology for convicting fatigue as the cause of the crash (Crummy et al., 2008; Filtness et al., 2015). Therefore, the identification of fatigue-related crashes is based on subjective evidences or indirect measurements. Table 2.1 lists the criterions for identifying fatigue-related crashes in United Kingdom, United States of America, Australia, and China.

Table 2.1 Fatigue-related crash determination criterion in researches

Source	Country	Criterions for identifying fatigue-related crash
Horne and Reyner (1995)	United Kingdom	(1) vehicle ran off the road and/or collided with another vehicle or object; (2) absence of skid marks or braking; (3) driver saw the point of run-off or the object hit prior to the crash; (4) witnesses reported lane drifting prior to the crash; (5) excluded are those instances where another cause may have been the primary factor (e.g. mechanical defect, speeding, excess alcohol, bad weather).
NCSCR/NHTSA (2001)	United States of America	(1) occurred late at night, early morning or mid-afternoon; (2) resulted in higher than expected severity; (3) involved a single vehicle leaving the roadway; (4) occurred on a high-speed road; (5) driver did not attempt to avoid the crash; (6) driver was the sole occupant in the vehicle.
ATSB operational definition (2006)	Australia	(1) includes single vehicle crashes that occurred during 'critical times' (00:00–06:00 and 14:00–16:00); (2) includes head-on collisions where neither vehicle was overtaking at the time; (3) excludes the following types of crashes: crashes occurred on roads with speed limits under 80 km/hour, crashes involved pedestrians, crashes involved unlicensed drivers, crashes involved drivers with high levels of alcohol (blood alcohol over 0.05g/100ml).
China's Public Security Department (2007)	China	(1) drove cars more than eight hours a day; (2) engaged in other work excessive physical exertion; (3) lacked of sleep that results in sleepy or lower reaction rate, so that the driver is having difficulty in assessing traffic conditions immediately and reacting accurately.

2.2 Misclassification

Among all causes, fatigue-related crashes are easily neglected or misclassified because of the difficulty in observing and identifying driver fatigue (Radun et al., 2013; Filtness et al., 2015). Due to the absence of solid definition and objective test, fatigue-related crash is hard to detect in a crash. For example, police officers may consider a crash to be fatigue-related, when the following conditions appear (Horne et al. 1995; Horne et al. 1999; NCSDR/NHTSA, 2001): occured during late night or mid-afternoon; single vehicle ran off the roadway; occured on a high-speed road; absence of skid marks or braking. Some fatigue-related crashes are determined even by eliminating other causes of crashes (e.g. speeding, drunk driving, etc.).

To assist the identification of fatigue in a crash, proxy measurements are developed aiming to improve reporting accuracy (Filtness et al., 2015). In Australia, ATSB (2006) has developed the proxy definition for fatigue/sleep-related crash, and five jurisdictions in Australia have already incorporated proxy definition into their reporting process. In Queensland, for example, fatigue can be considered as a contributor to a crash when it fitted the proxy definitions: single-vehicle crashes in more than 100 km/h speed zones which occured during midnight and in the afternoon, or where a vehicle ran out of roadway and the driver did not try to avoid the crash (Armstrong et al., 2013; Filtness et al., 2015). Although these surrogate definitions are based on experience or scientific research, they are criticized for too specific (Crummy et al., 2008; Armstrong et al., 2013) and may provide misleading instructions for police officers. A questionnaire-based study conducted in Australia by Crummy et al. (2008) found that only a small proportion of participants that actually had a fatigue/sleep-related crash were correctly identified by ATSB proxy definitions (ATSB, 2006).

Reliable and accurate records are essential for assessing the scope of fatigue-related crash problems, as well as monitoring and evaluating the effectiveness of intervention measures. A survey in Ontario showed that 56.6% of traffic police felt that they did not receive enough training to identify drivers who were fatigued or drowsy, or determined the role of fatigue in a crash (Robertson et al., 2009). Although several risk factors identified by prior research and public belief are believed to contribute to fatigue-related crashes, few works have been done to prove whether these factors are useful for police officers to identify fatigue-related crashes. By the same token, some of the factors believed to be associated with fatigue-related crashes are not helpful in judging whether a crash is fatigue-related and may even lead to incorrect classification of the cause of crashes.

2.3 Methodology for Statistical Modeling

Recently, great improvement has been achieved in statistical methodologies for dealing with crash data. However, important methodological challenges still exist. Based on the review by Mannering and Bhat (2014), several statistical methodology relevant issues should be discussed in road safety research. Those issues include unobserved heterogeneity, endogeneity, spatial and temporal correlation, etc. Here, two of them, endogeneity and spatial correlation, will be discussed in the content of fatigue-related crash analysis.

2.3.1 Endogeneity

In econometrics, endogeneity problem is said to occur if the independent variable is correlated with the error term. This correlation can be caused by several reasons: omitted variables, measurement error, and simultaneity in simultaneous equations models. Endogeneity induces estimation bias in statistical models and may eventually lead to mistaken conclusions.

In the context of road safety research, endogeneity is often observed in crash injury severity models. For example, the endogeneity of seat belt use in injury severity of traffic crash are discussed using different model structure. Eluru and Bhat (2007) addressed the endogeneity of seat belt use in injury severity analysis by a joint random coefficients binary-ordered logit model. Since seat belt non-users may be intrinsically unsafe drivers and are more likely to be involved in severe crashes. Thus, the unobserved factors (e.g. unsafe driving habits) would influence injury severity outcomes. de Lapparent (2008) used bivariate ordered probit model to modeled seat belt use and injury severity regarding different types of vehicle users (drivers, front passengers, and rear passengers). The results shown that drivers may tend to take more risks as the compensation of the efficiency of seat belt in reducing injury severity. More discussion on endogeneity of seat belt-use can be found in Evans (1996), Dee (1997), Derrig et al. (2000), Cohen and Einav (2003), and Abay et al. (2013).

Besides, other factors such as collision types, driver characteristics, road characteristics can be endogenous to crash severity. Ye et al. (2008) applied a joint random coefficient multinomial logit-ordered to examined the two-vehicle collision and crash severity. Results suggested that the unobserved factors contributing to head-on collisions were negatively associated with unobserved factors contributing to severe injuries. On the contrary, the unobserved factors contributing to rear end crashes had positively impacts on severe injuries. Lee and Abdel-Aty (2008) evaluated the endogeneity of passenger characteristics (presence, number, and age of passengers) with crash characteristics (e.g. injury severity) by bivariate ordered probit models. Kim and Washington (2006) applied a joint model to examine the relationship between angle crashes and left turn lanes occurred in the intersections of 38 counties in Georgia. The results confirmed the endogeneity of left turn lane presence in angle crash occurrence models.

From the prospect of model structure, several types of model structures are applied to addressed endogeneity problem. The joint random coefficient models were applied to deal with the endogeneity of independent variables (Kim & Washington, 2006; Eluru & Bhat, 2007; de Lapparent, 2008; Ye et al., 2008). In the studies of Hutchinson (1983, 1986), Ouyang et al. (2002) and Yamamoto and Shankar (2004), they addressed the endogeneity problem due to common unobserved factors for different individuals involved in the same crash by bivariate models. Moreover, multilevel model (Jones & Jorgensen, 2003; Lenguerrand et al., 2006; Kim et al., 2007; Huang et al., 2008) and copula-based model (Rana et al., 2010; Abay et al., 2013) were also introduced to deal with endogeneity in automobile crash model.

For fatigue-related crash, although it is not in agreement, fatigue driving and injury severity in the crash may share some common influential factors, including observed and unobserved factors. Radun and Radun (2009) claimed that there was no connection between crash severity and whether the driver was judged to have been fatigued. However, more studies believed there existed some kind of connection (Haworth, 1998; Zhang et al., 2016). Fatigue-related crashes were often severe for that drivers could not take evasive action under fatigue (Haworth, 1998). Some factors related to fatigue driving may impair driver performance, then affect injury severity. For example, some unobserved factors related to the driver's internal state and circadian cycle can also affect both fatigue propensity and driving performance (Williamson et al. (2011) has given a detail review on that). Unfortunately, this information is almost impossible to collect due to traumatic effects and emotional state change after the crash (Radun & Radun, 2009). Some drivers might not admit fatigue or falling asleep during driving concerning about insurance and legal consequences (Corfitsen, 1999). Therefore, those common factors are often neglected, which may lead to endogeneity problem and biased estimation when analyzing the relationship between fatigue driving and injury severity.

2.3.2 Spatial correlation

There has been several road safety researches accounting for spatial correlation. Researches try to model spatial patterns with spatial components in the content of road safety studies, in which spatial heterogeneity and spatial autocorrelation are the two major concerns (LeSage, 1999; Bhat, 2000; Loo & Anderson, 2016).

Spatial heterogeneity refers to variations or non-stationarity over spatial units (Fotheringham et al., 2002; LeSage & Pace, 2009). The usage of random parameter is a commonly applied method to address this issue in spatial modeling. The application of random parameter in statistical models can be found in researches examining the heterogeneity of road sections or intersections (Milton et al., 2008; Anastasopoulos & Mannering, 2009; Anastasopoulos et al., 2012; Wu et al., 2013; Venkataraman et al., 2013; Chen & Tarko, 2014) as well as regionals (Xu & Huang, 2015; Coruh et al., 2015; Truong et al., 2016). On the other hand, spatial autocorrelation is defined as the value of a variable at a given spatial unit affects the value at contiguous unit (Cliff & Ord, 1973). It originates from missing exogenous variables and inappropriate spatial aggregation of the underlying observational units (Anselin, 1988; Tiefelsdorf & Griffith, 2007; Wang et al., 2013). It is expected that county-level variables within the same city have similar properties. For example, variables such as transportation regulations and traffic flow, are often absent from traffic crash models, which are spatial correlated across neighborhood areas. Generally, there are two ways to handle spatial autocorrelation: to take it into account in model setting or remove the spatial correlation between observations. Several attempts have been made to develop models to take spatial autocorrelation into account. One can specify the spatial structure to model spatial autocorrelation by geographically weighted regression (GWR) (Quddus, 2008; Hadayeghi et al., 2010; Pirdavani et al., 2013). However, GWR is computation-intensive and requires knowledge about the correlation structure (Griffith, 2002). Misspecification of correlation structure can also induce estimation bias (McMillen, 2004).

On the other hand, spatial filtering technique is a relatively new method for dealing with spatial data (Griffith, 2000a; 2000b; 2007). By constructing synthetic variables that accounts for spatial autocorrelation, spatial filtering can deal with spatial autocorrelation in regression analysis. This approach has already been widely used in disease mapping (Johnson, 2004; Griffith, 2005). Actually, spatial filtering technique is closely related to GWR. Griffith (2008) noted that GWR can be viewed as a special case of spatial filtering and established an indirect linkage between them via including includes interaction terms between spatial filtering and attribute variables. Generally speaking, three types of spatial filtering are introduced in the literature: distance-based eigenvector spatial filtering (Dray et al., 2006), G-statistics spatial filtering (Getis, 1990; Ord & Getis, 1995), and eigenfunction spatial filtering (Griffith, 2000a; 2000b).

The principal coordinate analysis of neighbor matrices (PCNM) approach: the PCNM was originally proposed by (Borcard & Legendre, 2002). Dray et al. (2006) investigated the formal mathematical foundations and connections between PCNM and spatial autocorrelation structure functions. The original approach constructs a truncated distance matrix by calculating the pairwise Euclidean distance matrix given certain threshold value between the sampling spatial units. Then, the principal coordinate analysis (PCoA) is conducted to obtain principal coordinate by scaling each of eigenvectors of distance matrix. In Borcard and Legendre's (2002) approach, only eigenvectors with positive eigenvalues were selected. Dray et al. (2006) extend this approach by allowing negative eigenvalues that representing negative spatial autocorrelations.

The Getis's spatial filtering approach: the spatial correlation variable can be partitioned into a filtered nonspatial variable and a residual spatial variable (Getis & Griffith, 2002). Getis's spatial filtering approach is based on the difference between observed and expected local spatial statistics $G_i(d)$ within distance the d proposed by Getis and Ord (1992):

$$G_i(d) = \sum_{j=1}^{n} w_{ij}(d) x_j \Big/ \sum_{j=1}^{n} x_j \,,\ i \neq j \tag{2-1}$$

The spatial filtered variable x_i^* can be written as:

$$x_i^* = x_i \left[W_i / (n-1) \right] / G_i(d) \tag{2-2}$$

where x_i is the observed variables, W_i is the sum of the ith row of spatial weight matrix, and n represents the number of spatial units. $G_i(d)$ is the observed local spatial statistics value while $W_i/(n-1)$ represents the expected value of $G_i(d)$.

Therefore, the difference between x_i and x_i^* can be interpreted as the spatial component of variable X at spatial unit i. If there does not exist spatial correlation within d, $x_i - x_i^* = 0$. The positive value of $x_i - x_i^*$ indicates spatial autocorrelation among high value of variable X while the negative value of $x_i - x_i^*$ indicates spatial autocorrelation of lower value of X (Getis & Griffith, 2002). More discussion on Getis's spatial filtering approach can be found in Getis (1990) and Ord and Getis (1995).

The Griffith's eigenfunction spatial filtering approach: the Griffith's eigenfunction spatial filtering approach is based on the computational formula of Moran's I (MI) statistic (Griffith, 2000a; 2000b). Moran's I statistic is the most common used indicator of spatial autocorrelation. It is calculated as:

$$MI = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{2-3}$$

where n is the number of spatial units, x_i is the ith observed value of variable X at location i, $\bar{x}$ is the mean of X, and w_{ij} is the spatial weight of location i and j. This approach extracts orthogonal and uncorrelated components from the matrix representing spatial structure (Tiefelsdorf & Boots, 1995). These components can be viewed as independent map patterns, which represent the latent spatial correlation of georeferenced variable given spatial weight matrix (Patuelli et al., 2006). MI can be considered as a weighted sum of the eigenvalues of matrix $\left(I - 11^T / n\right)C\left(I - 11^T / n\right)$, where C is the binary spatial connectivity matrix, I is an $n \times n$ identity matrix, and 1 is the $n \times 1$ vector of ones (Tiefelsdorf & Boots, 1995). The Griffith's eigenfunction spatial filtering approach takes locational information into account by generating eigenvectors of given geographic weights matrix. Griffith's eigenvector spatial filtering approach enjoys more flexibility since it can be applied in non-linear models as well as dealing with both positive and negative autocorrelation. Moreover, the mutually orthogonality of eigenvectors makes it possible to be applied either individually as predictor variables or simultaneously to a regression system (Tiefeldorf & Griffith, 2007).

The spatial filtering approach has been successfully applied in many fields, such as disease mapping (Tiefelsdorf & Griffith, 2007), crime analysis (Helbich & Jokar Arsanjani, 2015), migration flow (Chun, 2008; Chun & Griffith, 2011), land use (Wang et al., 2013), real estate (Helbich & Griffith, 2016), and ecology and biogeography (Diniz-Filho & Bini, 2005; Griffith & Peres-Neto, 2006). However, the application of spatial filtering in road safety research context has not been found.

CHAPTER 3

Data Description

3.1 Introduction

This chapter introduces the data sets used in the dissertation. The study area is Guangdong province (109°45′E - 117°20′E, 20°09′N - 25°31′N), China. It is located in the southern part of China, including 21 cities, and 121 counties. By the end of year 2014, Guangdong is populated by 110 million residents over an area of 179,716.02 square kilometers. In 2012, there are 25,424 traffic crashes in Guangdong province, accounting for 12.81% (the highest among all 31 provinces in China) of the total number of crashes. The corresponding proportion of death, injury and direct property losses of Guangdong province accounts for 9.65%, 13.30% and 7.71% of the total number/amount, which is also the highest among all 31 provinces in China (Traffic Management Bureau, Ministry of Public Security, PRC, 2013).

Since the focus of this research is fatigue-related crash, the definition of fatigue-related crash in laws and regulations should be clarified. In China, the laws and regulations related to fatigue-related crash are as following:

- Law of The People's Republic of China on Road Traffic Safety, article 22 (The

Standing Committee of the National People's Congress, 2004; 2011 revised): no one shall force, instigate or connive at a driver to violate the road traffic safety laws and regulations or the driving requirements on motor vehicle safety to drive a motor vehicle.

- Implementation of the Road Traffic Safety Law of the People's Republic of China, article 62 (State Council, 2004; 2017 revised): drivers shall not drive more than four hours without rest breaks or rest breaks are less than 20 minutes.
- Provisions on the Application for and Use of Driving Licenses (Ministry of Public Security, 2013): drivers who drive large/medium passenger vehicles or road transport vehicles for dangerous goods more than four hours without rest breaks or rest breaks are less than 20 minutes will receive the harsher 12 demerit points; drivers who drive other types of vehicles except for large/medium passenger vehicles or road transport vehicles for dangerous goods more than four hours without rest breaks or rest breaks are less than 20 minutes will receive 6 demerit points.
- Regulation on the Road Traffic Safety of Guangdong Province, *article 59* (The Standing Committee of the Guangdong Provincial People's Congress, 2011): drivers who drive more than four hours without rest breaks or rest breaks are less than 20 minutes shall be ordered to make correction, given disciplinary warning or 200 yuan fine.

It is important to noted that in this study the definition of fatigue-related crash follows the definition by Guangdong Traffic Accident Dataset. A crash is defined as fatigue-related crash when fulfilled one of the following conditions in this study: (1) driving cars more than eight hours a day, (2) engaging in other work excessive physical exertion, and (3) lacking of sleep which results in sleepy or lower reaction rate, so that the driver is having difficulty in assessing traffic conditions immediately and reacting accurately.

3.2 Data Source

Three categories of data are used in this study: crash data, GIS data and macroscopic data.

3.2.1 Crash data

The crash data is extracted from Guangdong Traffic Accident Dataset (GTAD). GTAD is sourced from the Traffic Management Sector Specific Incident Case Data Report, the Road Traffic Accident Database of China’s Public Security Department. All police-recorded crash records occurred in Guangdong Province during 2005-2014, were filtered from the Traffic Accident Database. Five categories of variables are included in the crash data: driver characteristics, vehicle characteristics, road characteristics, environment characteristics and crash characteristics. Some selected variables from all five categories are listed in Table 3.1.

Table 3.1 List of variables in crash data

Categories	Variable
Personal Characteristics	Age; Gender, Occupation; Driving experience; Driving license; Nationality; Injury level; Injured position; Mean of transportation, etc.
Vehicle Characteristics	Vehicle type (e.g. large/medium/mini, cargo/passenger vehicle); Commercial condition; Loading condition; Safety condition; Ownership, etc.
Road Characteristics	Road type (e.g. highway, urban road); Intersection type; Separator; Signal; Road surface type and condition; Road alignment, etc.
Environment Characteristics	Date; Time in a day; Weather; Visibility; Terrain, Lightening condition, etc.
Crash Characteristics	Collision type (e.g. head-on, hit fixed object); Crash type (e.g. speeding, fatigue driving); First assessment of crash cause; Final assessment of crash cause; Number of casualties and loss in property; Crash location, etc.

During this period, a total number of 328,733 crashes are recorded and stored in the database. Among them, 277,924 crashed are motor vehicle violations, 10830 are non-motor vehicle violations, and 6,703 are pedestrian or passenger violations (Figure 3.1).

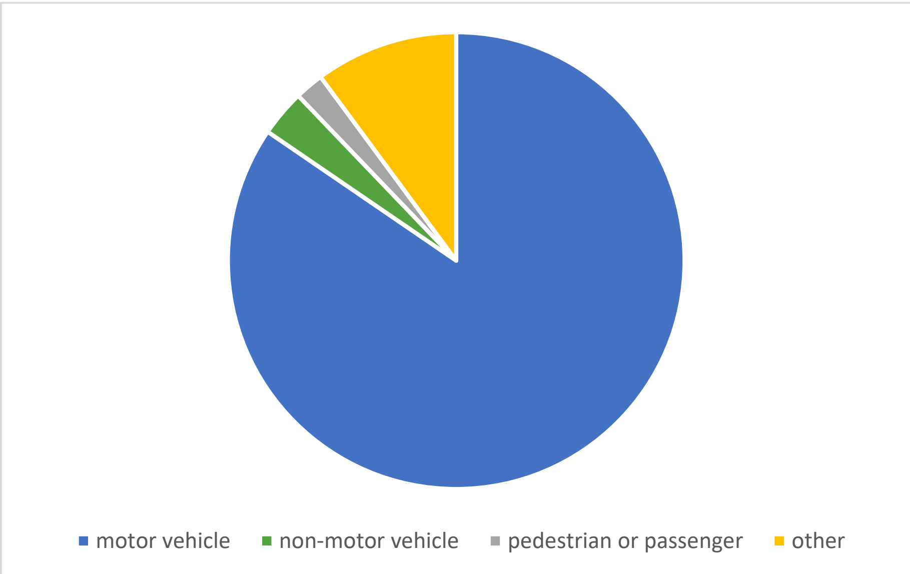

Figure 3.1 Total number of crash by violation

It should be noted that the injury severity of each individual involved in this data set is categorized into three levels:

- Fatal injury: a person who dies within 30 days of a crash as a result of injuries caused by that crash.
- Severe injury: a person who sustains severe injuries as a result of the road crash and who does not die as a result of those injuries within 30 days of the crash.
- Minor injury: a person who sustains minor injuries as a result of a road crash and who does not die as a result of those injuries with 30 days of the crash.

Accordingly, crash types based on injury level can be categorized as:

- Fatal crash: a crash for which there is at least one fatality.
- Injured crash: a non-fatal crash for at least one person sustains injury but no person is admitted to dies within 30 days of the crash.
- Property damage only crash: a crash resulting in property damage and no person is injured or dies within 30 days of the crash.

The following Figure 3.2 shows the number of crash by injury level. The majority number of crashes are injured crashes (75%), 19% crashes are fatal and 5% crashes are property loss only. It should be noted that the small number of property damage only crash is due to most of these crashes are not put into investigation procedure. Therefore, these crashes are not included in the crash dataset.

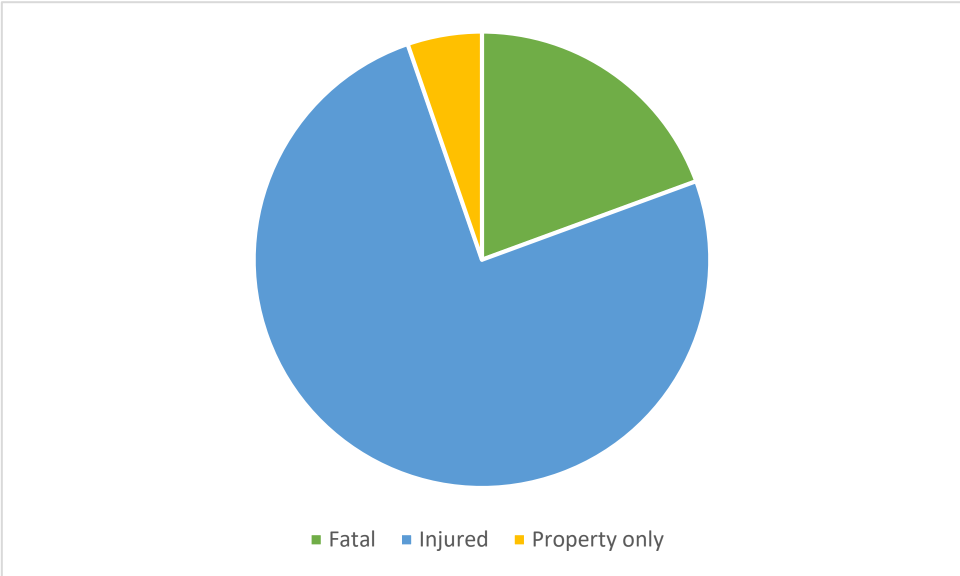

Figure 3.2 Number of crash by injury level

Figure 3.3 and Figure 3.4 show the proportion of collision types in all crash records and fatigue-related crash records. In Figure 3.3, the three major collision types among all the crash are side collision (43%), head-on collision (17%) and hitting pedestrian (13%). For fatigue-related crashes, the major collision types are rear-end collision (25%), side collision (23%) and hitting fixed object (19%). It should be noticed that in fatigue-related crash records, the proportion of single vehicle crash is much higher than the proportion in all crash records. In fatigue-related crash, the proportion of hitting fixed object crash (19%) and rollover (10%) are much higher than the proportion of hitting fixed object crash (6%) and rollover (2%) among all crash records.

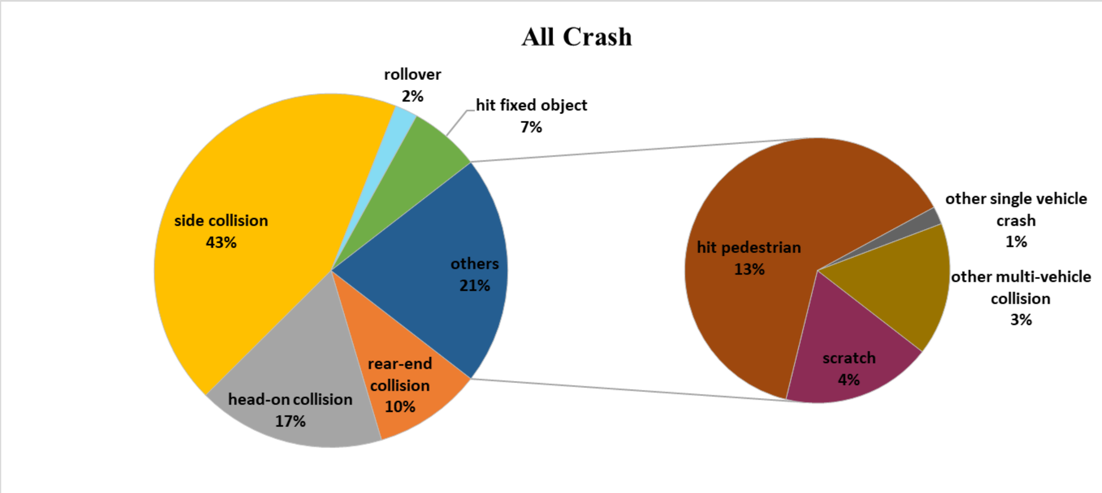

Figure 3.3 Proportion by collision types for all crash records

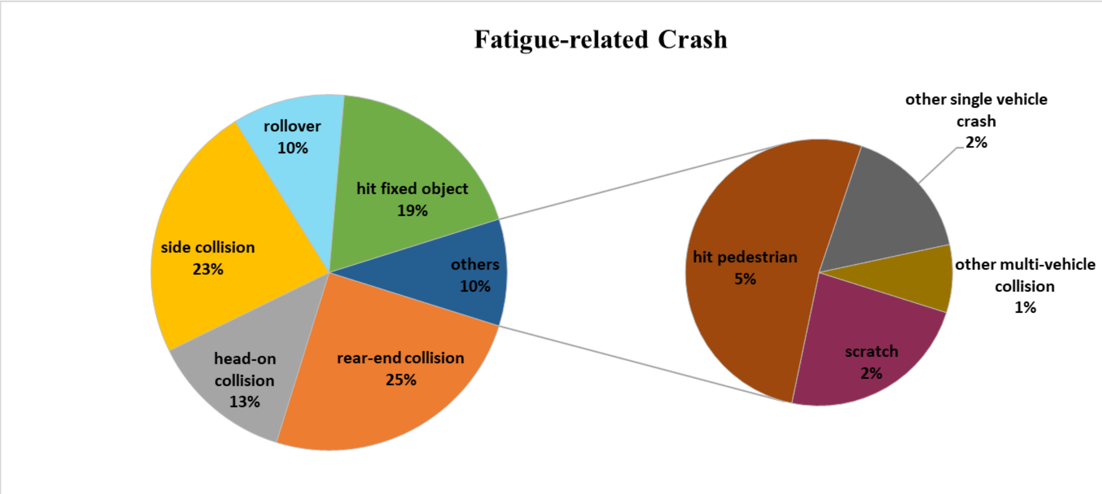

Figure 3.4 Proportion of collision types for fatigue-related crashes

As shown in Figure 3.5, illegal lane change leads to highest number of crash (11427 crashes). 8509 crashes were caused by speeding, 6155 crashed were caused by drunk driving while 1628 crashes were fatigue-related.

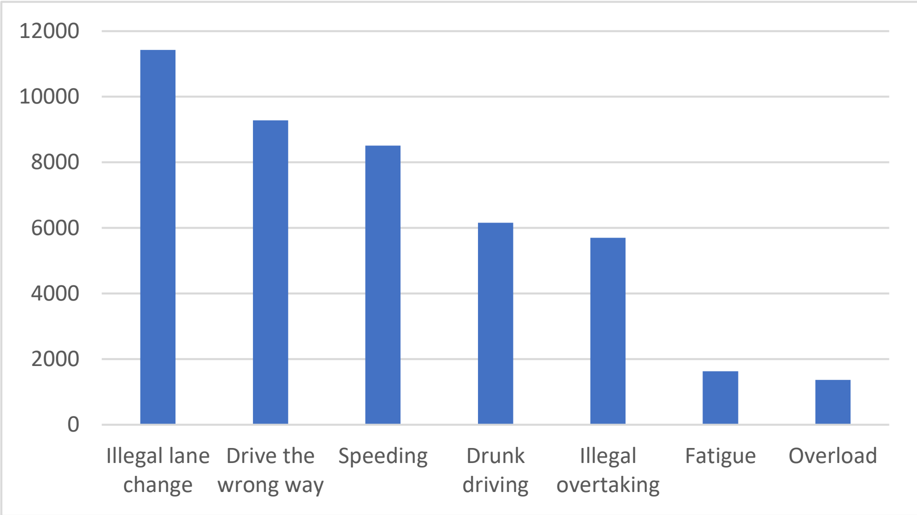

Figure 3.5 Total number of major traffic violations of motor vehicle driver

Figure 3.6 display the temporal trend of total crash number and fatigue-related crash number in Guangdong province during 2006-2014. The crash count in 2005 is not presented in Figure 3.6. due to the incompleteness of crash records in this year. It can be seemed in the figure that the total number of crash was decreasing during 2006-2012, then was increasing from 2012. For fatigue-related crash, it also shows similar trend as the change of total number of crash. The decrease of fatigue related crash number may be related to the raise of penalty for fatigue driving (Ministry of Public Security, 2013).

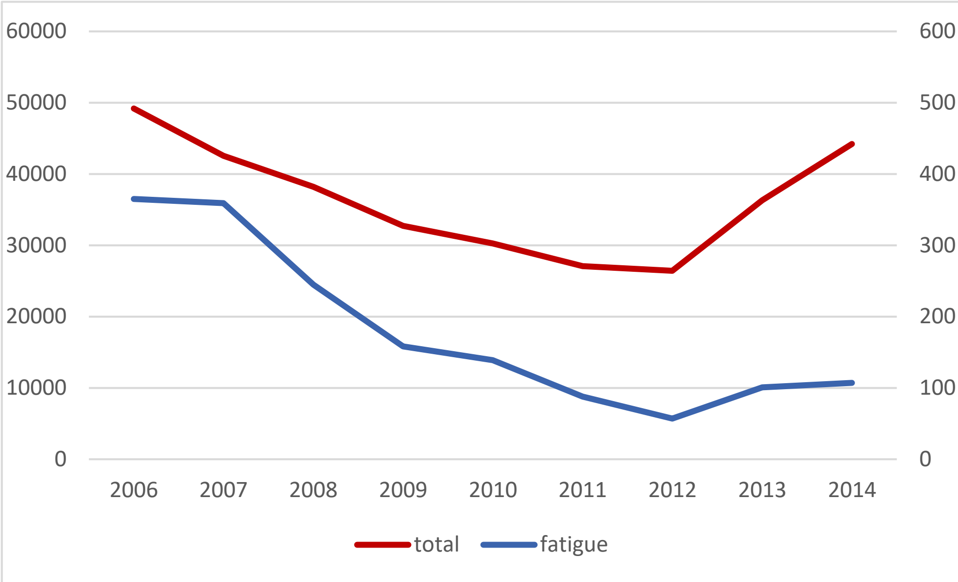

Figure 3.6 Temporal trend of total crash and fatigue-related crash from 2006-2014

3.2.2 GIS data

Road length and boundary information applied in this research are obtained from the Geospatial Database of the 1:1,000,000 Geological map of China. This database contains basic information about road and county boundary. The length of road of different level within the boundary of a county are summed up separately. Figure 3.7 shows the maps of highway, national road and provincial road in Guangdong province.

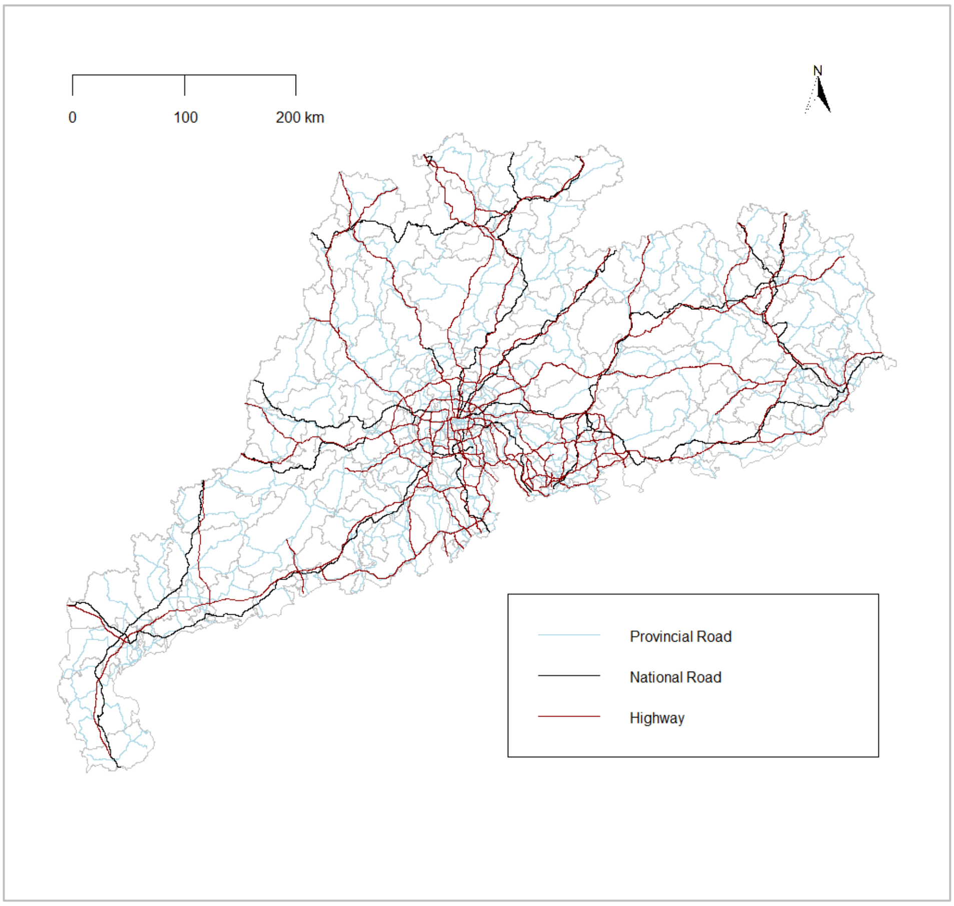

Figure 3.7 Map of roadway in Guangdong

Furthermore, a statistical descriptive analysis for four types of roads are considered: expressway, national road, provincial road, and urban expressway can be found in Table 3.2.

Table 3.2 Summary of total length for four types of road by county

Road types	N	Mean	SD	Min.	Max.
Highway (1000 km)	120	0.11	0.11	0.00	0.68
National road (1000 km)	120	0.06	0.05	0.00	0.22
Provincial road (1000 km)	120	0.16	0.12	0.01	0.80
Urban express way (1000 km)	120	0.01	0.03	0.00	0.23

3.2.3 Macroscopic data

The socio-economic data are collected from the Guangdong Statistical Yearbooks and Guangdong 1% Population Sampling Survey Data (Guangdong Statistical Bureau, 2017). The following figures display the distribution of highway, national road, provincial road and urban expressway in 120 counties in Guangdong and darker color means higher value. It can be seemed from Figure 3.8 and Figure 3.11 that highway and urban express way are clustering in the middle south part of Guangdong, in which several big cities are located. The distribution of national road and provincial road are more even comparing to highway and urban expressway (Figure 3.9 and Figure 3.10).

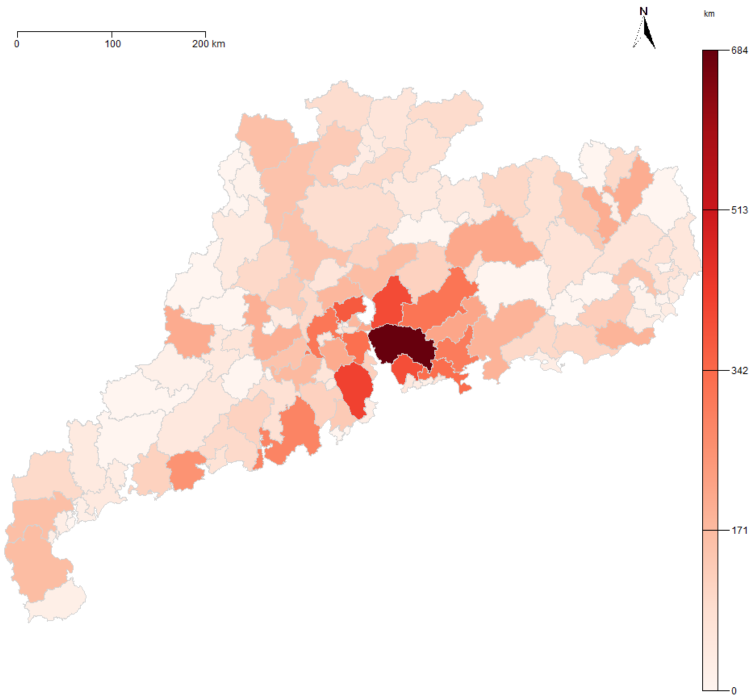

Figure 3.8 Total length of highway distribution (km)

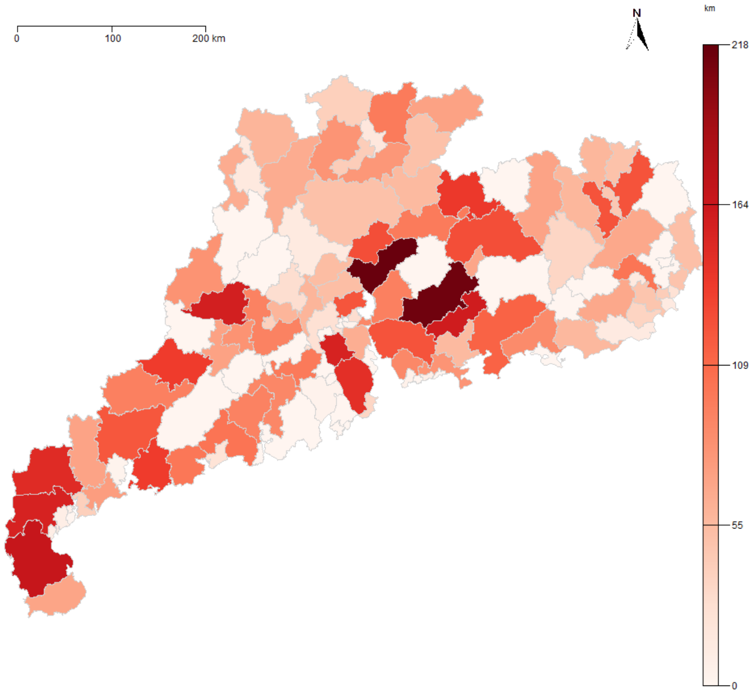

Figure 3.9 Total length of national road distribution (km)

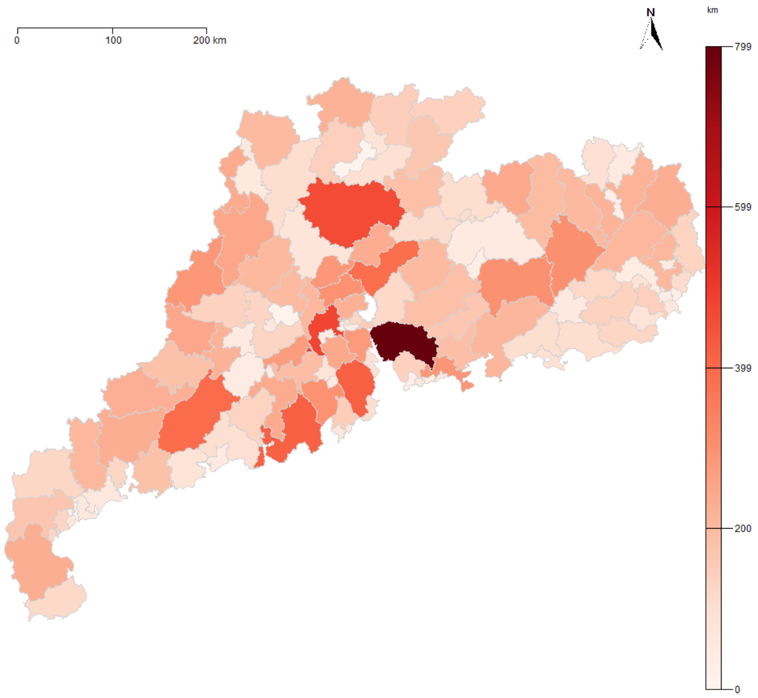

Figure 3.10 Total length of provincial road distribution (km)

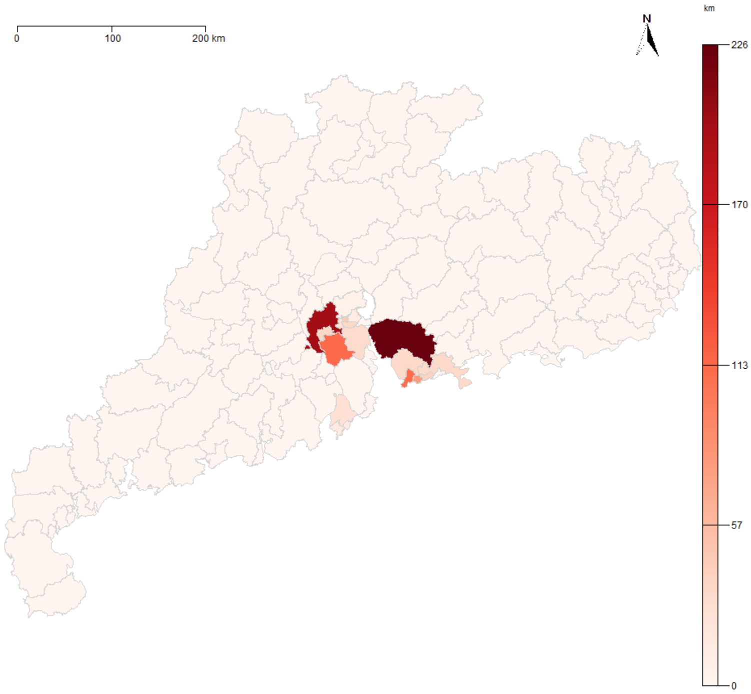

Figure 3.11 Total length of urban expressway distribution (km)

The distributions of other macroscopic variables are also presented. Population (Figure 3.12), transportation employees (Figure 3.15) as well as the number of public transport passengers (Figure 3.16) are also clustered in the middle south Guangdong. On the contrary, the proportion of youth (Figure 3.13) and elder people (Figure 3.14) are higher in other parts rather than middle south of Guangdong. The summary of those variables is presented in Table 3.3.

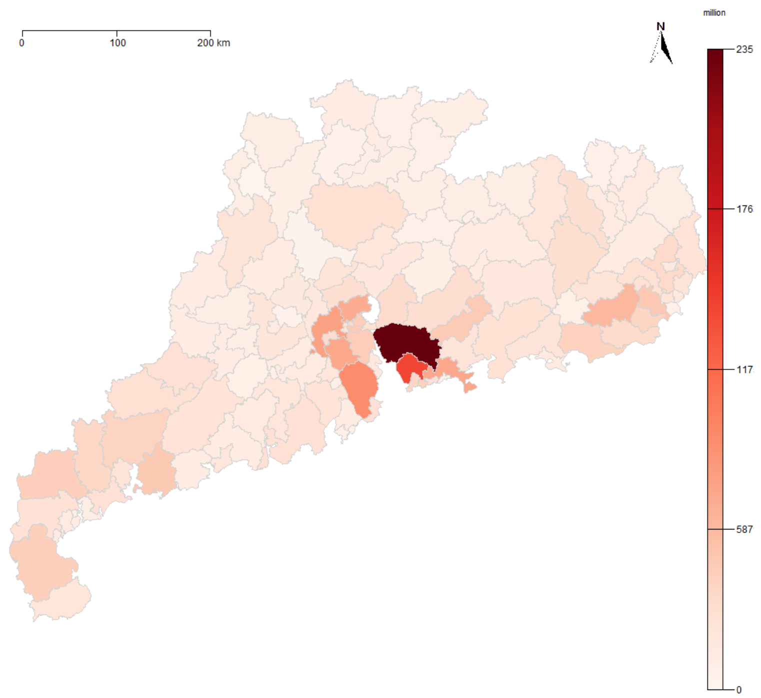

Figure 3.12 Population distribution (million)

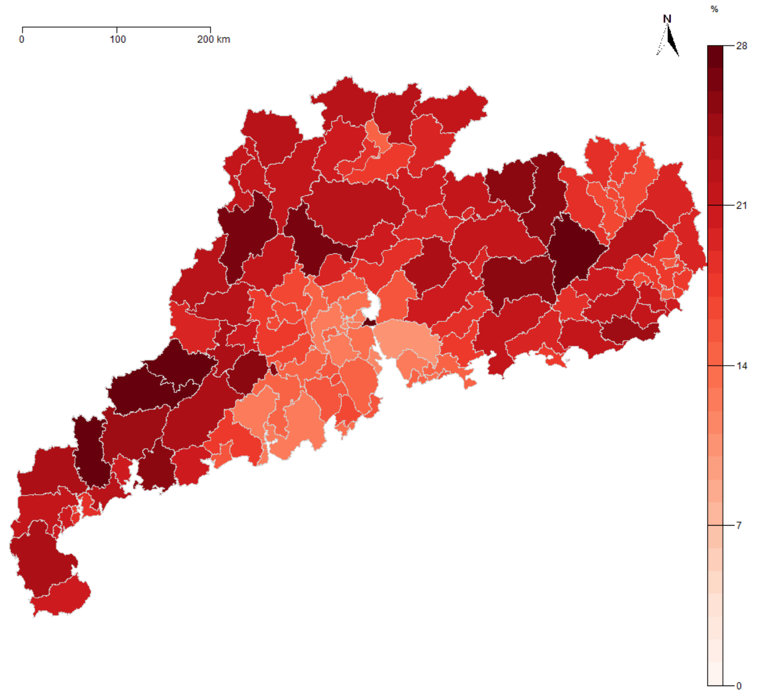

Figure 3.13 Proportion of population age from 0-15 years old distribution (%)

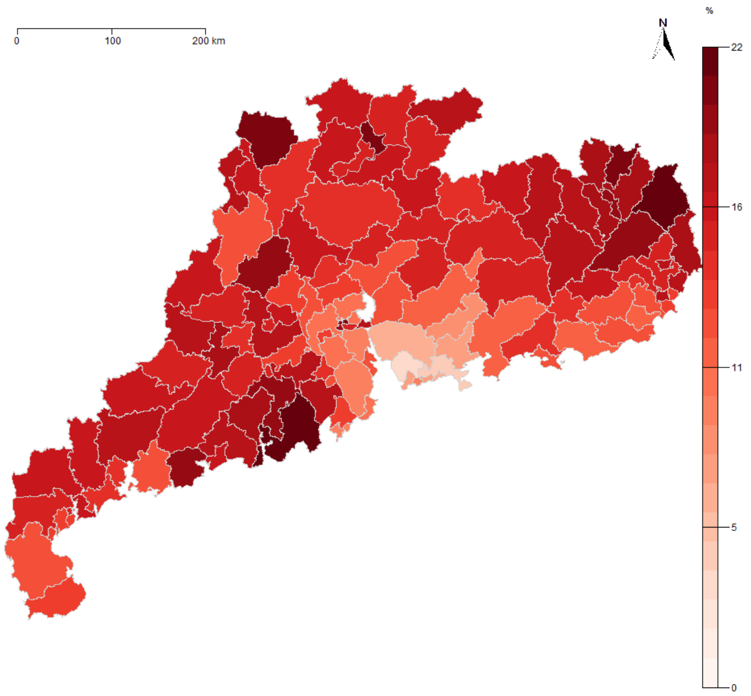

Figure 3.14 Proportion of population age elder than 60 years old distribution (%)

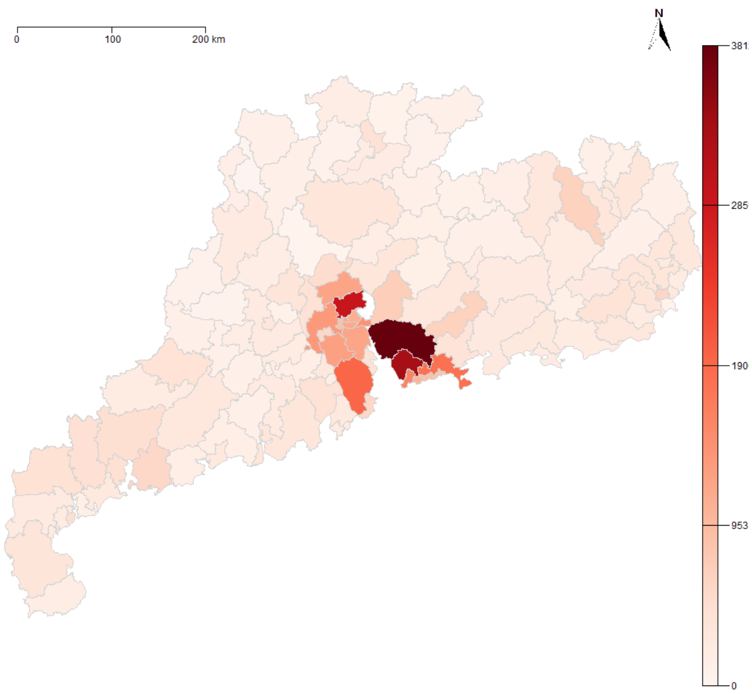

Figure 3.15 Number of transportation employee distribution

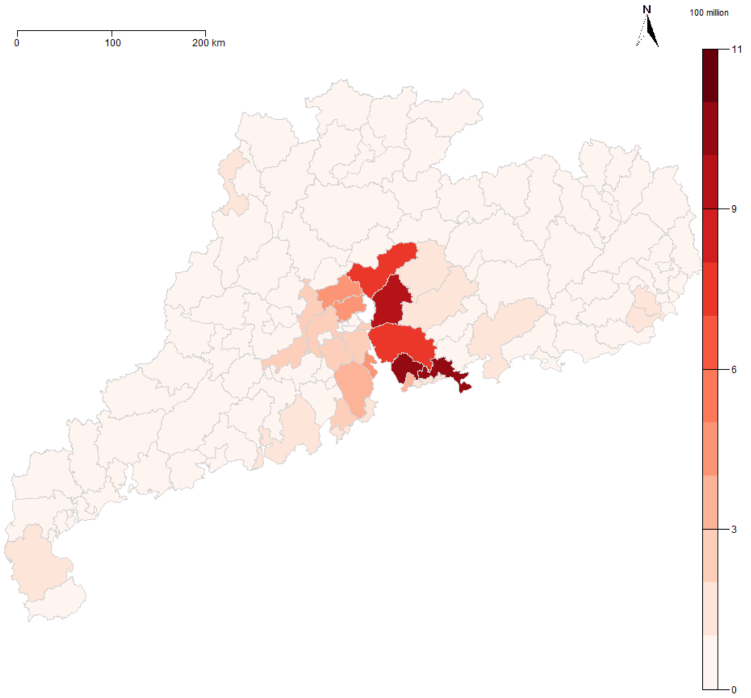

Figure 3.16 Number of public transport passenger distribution (100 million)

Table 3.3 Descriptive statistics of variables

Variable	Mean	SD	Min.	Max.
Number of population (Million people)	2.95	0.75	0.98	5.46
Proportion of people aged from 0-15	0.18	0.04	0.10	0.28
Proportion of people aged elder than 60	0.14	0.04	0.03	0.22
Average room per person	1.06	0.29	0.44	1.95
Proportion of employee in transportation industry	0.03	0.02	0.01	0.22
Number of passenger using bus and taxi (100 million)	0.81	1.91	0.01	11.48

CHAPTER 4

Understanding Factors Associated with Misclassification of Fatigue-related Crashes in Police Record

4.1 Introduction

Fatigued driving is a serious problem threatening road safety all around the world. Police records from different countries indicate a range from 1-4% incidence of fatigue/sleep-related crashes of all registered crashes (Traffic Management Bureau, Ministry of Public Security, PRC, 2008; Radun & Radun, 2009). However, several questionnaire-based surveys suggest the role of fatigue in a traffic crash is misestimated. National Sleep Foundation (2008) reported that approximately 32% of respondents in Sleep in America Poll had driven while fatigued at least once a month. In China, a survey conducted in Guangdong province in 2007 also showed that 9.3% vehicle drivers had the fatigue driving experience in the past 30 days (Yan et al., 2010). The difference between police reports and surveys implies that police reports could have significantly misestimated the harmfulness of fatigue in road safety. One of the possible reasons is police officers are not so alert to the presence of fatigue and have difficulties in identifying fatigue-related crashes (Robertson et al., 2009).

Among all causes of traffic crashes, fatigue-related crashes are easily neglected or misclassified due to the difficulty in observing and identifying driver fatigue (Radun et al., 2013; Filtness et al., 2015). No blood or breath test can be applied to quantify driver's fatigue level at crash scene (Pack et al. 1995; DaCoTA, 2012). As a result, there is currently no standard methodology for identifying fatigue as the cause of the crash (Crummy et al., 2008; Filtness et al., 2015) and defining fatigue-related crash largely relies on inferential evidence or experience. For example, police officers may consider a crash to be fatigue-related, when the following conditions appear (Horne et al., 1995; NCSDR/NHTSA, 1998; Horne et al., 1999): occur during late night or mid-afternoon; single vehicle run off the roadway; occur on a high-speed road; absence of skid marks or braking. Some fatigue-related crashes were determined even by eliminating other causes of crashes (e.g. speeding, drunk driving, etc.).

To assist identification of fatigue in a crash, proxy measurements are developed aiming to improve reporting accuracy of fatigue-related crashes (Filtness et al., 2015). In Australia, ATSB (2006) has developed the proxy definition for fatigue/sleep-related crash, and five jurisdictions in Australia have already incorporated proxy definition into their reporting process. In Queensland, for example, fatigue can be considered as a contributor to a crash when it fitted the proxy definition: single-vehicle crashes in more than 100 km/h speed zones which occur during midnight and in the afternoon, or where a vehicle runs out of roadway and the driver does not try to avoid the crash (Armstrong et al., 2013; Filtness et al., 2015). Although these proxy definitions are based on experience or scientific research, they are criticized for too specific (Crummy et al., 2008; Armstrong et al., 2013) and may provide misleading instructions for police officers. A questionnaire-based study conducted in Australia by Crummy et al. (2008) found that only a small proportion of participants that actually had a fatigue/sleep-related crash were correctly identified by ATSB proxy definitions (ATSB, 2006).

Reliable and accurate records are essential for assessing the scope of fatigue-related crash problems, monitoring and evaluating the effectiveness of intervention measures. A survey in Ontario showed that 56.6% of traffic police felt that they did not receive enough training to identify drivers who were fatigued or drowsy, or determined the role of fatigue in a crash (Robertson et al., 2009). Although several risk factors identified by prior research and public belief are believed to contribute to fatigue-related crashes, few works have been done to prove that whether these factors are useful for police officers to identify fatigue-related crashes. That is, some of the factors believed as associated with fatigue-related crashes are not helpful in judging whether a crash is fatigue-related and may even lead to incorrect classification of the cause of crashes. Therefore, in this study, we proposed an analysis framework based on existing crash data to identify factors that easily make fatigue-related crashes misclassified by police officers, examine the interactive effects of those factors, and provide better inference for determining fatigue-related crash by removing some misleading terms, which help to improve enforcement strategies.

4.2 Methodology

4.2.1 Objectives and research strategy

This study aims at investigating potential factors that hinder police officers' identification of fatigue-related crashes. However, some factors have not only individual effects but also combinatorial effects on the determining of fatigue-related crashes. Classic logistic regression model lacks appropriate criteria to incorporate interactions between independent variables when there are a large number of variables to be considered. Instead, ignoring interactions may cause biased estimation. Therefore, our strategies for this analysis are: (1) Association rule data mining technique is applied to identify important interactions between factors, which helps overcome the disadvantage of classic logistic regression model in selecting appropriate interactions; (2) Incorporating the interactions identified by association rules, binary logistic regression models are applied to find out factors that hinder police officers from correctly identifying fatigue-related crashes.

4.2.2 Association rule analysis

Regression models in road safety research focus on establishing and analyzing relationships between "dependent" and "independent". It is also important to take the correlation between "independent" variables into consideration since it may hamper the statistical analysis (Pande & Abdel-Aty, 2009). With the increasing number of independent variables, however, the number of interactions will grow at an accelerated rate. Thus, the methodology for identifying potential interaction among a large number of crash-related factors is needed. Association rule data mining technique can potentially identify relationships that are not well known from current research works and have been used in traffic safety research (Pande & Abdel-Aty, 2009; Montella et al., 2011; Das & Sun, 2014; Weng et al., 2016). Some studies have combined association rule data mining technique with logistic regression model for other purposes (Kamei et al., 2008; Shaharanee et al., 2009), but few of them use association rule analysis as a tool for selecting potential interactions among variables. Changpetch and Lin (2013) proposed a model selection method procedure for logistic (Changpetch & Lin, 2013a) and multinomial logit model (Changpetch & Lin, 2013b), which help to improve the classic model by considering potential interactions.

In this study, association rule analysis is performed using a priori algorithm according to the methodology introduced by Agrawal et al (1993). A rule is defined as an implication of the form "$A \to B$", where A is the antecedent (left-hand-side, LHS) and B is the consequent (right-hand-side, RHS). It is important to note that the rule should not be interpreted as a direct causation, but as associations between variables (Pande & Abdel-Aty, 2009; Montella et al., 2012). Three measures are commonly used in filtering rules: Support, Confidence and Lift. Support measures the frequency of LHS and RHS appearing in the dataset and is calculated as follows:

$$Support(A \to B) = P(AB) \tag{4-1}$$

where $P(AB)$ represents the probability of case containing A and B at the same time. Confidence determines how frequently RHS appears given that LHS occurs:

$$\begin{aligned} Confidence(A \rightarrow B) &= P(B \mid A) \\ &= P(AB)/P(A) \end{aligned} \tag{4-2}$$

where $P(A)$ is the probability of case containing A. Lift is a measure of the statistical dependence of the rule. A lift value which is smaller than one indicates negative independence between LHS and RHS, a value equal to one indicates independence, and a value which is greater than 1 indicates positive interdependence (Montella et al., 2012). Higher lift value indicates stronger associations. Lift is defined as follows:

$$\begin{aligned} Lift(A \rightarrow B) &= P(B \mid A)/P(B) \\ &= P(AB)/P(A)P(B) \end{aligned} \tag{4-3}$$

To make sure that the identified rules are reasonable and accurate, the minimum threshold values for these three indexes need to be specified. Since there are no clear criteria for choosing threshold values, different studies employed different threshold support and confidence values (Pande & Abdel-Aty, 2009; Montella, 2011; Montella et al., 2012; de Oña et al., 2013) based on the nature of the data (balanced or not) and sample size (small or large databases). For example, Pande and Abdel-Aty (2009) set 0.009 and 0.1 for them respectively. Thus, in this study the minimum threshold values for Support, Confidence and Lift are set as follows: $Support \geq 0.01$, $Confidence \geq 0.1$, and $Lift \geq 1.2$. It also needs to be emphasized that only rules with two items in the

LHS are selected for ease of interpretation. We firstly generate rules with non-fatigue to fatigue crash or fatigue to non-fatigue crash in RHS from all the generated rules. Then, all the selected rules are descending ordered by confidence, and the top ten are changed into interactions and incorporated into logistic regression models as inputs.

4.2.3 Binary logistic regression model

Two assessment results will be recorded for normal procedure crash records in the database. The one recorded by the police officer at the crash scene is denoted as on-site assessment. Normally, *on-site assessment* was determined by a quick check at the crash scene, surrounding environment and simply ask those who were present for facts of the crash. The other one recorded in the final report is denoted as *final assessment*, which is the assessment result after an in-depth investigation. After placing a case on file for investigation, more detail information of driving condition is also collected for investigation. Information such as skid marks and scrub marks, the condition of vehicle mechanical (e.g. brakes, steering, tires, and lights). Besides road surveillance video and driving records, police officers will obtain and examine marks from the crash scene, statements from drivers and witnesses, collect off-scene information, and do vehicle mechanical inspection. Based on that information, a crash reconstruction can be performed to examine the real cause of the crash. Therefore, based on more detailed and reliable evidence, the final assessment is believed to be the accurate assessment result for the crash. It needs to be emphasized that on-site assessment and final assessment can be different, and this inconsistency of crash cause enables us to analyze the factors hindering making correct judgments. And the simple procedure crash records with on-site assessment only are not included in this study due to lack of information for judging whether they have been correctly classified or not. Thus, based on the combinations of on-site assessment and final assessment, three types of records are presented in Table 4.1:

- **Non-fatigue** to **Fatigue** (**N-F crash**): crashes with non-fatigue as *on-site assessment* and fatigue as *final assessment*;
- **Fatigue** to **Non-fatigue** (**F-N crash**): crashes with fatigue as *on-site assessment* and non-fatigue as *final assessment*;
- **Fatigue** to **Fatigue** (**F-F crash**): the *on-site assessment* and *final assessment* are both fatigue.

Table 4.1 Summary of types of records by injury severity level

	Fatal and severe injury	Minor injury	Total
N - F crash	55	66	121
F - N crash	156	173	329
F - F crash	211	350	561
Total	422	589	

Similar to the terms widely used in medical screening (Stegeman et al., 2013), fatigue detection at the crash scene can be regarded as a test and the final result after full investigation as the real cause. Then, we can define the false negative fatigue-related crash as a fatigue-related crash that was misclassified into other causes at the crash scene. A false positive fatigue-related crash is defined as a crash believed to be fatigue-related but is actually not. Binary logistic regression models are employed to identify significant factors affecting false positive and false negative fatigue-related crash detection. The first model is established to identify factors related to false negative fatigue-related crash detection. The binary outcomes represented by a dummy variable (1 indicates false negative fatigue-related crash detection, 0 is correct fatigue-related crash detection), is used as dependent variables. The second model is built to identify factors affecting false positive fatigue-related crash detection. Similarly, the dependent variable is also dummy variable (1 indicates false positive fatigue-related crash detection, 0 is correct fatigue-related crash detection).

Furthermore, police officers may treat crashes differently under different injury severity level (e.g. if police suspect the crash may be related to fatigue driving, they tend to give an oral examination to determine whether the crash is fatigue-related level in less severe crashes (if the people involved are sober)). Thus, we separated our dataset into two groups based on their recorded injury severity level of the driver who is responsible for the crash: fatal and severe injury, and minor injury. We do not include property only crashes because part of them was applied to the simple procedure without further investigation and analysis based on selected sample will be biased.

4.3 Data

The data employed in this study were obtained from the Road Traffic Accident Database of China's Public Security Department (Zhang et al., 2016). All police-recorded fatigue crashes, relevant crash records occurred in Guangdong Province during 2005 - 2014, were filtered from the Traffic Accident Database. Only records in which the cause of a crash was convicted as fatigue-related and the involving driver who was fully or mainly responsible for the crash, were used in this study. According to definition by the database, a crash was defined as fatigue-related crash when fulfilled one of the following conditions: (1) driving cars more than eight hours a day, (2) engaging in other work excessive physical exertion, and (3) lack of sleep which results in sleepy or lower reaction rate, so that the driver is having difficulty in assessing traffic conditions immediately and reacting accurately. In this study, 1101 general procedure crash records were extracted from the database. Among them, 561 are F-F crashes, 121 are N-F crashes, and 329 are F-N crashes. Although fatigue-related crashes account for a small percentage of all crashes in our dataset, the percentage of incorrect detection of fatigue-related crashes is really high.

To focus on the meaningful analysis, several variables will be considered in this study guided by prior research. These variables were selected into the final models: crash characteristics (crash type), driver characteristics (driver's gender, age, and occupation), vehicle characteristics (vehicle type and insurance condition), roadway characteristics (road type, lane type, and road segment) and environmental characteristics (lighting condition and weather). In order to consider non-linear effects in logistic regression models, driver's age was categorized: age (⩽30, 31-40, 41-50, and ⩾51). The description of these variables is presented in Table 4.2.

Table 4.2 Descriptive statistics of variables

	Fatal and severe injury		Minor injury	
Variable	**Count**	**Percentage**	**Count**	**Percentage**
Crash characteristics				
Motor vehicle crash	233	55.21%	402	68.25%
Rollover	70	16.59%	75	12.73%
Hit fixed object	114	27.01%	88	14.94%
Driver characteristics				
Male	416	98.58%	553	93.89%
31-40 years old	156	36.97%	218	37.01%
41-50 years old	94	22.27%	108	18.34%
≥51 years old	21	4.98%	47	7.98%
Clerk	24	5.69%	44	7.47%
Worker	93	22.04%	99	16.81%
Farmer	71	16.82%	131	22.24%
Self-employed	57	13.51%	64	10.87%
Migrant worker	52	12.32%	37	6.28%
Unemployed	3	0.71%	10	1.70%
Vehicle characteristics				
Heavy/medium truck	111	26.30%	117	19.86%
Light truck	32	7.58%	33	5.60%
Large/medium bus	15	3.55%	13	2.21%
Passenger car	89	21.09%	104	17.66%
Have insurance	346	81.99%	495	84.04%
Roadway characteristics				
Expressway	99	23.46%	110	18.68%
Urban expressway	15	3.55%	18	3.06%
Urban road	51	12.09%	128	21.73%
Motor vehicle lane	340	80.57%	414	70.29%
Non-motorized vehicle lane	14	3.32%	35	5.94%
Mix lane	50	11.85%	122	20.71%
Intersection	46	10.90%	78	13.24%
Special road segment	35	8.29%	31	5.26%
Environmental characteristics				
Dark with street light	97	22.99%	125	21.22%

Dark without street light	165	39.10%	169	28.69%
Cloudy	58	13.74%	98	16.64%
Total number of crash	422		589	

4.4 Results

R software 3.3.1 was applied to generate association rules and estimate the models.

4.4.1 Association rule analysis

In this study, 'arules' package in R software is employed for computation of association rules (Hahsler et al., 2007). To find out the potential associations or patterns among the items, association rule analysis is conducted using N-F crash and F-N crash sample sets separately for two injury severity levels. The results are presented in Table 4.3. Then, these rules are converted into dummy variables. Suppose the rule being selected is "if $X_i = x_i$ and $X_j = x_j$ then $Y = y$", where x_i is the level of variable X_i, x_j is the level of variable X_j, and y is the level of response Y (Changpetch & Lin, 2013). The interaction between X_i and X_j is denoted as 1 if $X_i = x_i$ and $X_j = x_j$, and as 0 otherwise. For example, Rule 1 for N-F fatal and severe injured crash in Table 4.3 which is "Given a fatal and severe injured crash is N-F crash, the crash occurred in cloudy weather condition and the mainly responsible driver drove large/medium cargo vehicle" is converted into a dummy variables D1 (D1=1 if weather=cloudy and large/medium cargo vehicle; D1=0, otherwise).

Table 4.3 Rules for N-F crash and F-N crash by injury severity level

	Fatal and Severe Injury	Support	Confidence	Lift	Minor Injury Crash	Support	Confidence	Lift
				N - F crash				
1	"cloudy" & "large/medium cargo vehicle"	0.01	0.40	3.07	"light cargo vehicle" & "motorized vehicle lane"	0.01	0.29	2.60
2	"age 31-40" & "cloudy"	0.02	0.32	2.44	"age 41-50" & "expressway"	0.01	0.29	2.55
3	"motor vehicle crash" & "urban road"	0.01	0.32	2.42	"male driver" & "light cargo vehicle"	0.01	0.24	2.16
4	"night without street light" & "cloudy"	0.02	0.30	2.27	"motor vehicle crash" & "expressway"	0.03	0.22	1.97
5	"self-employed" & "night without street light"	0.02	0.27	2.07	"vehicle with insurance" & "light cargo vehicle"	0.01	0.22	1.95
6	"vehicle with insurance" & "cloudy"	0.03	0.25	1.92	"age 41-50" & "might without street light"	0.01	0.22	1.95
7	"motor vehicle crash" & "night without street light"	0.02	0.24	1.85	"expressway" & "motorized vehicle lane"	0.04	0.21	1.89
8	"age 41-50" & "passenger car"	0.01	0.23	1.77	"night without street light" & "cloudy"	0.01	0.20	1.78
9	"age 41-50" & "night without street light"	0.01	0.23	1.74	"male driver" & "expressway"	0.04	0.20	1.78
10	"age 31-40" & "night without street light"	0.04	0.22	1.69	"larger/medium cargo vehicle" & "expressway"	0.02	0.19	1.69
				F - N crash				
1	"hit fixed object" & "non-motorized vehicle lane"	0.01	0.86	2.32	"urban road" & "night without street light"	0.01	0.57	1.95
2	"urban road" & "cloudy"	0.01	0.83	2.25	"night with street light" & "rollover"	0.01	0.55	1.86
3	"age 31-40" & "non-motorized vehicle lane"	0.01	0.75	2.03	"urban road" & "intersection"	0.02	0.50	1.70
4	"urban road" & "intersection"	0.01	0.71	1.93	"migrant worker" & "motorized vehicle lane"	0.02	0.48	1.64
5	"intersection" & "night without street light"	0.02	0.70	1.89	"migrant worker" & "motor vehicle crash"	0.02	0.48	1.63
6	"male driver" & "non-motorized vehicle lane"	0.02	0.69	1.87	"intersection" & "night with street light"	0.02	0.48	1.62
7	"passenger car" & "night without street light"	0.02	0.64	1.74	"mixed lane" & "hit fixed object"	0.02	0.47	1.61
8	"urban road" & "night without street light"	0.02	0.64	1.72	"migrant worker" & "urban road"	0.01	0.47	1.59
9	"motor vehicle crash" & "urban expressway"	0.01	0.63	1.69	"clerk" & "night with street light"	0.01	0.46	1.57
10	"passenger car" & "intersection"	0.01	0.63	1.69	"worker" & "large/medium cargo vehicle"	0.01	0.46	1.57

4.4.2 Binary logistic model

The independent variables are selected based on their significance and model fitness. To avoid neglecting variable problem, a conservative selection strategy was employed in the current study. First, all the variables were tested in the basic models. If no variable is significant in a particular group, log-likelihood ratio test would be conducted to compare models with or without those variables. Information criteria were also compared for the same purpose. If log-likelihood ratio test cannot reject the null hypothesis and information criteria also showed better fit in the model without insignificant variables, those variables would be removed from the basic model. At the final results, variables with 90% or higher levels of significance are kept in final results for examining more possible impact factors given relatively smaller sample size. The results of logistic regression models for factors associating with false negative and false positive fatigue-related crash detection on two injury levels are presented in Table 4.4 and Table 4.5. For fatal and severe crashes, seven variables (include one interaction) are found to be significant for false negative fatigue-related crash detection and eight variables (include one interaction) are significant impact factors for false positive fatigue-related crash detection at the 90% level. Among minor injury crash, nine significant variables were identified for false negative fatigue-related crash detection and eight variables (include two interactions) are found to be significant for false positive fatigue-related crash detection. More detailed discussion of results will be presented in next section.

Table 4.4 Factors associating with false negative fatigue-related crash detection

	Fatal and severe injury			Minor injury		
	OR	95% CI		OR	95% CI	
Crash Characteristics						
Motor vehicle crash	0.334	0.075	1.485	0.422	0.104	1.709
Rollover	0.080***	0.014	0.472	0.248*	0.048	1.289
Hit fixed object	0.196**	0.041	0.930	0.469	0.096	2.287
Driver characteristics						
Male						
31 ~ 40 years old	3.536***	1.488	8.400	0.495**	0.251	0.974
41 ~ 50 years old	3.463**	1.319	9.096	0.860	0.390	1.897
≥ 51 years old	4.673**	1.179	18.520	0.380	0.098	1.478
Clerk				0.248**	0.066	0.939
Worker				0.301**	0.110	0.818
Farmer				0.876	0.397	1.932
Self-employed				0.469	0.175	1.258
Migrant worker				0.286	0.064	1.285
Vehicle-specific characteristics						
Large/Medium cargo vehicle				1.683	0.664	4.265
Light cargo vehicle				4.417***	1.439	13.550
Large/Medium passenger vehicle				8.054**	1.618	40.100
Light passenger car				1.420	0.591	3.412
Vehicle with insurance				0.475*	0.204	1.11
Roadway characteristics						
Expressway	1.054	0.472	2.354	1.761	0.772	4.017
Urban expressway	1.244	0.231	6.708	5.651**	1.356	23.550
Urban road	2.428*	0.894	6.593	1.150	0.497	2.661
Special road segment				0.134*	0.016	1.117
Intersection				0.509	0.201	1.286
Motorized vehicle lane						
Non-motorized vehicle lane						
Mixed lane						
Interaction						
Cloudy & Night without street light	6.667***	2.069	21.490			
Constant	0.314	0.069	1.423	1.223	0.229	6.527

*Note: *** Statistically significant at 1% level; ** Statistically significant at 5% level; * Statistically significant at 10% level*

Table 4.5 Factors associating with false positive fatigue-related crash detection

	Fatal and severe injury			Minor injury		
	OR	95% CI		OR	95% CI	
Crash Characteristics						
Motor vehicle crash	0.588	0.157	2.205			
Rollover	0.248*	0.061	1.004			
Hit fixed object	0.227**	0.057	0.902			
Driver characteristics						
Male driver				0.269***	0.122	0.593
31 ~ 40 years old				0.632**	0.406	0.986
41 ~ 50 years old				0.910	0.532	1.556
≥ 51 years old				0.500*	0.231	1.085
Clerk	0.640	0.214	1.912	0.603	0.261	1.393
Worker	1.975**	1.010	3.859	0.476**	0.245	0.921
Farmer	1.518	0.747	3.085	1.588*	0.964	2.617
Self-employed	0.584	0.268	1.274	1.145	0.598	2.193
Migrant worker	1.671	0.774	3.607	1.921*	0.900	4.099
Vehicle-specific characteristics						
Large/Medium cargo vehicle						
Light cargo vehicle						
Large/Medium passenger vehicle						
Light passenger car						
Vehicle with insurance	0.412***	0.232	0.733			
Roadway characteristics						
Expressway	2.088**	1.116	3.908			
Urban expressway	1.652	0.466	5.855			
Urban road	3.127***	1.327	7.369			
Special road segment						
Intersection						
Motorized vehicle lane						
Non-motorized vehicle lane	5.334***	1.558	18.260			
Mixed lane	1.387	0.673	2.861			
Interaction						
Passenger car & Night without street light	4.372*	0.921	20.750			
Urban road & Intersection				3.309**	1.234	8.868
Worker & Large/medium cargo vehicle				4.863**	1.340	17.650
Constant	1.885	0.468	7.599	1.924	0.819	4.521

*Note: *** Statistically significant at 1% level; ** Statistically significant at 5% level; * Statistically significant at 10% level*

4.4.3 Model evaluation

We conducted likelihood ratio tests to compare the overall fitness between models with interactions and without interactions. The likelihood ratio is calculated as follow:

$$LR = -2\left(LL_r - LL_u\right) \tag{4-4}$$

where LL_r represents the log-likelihood at convergence of restricted model (model without interactions) and LL_u is the log-likelihood at convergence of unrestricted model (model with interactions). Under the null hypothesis that the coefficient of interaction is equal to zero, LR statistic is chi-square distributed with degree of freedom equal to the number of interactions. The results of LR statistic and several goodness-of-fit statistics are shown in Table 4.6.

The results indicate that models with interactions outperformed the models without interactions in both fatal and severe injury crash sample and minor injury crash sample. As shown in Table 4.6, the p-value of Hosmer-Lemeshow Test (Hosmer & Lemeshow, 1980) for all four models is greater than 0.05 that show no evidence of poor fit.

Table 4.6 Goodness-of-fit measures

	N-F Crash		F-N Crash	
	Fatal and severe	Minor	Fatal and severe	Minor
	With Interactions			
Number of coefficients	11	22	16	12
McFadden pseudo-R2	0.132	0.128	0.105	0.057
Log-likelihood at convergence	-117.781	-158.768	-223.993	-313.076
Log-likelihood at null	-135.565	-181.971	-250.248	-331.965
Akaike Information Criteria (AIC)	267.150	361.535	479.987	650.152
Bayesian Information Criteria (BIC)	296.853	450.210	542.472	701.267
Hosmer-Lemeshow test (p-value)	0.376, g=13	0.833, g=24	0.401, g=18	0.977, g=14
	Without Interactions			
Number of coefficients	10	22	15	10
McFadden pseudo-R2	0.095	0.128	0.097	0.040
Log-likelihood at convergence	-122.698	-158.768	-225.878	-318.545
Log-likelihood at null	-135.565	-181.971	-250.248	-331.965
Akaike Information Criteria (AIC)	265.397	361.535	481.757	657.091
Bayesian Information Criteria (BIC)	301.232	450.210	540.338	699.687
Hosmer-Lemeshow test (p-value)	0.467, g=12	0.833, g=24	0.612, g=17	0.935, g=12

4.5 Discussion

4.5.1 False negative fatigue-related crash detection

With regards to crash type, the odds for rollover or hitting fixed object fatigue-related crash with fatal and severe injury being misclassified are lower than being correctly classified (OR=0.080 and OR=0.196, respectively). Some researchers had already pointed out that single car crashes were closely associated with driver fatigue (Radun et al. 2009; Armstrong et al. 2008), and hitting fixed object and rollover were two major types of single vehicle crashes. A statistic from Australia has shown that hitting fixed object crashes and rollover crashes accounted for 54% and 28% of all fatigue-related crashes during 2005 - 2009 in South Australia (Government of South

Australia, 2010). In line with those findings, these two types of crashes have lower odds of being misclassified into non-fatigue crashes than other types of crashes. In serious injury crashes, lacking witnesses makes it almost impossible to observe the syndrome of fatigue from a dead body and determining crash causes based on some specific types of crash seems to be a useful and effective way. However, rollover was found to be significant only at 90% level (OR=0.248) and hitting fixed object is not significant for minor injury crashes. In a minor injured crash, besides of crash types, police officers might also ask the witness for relevant information about driver fatigue condition to assist their judgment.

Driver's age and occupation are found to have a significant influence on false negative fatigue-related detection for serious injured crashes. The odds of false negative fatigue-related crash detection for drivers in age groups of 31-40 (OR=3.536), 41-50 (OR=3.463), ≥51 (OR=4.673) are higher than younger drivers (≤30). Young drivers were believed to be frequently involved in fatigue-related crashes because of their lifestyle (Horne & Reyner, 1995; Maycock, 1996; McKernon, 2009). Moreover, some practical guidance for identifying for police officers also suggested young driver to be one of the high-risk groups. On the other hand, compared to the young drivers, drivers in other age groups may not attract sufficient attention. Table 4.2 shows that drivers who are 31 - 40 years old and 41 - 50 years old occupied 37.0% and 22.3% of all fatigue-related fatal and severe injured crashes, and 37.0% and 18.3% for minor injured crashes. Without enough attention, fatigue-related crashes in which the responsible drivers who are older than 31 years old, have high odds for being misclassified into non-fatigue crashes than younger drivers. On the contrary, among minor injury crashes, the odds of a fatigue-related crash involving drivers aged between 31 - 40 being misclassified into non-fatigued crashes is lower than younger drivers (OR=0.495). For clerk and normal worker, the odds of being misclassified into non-fatigue-related crahses are 0.248 and 0.301 the odds for other occupation in minor injured crashes. Studies showed that shift workers and commercial vehicle drivers were more likely to drive under fatigue (Morrow & Crum, 2004; Philip, 2005). According to the coding rules of this database, both professional drivers and shift workers were coded as "worker" or "clerk". If the driver belongs in these two types of occupations and there is no any other extra information for determining crash cause, police officers may easily connect them to fatigue driving. Thus, they are less likely to be misclassified into other causes.

The odds for both large/medium passenger vehicle and light cargo vehicle involving fatigue-related crashes being assigned to other cause are approximately 8.05 times and 4.42 times the odds for other types of vehicles in minor injured crashes (OR=8.054 and OR=4.417, respectively). Even though large/medium cargo or passenger vehicles are recommended to install driving recorders by road management authorities, not all of them would actually install them since it is not mandatory, especially for privately-owned cargo vehicles. Driving records can help to identify the cause of crashes. However, police officers may not have enough time to check them at crash scene due to complaints from drivers and passengers. Therefore, a more common way for officers is to ask the drivers how long they had driven or whether they felt fatigued or sleepy at the crash scene to determine whether fatigue involved in the crash

(Robertson et al., 2009). For fatigue-related crashes involving light cargo vehicles, they are also easy to be assigned to other cause of the crash. This may be partly due to the difficulty in proving fatigue driving behavior without driving recorders as well as a similar working pattern as heavy cargo vehicles. Previous research noticed that the long and monotonous journey made heavy cargo vehicles more likely to involve in fatigue-related crashes (Summala & Mikkola, 1994; Chang & Mannering, 1999). The combination effects make fatigue-related crash involving those vehicles easily be misclassified.

The odds of minor injured fatigue-related crashes took place on urban expressways being misclassified into other cause is 5.65 times higher than correctly classified (OR=5.651). These crashes often occurred on expressways because of the average trip length and high-speed limit (Pack et al., 1995; Diamantopoulou et al. 2003). Expressway has been widely considered to be of high risk of fatigue-related crash (NCSDR/NHTSA, 2001). However, less attention has been paid to urban expressways which have similar road condition with expressway, that driving on them also easily leads to driver fatigue (Li et al., 2010). With the rapid urbanization in Guangdong, more and more urban expressways were built to serve the city traffic. Therefore, police officers have lower sensitivity and fail to correctly identify driver fatigue when crashes occurred on urban expressways for minor injured crashes, which make them easily being misclassified.

4.5.2 False positive fatigue-related crash detection

The odds of fatal and severe injured non-fatigue hitting fixed object crash being misclassified into the fatigue-related crash is lower than other types of crashes (OR=0.227). Rollover crashes are also found significant at 90% level. This finding is similar to false negative fatigue-related crash detection indicating that crash types is a good indicator for identifying fatigue-related crashes.

Some driver characteristics significantly contribute to false positive fatigue-related crash detection. Crashes involving drivers whose occupation were categorized as "worker" have higher odds for false positive fatigue detection in fatal and severe injury crashes (OR=1.975) than other occupations. A survey conducted by police officers in Ontario confirmed that approximately 61% of them believed that night or shift workers tended to involve in fatigue-related crashes (Robertson et al., 2009). This image of shift workers is in line with some previous research indicating that workers with non-fixed working schedule were more likely to have sleep problems (Marcus & Loughlin, 1996; McCartt et al., 1996; Dalziel & Job, 1997), which contributed to fatigue-related crashes (Connor et al., 2001a). Thus, these crashes are easily being considered as fatigue-related. For minor injury crashes, the odds of workers involving crashes being false positive fatigue detected is lower than other occupations (OR=0.476) even if police officers can ask those drivers about the fatigue condition at the crash scene. Moreover, farmers (OR=1.588, significant at 90% level) and migrant workers (OR=1.921) have higher odds of being misclassified into other cause in minor injured crashes. Some common features are shared by farmers and migrant workers: low salary, non-fixed working schedule, and low social status. Drivers with these features are widely thought to be related to fatigue driving, and this stereotype can influence the judgment of police officers on fatigue-related crash detection. Therefore, police officers tend to believe that they are more likely to involve in traffic violation related to fatigue. The odds of male drivers for false positive fatigue detection is significantly lower than the odds of female drivers (OR=0.269). While the male driver was believed were more at risk of driving while fatigued (Robertson et al., 2009; Horne & Reyner, 1995), they also found to be at high risk of other violations. As a result, they were not easily to misclassify into fatigue-related crashes. Drivers' age (31 - 40, OR=0.632) also shows similar results that the odds of false positive fatigue-related crash detection is lower for them compared to young drivers, which is similar to the previous discussion.

The odds of a vehicle with insurance for false positive fatigue-related crash detection is lower than vehicles without insurance for fatal and severe injury crash (OR=0.421). According to the insurance claim process, investigators from insurance companies need to do site investigation which may help police officer to determine the crash cause better.

As for roadway characteristics, expressways and urban roads have higher odds of false positive fatigue-related crash detection in fatal and severe injury crashes (OR=2.088 and OR=3.127) since many researches discussed the relationship between urban expressways and fatigue-related crash detection and their potential danger may be overemphasized. In addition, there are dozens of monitoring facilities on expressways and urban roads, it is still difficult to identify whether the cause of crashes is fatigue-related at the crash scene immediately. Crashes occurred in non-motorized vehicle lanes have higher odds of false positive fatigue-related crash detection (OR=5.334) since most of the fatigue-related crashes since most of the fatigue crash are vehicle-related.

4.5.3 Interactions

Even though the effect of some individual factors may not have significant impacts, their combination with other factors did influence police officers' judging on determining fatigue-related crashes. For example, lightening condition and weather do not show significant influence for failing to recognize fatigue-related crashes. But driving at night without street light in a cloudy day was identified to contribute to false negative fatigue-related crash detection in serious crashes (OR=6.667). Therefore, when a serious crash take place at night without street light in a cloudy day, the police officer should carefully consider fatigue might be one of the causes of crashes.

Some other interactive factors are identified to hamper the judgment of police officers. A crash involving passenger car during night time without street lighting are found more likely to be false positive fatigue-related crash detection for fatal and severe injured crashes (OR=4.372, significant at 90% level). For minor injured crash, crashes occurred on interactions of urban roads (OR=3.309) is easy to be considered as fatigue-related crashes when fatigue is actually not the primary cause. In addition, non-fatigue crashes in which driver is labeled as "worker" that drives trucks or other large size cargo vehicles, are more likely to be considered as fatigue-related (OR=4.836). Commercial truck drivers are commonly believed to be associated with fatigue driving, thus, they are also more easily to be mistakenly believed to fatigue driving. One possibility is commercial large/medium cargo vehicle drivers are more skillful that they have enough experience and ability to avoid fatigue-related crashes. Therefore, fatigue may not be the major cause of crashes that have these three interactive features, and other possible causes of crashes should be considered.

4.6 Conclusions and Practical Applications

Due to lack of proper criteria, the identification of fatigue-related crash by police officers largely depend on inferential evidence and their own experience and may even lead to incorrect classification of the cause of crashes. Even though some risk factors identified were believed to contribute to fatigue-related crashes, less research has been done to prove that whether these factors are helpful for fatigue-related crash identification. The purpose of this study is to find out factors affecting police officers' judgment when dealing with fatigue-related crashes.

The results show that single vehicle rolling over or hitting fixed object crashes are good indicators for determining fatigue involvement. Crashes that include two or more vehicles have not been found to have significant influence both misclassification types since fatigue could be one of the causes of crashes that have not been noticed. Driving light cargo vehicle, driving large/medium passenger vehicle and urban expressway should attract more attention on determining whether a crash is fatigue-related.

Moreover, some stereotypes should be abandoned. Some occupations (e.g. workers, farmer and migrant workers) should not be labeled as "high risk of fatigue driving" when investigating cause of crashes. Expressways and urban roads are also easy to be viewed as high-risk places for fatigue-related crashes. These images will hinder the judgment of fatigue detection whey there are used as an evidence for convicting a crash is fatigue-related. Fatigue should be considered as a possibility rather than a conclusion for cause of the crashes. We also recognized some interactive effects between variables that may also affect fatigue-related crash detection. However, it should be re-emphasized that significant combinatorial effects of factors in this study may only reveal the characteristics of a small subset of fatigue-related crashes. More rules can be generated based on this dataset and it would be better to be used as additional information for training police officers to identify fatigue-related crashes correctly.

Based on the findings from this study, some countermeasures should be considered to improve the fatigue-related crash detection. A clear and easy to implement fatigue definition is an essential solution to this problem. However, up until now, we still do not have a completely general method to quantify driver fatigue and to determine what kind of fatigue level should be considered as fatigue driving. Finding an appropriate definition and quantification method for driver fatigue is one of the challenges in fatigue research during the coming years. In the current stage, more partial countermeasures are needed to improve fatigue-related crash detection. First of all, raising fatigue driving violation penalty for passenger vehicles can prevent drivers from fatigue driving as well as stimulate police officer to put more attention on identifying the involvement of fatigue in a crash. Secondly, providing training for identifying fatigue in traffic crashes can be beneficial for more police officers and give them a better understanding of fatigue considering the experience of Finland (Radun & Radun, 2013). In this case, misleading factors in identifying fatigue-related crash should be addressed in training process. Moreover, useable and reliable vehicle-based fatigue measurement devices should be encouraged. These devices not only can be used in monitoring drivers' behaviors and the level of driver fatigue by placing sensors on the steering wheels and acceleration pedals, but also provide useful information for traffic police officers to determine the role of fatigue in a crash (Liu et al., 2009; Sahayadhas et al., 2012). Although information such as the length of time spent on driving, detail previous work, sleeping condition and rest schedules of the drivers involved can be found in investigation reports of final assessment, it should be recorded as a necessary part in standard crash investigation procedure.

There are several limitations which need to be acknowledged. Firstly, for the purpose of identifying factors contributing to the detection of fatigue-related crashes, N-F crashes and F-N crashes are compared with F-F crashes in our study. However, we have not examined whether these factors also influence police officers' judgment on other types of crash. To understand whether those factors are unique for fatigue-related crashes, comparison studies of factors contributing to misclassification of other types of crashes should be conducted in the future. Furthermore, the misclassification scale of property only fatigue-related crash may be underestimated. Filtness et al. (2015) also mentioned that identifying fatigue among less serious crashes may be inaccurate.

Moreover, due to the difficulties in proving fatigue as a contributor in a crash, it is possible that not all fatigue-related crashes have been detected even after full investigation process. Some fatigue-related crashes cannot be detected even after in-depth investigation since there is no extra information to evaluate the scale of miscoding problem. Thus, additional data should be collected (e.g. self-report fatigue questionnaire) for better assessing fatigue identification. Several important variables should be considered such as time of day and pre-crash activity. These variables can provide valuable information for identify fatigue-related crash. Unfortunately, they are not included in our dataset.

CHAPTER 5

The Effect of Fatigue Driving on Injury Severity Considering Endogeneity

5.1 Introduction

Fatigue driving was identified as one of the four most risky driving-related behaviors, especially in fatal traffic crashes (Fernandes et al., 2010) and represented a significant social and economic cost to the community. Despite extensive body of research addressing the harmfulness of fatigue driving on road safety, it has not attracted enough attention. Drivers were less concerned about fatigued driving than other traffic safety issues (Vanlaar et al., 2008). Studies from different countries showed that many people still drove when they felt fatigue (Beirness et al., 2005; Nordbakke & Sagberg, 2007; Tefft, 2010). Besides drivers, public are also not fully aware of the potential risk of fatigue driving because it is difficult to evaluate its effect accurately. For example, fatigue could be resolved after a period of rest (Karrer et al., 2004), this feature made it hard to detect and identify after crashes occurred. When other risky driving behaviors are involved, it is even harder to tell what the major contributor is and may lead to misclassification of the cause of crash (Horne & Reyner, 1995; Philip et al., 2005; Armstrong et al., 2008). In addition, police also tended to assign the cause of crash to current interest (Ogden & Moskowitz, 2004).

Several studies have examined the relationship between fatigue driving and traffic injury severity from different aspects. However, fatigue driving and injury severity in traffic crashes may share some observed common influential factors (e.g. road types). There are also some unobserved factors between fatigue driving and injury severity. The connection between sleep disorder, fatigue and traffic injury severity were discussed by many researchers (Akerstedt et al., 2001; Horne & Reyner, 2001; Philip et al., 2003; Stutts et al., 2003). Ignoring the impact of these common factors will lead to endogeneity problem and incorrect conclusion. This study contributes toward current fatigue driving research by applying a bivariate endogenous binary-ordered probit model framework to examine the relationship between fatigue driving propensity and fatal injury propensity in a crash considering the potential endogeneity of fatigue driving. Considering the potential systematic differences between commercial and non-commercial vehicle drivers, this model also identifies the observed common factors of fatigue driving and injury severity for two groups of drivers and makes a comparison. This result may help better understand how those factors affect fatigue driving propensity and injury severity and contributes to more efficient policy for preventing the harmfulness of fatigue-related crashes. The analysis includes several types of factors, including driver characteristics, vehicle characteristics, road characteristics, environmental characteristics, and collision characteristics.

5.2 Literature Review

Fatigue is a gradual and cumulative process closely related to deterioration of performance efficiency like driving performance (Haworth, 1998; Rajaratnam & Arendt ,2001; Philip et al., 2005), and could be induced by repetitive and monotonous activities like driving long distances (Stutts et al., 1999). Research pointed out that fatigue was not a strictly monotone decreased progress (Karrer et al., 2004), but an interaction between deactivation and compensation processes, resulting in variability of performance (Dinges & Kribbs, 1991).

As for the influential factors related to fatigue driving, prior studies basically focused on four categories: driver characteristics, road characteristics, environmental characteristics and vehicle characteristics. Considering driver characteristics, male drivers were at high risk of fatigue driving for the reason that males were more likely to drive for a longer time (Fernandes et al., 2010; Amstrong et al., 2011). In Amstrong et al. (2008)'s study, it was found that drivers aged 17-24 years were more likely to be involved in a fatigue-related crash. However, the influence of age is much more complicated and there exist different behavior patterns between young drivers and elder drivers. Young drivers frequently committed their fatigue-related offenses during early morning and night-time hours (Horne & Reyner, 1995; Pack et al., 1995; Maycock, 1996; Horne & Reyner, 2001) while elder drivers mostly in the afternoon (Summala & Mikkola, 1994). In addition, the motivation for driving while fatigue for young drivers might be their overestimation of capabilities (Gregersen & Bjurulf, 1996) and miscalculation of the cost of consequence (Fernandes et al., 2010).

For road characteristics and environmental characteristics, driving on different types of road can lead to similar consequence. Both high-demand and low-demand road condition could induce driver fatigue (Oron-Gilad et al., 2008; Zhao & Rong, 2013). Dyani (2007) divided driver fatigue into two groups: passive fatigue and active fatigue. Passive fatigue was defined closely related to underload, which has been confirmed by simulated driving studies in monotonous condition (Desmond & Hancock, 2001; Thiffault & Bergeron, 2003). Active fatigue was defined related to overload of driver. For example, poor road condition (Arnold et al., 1997), complex traffic conditions and road environments (Pilcher & Huffcutt, 1996) required more attention and could easily induce physical and mental fatigue. Time of day was mentioned by several fatigue-related studies. Folkard (1997) has reviewed several researches that studied the relationship between road safety and time of day. It was widely believed that time of day were closely related to human rhythms, which was identified as an important factor affecting driver fatigue (Haworth, 1998; Philip et al., 2005). Horne and Reyner (2001) found that 02:00-06:00 and 14:00-16:00 is time period associated with higher probability of fatigue. Haworth (1998) also pointed out that nighttime is significant contributor of fatigue-related crashes. Light level (Sullivan & Flannagan, 2002) and season were also identified to play important role (Radun & Radun, 2009).

Nevertheless, fatigue-related crashes are severe among commercial vehicle drivers. Statistics from Europe pointed out that approximately 20% of commercial vehicle crashes were related to driver fatigue (European Transport Safety Council, ETSC, 2001). The causes of fatigue varied since fatigue could be developed while on the job with regular sleep patterns or arrived at work already fatigued with irregular sleep patterns (Young & Hashemi, 1996). Commercial vehicle drivers suffered from sleep restriction (Hanowski et al., 2007) and were under great work pressure, which made them vulnerable to fatigue-related crashes. Specifically, drivers in developing countries are more likely to drive while fatigue for financial reasons (Mock et al., 1999; Nantulya & Muli-Musiime, 2001). Surveys conducted among truck and taxi drivers in Beijing, China, showed that driver fatigue was prevalent and the most important reason was prolonged driving time (Meng et al., 2015).

Even though it is not in agreement, fatigue driving and injury severity in the crash may share some common influential factors, including observed and unobserved factors. Radun and Radun (2009) claimed that there was no connection between crash severity and whether the driver was judged to have been fatigued. However, more studies believed there existed some kind of connection (Haworth, 1998; Zhang et al., 2016). Fatigue-related crashes were often severe that drivers could not take evasive action under fatigue (Haworth, 1998). Some factors related to fatigue driving may impair driver performance, then affect injury severity. For example, some unobserved factors related to the driver’s internal state and circadian cycle can also affect both fatigue propensity and driving performance (Williamson et al. (2011) has given a detail review on that). Unfortunately, this information was almost impossible to collect due to traumatic effects and emotional state change after the crash (Radun & Radun, 2009). Some drivers might not admit fatigue or falling asleep during driving concerning about insurance and legal consequences (Corfitsen, 1999). Therefore, those common factors were often neglected, which may lead to endogeneity problem and biased estimation when analyzing the relationship between fatigue driving and injury severity.

5.3 Econometric Framework

5.3.1 Model structure

In fatigue-related crashes, drivers who are more likely to be involved in fatigue-related crashes and injury severity can be correlated, which may cause endogenous problem. In econometrics, endogeneity problem is said to occur if the independent variable is correlated with the error term. This correlation can be caused by several reasons: omitted variables, measurement error, and simultaneity in simultaneous models. Endogeneity induces estimation bias in statistical models and may eventually lead to mistaken conclusions. To take into account the potential endogeneity of fatigue driving, we apply a bivariate endogenous binary-ordered probit model in the current paper. Bivariate endogenous binary-ordered probit model is a hierarchical model system of two equations that can be used to model two response variables simultaneously, and addresses endogeneity problem. This model addresses endogeneity by considering error correlations among two equations that capture the relationships among endogenous variable, exogenous variables and error term (for further discussion, see Greene, 2007; Fernandez-Antolin et al., 2014).

Let i ($i = 1, 2, ..., N$) be an index representing drivers and k ($k = 1, 2, ..., K$) be indices representing ordinal categories of injury severity sustained by driver i in the crash. Suppose y_i is the observed injury severity level and y_i^* represents latent injury severity propensity of driver i in the crash. Thus, the latent propensity y_i^* is mapped to the actual injury severity level y_i by threshold ψ_k ($\psi_0 = -\infty$ and $\psi_K = \infty$, $\psi_0 < \psi_1 < \cdots < \psi_K$) as the following equations:

$$y_i^* = \alpha' x_i + \theta\, fatig_i + v_i, \tag{5-1}$$

and

$$y_i = k, \ \text{ if } \psi_{k-1} \le y_i^* < \psi_k \tag{5-2}$$

where x_i is an $M \times 1$ column vector of variables that influences y_i^* (not including a constant) and $fatig_i$ is a dummy variable indicating whether driver i is convicted as fatigue driving or not. α represents an $M \times 1$ coefficient vector of x_i and θ is the coefficient of $fatig_i$. v_i is the error term assumed to be identically and independently across driver i.

However, $fatig_i$ included in Eq. (5-1) may be endogenous. Therefore, we specify here:

$$fatig_i^* = \beta' z_i + \omega_i, \tag{5-3}$$

$$fatig_i = \begin{cases} 1, & \text{if } fatig_i^* \ge 0 \\ 0, & \text{otherwise} \end{cases} \tag{5-4}$$

This equation represents the latent fatigue driving propensity $fatig_i^*$ of driver i. $fatig_i$ is the actual observed fatigue driving behavior by driver i, and z_i is an $L \times 1$ column vector of independent variables (including a constant) influencing fatigue driving propensity $fatig_i^*$. β is an $L \times 1$ coefficient vector of z_i. x_i and z_i can share some common variables, representing the common observed influential factors between fatigue driving propensity and fatal injured propensity. ω_i represents the random components that capture all unobserved factors.

Still, there could be unobserved correlation between injury severity and fatigue driving. To capture the unobserved correlation, without losing generality we assume that v_i and ω_i form a bivariate normal distribution. In particular, the probability is given as:

$$\begin{aligned}
&\text{Prob}\left(y_i = k,\ fatig_i = j \mid x_i, z_i\right) \\
&= \text{Prob}\left(\psi_{k-1} \le y_i^* < \psi_k,\ \tau_{j-1} \le fatig_i^* < \tau_j\right) \\
&= \begin{pmatrix}
\Phi_2\left(\psi_k - (\alpha' \mathrm{x}_i + \theta\, fatig_i),\ \tau_j - \beta' \mathrm{z}_i; \rho\right) \\
- \Phi_2\left(\psi_{k-1} - (\alpha' \mathrm{x}_i + \theta\, fatig_i),\ \tau_j - \beta' \mathrm{z}_i; \rho\right) \\
- \Phi_2\left(\psi_k - (\alpha' \mathrm{x}_i + \theta\, fatig_i),\ \tau_{j-1} - \beta' \mathrm{z}_i; \rho\right) \\
+ \Phi_2\left(\psi_{k-1} - (\alpha' \mathrm{x}_i + \theta\, fatig_i),\ \tau_{j-1} - \beta' \mathrm{z}_i; \rho\right)
\end{pmatrix}
\end{aligned} \tag{5-5}$$

where $\Phi_2(\cdot)$ is the standard bivariate normal cumulative distribution function. τ_j and τ_{j-1} $(j = 0,1)$ represent thresholds for mapping the latent variable $fatig_i^*$ to the observed variable $fatig_i$ in Eq. (5-4). Specifically, in the binary probit model we set $\tau_{-1} = -\infty$, $\tau_0 = 0$, and $\tau_1 = \infty$. ρ measures the correlation between disturbances in the equations, which measures correlation between injury severity and fatigue driving propensity after the influence of fatigue is accounted in injury severity function. If this correlation between fatigue driving propensity and fatal injury propensity is ignored when actually exists, it could lead to inconsistent estimation of the effect of fatigue on injury severity. We also introduce a univariate endogenous binary-ordered probit model in which we assume $\rho = 0$, neglecting the correlation between v_i and w_i for comparison purpose.

5.3.2 Model estimation

The log-likelihood function is given by:

$$LL = \sum_{i=1}^{N} \ln \begin{pmatrix} \Phi_2\left(\psi_k \ -\left(\alpha' x_i + \theta\, fatig_i\right),\ \tau_j \ -\left(\beta' z_i\right); \rho\right) \\ -\ \Phi_2\left(\psi_{k-1} -\left(\alpha' x_i + \theta\, fatig_i\right),\ \tau_j \ -\left(\beta' z_i\right); \rho\right) \\ -\ \Phi_2\left(\psi_k \ -\left(\alpha' x_i + \theta\, fatig_i\right),\ \tau_{j-1} -\left(\beta' z_i\right); \rho\right) \\ +\ \Phi_2\left(\psi_{k-1} -\left(\alpha' x_i + \theta\, fatig_i\right),\ \tau_{j-1} -\left(\beta' z_i\right); \rho\right) \end{pmatrix} \qquad (5\text{-}6)$$

The corresponding parameters α, β, θ, ψ_k, and ρ are estimated simultaneously by maximizing the log-likelihood function of Equation (5-6). R software (version 3.3.1) is used for estimation in this study.

5.4 Data

5.4.1 Data source

The Guangdong Traffic Accident Dataset (GTAD) is sourced from the Traffic Management Sector Specific Incident Case Data Report, the Road Traffic Accident Database of China’s Public Security Department. A total of 38,564 crash records during 2006-2011 are applied in this study. The data we used in this study were drawn from police-reported crashes in 21 cities across Guangdong Province, and compiled from a sample of crashes that involve at least one motor vehicle and resulting in property damage, injury, or death.

Several crash-related attributes are collected for each record in GTAD, including driver characteristics, vehicle characteristics, road characteristics, environmental characteristics, and crash characteristics. The injury severity of each individual involved in the crash is categorized into four ordinal levels: (1) No injury, (2) Minor injury, (3) Serious injury, and (4) Fatal injury.

This study mainly focuses on drivers who were chiefly responsible for the occurrence of crash that was convicted to be fatigue-related. The reason is that crash and personal information is better recorded. The definition for fatigue driving in GTAD is defined as fulfilling one of the following conditions: (1) Driving cars more than eight hours a day, (2) Engaging in other work with excessive physical exertion, and (3) Lacking of sleep which results in sleepy or weakness of limbs, so that the driver is having difficulty in assessing traffic conditions immediately and reacting accurately. Normally, the police officer would interview the involved parties and witnesses, and check the driving records to identify the cause of crash. Technical reconstruction is also helpful for determining the cause of crash by studying testimony of witnesses and physical evidence, especially in serious crashes. Fatigue-related crashes defined by this definition constitute 6.5% of all crashes in GTAD dataset. The distribution of fatigue driving and injury severity across observations is presented in Table 5.1. Overall, the descriptive statistics in Table 5.1 indicate a substantially higher percentage of fatal fatigue-related crashes (13.2%) than non-fatigue-related crashes (6.5%).

Table 5.1 Number of fatigue related crashes by injury level

Injury Severity	Fatigue Driving		All (%)
	No (%)	Yes (%)	
All			
No Injury	25070	173	25243
	(65.5)	(56.9)	(65.5)
Minor Injury	8156	64	8220
	(21.3)	(21.0)	(21.3)
Serious Injury	2559	27	2586
	(6.7)	(8.9)	(6.7)
Fatal Injury	2475	40	2515
	(6.5)	(13.2)	(6.5)
Total	38260	304	38564
	(100)	(100)	(100)
Commercial			
No Injury	8297	96	8393
	(86.0)	(57.5)	(85.5)
Minor Injury	725	34	759
	(7.5)	(20.3)	(7.7)
Serious Injury	227	15	242
	(2.4)	(9.0)	(2.5)
Fatal Injury	400	22	422
	(4.1)	(13.2)	(4.3)
Total	9650	167	9817
	(100)	(100)	(100)
Non-commercial			
No Injury	16773	77	16850
	(58.6)	(56.2)	(58.6)
Minor Injury	7431	30	7461
	(26.0)	(21.9)	(25.9)
Serious Injury	2332	12	2344
	(8.1)	(8.8)	(8.2)
Fatal Injury	2075	18	2093
	(7.3)	(13.1)	(7.3)
Total	28610	137	28747
	(100)	(100)	(100)

5.4.2 Variables

Five types of variables were considered in the empirical analysis. ***Driver characteristics*** include: driver's age (≤25, 26-35, 36-45, 46-55, 56-65, and ≥66 years old), driver's gender, driving experience (≤2 years), and whether the driver has a valid driving license. ***Vehicle characteristics*** include: whether the vehicle has insurance. Other vehicle characteristics, such as vehicle speed just before collision, could not be included because of the absence of data in the GTDA. ***Road characteristics*** include: road type (whether the crash occurred on express way or urban roads), isolated lanes (whether the road has separated lanes for motorized and non-motorized vehicles), and terrain (mountain area). ***Environmental characteristics*** include: time of day represented in three categories (early morning (00:00-06:59), morning peak hours (07:00-08:59), and afternoon peak hours (17:00-19:59)) and lighting conditions (dark with street lights and dark without street lights). ***Crash characteristics*** include: collision type (head-on collision, rear-end collision, sideway collision). Variable description is presented in Table 5.2.

Firstly, a general model including all the variables suggested by prior studies and intuitiveness considerations are applied. Then, variables are chosen based on a systematic process of removing statistically insignificant variables and combining variables when their effects were not significantly different. Furthermore, continuous variables, such as driver' age and time of day, were converted into dummy variables and different ranges are also tested.

Table 5.2 Variable description

Variables	Description	Mean
Driver characteristics		
Driver's gender		
Male	Male=1; Female=0	0.947
Driver's age		
≤25	≤ 25=1; Others=0	0.209
26-35	26-35=1; Others=0	0.355
36-45	36-45=1; Others=0	0.296
46-55	46-55=1; Others=0	0.106
56-65	56-65=1; Others=0	0.028
≥66	≥66=1; Others=0	0.006
Driving experience		
≤2 years	≤2 years=1; Others=0	0.135
Driving license		
Not valid	Not valid=1; Others=0	0.280
Vehicle characteristics		
Insurance		
Yes	Insurance=1; Others=0	0.779
Road characteristics		
Road type		
Express way	Express way=1; Others=0	0.046
Urban road	Urban road=1; Others=0	0.409
Isolated lanes		
Yes	Isolated lanes=1; Others=0	0.397
Terrain		
Mountain	Mountain=1; Others=0	0.600
Environmental characteristics		
Lighting condition		
Dark with street light	Dark with street light=1; Others=0	0.261
Dark without street light	Dark without street light=1; Others=0	0.173
Time of day		
00:00-06:59	00:00-06:59=1; Others=0	0.155
07:00-08:59	07:00-08:59=1; Others=0	0.087
17:00-19:59	17:00-19:59=1; Others=0	0.171
Crash characteristics		
Head-on	Head-on collision=1; Others=0	0.224
Sideway	Sideway collision=1; Others=0	0.421
Rear-end	Rear-end collision=1; Others=0	0.119

5.5 Estimation Results and Discussion

In this study, two different models were estimated: (1) Bivariate Endogenous Binary-Ordered Probit model, and (2) Univariate Endogenous Binary-Ordered Probit model (by assuming $\rho = 0$ as noted earlier). There could be potential systematic differences between commercial and non-commercial vehicle drivers (i.e. driving skill, driving time), we compare the factors associated with fatigue driving propensity and fatal injury propensity between them. Variables considered in the models at the very beginning of fatigue and injury severity function are listed in Table 5.3, and all the variables that were included in fatigue function also being included in injury severity function. In Table 5.4 and Table 5.5, we present the results of both models for commercial and non-commercial vehicle drivers, and only significant variables (at 95% significant level) will be listed and discussed in the following parts.

Table 5.3 Variable selection for fatigue model and injury severity model

	Fatigue		**Injury severity**	
Variable	Commercial	Non-commercial	Commercial	Non-commercial
Driver characteristics				
Driver's gender	√	√	√	√
Driver's age	√	√	√	√
Driving experience	√	√	√	√
Driving license	√	√	√	√
Vehicle characteristics				
Vehicle type	√	√	√	√
Insurance	√	√	√	√
Road characteristics				
Road type	√	√	√	√
Isolated lanes	√	√	√	√
Terrain	√	√	√	√
Environmental characteristics				
Lighting condition	√	√	√	√

Time of day	√	√	√	√
Crash characteristics				
Collision type			√	√

5.5.1 Measures of fit

Before discussing the estimation results, likelihood ratio test is conducted to compare bivariate and univariate model. The test statistic is given as

$$LR = -2 \times \left(llk_{nc} - llk_{c} \right) \tag{5-8}$$

where llk_{nc} is the log-likelihood at convergence of bivariate model, and llk_{c} is the log-likelihood at convergence of the models estimated on univariate model. The LR statistic for commercial and non-commercial vehicle drivers is 4.38 and 3.66, which reject the null hypothesis of $\rho = 0$ at $p < 0.05$ and $p < 0.1$, respectively. It should be noted that, in this case, ρ is conservatively retained in non-commercial vehicle driver sample since the correlation does significantly change the coefficient of fatigue in the model. This result indicates that correlation due to unobserved factors between injury severity and fatigue driving propensity is significant, and model estimation without considering this correlation may result in inefficient parameter estimates (Yamamoto & Shankar, 2004).

We also conduct information criteria to compare model performance. Both of the value of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) decline for commercial and non-commercial vehicle drivers by including the correlation, which also suggests that the proposed model is more efficient.

5.5.2 Estimation results

The results in Table 5.4 and Table 5.5 indicate that fatigue has significant impacts on fatal injury propensity for both commercial and non-commercial vehicle drivers. The coefficient of fatigue in the bivariate endogenous model is larger than in univariate model. The estimated coefficient of fatigue driving among commercial drivers is 0.984 in bivariate model and 0.291 in univariate model. The impact of fatigue driving on fatal injury propensity is underestimated by 0.693 in the latter. For non-commercial vehicle drivers, the coefficient of fatigue driving in bivariate model is 0.895 while in univariate model is 0.234 and the gap is 0.661. The impact of fatigue driving on injury severity in a crash on both groups are underestimated and the gaps between these two groups are similar. Larger coefficient of fatigue driving indicates higher risk of involving in fatal injured crash. Results suggest that commercial vehicle drivers are somewhat more risky when driving under fatigue than non-commercial vehicle drivers. Commercial vehicle drivers often drove a high number of miles (National Sleep Foundation, 2009), and some of them tend to break the rules about duty and rest hours for pursuing more profit (Radun & Radun, 2009). Thus, they are more likely to lose focus or even fall asleep at the wheel, which may lead to severe crashes.

This study also identifies the observed common factors of fatigue driving propensity and fatal injury propensity. In summary, the observed common factors for commercial vehicle drivers are: insurance, road types, and terrain. For non-commercial vehicle drivers, the observed common factors are: insurance and road types. More detail discussions of estimation results by groups are as following:

Driver characteristics: although gender, age and driving experience do not show significant impact on fatigue driving propensity, they do influence the driver's propensity of fatal injury in the crash. Non-commercial vehicle drivers who is over 45 years old are more likely to be fatal injured in the crash while male, young non-commercial vehicle drivers are found less likely. This result is consistent with previous findings (Hatfield et al., 2005; McConnell, 2003). Less experienced drivers (≤ 2 years) are found more likely to be more severe injured than those with more driving experienced. Less experienced drivers with dynamic driving style are with more risk in the monotonous setting than experienced and calm drivers (Stutts et al., 2003; Karrer et al., 2004). However, these effects are not significant for commercial vehicle drivers due to small variation in commercial driver group. Drivers without a valid driving license are significantly more likely to be severe injured for both commercial and non-commercial vehicle drivers.

Vehicle characteristics: the impact of factors associated with vehicle itself shows contrast effects on fatigue driving propensity and fatal injury propensity. For both groups of drivers, driving vehicles with insurance are less likely to be fatally injured in the crash. Insurance lowers the monetary loss of crashes. Nevertheless, monetary compensation can never compensate for losing one's life. Therefore, drivers will pay enough attention to preventing themselves from fatal crashes. On the other hand, they might let their defenses down under the circumstances which they thought to be not serious. Thus, light-injured crashes are more likely to happen. In addition, non-commercial vehicle with insurance presents higher risk of fatigue driving while the impact for commercial vehicle is not significant. This finding may be related to different penalties for commercial drivers and non-commercial vehicle drivers when conducting fatigue driving. According to the Road Traffic Safety Law of the People's Republic of China and local traffic regulations, commercial vehicle drivers will have their driving licenses endorsed with at least six penalty points even lose their driving licenses once caught fatigue driving. However, for non-commercial vehicle drivers, fatigue driving will only incur traffic tickets without losing any points on their driving license.

Road characteristics: driving on express way is at high risk of fatigue driving and fatal injury for commercial and non-commercial vehicle drivers. Express ways are mostly monotonous and of high speed. Driving on them can be regarded as a repetitive activity which requires sustained attention and can easily lead to fatigue (McCartt et al., 2000; Thiffault & Bergeron, 2003). On the contrast, driving on urban road or mountain area is less likely to fatigue driving as well as sustain fatal injured. The lower propensity of fatigue driving may be the result of high rate of environmental stimulation and continuous changes in the driving scenery (Mavjee & Horne, 1994; Horne & Reyner, 1999), which help to maintain driver's attention persistently. The impact of driving in mountain area on fatigue driving propensity is not significant for non-commercial vehicle drivers. Isolated lanes show no significant impact on fatigue driving propensity, however, its impact on injury severity differs between commercial and non-commercial vehicle drivers. For commercial vehicle drivers, driving on isolated lanes is less likely to sustain fatal injury, but is more likely for non-commercial vehicle drivers.

Environmental characteristics: The effect of time period on fatigue driving propensity also shows different patterns. During midnight to early morning (00:00-06:59), both commercial and non-commercial vehicle drivers are more likely to fatigue driving compared to other time period in a day. For example, 75% of fatigue-related crashes occurred between 02:00 and 08:00 in 107 heavy truck crashes reviewed (National Transportation Safety Board, U.S., 1995). Morning peak hour (07:00-08:59) only affects the fatigue driving propensity of commercial drivers. It is still not clear whether this result is due to sleep loss or other reasons. Driving at night significantly contributes to more severe injury crashes for both commercial and non-commercial vehicle drivers, but the propensity of fatal injured crashes declines following the installation of street lights. This finding is also consistent with several previous studies (Elvik, 1995; Owens & Sivak, 1996; Plainis et al., 2006).

Crash characteristics: since collision type does not affect fatigue driving behavior, this variable is only considered in injury severity function. The result indicates that commercial vehicle drivers are more likely to fatal injured when involved in rear-end collision while sideway collision and head-on collision is less likely to fatal injured. Some commercial vehicles have larger size and are heavier than the other passenger vehicles with which they share the roads, and the stopping distance for them is much longer. Thus, large and heavy commercial vehicles involving in rear-end crashes may be due to their inability to stop immediately, and cause severe injuries. For sideway and head-on collision, commercial vehicle drivers can reduce the harmfulness of collision by taking sudden turns or other protecting behaviors based on their experience.

However, for non-commercial vehicle drivers, both head-on and rear-end collision have higher propensity of fatal injury, which may be related to lacking experience in handling emergencies on road compared to commercial vehicle drivers. The impact of side collision is not significant for non-commercial vehicle drivers.

Table 5.4 Estimation result of commercial vehicle driver sample

Variables	Correlated		Uncorrelated	
	Coef.	SE	Coef.	SE
	Fatigue Driving Propensity			
Road type				
Express way	0.368^{***}	0.080	0.400^{***}	0.079
Urban road	-0.718^{***}	0.152	-0.702^{***}	0.151
Terrain				
Mountain	-0.179^{**}	0.072	-0.178^{**}	0.072
Time of day				
00:00-06:59	0.771^{***}	0.079	0.740^{***}	0.079
07:00-08:59	0.466^{***}	0.114	0.457^{***}	0.114
Intercept	-2.379^{***}	0.068	-2.381^{***}	0.068
	Injury Severity Propensity			
Fatigue	0.984^{***}	0.312	0.291^{***}	0.097
Driving license				
Not valid	0.291^{***}	0.064	0.293^{***}	0.065
Insurance				
Yes	-0.192^{***}	0.064	-0.193^{***}	0.064
Road type				
Express way	0.850^{***}	0.056	0.883^{***}	0.054
Urban road	-0.129^{***}	0.044	-0.133^{***}	0.044
Isolated lanes				
Yes	-0.085^{**}	0.040	-0.084^{**}	0.040
Terrain				
Mountain	-0.217^{***}	0.034	-0.223^{***}	0.034
Lighting condition				
Dark with street light	0.131^{***}	0.048	0.136^{***}	0.048
Dark without street light	0.243^{***}	0.040	0.252^{***}	0.040
Collision type				
Head-on	-0.142^{***}	0.050	-0.140^{***}	0.051
Side	-0.377^{***}	0.047	-0.376^{***}	0.047

Rear-end	0.276***	0.044	0.280***	0.044
ρ	-0.313**	0.132		
Cut1	0.925***	0.071	0.921***	0.071
Cut2	1.450***	0.072	1.449***	0.072
Cut3	1.719***	0.074	1.720***	0.074
log-likelihood	-5491		-5493	
AIC	11025		11028	
BIC	11057		11059	
N	9816		9816	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5.5 Estimation result of non-commercial vehicle driver sample

	Correlated		**Uncorrelated**	
Variables	Coef.	SE	Coef.	SE
	Fatigue Driving Propensity			
Insurance				
Yes	0.351***	0.090	0.351***	0.089
Road type				
Express way	0.595***	0.108	0.604***	0.108
Urban road	-0.267***	0.068	-0.268***	0.068
Time of day				
00:00-06:59	0.419***	0.071	0.384***	0.070
Intercept	-2.896***	0.086	-2.891***	0.086
	Injury Severity Propensity			
Fatigue	0.895**	0.364	0.234**	0.101
Driver's gender				
Male	-0.200***	0.028	-0.199***	0.028
Driver's age				
26-35	-0.170***	0.016	-0.170***	0.016
46-55	0.238***	0.023	0.238***	0.023
56-65	0.500***	0.037	0.500***	0.037
⩾65	0.606***	0.075	0.606***	0.075
Driving experience				
≤ 2 years	0.153***	0.022	0.153***	0.022
Driving license				
Not valid	0.703***	0.017	0.703***	0.017
Insurance				
Yes	-0.281***	0.017	-0.280***	0.017

Road type				
Express way	0.370^{***}	0.050	0.388^{***}	0.050
Urban road	-0.250^{***}	0.015	-0.251^{***}	0.015
Isolated lanes				
Yes	0.060^{***}	0.016	0.060^{***}	0.016
Terrain				
Mountain	-0.135^{***}	0.015	-0.135^{***}	0.015
Lighting condition				
Dark with street light	0.076^{***}	0.017	0.077^{***}	0.017
Dark without street light	0.181^{***}	0.020	0.182^{***}	0.020
Collision type				
Head-on	0.140^{***}	0.017	0.140^{***}	0.017
Rear-end	0.180^{***}	0.024	0.180^{***}	0.024
ρ	-0.234^{*}	0.122		
Cut1	0.053	0.036	0.052	0.036
Cut2	0.960^{***}	0.036	0.960^{***}	0.036
Cut3	1.442^{***}	0.037	1.442^{***}	0.037
log-likelihood	-28726		-28728	
AIC	57504		57506	
BIC	57541		57541	
N	28748		28748	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.5.3 Marginal effects

The coefficient estimates do not provide the magnitude of impacts on probability in each injury level. Thus, we calculate the marginal effect of variables, which directly influence fatal injury propensity, on each injury level for both commercial and non-commercial vehicle drivers. Considering all variables in this model are dummy variables, we compute probabilities by setting the variable to one and then zero, and take the difference. That is,

$$ME = \text{Prob}(y_i = k \mid m_i = 1) - \text{Prob}(y_i = k \mid m_i = 0) \qquad (5\text{-}7)$$

where ME is the marginal effect of dummy m_i on injury level k. The marginal effect can be interpreted as the change of probability due to the change in variable from zero to one. The results of bivariate and univariate models for commercial and non-commercial vehicle drivers are presented in Table 5.6 and Table 5.7. (These tables only present the marginal effect for significant variables identified in earlier discussion).

Some important features should be addressed here. First, this study shows that the marginal effect of fatigue driving on fatal injury for commercial vehicle driver and non-commercial driver is 12.9% and 17.9%. That is, the occurrence of fatigue driving will increase the probability of fatal injury in a crash by 12.9% for commercial vehicle driver and 17.9% for non-commercial vehicle drivers. Moreover, comparing marginal effect in bivariate and univariate model, we find that the estimated impact of fatigue is much lower without considering correlation. Ignoring the correlation of unobserved factors may lead to underestimation of the harmfulness of fatigue driving behavior. In our study, the harmfulness of fatigue is underestimated by 10.5% and 14.6% for commercial and non-commercial vehicle driver, respectively. Second, for commercial vehicle drivers, other major risk factors of fatal injury include express way, not valid driving license, and rear-end collision. For non-commercial vehicle, elder driver (aged ≥66 years old and 56-65 years old), not valid driving license, and express way are the most significant contributors. Third, side collision, driving in mountain area, and insurance are recognized as the three most influencing factors for commercial vehicle drivers to survive in a crash while for non-commercial vehicle drivers are insurance, urban road, and male.

Table 5.6 Marginal effect for commercial vehicle driver sample

	Bivariate Binary-Ordered Probit				Univariate Binary-Ordered Probit			
	Y = 1	**Y = 2**	**Y = 3**	**Y = 4**	**Y = 1**	**Y = 2**	**Y = 3**	**Y = 4**
Fatigue Driving	-0.256	0.086	0.041	0.129	-0.060	0.025	0.010	0.024
Driving License								
Not valid	-0.058	0.025	0.010	0.023	-0.060	0.025	0.010	0.024
Insurance								
Yes	0.037	-0.016	-0.006	-0.014	0.038	-0.016	-0.006	-0.015
Road type								
Express way	-0.225	0.084	0.037	0.087	-0.237	0.087	0.039	0.094
Urban road	0.037	-0.009	-0.004	-0.008	0.036	-0.009	-0.003	-0.007
Isolated lanes								
Yes	0.015	-0.007	-0.003	-0.006	0.015	-0.007	-0.003	-0.006
Terrain								
Mountain	0.054	-0.022	-0.008	-0.018	0.046	-0.018	-0.007	-0.015
Lighting condition								
Dark with street light	-0.024	0.011	0.004	0.009	-0.026	0.011	0.004	0.010
Dark without street light	-0.046	0.021	0.008	0.017	-0.049	0.022	0.009	0.019
Collision type								
Head-on	0.024	-0.011	-0.004	-0.009	0.024	-0.011	-0.004	-0.009
Sideway	0.063	-0.031	-0.011	-0.022	0.065	-0.031	-0.011	-0.022
Rear-end	-0.054	0.024	0.009	0.020	-0.056	0.025	0.010	0.021

Table 5.7 Marginal effect for non-commercial vehicle driver sample

	Bivariate Binary-Ordered Probit				Univariate Binary-Ordered Probit			
	Y = 1	**Y = 2**	**Y = 3**	**Y = 4**	**Y = 1**	**Y = 2**	**Y = 3**	**Y = 4**
Fatigue Driving	-0.303	0.051	0.073	0.179	-0.081	0.028	0.020	0.033
Driver's gender								
Male	0.069	-0.025	-0.017	-0.027	0.069	-0.025	-0.017	-0.027
Driver's age								
26-35	0.058	-0.024	-0.014	-0.020	0.058	-0.024	-0.014	-0.020
46-55	-0.083	0.030	0.021	0.032	-0.083	0.030	0.021	0.032
56-65	-0.175	0.051	0.044	0.080	-0.175	0.051	0.044	0.080
≥65	-0.211	0.054	0.053	0.105	-0.211	0.054	0.053	0.105
Driving experience								
Less than 2 years	-0.053	0.020	0.013	0.020	-0.052	0.020	0.013	0.020
Driving License								
Not valid	-0.259	0.098	0.067	0.093	-0.258	0.098	0.067	0.093
Insurance								
Yes	0.097	-0.039	-0.025	-0.036	0.097	-0.039	-0.025	-0.037
Road type								
Express way	-0.140	0.038	0.031	0.055	-0.143	0.038	0.032	0.057
Urban road	0.089	-0.035	-0.021	-0.029	0.089	-0.035	-0.021	-0.029
Isolated lanes								
Yes	-0.021	0.008	0.005	0.007	-0.020	0.008	0.005	0.007
Terrain								
Mountain	0.046	-0.018	-0.011	-0.017	0.046	-0.018	-0.011	-0.017
Lighting condition								
Dark with street light	-0.026	0.010	0.006	0.009	-0.026	0.010	0.006	0.010
Dark without street light	-0.063	0.024	0.016	0.024	-0.063	0.024	0.016	0.024
Collision type								
Head-on	-0.048	0.019	0.012	0.018	-0.049	0.019	0.012	0.018
Rear-end	-0.062	0.023	0.015	0.024	-0.063	0.023	0.016	0.024

5.6 Conclusions and Practical Applications

Several studies have examined the relationship between driver fatigue and traffic injury severity from different aspects. However, some of factors that affect driver's fatigue propensity also have influence on driver's injury severity in a crash, including observed and unobserved factors. Ignoring the impact of these common factors will lead to endogeneity problem and incorrect conclusion. Based on 38,564 crash records during 2006-2011, we conduct an empirical analysis to examine the relationship between fatigue driving propensity and fatal injury severity by comparing bivariate and univariate endogenous binary-ordered probit model. Five types of factors are included. It is essential to quantify the impact of these characteristics on injury severity by calculating marginal effect, so that measures to prevent or reduce harmfulness of fatigue driving can be identified and implemented.

The result reveals a substantial and significant negative error correlation between fatigue driving propensity and fatal injury propensity, which lends strong support for endogeneity of fatigue driving propensity. The influence of fatigue driving on injury severity is significantly underestimated if ignoring the unobserved correlation between fatigue driving behavior and crash injury severity propensity. This study also compares the difference in risk factors of fatigue driving behavior and crash-related injury between commercial vehicle drivers and non-commercial drivers. Some common observed influential factors are identified. For instance, driving on express way not only contribute to higher fatal injury propensity but also high fatigue driving propensity. Measures aiming at preventing driver fatigue such as light signals or signs may also help to reduce injury severity in the crash. It is also found in the paper that factors show different impacts on them. Driver's gender and age has significant influence on fatigue driving propensity of non-commercial vehicle driver, but this influence is not significant on commercial vehicle driver.

It should arouse the attention of researchers that the harmfulness of driver fatigue on traffic crash injury severity is larger than expected due to neglecting of the endogeneity of fatigue. Furthermore, correctly understanding the impact of fatigue-related crash is considered to be essential to the development and design of countermeasures aimed at reducing the hazard of fatigue crash. Different impact factors identified between commercial and non-commercial vehicle drivers in this study should be addressed. Some factors have similar impacts for both commercial and non-commercial vehicle drivers (e.g. road types and lighting conditions), but some factors have not (e.g. collision types). Thus, developing effective measures to reduce fatigue-related crash occurrence and its injury severity should take into account those differences. Moreover, according to our findings, police makers should also consider installing driver fatigue prevention devices (e.g. deceleration strip or warning signs) on express ways since those devices help reducing driver fatigue as well as injury severity.

With respect to fatigue driving behavior, our results suggest that fatigue is important in reducing the likelihood of fatal injury. However, one of the major limitations of this research is the sparseness of fatigue-related crash in this dataset for the reason that small sample size may influence model estimation. It is essential to address endogenous in the model since endogeneity would cause inconsistent estimation. And this model can apply for any crash type when there is potential endogenous dependent variable. Therefore, it is reasonable to introduce this model in our analysis given the data limitation. In addition, the number of observations of some variables are small that may also limit the ability of determining effects precisely. And the vague and broad definition of fatigue may also cause misclassification problem and reduce the accuracy of our data analysis. This paper also does not consider the potential confounding effects of driving mileage, driver's health condition, drug use, which could affect both fatigue and risk of crash (Connor et al., 2001a), due to the limitation of data. To deal with this problem, more detail and complete data are needed. Interaction effects or non-linear effects of variables and heterogeneity of drivers, which may also have significant impacts on injury severity and fatigue driving propensity, are not considered in this study for the first attempt since the focus is on the endogeneity of fatigue driving. Those problems will be discussed in our future studies.

CHAPTER 6

Identifying Factors Contributing to County-level Fatigue-related Crash Considering Spatial Correlation

6.1 Introduction

Traffic crash has become a major threat all around the world and fatigue-related crashes have long been the topic of discussion in road safety research (Horne & Reyner, 1995a; Horne & Reyner, 1995b; Philip et al., 2001; Akerstedt & Kecklund, 2001; Connor et al., 2001b; Zhang et al., 2016). However, traffic crashes do not occur randomly and the spatial correlation involves in the processes and events leading to the traffic crash (Loo & Anderson, 2016). Different from other types of automobile crash, fatigue-related crash may be more closely related to the social and economic development level. For example, one area of high commercial transportation demand is more likely to have fatigue-related crash. It is important to evaluate the relationship between variables related to social and economic development and fatigue-related crashes.

Spatial correlation has long been discussed in transportation researches (Quddus, 2008; Huang et al., 2010). It originates from missing exogenous variables and inappropriate spatial aggregation of the underlying observational units (Anselin, 1988; Tiefelsdorf & Griffith, 2007; Wang et al., 2013). For example, variables such as transportation regulations and traffic flow, are often absent from traffic crash models, which may lead to spatial correlation across the neighborhood areas. In addition, the presence of spatial autocorrelation violates the assumption of independence in most conventional statistical models; ignoring spatial autocorrelation will lead to biased parameter estimation and standard errors (Wakefield, 2003).

Generally, there are two ways to handle spatial correlation: to take it into account in model setting or remove it from observations. Several attempts have been made to develop models that take spatial correlation into account, such as geographically weighted regression models (Quddus, 2008; Hadayeghi et al., 2010; Wang et al., 2012; Pirdavani et al., 2013). On the other hand, eigenvector spatial filtering is a relatively new method to deal with spatial correlation method (Griffith, 2000a; 2000b; 2007). Spatial filtering approach avoids the complex calculation of geographically weighted regression models and extracts the unobserved map patterns from the spatial structure in matrix form. The application of spatial filtering technique has been discussed and applied several studies (Cliff & Ord, 1981; Getis, 1990; Getis & Anselin, 1995; Griffith, 2002; Griffith, 2004).

This study aims to evaluate factors contributing to county-level fatigue-related crash frequency in Guangdong, China. The spatial filtering technique is used to capture the unobserved spatial correlation among studied areas. With the filtered spatial components, a semi-parametric Poisson model is developed to examine the impacts of both road and macroscopic variables on fatigue-related crash frequency. Since the spatial filtering process is based on the decomposition of spatial neighbor matrix, the spatial components can be included in the model as exogenous variables. In this way, the importance of spatial effects as well as macroscopic variables in explaining fatigue-related crash occurrence in Guangdong can be assessed. This study can provide some insights for regional policy evaluation and strategic planning in preventing fatigue-related crashes.

6.2 Methodology

6.2.1 Spatial neighbor matrix

The neighboring connectivity spatial unit i and j can be defined by a spatial neighbor matrix W. This matrix is a square symmetric $n \times n$ matrix with the (i, j) element $w_{i,j} = 1$ if units i and j are considered to be neighbors, and $w_{i,j} = 0$ otherwise. There are many ways for constructing the spatial neighbor matrix (LeSage, 1998):

Rook contiguity: two units are considered to be spatial neighbors if they share a common boundary only (or the boundary is longer than a given distance).

Bishop contiguity: two units are considered to be spatial neighbors if they share a common vertex only (or the boundary is shorter than a given distance).

Queen contiguity: two units are considered to be spatial neighbors if they share a common boundary or vertex. Queen contiguity can be viewed as the combination of Rook and Bishop contiguity.

Other than the above "first-order" measures of contiguity, one can extend the spatial neighbor definition to "second-order" contiguity. By the same token, the neighbors of first-order neighbor units can be viewed as second-order neighbors. In addition, distance-based measures of neighbors are also commonly applied. In practice, distance-based spatial weights matrices are popular since it allows us to take various forms of distance influence into account. However, distance-based spatial weight matrices do not always guarantee better fitness. El-Basyouny and Sayed (2001) mentioned that the spatially correlation may be simply due to the omission of variables. Therefore, in the case of omission of variables, testing the influence of different neighboring structures is more promising than constructing complex spatial structure.

The construction of neighbor matrix W depends on the problem and is somehow artificial due to lacking of information for specifying neighbors. Tsai et al. (2009) and Tsai (2012) found that first-order Queen contiguity was more appropriate when areas were highly irregular in shape and size. Griffith (1996) and Griffith and Lagona (1998) also mentioned that small number of neighbors was preferred after comparing different spatial weight matrix. Given the diverse size of spatial units in our dataset, binary contiguity is employed in this study. Rook and Queen contiguity structure are evaluated and the comparison of the basic information between them are listed in Table 6.1.

Table 6.1 The comparison of Queen and Rook contiguity

	Queen	**Rook**
Number of neighbors	120	120
Number of non-zero links	560	556
Average number of links	4.67	4.63

The results show that Rook and Queen contiguity are quite similar in light of number of non-zero links. There is no formal guidance for choosing a proper spatial matrix (Anselin, 2002; Paez & Scott, 2005). Therefore, both of the them will be used in spatial filtering and the final choice of spatial neighbor matrix is determined by optimal performance of the model (Getis & Aldstadt, 2004; Dray et al., 2006).

6.2.2 Eigenvector spatial filtering approach

This study follows the spatial filtering approach proposed by (Griffith, 2000a; 2000b). This approach also known as Moran eigenvector spatial filtering approach since it is based on the computational formula of MI statistic (Griffith, 2000a; Patuelli et al., 2006). The Moran eigenvector filtering approach relax the restriction of Getis's spatial filtering approach that can deal with both positive and negative spatial autocorrelation. In addition, it can be used either individually in a model as predicting variables or simultaneously in a regression system (Tiefelsdorf & Griffith, 2007).

The Moran spatial filtering approach eigenvectors $\{E_1,...,E_n\}$ are extracted from the transformed spatial neighbor matrix

$$\{E_1,\dots,E_n\} \equiv \text{evec}\left[M\,\tfrac{1}{2}\left(W+W^T\right)M\right] \tag{6-1}$$

where W represents the spatial neighbor matrix. M is a symmetric and idempotent projection matrix

$$M = I - 1\left(1^T 1\right)^{-1} 1^T \tag{6-2}$$

where I is an $n\times n$ identity matrix and 1 is an $n\times 1$ vector of ones. In this way, the spatial correlation can be expressed by independent and uncorrelated eigenvectors (Tiefelsdorf & Griffith, 2007). The eigenvectors extracted are arranged in an ascending order. The first eigenvector E_1 is the one whose numerical values generate the largest MI statistic among all eigenvectors of the transformed matrix. Similarly, the second eigenvector E_2, which is uncorrelated with E_1, is the set of numerical values that maximize the MI statistic. The process continues until the nth eigenvectors have been generated (Patuelli et al., 2006). By employing the suitable orthogonal and uncorrelated map patterns representing by eigenvectors to Poisson model, the spatial correlation present in the residuals can be removed (Dray et al., 2006).

To determine the suitable and parsimonious number of eigenvectors, a subset of representing eigenvectors is extracted and selected by a stepwise procedure. To avoid inflated sets of eigenvectors, Griffith (2003) suggested that a restriction is needed for searching over eigenvectors with moderate to high spatial autocorrelation. Suppose the first eigenvector E_1 express the highest level of spatial correlation. Then, it is possible to assume there exists an eigenvalue e_i corresponding to eigenvector E_i that $e_i \geq \gamma e_1$, which indicates E_i contains at least weakly spatial correlation given certain level γ. In this study, $\gamma = 0.5$ is used according to some prior researches (Griffith, 2003; Patuelli et al., 2006).

The selected Q eigenvectors are included into the Poisson model as following:

$$\mu = \exp\left[X\beta + \sum_{q=1}^{Q} E_q \delta_q \right] \tag{6-3}$$

where X is the exogenous explanatory variables (i.e. road length and macroscopic variables in this study) and β is the corresponding coefficients. E_q is the selected eigenvectors and δ_q represents the corresponding coefficient. However, there can be too many eigenvectors, the stepwise selection is used to choose appropriate variables. The stepwise selection can be done by maximizing model accuracy, which not only suitable for linear model but also for generalized linear models (Murakami & Griffith, 2005). The model accuracy is evaluated by three indexes: log-likelihood ratio test (LR), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC).

6.2.3 Proportion of reduction in variance

The literature suggests that various socio-demographic variables such as population, employment, poverty, economic activity affect traffic casualties (Graham & Glaister, 2003; Noland & Quddus, 2004a; Aguero-Valverde & Jovanis, 2006; Kim et al., 2006). Therefore, to evaluate the overall explanatory power of macroscopic variables, the proportion of reduction in variance (PRV, also called explained variance) is introduced (Raudenbush & Bryk, 2002; Wang et al., 2017). PRV was calculated by the following function:

$$PRV = \frac{\tau_0^2 - \tau_1^2}{\tau_0^2} \tag{6-4}$$

where τ_0^2 represent the variance of error term in the model which did not contain target variables while τ_1^2 is the variance of error term of model containing target variables. PRV should be bounded between 0 to 1, and high PRV value indicates strong explanatory power of variables on the occurrence of crash.

6.3 Data

6.3.1 Study area

The study area is Guangdong province (109°45′E - 117°20′E, 20°09′N - 25°31′N), China. It is located in the southern part of China, including 21 cities, and 121 counties. By the end of year 2014, Guangdong is populated by 110 million residents over an area of 179,716.02 square kilometers. It should be noticed that Nan'ao is excluded in this study since it is the only island and is not connected to any other county in Guangdong. Finally, totally 120 administrative counties are discussed in this study.

6.3.2 Data collection

The data used in this study were extracted from different sources. Crash data were collected from the Traffic Management Sector Specific Incident Case Data Report, the Road Traffic Accident Database of China's Public Security Department. Fatigue-related crash records during 2006-2014 were extracted and aggregated by counties in this study. In total, 2342 fatigue-related crashes were recorded in the database.

Road length were calculated from the Geospatial Database of the 1:1,000,000 Geological map of China. This database contains basic information about road and county boundary. The length of road of different level within the boundary of a county are summed up separately. Four types of roads are considered: expressway, national road, provincial road, and urban expressway.

To investigate the relationships between the occurrence of fatigue-related crash and macroscopic variables, the macroscopic data for county-level units in Guangdong were collected from the Guangdong Statistical Yearbook and Guangdong 1% Population Sampling Survey Data (Guangdong Statistical Bureau, 2017). It is noted that some variables are only available for city-level, they were adjusted by the land area assuming the distributed evenly given the same city.

6.3.3 Variables

The total number of fatigue-related crash is selected as the dependent variables. Various explanatory variables aggregated at county-level are evaluated. Generally, the variables can be grouped into two categories: road characteristics and macroscopic characteristics.

Several road characteristics (e.g. roadway density, functional classification, speed limit, number of lanes, etc.) have been discussed by previous studies (Noland & Quddus, 2005; Quddus, 2008; Hadayeghi et al., 2010; Huang et al., 2010; Abdel-Aty et al., 2011). The impacts of different road types on fatigue-related crash are discussed. Four variables are included in the model: the total length of highway, national road, provincial road and urban expressway.

Regarding the socio-economic variables, Wang et al. (2017) demonstrated that macroscopic variables can be considered as surrogate of individual behaviors. Population density (de Guevara et al., 2004; Permpoonwiwat & Kotrajaras, 2012), gender and age (MacNab, 2004; Aguero-Valverde & Jovanis, 2006; Quddus, 2008;

Hadayeghi et al., 2010; Huang et al., 2010; Li et al., 2013), employment (Hadayeghi et al., 2010, Huang et al., 2010; Pulugurtha et al., 2013; Xu et al., 2014) were found to has significant influence on the crash occurrence (Loukaitou-Sideris et al., 2007; Wier et al., 2009). The population number, proportion of population younger or equal to 15 years old, proportion of population equal to or older than 65 years old, total employment number, proportion of unemployment by various industry, average room per person, total commuters by public transportation modes (including bus and taxi), and distance to the capital city in Guangdong.

However, traffic characteristics is another important category of variables that has been widely discussed in traffic crash analysis. For example, Vehicle Miles Travelled (VMT) (Dong et al., 2015; Cai et al., 2017) and proportion of heavy vehicle mileage (Cai et al., 2017) are used as a measurement of exposure of traffic. Unfortunately, those data are not available for our data set and will not discussed in this study. It should be noted that more variables were tested, however, only the descriptive statistics for final selected variables are shown in Table 6.2.

Table 6.2 Descriptive statistics of variables

Variable	**Definition**	**Mean**	**SD**	**Min.**	**Max.**
Dependent variable					
Total	Number of total fatigue-related crash	19.52	60.75	0.00	397.00
Independent variable					
Highway	The length of highway (1000 km)	0.11	0.11	0.00	0.68
National	The length of national road (1000 km)	0.06	0.05	0.00	0.22
Provincial	The length of provincial road (1000 km)	0.16	0.12	0.01	0.80

Urbanexp	The length of urban expressway (1000 km)	0.01	0.03	0.00	0.23
Pop	Number of population (Million people)	2.95	0.75	0.98	5.46
Age0014	Proportion of people aged from 0 - 15	0.18	0.04	0.10	0.28
Age6000	Proportion of people aged elder than 60	0.14	0.04	0.03	0.22
Ave_Room	Average room per person	1.06	0.29	0.44	1.95
Trans_emp	Proportion of employee in transportation industry	0.03	0.02	0.01	0.22
Distance	Distance to Guangzhou (the capital city of Guangdong) (km)	171.36	121.40	0.00	366.06
Passenger_pub	Number of passenger using bus and taxi (100 million)	0.81	1.91	0.01	11.48

6.4 Result and Discussion

The study investigates the relationship between different contributing factors and county-level fatigue-related crash frequency using basic Poisson model and spatial filtering Poisson model. Two different spatial contiguity structures are applied in spatial filtering process. The extracted eigenvectors are selected by a stepwise procedure and included in the Poisson model as the surrogate measurement of spatial correlation. Those models are denoted as Basic Poisson (BP) model, Rook Spatial Filtering Poisson (RSFP) model and Queen Spatial Filtering Poisson (QSFP).

The explanatory variables selected into the final model setting are based on theoretically important and their significant level. Also, the variance inflation factor (VIF) is also calculated to check the multicollinearity between variables. Generally, as a rule of thumb, VIF>10 is considered to have high multicollinearity (Kutner, et al., 2004). Variables with VIF>10 will be removed from the regression model to avoid multicollinearity. R software (version 3.4.3) is used for all the estimation in this study.

6.4.1 Model goodness-of-fit and comparison

Then the goodness-of-fit statistics for the BP, RSFP and QSFP are shown and compared in Table 6.3. Results indicate that incorporating the filtered spatial components in fatigue-related crash Poisson model leads to a considerable improvement in the overall model fitness. Both RSFP and QSFP include 15 eigenvectors, which is selected by the stepwise procedure. And the selected eigenvalues explain 60.11% and 60.34% of positive spatial correlations in RSFP and QSFP, respectively. As shown in Table 6.3, the likelihood ratio (LR) test comparing the RSFP models (LR = 3633.484) and QSFP model (LR = 3409.168) with the base Poisson model are both significance at the 0.01 level. Moreover, the AIC and BIC also prefer RSF Poisson and BIC Poisson rather than the base Poisson model. These results suggest that the filtered spatial components should not be omitted in fatigue-related crash frequency analysis.

However, comparing the goodness-of-fit results between RSFP model and QSFP model, Rook contiguity structure can better explain the spatial patterns of fatigue-related crashes than Queen contiguity structure. Although these two types of contiguity structure show in number of links, common vertex may not be a good measure of spatial connection of fatigue-related crashes based on our results.

Table 6.3 Goodness-of-fit statistics

Goodness-of-fit	BP	RSFP	QSFP
Number of spatial units	120	120	120
Number of selected eigenvector	0	15	15
Degrees of freedom	12	27	27
Adjusted R-square	0.222	0.590	0.567
Log-likelihood at convergence	-3839.35	-2022.608	-2134.766
AIC	7702.701	4099.217	4323.533
BIC	7736.151	4174.479	4398.795

6.4.2 Parameters and marginal effects

Several factors are found to have significant impacts on fatigue-related crash occurrence. The estimation results for BP, RSFP and QSFP are presented in Table 6.4 and Table 6.5. The significant level of most of the variables is the same among those three models, except for the length of urban expressway and total number of public transport commuters. According to the discussion above, RSFP gives the better model fitness. Therefore, the following discussions on the estimated coefficients and marginal effects are based on results of RSFP model. The marginal effects calculated based on RSFP are also presented in the final column of Table 6.4.

The road length variables are found to have inconsistent effects. The total length of highway has negative impact on occurrence of fatigue-related crashes while national road and provincial road has positive impact on fatigue-related crash occurrence. The marginal effect of highway is -9.138 while the marginal effect of national road and provincial road is 3.650 and 4.437, respectively. It is possible that commercial vehicles are one of the high-risk groups in fatigue-related crash; they tend to choose lower level road (national and provincial road) rather than highway considering travel cost.

Therefore, the total length of national and provincial road contributes to fatigue-related crash occurrence.

Population is positively associated with fatigue-related crash occurrence with marginal effect of 0.491. This finding it consistence to other researches analyzing crash occurrence or casualties. Larger population was identified to be closely related with higher crash exposure opportunity (Pulugurtha & Sambhara, 2011; Mitra & Washington, 2012; Wang et al., 2017), which is true not only for fatigue-related crash but also other types of crash. Moreover, the proportion of population elder or equal to 60 years old and the proportion of population younger or equal to 15 years old also have significant negative impact on the occurrence of fatigue-related crash. The marginal effect for them are -13.306 and -9.243, respectively. Some researchers claimed that young and old people are of relatively high risk in traffic crashes (Huang et al., 2010; Aguero-Valverde, 2013; Wang et al., 2017). However, the result suggests that higher proportion of young or old people are associated with lower fatigue-related crash occurrence. This may be an expected result in the context of fatigue-related crash given that few of people in these two groups will drive on road. Also, fatigue-related crashes are often observed among professional driver, who's age obviously not belongs to either groups.

Table 6.4 Parameters estimation results

	BP			RSFP			QSFP			Marginal
Variables	**Coef.**	**S.E.**	**P-value**	**Coef.**	**S.E.**	**P-value**	**Coef.**	**S.E.**	**P-value**	
Intercept	2.508	0.272	0.000	-1.812	0.646	0.005	-1.497	0.589	0.011	-
Highway	-3.783	0.372	0.000	-21.286	0.944	0.000	-19.709	0.803	0.000	-9.138
National	5.495	0.532	0.000	8.503	0.708	0.000	8.383	0.689	0.000	3.650
Province	3.798	0.297	0.000	10.335	0.566	0.000	7.572	0.495	0.000	4.437
Urbanexp	5.286	0.718	0.000	1.064	1.282	0.407	6.434	1.19	0.000	-
Log(pop)	0.199	0.047	0.000	1.145	0.094	0.000	1.141	0.088	0.000	0.491
Trans_emp	2.371	0.977	0.015	9.05	1.305	0.000	7.992	1.335	0.000	3.885
Ave_Room	0.889	0.118	0.000	-1.86	0.231	0.000	-2.575	0.233	0.000	-0.798
Log(distance)	0.365	0.025	0.000	1.783	0.086	0.000	1.555	0.074	0.000	0.765
Age0014	-6.219	1.034	0.000	-21.531	1.887	0.000	-9.967	1.516	0.000	-9.243
Age6000	-17.642	1.119	0.000	-30.995	1.959	0.000	-26.45	1.869	0.000	-13.306
Passenger_pub	-0.139	0.019	0.000	-0.044	0.034	0.189	-0.116	0.045	0.009	-

Table 6.5 Estimation results of eigenvectors

RSFP				QSFP			
eigenvector	Coef.	S.E.	P-value	eigenvector	Coef.	S.E.	P-value
r_e1	17.435	1.422	0.000	q_e1	17.503	1.326	0.000
r_e2	0.699	0.761	0.359	q_e2	-13.175	0.915	0.000
r_e3	-31.369	1.588	0.000	q_e3	-16.934	0.894	0.000
r_e4	-18.206	0.976	0.000	q_e4	-2.202	0.619	0.000
r_e5	-18.068	0.910	0.000	q_e5	-14.420	0.864	0.000
r_e6	-0.615	0.571	0.282	q_e6	4.165	0.557	0.000
r_e7	6.055	0.562	0.000	q_e7	17.391	0.932	0.000
r_e8	39.790	1.818	0.000	q_e8	23.898	1.160	0.000
r_e9	-10.523	0.801	0.000	q_e9	-5.563	0.559	0.000
r_e10	-6.760	0.500	0.000	q_e10	-4.823	0.534	0.000
r_e11	-13.626	0.577	0.000	q_e11	-12.423	0.611	0.000
r_e12	7.821	0.651	0.000	q_e12	-6.179	0.536	0.000
r_e13	1.991	0.412	0.000	q_e13	1.979	0.462	0.000
r_e14	-7.537	0.514	0.000	q_e14	3.377	0.373	0.000
r_e15	13.833	0.560	0.000	q_e15	-15.522	0.697	0.000

RSFP				QSFP			
eigenvector	Coef.	S.E.	P-value	eigenvector	Coef.	S.E.	P-value
r_e1	17.435	1.422	0.000	q_e1	17.503	1.326	0.000
r_e2	0.699	0.761	0.359	q_e2	-13.175	0.915	0.000
r_e3	-31.369	1.588	0.000	q_e3	-16.934	0.894	0.000
r_e4	-18.206	0.976	0.000	q_e4	-2.202	0.619	0.000
r_e5	-18.068	0.910	0.000	q_e5	-14.420	0.864	0.000
r_e6	-0.615	0.571	0.282	q_e6	4.165	0.557	0.000
r_e7	6.055	0.562	0.000	q_e7	17.391	0.932	0.000
r_e8	39.790	1.818	0.000	q_e8	23.898	1.160	0.000
r_e9	-10.523	0.801	0.000	q_e9	-5.563	0.559	0.000
r_e10	-6.760	0.500	0.000	q_e10	-4.823	0.534	0.000
r_e11	-13.626	0.577	0.000	q_e11	-12.423	0.611	0.000
r_e12	7.821	0.651	0.000	q_e12	-6.179	0.536	0.000
r_e13	1.991	0.412	0.000	q_e13	1.979	0.462	0.000
r_e14	-7.537	0.514	0.000	q_e14	3.377	0.373	0.000
r_e15	13.833	0.560	0.000	q_e15	-15.522	0.697	0.000

Regarding employment, the number of employees or employment density has already been found to be a significant variable for predicting crashes (Noland & Quddus, 2003; Wier et al., 2009). Given that fatigue-related crashes are commonly found among professional drivers, the percentage of employee in transportation is used as a measurement of exposure. The result also shows that higher proportion of employee in transportation industry has positive impact on fatigue-related crash occurrence, which confirms that fatigue-related crash is closely related to professional drivers. The marginal effect for proportion of transportation employee is 3.885. That is, a unit increase in proportion of transportation employee will lead to the increase of frequency of fatigue-related crash by 3.885.

The average room per person, as the measurement of resident density, has a pronounced and stronger negative effect on the occurrence of fatigue-related crashes. The marginal effect for it is -0.798. The possible reason is the traffic within commercial area may increase the risk of fatigue-related crashes. Hence, high residential density has relatively lower risk of crash. Similar discussion can be found in Noland and Quddus (2004). Another interesting result can be found in the variable of distance to capital city. This variable measures the geographic connection as well as economic connections within Guangdong. Guangzhou, the capital city of Guangdong province, is surrounded by several notable cities (e.g. Shenzhen). This area, also known as “the Pearl River Delta (PRD)”, is the economic hub in Guangdong. Due to the economic clustering effect, areas which are far from PRD may be relatively less developed. As a general measurement, the results indicate that the overall development level (represent by distance to capital city) has a negative impact on fatigue-related crash occurrence with the marginal effect of 0.765 due to, for example, the relatively poor transportation infrastructure.

6.4.3 Contribution of variables

Table 6.5 shows the contribution of macroscopic components as well as the filtered spatial components. The PRV of macroscopic variables in fatigue-related crash frequency model is 7.69%. It means that 7.69% of unobserved variation can be explained by the omission of macroscopic variables. This result is similar to Wang et al. (2017), in which the motor vehicle crash-frequency model with macroscopic variables has PRV in a range of 2.67-7.98% for different buffer width. Interestingly, the PRV of spatial component is 44.73%, which is much larger than the PRV of macroscopic variables in fatigue-related crash occurrence. This finding also confirms the existence of spatial correlation in fatigue-related crash occurrence. Therefore, it is necessary to take this issue into account when analyze fatigue-related crash frequency to avoid estimation bias caused by the spatial correlated error term.

Table 6.6 Contribution of variables

Components	**PRV**
Macroscopic variables	7.69%
Spatial components	44.73%

6.5 Conclusion

This chapter examines factors contributing to fatigue-related crash occurrence considering spatial correlation using the fatigue-related crash frequency data of 120 counties in Guangdong, China. By the eigenvector spatial filtering approach, the unobserved spatial correlation can be extracted from the error term. These eigenvectors represent the orthogonal and uncorrelated map patterns given spatial structure and can be embedded into a semiparametric statistical framework. Then, the suitable and parsimonious subset of eigenvectors are determined by the stepwise procedure while the spatial contiguity matrix is selected based on the optimal model selection. The model fitness suggests Rook contiguity matrix better explains the neighbor structure of county-level fatigue-related crash occurrence in Guangdong.

Several variables are evaluated in the model. Based on the result, the total length of national, provincial and urban expressway is found to positively related to fatigue-related crash occurrence while the total length of highway is negatively related to fatigue-related crash occurrence. For macroscopic variables, higher proportion of young and elder people has negative impacted on the fatigue-related crashes. Other macroscopic variables, for example population and employment, are found to be have positive influence on the occurrence of fatigue-related crash. The results suggest that special attention should be pay for transportation industry employees since they are related to high risk of fatigue-related crash.

Besides, the effects of omitted spatial components and macroscopic variables in the model evaluating fatigue-related crash are examined, and the PRV for spatial components and macroscopic are calculated. Macroscopic variables covered 7.69% of the unobserved variation in the error term while the filtered spatial components explain 44.73%. Thus, this finding indicates the omission of those variables will leave those correlation in the error term, which violates the independent assumption in many conventional models.

However, there are still some limitations that need to be discussed in the future studies. Poisson model is criticized for the equality of its mean and variance assumption, which is not appropriate to deal with overdispersion data (i.e. variance is larger than the mean). In this case, negative binomial should be considered. In addition, the data applied is aggregated by county, which is an administrative unit and might not matched the real crash diffusion process. Large territory of one county may be heterogeneous. Several important variables such as traffic volume, are not considered due to lacking of data source. More spatial contiguity matrices should be tested and evaluated the sensitivity of the results.

CHAPTER 7

Conclusions and Future Work

7.1 Summary

This dissertation mainly focused on the fatigue-related crashes. The main objectives of this study are to (1) examine possible reasons for neglecting the harmfulness of fatigue-related crash, and (2) identify factors contributing to the occurrence of fatigue-related crash as well as severe outcome in a crash.

In the micro-level analysis, it is found that both misclassification and endogeneity can lead to the underestimation of driver fatigue harmfulness. Chapter 4 contributes to safety literature by introducing an analysis framework based on existing police recorded data to identify factors that easily make fatigue-related crashes misclassified by police officers, and examining the interactive effects of those factors. Some variables such as road types, collision types, and vehicle types, and their interactions are identified to have significant impacts on fatigue-related crash detection. In Chapter 5, another possible reason for underestimating the harmfulness of fatigue is explored. Ignoring the common factors can lead to endogeneity problem and result in biased parameter estimation. The factors contributing to fatigue crash propensity and its consequent severity are different between commercial and non-commercial vehicle drivers. The results show that the influence of fatigue driving on injury severity is significantly underestimated if the endogeneity of fatigue driving on fatal injury propensity is ignored.

For the perspective of macro-level analysis, chapter 6 investigates the fatigue-related crash from macro-level. County-level fatigue-related crash occurrence is influenced by several macroscopic factors. To capture the unobserved spatial correlation, a spatial filtering technique is applied. With the filtered spatial components, a semi-parametric Poisson model is developed to evaluating the impacts of both roads and macroscopic variables on fatigue-related crash frequency. The calculation results indicate the filtered spatial components and macroscopic variables explained more than half of the unobserved variation in the error term.

Some conclusions can be drawn based on the discussions in this research. First of all, single vehicle rolling over or hitting fixed object crashes are good indicators for determining fatigue-related crashes. From chapter 3, the proportions of single vehicle rolling over and hitting fixed object crash in fatigue-related crashes are higher than all crashes. The results in chapter 4 also show that these two types of crashes are related to lower propensity of misclassification of fatigue-related crashes.

Workers and migrant workers who were labeled as “high-risk of fatigue driving group” by previous research or the guidance of detecting fatigue-related crash for police officers, turn out to be misleading when detecting fatigue-related crashes. As well, expressway and urban expressway are also contributing to fatigue-related crash misclassification.

Higher number of total length of national and provincial road is positively correlated with higher number of fatigue-related crashes. Comparing to expressway, cargo vehicle drivers prefer national and provincial roads due to lower cost. Normally, these roads lack surveillance and rest stop, which can contribute to driver fatigue.

Factors such as vehicle insurance and road types not only affect fatal injury propensity, but also fatigue driving propensity. Road types and road lighting conditions show similar impacts for both commercial and non-commercial vehicle drivers. On the contrary, the impact of collision types on fatigue driving propensity and fatal injury propensity differs between commercial and non-commercial vehicle drivers.

Large and medium passenger vehicle driver will receive severe punishments if they are convicted as fatigue driving according to the current transportation laws and regulations. Raise of penalty caused a decrease in fatigue-related crashes as shown in chapter 3. However, punishments for other types of drivers are relatively mild even though these drivers are also likely to drive under fatigue (e.g. cargo vehicle drivers).

According to conclusions above, countermeasures are proposed to better deal with fatigue-related crashes. First of all, on way to prevent fatigue-related crashes is providing training for police officers. The current laws or regulations about fatigue-related crash are too vague to implement that the police and other enforcement authorities need appropriate knowledge to detect driver fatigue. Some features are good indicators for detecting fatigue-related crash, which should be highlighted. For example, it should be mentioned in the guidance and training process for police officers that if a crash is a single vehicle rolling over or hitting fixed object crash that they should realize this crash may be fatigue-related. In this way, the routine investigation of driver fatigue (e.g. asking the rest schedule of driver, self-estimation survey of fatigue level) should be conducted in the investigation procedure. At the same time, misleading features in identifying fatigue-related crashes such as workers, immigration workers, and expressway, should also be addressed in the training process.

Besides training, reliable and detailed records are also essential to identify driver fatigue. In the current traffic crash investigation procedure, it is suggested to collect information related to driver fatigue such as driving duration, health condition, etc. However, this information is not a compulsory part that sometimes it is neglected. Since the information of driving duration, rest time, sleep condition as well as self-evaluation of fatigue condition is vital for determining driver fatigue, it should be included in routine traffic injury investigation procedural compulsorily and displayed in the final crash investigation report. Also, usable and reliable vehicle-based fatigue measurement devices should be encouraged. These devices not only can be used in monitoring drivers' behaviors and the level of driver fatigue by placing sensors on the steering wheels and acceleration pedals, but also provide useful information for traffic police officers to determine the role of fatigue in a crash.

This research also suggests that national road and provincial road should attract more attention. There is only limited number of parking places (normally gas station) along national roads and provincial roads in which drivers can take a break during their trips. More rest stops are needed to prevent fatigue driving, especially in Summer. Also, setting checkpoints at the entrances and exits of national roads and provincial roads to inquire driver fatigue condition are also helpful to prevent fatigue-related crashes.

Finally, raising fatigue driving violation penalties for cargo vehicles can prevent drivers from fatigue driving as well as stimulate police officers to put more attention on identifying the involvement of fatigue in a crash. Moreover, the findings also suggested that different impact factors are identified between commercial and non-commercial vehicle drivers due to their different working patterns. For commercial vehicle driver, their employers should take the responsibility to secure that their employees are adequately informed and educated to stay vigilant while driving.

7.2 Future Work

First and foremost, a clear and easy-to-implement fatigue definition is essential to evaluate and make countermeasure to prevent fatigue-related crashes. Therefore, finding an appropriate definition and quantification method for driver fatigue is one of the biggest challenges in fatigue research during the coming years. More efforts should be invested in finding solid indexes for measuring driver fatigue.

Additionally, it is possible that not all fatigue-related crashes have been detected even after full investigation process. Some fatigue-related crashes cannot be detected, especially among property only crashes. Future study needs to investigate the impact of fatigue in property only crash.

Moreover, factors contributing to the detection of fatigue-related crashes in the dissertation, however, whether these factors also influence police officers' judgment on other types of crash. To understand whether those factors are unique for fatigue-related crashes, comparison studies of factors contributing to misclassification of other types of crashes should be conducted in the future.

Although the influence of endogeneity and spatial correlation on fatigue-related crash model has been discusses, other statistical issues which may also affect the estimation results, should be considered in the future studies. For example, heterogeneity of drivers, which may also have significant impacts on injury severity and fatigue driving propensity, should be discussed in our future studies.

BIBLIOGRAPHY

Abay, K. A., Paleti, R., & Bhat, C. R. (2013). The joint analysis of injury severity of drivers in two-vehicle crashes accommodating seat belt use endogeneity. *Transportation research part B: methodological*, *50*, 74-89.

af Wåhlberg, A. E., Dorn, L., & Kline, T. (2010). The effect of social desirability on self-reported and recorded road traffic accidents. *Transportation Research Part F: Traffic Psychology and Behaviour*, *13*(2), 106-114.

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Acm sigmod record* (Vol. 22, No. 2, pp. 207-216). ACM.

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

Åkerstedt, T., & Kecklund, G. (2001). Age, gender and early morning highway accidents. *Journal of sleep research*, *10*(2), 105-110.

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

Anastasopoulos, P. C., Mannering, F. L., Shankar, V. N., & Haddock, J. E. (2012). A study of factors affecting highway accident rates using the random-parameters tobit model. *Accident Analysis & Prevention*, *45*, 628-633.

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical analysis*, *20*(1), 1-17.

Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural economics*, *27*(3), 247-267.

Armstrong, K. A., Smith, S. S., Steinhardt, D. A., & Haworth, N. L. (2008). Fatigue crashes happen in urban areas too: characteristics of crashes in low speed urban areas. 2008 Road Safety Conference.

Armstrong, K., Filtness, A. J., Watling, C. N., Barraclough, P., & Haworth, N. (2013). Efficacy of proxy definitions for identification of fatigue/sleep-related crashes: An Australian evaluation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *21*, 242-252.

Armstrong, K., Obst, P., Livingstone, K., & Haworth, N. (2011). Investigation of Differences in Crash Characteristics between Males and Females Involved in Fatigue-Related Crashes or Close-Call events. *Women's Issues in Transportation*, 26.

Arnold, P. K., Hartley, L. R., Corry, A., Hochstadt, D., Penna, F., & Feyer, A. M. (1997). Hours of work, and perceptions of fatigue among truck drivers. *Accident Analysis & Prevention*, *29*(4), 471-477.

Australian Transport Council (ATC) (2011). *National Road Safety Strategy 2011-2020.*

Australian Transport Safety Bureau (ATSB) (2006). *Characteristics of fatal road crashes during national holiday periods*. Canberra: Commonwealth of Australia.

Beirness, D. J., Simpson, H. M., & Desmond, K. (2005). *The road safety monitor 2004: Drowsy driving*.

Board, S. (1999). Evaluation of US Department of Transportation Efforts in the 1990s to Address Operator Fatigue. *Safety Report NTSB/SR-99/01. Washington, DC*.

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*(1-2), 51-68.

Canadian Council of Motor Transport Administrator (CCMTA) (2010). *Canadian fatigue related collisions: Fatality estimates 2000-2005*.

Chang, L. Y., & Mannering, F. (1999). Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents. *Accident Analysis & Prevention, 31*(5), 579-592.

Changpetch, P., & Lin, D. K. (2013a). Model selection for logistic regression via association rules analysis. *Journal of Statistical Computation and Simulation, 83*(8), 1415-1428.

Changpetch, P., & Lin, D. K. (2013b). Selection of multinomial logit models via association rules analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *5*(1), 68-77.

Chen, E., & Tarko, A. P. (2014). Modeling safety of highway work zones with random parameters and random effects models. *Analytic methods in accident research*, *1*, 86-95.

Chun, Y. (2008). Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems*, *10*(4), 317-344.

Chun, Y., & Griffith, D. A. (2011). Modeling network autocorrelation in space–time migration flow data: an eigenvector spatial filtering approach. *Annals of the Association of American Geographers*, *101*(3), 523-536.

Cliff, A. D., & Ord, J. K. (1973). *Spatial Autocorrelation* (Pion, London).

Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: models & applications*. Taylor & Francis.

Cohen, A., & Einav, L. (2003). The effects of mandatory seat belt laws on driving behavior and traffic fatalities. *Review of Economics and Statistics*, *85*(4), 828-843.

Connor, J., Whitlock, G., Norton, R., & Jackson, R. (2001b). The role of driver sleepiness in car crashes: a systematic review of epidemiological studies. *Accident Analysis & Prevention*, *33*(1), 31-41.

Connor, J., Norton, R., Ameratunga, S., Robinson, E., Wigmore, B., & Jackson, R. (2001a). Prevalence of driver sleepiness in a random population-based sample of car driving. *Sleep*, *24*(6), 688-694.

Corfitsen, M. T. (1999). "Fatigue" among young male night-time car drivers: is there a risk-taking group?. *Safety science*, *33*(1-2), 47-57.

Coruh, E., Bilgic, A., & Tortum, A. (2015). Accident analysis with aggregated data: The random parameters negative binomial panel count data model. *Analytic methods in accident research*, *7*, 37-49.

Crummy, F., Cameron, P. A., Swann, P., Kossmann, T., & Naughton, M. T. (2008). Prevalence of sleepiness in surviving drivers of motor vehicle collisions. *Internal medicine journal*, *38*(10), 769-775.

DaCoTA (2012). Fatigue. Deliverable 4.8h of the EC FP7 Project DaCoTA. http://www.dacota-project.eu/Deliverables/Webtexts/Fatigue.pdf

Dalziel, J. R., & Job, R. S. (1997). Motor vehicle accidents, fatigue and optimism bias in taxi drivers. *Accident Analysis & Prevention*, *29*(4), 489-494.

Das, S., & Sun, X. (2014). Investigating the pattern of traffic crashes under rainy weather by association rules in data mining. In *Transportation Research Board 93rd Annual Meeting* (No. 14-1540).

de Lapparent, M. (2008). Willingness to use safety belt and levels of injury in car accidents. *Accident Analysis & Prevention*, *40*(3), 1023-1032.

de Oña, J., López, G., & Abellán, J. (2013). Extracting decision rules from police accident reports through decision trees. *Accident Analysis & Prevention*, *50*, 1151-1160.

Dee, T. S. (1998). Reconsidering the effects of seat belt laws and their enforcement status. *Accident Analysis & Prevention*, *30*(1), 1-10.

Department for Transport (2016). *Reported Road Casualties Great Britain (2015) Annual Report.* Department for Transport, London.

Derrig, R. A., Segui-Gomez, M., Abtahi, A., & Liu, L. L. (2002). The effect of population safety belt usage rates on motor vehicle-related fatalities. *Accident Analysis & Prevention*, *34*(1), 101-110.

Desmond, P. A. & Hancock, P. A. (2001). *Active and passive fatigue states. Stress, workload and fatigue.*

Diamantopoulou, K., Hoareau, E., Oxley, P. E., & Cameron, M. H. (2003). *The Feasibility of Identifying Speeding-Related & Fatigue-Related Crashes in Police-Reported Mass Crash Data*. (197 ed.) Clayton Vic Australia: Monash University.

Dinges, D. F. (1995). An overview of sleepiness and accidents. *Journal of sleep research*, *4*(s2), 4-14.

Dinges, D. F., & Kribbs, N. B. (1991). Performing while sleepy: effects of experimentally-induced sleepiness.

Diniz-Filho, J. A. F., & Bini, L. M. (2005). Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography*, *14*(2), 177-185.

Dobbie, K. (2002). *Fatigue-related crashes: An analysis of fatigue-related crashes on Australian roads using an operational definition of fatigue* (No. OR 23).

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

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3-4), 483-493.

Dyani, J. S., 2007. *Active versus passive fatigue in simulated driving*. University of Cincinnati.

El-Basyouny, K. & Sayed. T. (2009). Urban Arterial Accident Prediction Models with Spatial Effects. *Transportation Research Record*, Vol. 2102, No. -1, pp. 27-33.

Eluru, N., & Bhat, C. R. (2007). A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis & Prevention*, *39*(5), 1037-1049.

Elvik, R. (1995). Meta-analysis of evaluations of public lighting as accident countermeasure. *Transportation Research Record*, *1485*, 112-123.

European Transport Safety Council (2001). *The role of driver fatigue in commercial road transport crashes.*

Evans, L. (1996). Safety-belt effectiveness: the influence of crash severity and selective recruitment. *Accident Analysis & Prevention*, *28*(4), 423-433.

Fadhli, M. M., Mohamed, N., Othman, H., Sarani, R., & Voon, W. S. (2008). *Prevalence of fatigue among commercial bus drivers in Malaysia* (No. MRR 06/2008).

Fernandes, R., Hatfield, J., & Job, R. S. (2010). A systematic investigation of the differential predictors for speeding, drink-driving, driving while fatigued, and not wearing a seat belt, among young drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *13*(3), 179-196.

Fernandez-Antolin, A., Stathopoulos, A., & Bierlaire, M. (2014). Exploratory analysis of endogeneity in discrete choice models. In *14th Swiss Transport Research Conference* (No. EPFL-CONF-202581).

Feyer, A. M. (2001). *Fatigue: time to recognise and deal with an old problem: It's time to stop treating lack of sleep as a badge of honour*.

Filtness, A. J., Armstrong, K. A., Watson, A., & Smith, S. S. (2017). Sleep-related crash characteristics: Implications for applying a fatigue definition to crash reports. *Accident Analysis & Prevention, 99*, 440-444.

Folkard, S. (1997). Black times: temporal determinants of transport safety. *Accident Analysis & Prevention*, *29*(4), 417-430.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships/cA. Stewart Fotherington, Chris Brunsdon, and Martin Charlton*. Wiley.

Geti, S. A., Anselin, L., & Florax, R. J. G. M. (1995). Spatial filtering in a regression framework. *New directions in spatial econometrics. Berlin: Springer*. *FREDERIC CARLUER*, *292*.

Getis, A. (1990). Screening for spatial dependence in regression analysis. *Regional Science, 69*(1), 69-81.

Getis, A., & Aldstadt, J. (2004). On the specification of the spatial weights matrix. *Geographical Analysis, 35*.

Getis, A., & Griffith, D. A. (2002). Comparative spatial filtering in regression analysis. *Geographical analysis*, *34*(2), 130-140.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical analysis*, *24*(3), 189-206.

Government of South Australia (2010). *Fatigue-related Crashes in South Australia 2010*, from http://www.dpti.sa.gov.au/__data/assets/pdf_file/0010/51022/Fatigue_Fact_Sheet_2010.pdf

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

Greene, W.H., 2012. *Econometric Analysis*. 7th Edition, Pearson Education, Inc., Boston.

Gregersen, N. P., & Bjurulf, P. (1996). Young novice drivers: Towards a model of their accident involvement. *Accident Analysis & Prevention*, *28*(2), 229-241.

Griffith, D. A. (1996). Some guideline for specifying the geographic weights matrix contained in spatial statistical models. *Practical handbook of spatial statistics*, 65-82.

Griffith, D. A. (2000a). A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems*, *2*(2), 141-156.

Griffith, D. A. (2000b). Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra and its Applications*, *321*(1-3), 95-112.

Griffith, D. A. (2002). A spatial filtering specification for the auto-Poisson model. *Statistics & probability letters*, *58*(3), 245-251.

Griffith, D. A. (2003). *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. Springer Science & Business Media.

Griffith, D. A. (2004). A spatial filtering specification for the autologistic model. *Environment and Planning A*, *36*(10), 1791-1811.

Griffith, D. A. (2005). A comparison of six analytical disease mapping techniques as applied to West Nile Virus in the coterminous United States. *International Journal of Health Geographics*, *4*(1), 18.

Griffith, D. A. (2007). Spatial structure and spatial interaction: 25 years later. *The Review of Regional Studies*, *37*(1), 28.

Griffith, D. A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, *40*(11), 2751-2769.

Griffith, D. A., & Lagona, F. (1998). On the quality of likelihood-based estimators in spatial autoregressive models when the data dependence structure is misspecified. *Journal of Statistical Planning and Inference*, *69*(1), 153-174.

Griffith, D. A., & Peres-Neto, P. R. (2006). Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, *87*(10), 2603-2613.

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

Hahsler, M., Grün, B., & Hornik, K. (2007). Introduction to arules–mining association rules and frequent item sets. *SIGKDD Explor*, *2*(4), 1-28.

Hanowski, R. J., Hickman, J., Fumero, M. C., Olson, R. L., & Dingus, T. A. (2007). The sleep of commercial vehicle drivers under the 2003 hours-of-service regulations. *Accident Analysis & Prevention*, *39*(6), 1140-1145.

Hartley LR (2000). Murdoch University submission to the House of Representatives Inquiry into Managing Fatigue in Transport.

Hartley, L., & Arnold, P. K. (1995). *Subjective and Objective Measures of Fatigue when Driving*. Institute for Research in Safety and Transport, Division of Psychology, Murdoch University.

Hatfield, J., Murphy, S., Kasparian, N., & Job, R. (2005). Risk perceptions, attitudes and behaviors regarding driver fatigue in NSW youth: The Development of an Evidence-Based Driver Fatigue Educational Intervention Strategy. *Report to the Motor Accidents Authority of NSW*.

Haworth, N. (1998). Fatigue and fatigue research: The Australian experience. *Accident Research Centre, Monash University*.

Helbich, M., & Griffith, D. A. (2016). Spatially varying coefficient models in real estate: Eigenvector spatial filtering and alternative approaches. *Computers, Environment and Urban Systems*, *57*, 1-11.

Helbich, M., & Jokar Arsanjani, J. (2015). Spatial eigenvector filtering for spatiotemporal crime mapping and spatial crime analysis. *Cartography and Geographic Information Science*, *42*(2), 134-148.

Horne, J. A., & Reyner, L. A. (1995). Sleep related vehicle accidents. *Bmj*, *310*(6979), 565-567.

Horne, J., & Reyner, L. (1999). Vehicle accidents related to sleep: a review. *Occupational and environmental medicine*, *56*(5), 289-294.

Horne, J., & Reyner, L. (2001). Sleep-related vehicle accidents: some guides for road safety policies. *Transportation research part F: traffic psychology and behaviour*, *4*(1), 63-74.

Hosmer, D. W., & Lemesbow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in statistics-Theory and Methods*, *9*(10), 1043-1069.

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

Huang, H., Chin, H. C., & Haque, M. M. (2008). Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis. *Accident Analysis & Prevention*, *40*(1), 45-54.

Hutchinson, T. P. (1983). A bivariate normal model for intra-accident correlations of driver injury with application to the effect of mass ratio. *Accident Analysis & Prevention*, *15*(3), 215-224.

Hutchinson, T. P. (1986). Statistical modelling of injury severity, with special reference to driver and front seat passenger in single-vehicle crashes. *Accident Analysis & Prevention*, *18*(2), 157-167.

Johnson, G. D. (2004). Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modelling. *International Journal of Health Geographics*, *3*(1), 29.

Jones, A. P., & Jørgensen, S. H. (2003). The use of multilevel models for the prediction of road accident outcomes. *Accident Analysis & Prevention*, *35*(1), 59-69.

Kamei, Y., Monden, A., Morisaki, S., & Matsumoto, K. I. (2008). A hybrid faulty module prediction using association rule mining and logistic regression analysis. In *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*(pp. 279-281). ACM.

Karrer, K., Vöhringer-Kuhnt, T., Baumgarten, T., & Briest, S. (2004). The role of individual differences in driver fatigue prediction. In *Third International Conference on Traffic and Transport Psychology, Nottingham, UK.*

Kim, D. G., & Washington, S. (2006). The significance of endogeneity problems in crash models: an examination of left-turn lanes in intersection crash models. *Accident Analysis & Prevention*, *38*(6), 1094-1100.

Kim, D. G., Lee, Y., Washington, S., & Choi, K. (2007). Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis & Prevention*, *39*(1), 125-134.

Kim, K., Brunner, I., & Yamashita, E. (2006). Influence of land use, population, employment, and economic activity on accidents. *Transportation Research Record: Journal of the Transportation Research Board*, (1953), 56-64.

Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied linear regression models*. McGraw-Hill/Irwin.

Ladron de Guevara, F., Washington, S., & Oh, J. (2004). Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona. *Transportation Research Record: Journal of the Transportation Research Board*, (1897), 191-199.

Lee, C., & Abdel-Aty, M. (2008). Presence of passengers: does it increase or reduce driver's crash potential?. *Accident Analysis & Prevention*, *40*(5), 1703-1712.

Lenguerrand, E., Martin, J. L., & Laumon, B. (2006). Modelling the hierarchical structure of road crash data—Application to severity analysis. *Accident Analysis & Prevention*, *38*(1), 43-53.

LeSage, J. P. (1999). *The theory and practice of spatial econometrics.* University of Toledo. Toledo, Ohio, 28, 33.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics.* Chapman and Hall/CRC.

Li, D. H., Liu, Q., Yuan, W., & Liu, H. X. (2010). Relationship between fatigue driving and traffic accident. *Journal of traffic and transportation engineering (Xi'an, Shaanxi)*, *10*(2), 104-109.

Li, Z., Wang, W., Liu, P., Bigham, J. M., & Ragland, D. R. (2013). Using geographically weighted Poisson regression for county-level crash modeling in California. *Safety science*, *58*, 89-97.

Liu, C. C., Hosking, S. G., & Lenné, M. G. (2009). Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of safety research*, *40*(4), 239-245.

LOO, B. P., & Tsui, M. K. (2005). Temporal and spatial patterns of vehicle-pedestrian crashes in busy commercial and shopping areas: a case study of Hong Kong. *Asian Geographer*, *24*(1-2), 113-128.

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

MacNab, Y. C. (2004). Bayesian spatial and ecological models for small-area accident and injury analysis. *Accident Analysis & Prevention*, *36*(6), 1019-1028.

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

Marcus, C. L., & Loughlin, G. M. (1996). Effect of sleep deprivation on driving safety in housestaff. *Sleep*, *19*(10), 763-766.

Mavjee, V., & Home, J. A. (1994). Boredom effects on sleepiness/alertness in the early afternoon vs. early evening and interactions with warm ambient temperature. *British Journal of Psychology*, *85*(3), 317-333.

Maycock, G. (1996). Sleepiness and driving: the experience of UK car drivers. *Journal of Sleep Research*, *5*(4), 229-231.

McCartt, A. T., Ribner, S. A., Pack, A. I., & Hammer, M. C. (1996). The scope and nature of the drowsy driving problem in New York State. *Accident Analysis & Prevention*, *28*(4), 511-517.

McCartt, A. T., Rohrbaugh, J. W., Hammer, M. C., & Fuller, S. Z. (2000). Factors associated with falling asleep at the wheel among long-distance truck drivers. *Accident Analysis & Prevention*, *32*(4), 493-504.

McConnell, C. F., Bretz, K. M., & Dwyer, W. O. (2003). Falling asleep at the wheel: a close look at 1,269 fatal and serious injury-producing crashes. *Behavioral sleep medicine*, *1*(3), 171-183.

McKernon, S. (2009). *A literature review on driver fatigue among drivers in the general public*. Land Transport New Zealand.

McMillen, D. P. (2004). Geographically weighted regression: the analysis of spatially varying relationships.

Meng, F., Li, S., Cao, L., Li, M., Peng, Q., Wang, C., & Zhang, W. (2015). Driving fatigue in professional drivers: a survey of truck and taxi drivers. *Traffic injury prevention*, *16*(5), 474-483.

Milton, J. C., Shankar, V. N., & Mannering, F. L. (2008). Highway accident severities and the mixed logit model: an exploratory empirical analysis. *Accident Analysis & Prevention*, *40*(1), 260-266.

Mitler, M. M., Carskadon, M. A., Czeisier, C. A., Dement, W. C., Dinges, D. F., & Graeber, R. C. (1988). Catastrophes, sleep, and public policy: consensus report. *Sleep*, *11*(1), 100-109.

Mitra, S., & Washington, S. (2012). On the significance of omitted variables in intersection crash modeling. *Accident Analysis & Prevention*, *49*, 439-448.

Mock, C., Amegashie, J., & Darteh, K. (1999). Role of commercial drivers in motor vehicle related injuries in Ghana. *Injury Prevention*, *5*(4), 268-271.

Montella, A., Aria, M., D'Ambrosio, A., & Mauriello, F. (2012). Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. *Accident Analysis & Prevention*, *49*, 58-72.

Montella, A., Aria, M., D'Ambrosio, A., & Mauriello, F. (2011). Data-mining techniques for exploratory analysis of pedestrian crashes. *Transportation Research Record: Journal of the Transportation Research Board*, (2237), 107-116.

Morrow, P. C., & Crum, M. R. (2004). Antecedents of fatigue, close calls, and crashes among commercial motor-vehicle drivers. *Journal of Safety Research*, *35*(1), 59-69.

Murakami, D., & Griffith, D. A. (2015). Random effects specifications in eigenvector spatial filtering: a simulation study. *Journal of Geographical Systems*, *17*(4), 311-331.

Nantulya, V. M., & Muli-Musiime, F. (2001). Kenya: Uncovering the social determinants of road traffic accidents. *Challenging inequities in health: from ethics to action*, 211-25.

National Sleep Foundation (2008). *Sleep in America Poll - Sleep, Performance and Workplace*. Sleep Health: Journal of the National Sleep Foundation, 1 (2), e7.

National Sleep Foundation (2009). *Facts About Drowsy Driving*.

National Transportation Safety Board (1995). *Factors that affect fatigue in heavy truck accidents.* Safety Study 95/01. National Transportation Board, Washington, DC.

NCSCR/NHTSA Expert panel on driver fatigue and sleepiness (2001). *Drowsy driving and automobile crashes*.http://www.nhtsa.dot.gov/people/injury/drowsy_driving1/drowsy.html

Noland, R. B., & Quddus, M. A. (2004a). A spatially disaggregate analysis of road casualties in England. *Accident Analysis & Prevention*, *36*(6), 973-984.

Noland, R. B., & Quddus, M. A. (2004b). Improvements in medical care and technology and reductions in traffic-related fatalities in Great Britain. *Accident Analysis & Prevention*, *36*(1), 103-113.

Noland, R. B., & Quddus, M. A. (2005). Congestion and safety: A spatial analysis of London. *Transportation Research Part A: Policy and Practice*, *39*(7-9), 737-754.

Nordbakke, S., & Sagberg, F. (2007). Sleepy at the wheel: Knowledge, symptoms and behaviour among car drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *10*(1), 1-10.

Ogden, E. J., & Moskowitz, H. (2004). Effects of alcohol and other drugs on driver performance. *Traffic injury prevention*, *5*(3), 185-198.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis*, *27*(4), 286-306.

Oron-Gilad, T., Ronen, A., & Shinar, D. (2008). Alertness maintaining tasks (AMTs) while driving. *Accident Analysis & Prevention*, *40*(3), 851-860.

Ouyang, Y., Shankar, V., & Yamamoto, T. (2002). Modeling the simultaneity in injury causation in multivehicle collisions. *Transportation Research Record: Journal of the Transportation Research Board*, (1784), 143-152.

Owens, D. A., & Sivak, M. (1996). Differentiation of visibility and alcohol as contributors to twilight road fatalities. *Human Factors*, *38*(4), 680-689.

Pack, A. I., Pack, A. M., Rodgman, E., Cucchiara, A., Dinges, D. F., & Schwab, C. W. (1995). Characteristics of crashes attributed to the driver having fallen asleep. *Accident Analysis & Prevention*, *27*(6), 769-775.

Páez, A., & Scott, D. M. (2005). Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, *61*(1), 53-67.

Pande, A., & Abdel-Aty, M. (2009). Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. *Safety science*, *47*(1), 145-154.

Patuelli, R., Griffith, D. A., Tiefelsdorf, M., & Nijkamp, P. (2006). The use of spatial filtering techniques: the spatial and space-time structure of German unemployment data.

Permpoonwiwat, C. K., & Kotrajaras, P. (2012). Pooled time-series analysis on traffic fatalities in Thailand. *World Review of Business Research*, *2*(6), 170-182.

Philip, P. (2005). Sleepiness of occupational drivers. *Industrial health*, *43*(1), 30-33.

Philip, P., Sagaspe, P., Taillard, J., Moore, N., Guilleminault, C., Sanchez-Ortuno, M., ... & Bioulac, B. (2003). Fatigue, sleep restriction, and performance in automobile drivers: a controlled study in a natural environment. *Sleep*, *26*(3), 277-280.

Pilcher, J. J., & Huffcutt, A. I. (1996). Effects of sleep deprivation on performance: a meta-analysis. *Sleep*, *19*(4), 318-326.

Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., & Wets, G. (2013). Evaluating the road safety effects of a fuel cost increase measure by means of zonal crash prediction modeling. *Accident Analysis & Prevention*, *50*, 186-195.

Plainis, S., Murray, I. J., & Pallikaris, I. G. (2006). Road traffic casualties: understanding the night-time death toll. *Injury Prevention*, *12*(2), 125-138.

Pulugurtha, S. S., & Sambhara, V. R. (2011). Pedestrian crash estimation models for signalized intersections. *Accident Analysis & Prevention*, *43*(1), 439-446.

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

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

Radun, I., & Radun, J. E. (2009). Convicted of fatigued driving: Who, why and how?. *Accident Analysis & Prevention*, *41*(4), 869-875.

Radun, I., Ohisalo, J., Radun, J., Wahde, M., & Kecklund, G. (2013). Driver fatigue and the law from the perspective of police officers and prosecutors. *Transportation research part F: traffic psychology and behaviour*, *18*, 159-167.

Radun, I., Radun, J., & Ohisalo, J. (2009). Fatigued driving and the law. Drivers punished of fatigued driving. *Liikenneturvan Tutkimusmonisteita*, (107/2009).

Rajaratnam, S. M., & Arendt, J. (2001). Health in a 24-h society. *The Lancet*, *358*(9286), 999-1005.

Rana, T., Sikder, S., & Pinjari, A. (2010). Copula-based method for addressing endogeneity in models of severity of traffic crash injuries: application to two-vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board*, (2147), 75-87.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Robertson, R., Holmes, E., & Vanlaar, W. (2009). *The facts about fatigued driving in Ontario: A guidebook for police*.

Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors*, *12*(12), 16937-16953.

Shaharanee, I. N. M., Hadzic, F., & Dillon, T. S. (2009, December). Interestingness of association rules using symmetrical tau and logistic regression. In *Australasian Joint Conference on Artificial Intelligence* (pp. 422-431). Springer, Berlin, Heidelberg.

Guangdong Statistical Bureau (2017). *Guangdong statistical year book, Beijing: China Statistics Press.*

Stegeman, I., Wijkerslooth, T. R., Stoop, E. M., Leerdam, M., Ballegooijen, M., Kraaijenhagen, R. A., ... & Bossuyt, P. M. (2013). Risk factors for false positive and for false negative test results in screening with fecal occult blood testing. *International journal of cancer*, *133*(10), 2408-2414.

Stutts, J. C., Wilkins, J. W., & Vaughn, B. V. (1999). Why do people have drowsy driving crashes. *Input from drivers who just did*, (202/638), 5944.

Stutts, J. C., Wilkins, J. W., Osberg, J. S., & Vaughn, B. V. (2003). Driver risk factors for sleep-related crashes. *Accident Analysis & Prevention*, *35*(3), 321-331.

Sullivan, J. M., & Flannagan, M. J. (2002). The role of ambient light level in fatal crashes: inferences from daylight saving time transitions. *Accident Analysis & Prevention*, *34*(4), 487-498.

Summala, H., & Mikkola, T. (1994). Fatal accidents among car and truck drivers: effects of fatigue, age, and alcohol consumption. *Human factors*, *36*(2), 315-326.

Tefft, B. C. (2010). Asleep at the wheel: The prevalence and impact of drowsy driving.

The Ministry of Transport, New Zealand (2017). *Fatigue*.

Thiffault, P., & Bergeron, J. (2003). Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention*, *35*(3), 381-391.

Tiefelsdorf, M., & Boots, B. (1995). The exact distribution of Moran's I. *Environment and Planning A*, *27*(6), 985-999.

Tiefelsdorf, M., & Griffith, D. A. (2007). Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A*, *39*(5), 1193-1221.

Traffic Management Bureau, Ministry of Public Security, PRC (2008). *China road traffic accidents annual statistical report*.

Traffic Management Bureau, Ministry of Public Security, PRC (2012). *China road traffic accidents annual statistical report*.

Traffic Management Bureau, Ministry of Public Security, PRC (2013). *China Road Traffic Accidents Annual Statistical Report.*

Truong, L. T., Kieu, L. M., & Vu, T. A. (2016). Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam. *Accident Analysis & Prevention*, *94*, 153-161.

Tsai, P. J. (2012). Application of Moran’s test with an empirical Bayesian rate to leading health care problems in Taiwan in a 7-year period (2002–2008). *Global journal of health science*, *4*(5), 63.

Tsai, P. J., Lin, M. L., Chu, C. M., & Perng, C. H. (2009). Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health*, *9*(1), 464.

Vanlaar, W., Simpson, H., Mayhew, D., & Robertson, R. (2008). Fatigued and drowsy driving: A survey of attitudes, opinions and behaviors. *Journal of Safety Research*, *39*(3), 303-309.

Venkataraman, N., Ulfarsson, G. F., & Shankar, V. N. (2013). Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type. *Accident Analysis & Prevention*, *59*, 309-318.

Wakefield, J. (2003). Sensitivity analyses for ecological regression. *Biometrics*, *59*(1), 9-17.

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

Wang, Y., Kockelman, K. M., & Wang, X. C. (2013). Understanding spatial filtering for analysis of land use-transport data. *Journal of Transport Geography*, *31*, 123-131.

Weng, J., Zhu, J. Z., Yan, X., & Liu, Z. (2016). Investigation of work zone crash casualty patterns using association rules. *Accident Analysis & Prevention*, *92*, 43-52.

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

Williamson, A., Lombardi, D. A., Folkard, S., Stutts, J., Courtney, T. K., & Connor, J. L. (2011). The link between fatigue and safety. *Accident Analysis & Prevention*, *43*(2), 498-515.

World Health Organization (2015). *Global status report on road safety 2015*. Geneva: World Health Organization, Department of Violence and Injury Prevention and Disability.

Wu, Z., Sharma, A., Mannering, F. L., & Wang, S. (2013). Safety impacts of signal-warning flashers and speed control at high-speed signalized intersections. *Accident Analysis & Prevention*, *54*, 90-98.

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

Xu, P., Huang, H., Dong, N., & Abdel-Aty, M. (2014). Sensitivity analysis in the context of regional safety modeling: identifying and assessing the modifiable areal unit problem. *Accident Analysis & Prevention*, *70*, 110-120.

Yamamoto, T., & Shankar, V. N. (2004). Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects. *Accident Analysis & Prevention*, *36*(5), 869-876.

Yan, Q.-H., Ma, W.J., & Xu, Y.J. (2010). *Incidence of unsafe driving behaviors and its related risk factors*. Chinese Journal of Public Health, 8, 033.

Ye, X., Pendyala, R. M., Al-Rukaibi, F. S., & Konduri, K. (2008). *Joint model of accident type and severity for two-vehicle crashes* (No. 08-2323).

Young, S. L., & Hashemi, L. (1996). Fatigue and trucking accidents: two modes of accident causation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 40, No. 18, pp. 952-956). Sage CA: Los Angeles, CA: SAGE Publications.

Zhang, G., Yau, K. K., Zhang, X., & Li, Y. (2016). Traffic accidents involving fatigue driving and their extent of casualties. *Accident Analysis & Prevention*, 87, 34-42.

Zhao, X., & Rong, J. (2013). The relationship between driver fatigue and monotonous road environment. In *Computational Intelligence for Traffic and Mobility* (pp. 19-36). Atlantis Press, Paris.