REAR-END CRASH MODELS ON THAI HIGHWAY

Thanapong Champahom

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Civil, Transportation and Geo-resources Engineering

Suranaree University of Technology

Academic Year 2019

แบบจำลองอุบัติเหตุชนท้าย บนถนนหลวงประเทศไทย

นายธนพงษ์ จำปาหอม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมโยธา ขนส่ง และทรัพยากรธรณี
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2562

REAR-END CRASH MODELS ON THAI HIGHWAY

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

____

(Assoc. Prof. Dr. Thaned Satiennam)

Chairperson

____

(Prof. Dr. Vatanavongs Ratanavaraha)

Member (Thesis Advisor)

____

(Assoc. Prof. Dr. Wichuda Satiennam)

Member

____

(Asst. Prof. Dr. Buratin Khampirat)

Member

____

(Asst. Prof. Dr. Sajjakaj Jomnonkwao)

Member

____

(Assoc. Prof. Flt. Lt. Dr. Kontorn Chamniprasart)

Vice Rector for Academic Affairs

and Internationalization

____

(Assoc. Prof. Dr. Pornsiri Jongkol)

Dean of Institute of Engineering

ธนพงษ์ จำปาหอม : แบบจำลองอุบัติเหตุชนท้าย บนถนนหลวงประเทศไทย (REAR-END CRASH MODELS ON THAI HIGHWAY) อาจารย์ที่ปรึกษา : ศาสตราจารย์ ดร.วัฒนวงศ์ รัตนวราห, 200 หน้า.

อุบัติเหตุทางถนนของประเทศไทยเป็นปัญหามาอย่างยาวนาน ในปัจจุบันทั้งภาครัฐ และเอกสารได้มีการมุ่งเน้นรณรงค์ และหาแนวทางแก้ไขอย่างสม่ำเสมอ เมื่อพิจารณาสถิติการเสียชีวิตจากอุบัติเหตุบนถนนหลวงพบว่า การชนท้ายเป็นรูปแบบการชนที่มีผู้เสียชีวิตสูงเป็นอันดับหนึ่ง ดังนั้นการมุ่งเน้นที่จะหาแนวทางในจำนวนอุบัติเหตุทางถนนรูปแบบชนท้าย และลดอัตราการเสียชีวิตจากการชนท้าย จึงเป็นสิ่งจำเป็นอย่างยิ่ง จากการทบทวนการงานวิจัยพบว่า มิติที่สำคัญที่เกี่ยวข้องกับการชนท้าย ซึ่งหากทำการศึกษาจะสามารถลดจำนวนและลดอัตราการเสียชีวิตอุบัติเหตุชนท้ายได้ ผลการทบทวนพบว่ามี 6 มิติที่สำคัญ จึงแสดงการศึกษาออกเป็น 6 การศึกษา ได้แก่

การศึกษาที่ 1: การสร้างแบบจำลองเพื่อวิเคราะห์หาปัจจัยที่ส่งผลต่อขนาดของการชนท้าย (Rear-end crash size) ซึ่งถูกชี้วัดด้วยจำนวนผู้บาดเจ็บ และจำนวนผู้เสียชีวิต อยู่ในรูปแบบของตัวแปรแฝง (Latent variables) และใช้แบบจำลองสมการเชิงโครงสร้างในการวิเคราะห์ความสัมพันธ์ดังกล่าว (Structural Equation Modeling)

การศึกษาที่ 2: การชนท้ายบริเวณทางแยกมีโอกาสเกิดขึ้นสูง เนื่องจากยานพาหนะต้องลดความเร็วเพื่อเข้าสู่ทางแยก ดังนั้นงานวิจัยนี้จึงมุ่งเน้นเพื่อหาแนวทางการลดจำนวนการชนท้ายบนทางแยกลง โดยผ่านการแนะนำแนวทางให้หน่วยงานที่เกี่ยวข้อง ปรับปรุงลักษณะทางกายภาพถนนรวมไปถึงแนวทางการรณรงค์ ข้อมูลที่ใช้วิเคราะห์เป็นกรณีอุบัติเหตุที่เกิดขึ้นบนถนนหลวงตั้งแต่ ปี 2011-2015 และถูกวิเคราะห์ classification and regression tree (CRT) โดยกำหนดตัวแปร target เป็น การชนทางบนทางแยก/การชนทางนอกทางแยก

การศึกษาที่ 3: ได้ประยุกต์ใช้ วิธีอธิบายการชักจูงเสมือน (Quasi-Induced Exposure) ซึ่งเป็นแนวทางหนึ่งในการศึกษาอุบัติเหตุชนท้ายเพื่อที่จะมุ่งเน้นในการลดจำนวนอุบัติเหตุชนท้ายและความรุนแรงลงได้ โดยผลลัพธ์ของวัตถุประสงค์นี้ สามารถนำไปใช้เป็นแนวทางในการฝึกฝนผู้ขับขี่ให้ตระหนักถึงความรุนแรงของอุบัติเหตุชนท้ายเพิ่มมากขึ้น

การศึกษาที่ 4:ความแตกต่างระหว่างถนนในเขตเมืองและนอกเมือง เนื่องจากคุณลักษณะถนนรวมไปถึงลักษณะของกระแสจราจร ระหว่างถนนในเขตเมืองและนอกเมืองมีความแตกต่างกัน การเกิดอุบัติเหตุชนท้ายย่อมมีความแตกต่างกันตามไปด้วย วัตถุประสงค์นี้มุ่งเน้นที่จะศึกษาหาความแตกต่างระหว่างอุบัติเหตุชนท้ายที่เกิดขึ้นบนถนนเขตเมืองและนอกเมืองเมื่อระบุนโยบายลดความรุนแรงที่แตกต่างกัน การศึกษานี้ได้ประยุกต์ใช้ Measurement of Invariance เพื่อเปรียบเทียบ

ความแตกต่างระหว่างการชนท้ายในเขตเมืองและเขตนอกเมือง ผลการศึกษาพบว่าทั้งสองแบบจำลองมีความแตกต่างกัน โดยเฉพาะอย่างยิ่งปัจจัยด้าน crash type และ vehicle involvement

การศึกษาที่ 5: การชนท้ายเป็นประเภทหนึ่งของ road accident ซึ่งได้มีการศึกษามาแล้วมากมาย ปัจจัยหนึ่งที่ค่อนข้างส่งผลต่อโอกาสการเสียชีวิตจากการชนท้ายคือพื้นที่ของถนน ณ จุดเกิดเหตุ ซึ่งได้ถูกจำแนกเป็นถนนในเมือง และนอกเมือง โดยมีความแตกต่างกันอย่างเห็นได้ชัด เช่น ความเร็ว จำนวนของทางแยก ประเภทรถ เป็นต้น อย่างไรก็ตามยังไม่เคยมีการศึกษาใดเปรียบเทียบการชนท้ายที่เกิดขึ้นระหว่างในเมืองและนอกเมืองมาก่อน ดังนั้น การศึกษานี้จึงมุ่งเน้นที่จะเปรียบเทียบปัจจัยที่ส่งผลต่อโอกาสการเสียชีวิตของการชนท้าย ที่แตกต่างกันระหว่าง 2 roadways. ด้วยแนวคิดบนพื้นฐานเชิงพื้นที่ จึงได้ประยุกต์เอา hierarchical logistic models มาใช้ โดยกำหนดให้ การประมาณค่าพารามิเตอร์แปรเปลี่ยนตาม road segment. เพิ่มเติม ได้เปรียบเทียบแบบจำลองที่มี coefficient with multilevel correlation และ coefficient without multilevel correlation ดังนั้น จึงมีทั้งหมด 4 แบบจำลอง ข้อมูลที่ใช้ในการศึกษาเป็นการชนท้ายที่เกิดขึ้นบนถนนหลวงในประเทศไทย ตั้งแต่ ปี 2011 - 2015 ผลการศึกษาพบว่าทิศทางของค่าพารามิเตอร์ของแบบจำลองในเมืองและนอกเมืองเป็นไปในทิศทางเดียวกัน

การศึกษาที่ 6: ในปัจจุบันการประยุกต์ใช้แบบจำลองทางสถิติขั้นสูงเพื่อทำนายความถี่ของจำนวนอุบัติเหตุได้ถูกนำมาใช้มากยิ่งขึ้น ซึ่งสามารถทำให้คาดการณ์แบบจำลองได้แม่นยำมากยิ่งขึ้น การศึกษานี้มุ่งเน้นที่จะเติมเต็มการประยุกต์เอาแบบจำลองทางสถิติหาความสัมพันธ์ระหว่าง explanatory variable และความถี่ในการชนท้าย. ข้อมูลที่ใช้ในการศึกษาเป็นการชนท้ายที่เกิดขึ้นบนถนนหลวงทั่วประเทศไทย ในปี 2011-2018 และได้กระจายลงตาม segment ที่มีลักษณะทางกายภาพถนนเหมือนกัน และ spatial correlation ให้แปรผลตามเขตรับผิดชอบของหน่วยงานกรมทางหลวง สำหรับการพัฒนาแบบจำลองมี 4 แบบจำลอง เริ่มจาก Poisson regression model, Negative binomial model, Zero-inflated negative binomial model และ Spatial zero-inflated negative binomial model (SZINB). ผลการเปรียบด้วยด้วย AIC พบว่า SIZNB มีค่าต่ำที่สุดแสดงให้เห็นว่าแบบจำลองนี้เหมาะสมกับข้อมูลมากที่สุด.

สาขาวิชา วิศวกรรมขนส่ง ลายมือชื่อนักศึกษา Thanapong Ch

ปีการศึกษา 2562 ลายมือชื่ออาจารย์ที่ปรึกษา____________

THANAPONG CHAMPAHOM : REAR-END CRASH MODELS ON THAI HIGHWAY. THESIS ADVISOR : PROF. VATANAVONGS RATANAVARAHA, Ph.D., 200 PP.

REAR-END CRASH/THAI HIGHWAY/CRASH SEVERITY/CRASH FREQUENCY/COUNT MODEL

The objective to find a way for reducing the number of road accidents and the death rate caused by rear-end collisions is absolutely necessary. According to the literature review, it was found that the important dimensions related to rear-end collisions which, if studied, will be able to reduce the number and death rate from rear-end accidents. The review results showed that there are 6 important dimensions. This study is therefore divided the studies into 6 dimensions which can be summarized as follows:

Study 1: Structural equation modeling (SEM) was used to be the tool for analyzing the factors affecting the injuries in the rear-end collision. After the acknowledgment of those factors, the involved organizations should play an important role in the road design and maintenance as well as the driver's training. The obtained results can be taken to reduce the severity of injuries.

Study 2: The rear-end is in accordance with many researches pointing out that the collisions are likely to highly occur at the intersection. Therefore, this research focuses on seeking for the ways decreasing the number of rear-end collisions at the intersection through the guidelines for relevant organizations to improve such as physical features of roads as well as the promotion methods.

Study 3: (1) a model which indicates the causes of rear-end crashes by applying Quasi-Induced Exposure to at-fault driver characteristics; (2) a determined model which studies fatal crashes. Results. Predictor variables in the model of at-fault and not-at-fault drivers found that driver age is most significant, for the mode of fatality, the use of safety equipment was found to be of most importance.

Study 4: Due to the various characteristics of urban and rural areas which possibly result in the different severity of rear-end collisions. Therefore, this study has focused on the comparison of models showing the severity of rear-end collisions between urban and rural areas.

Study 5: This study focuses on comparing different factors affecting the likelihood of rear-end crash fatality between rural and urban roadways. The significant variables in both models are the factors of traffic lane number, driver's seat belt use, and the incident time. In conclusion, this study can help fulfill another perspective of rear-end crashes encouraging policy makers to apply for safety policy decisions

The objective 6: Rear-end crash is a type of road traffic accidents that often occur. There were 4 models starting with Poisson regression model, Negative binomial model, Zero-inflated negative binomial model and spatial zero-inflated negative binomial model (SZINB). The model results found that SIZNB was the model that suit data most.

School of Transportation Engineering Student's Signature Thanapong

Academic Year 2019 Advisor's Signature

ACKNOWLEDGEMENTS

The author would like to pay great respects to persons, groups of people who give good advice and help me both in academic and research work as mentioned illustrations:

First and foremost, I would like to thank my thesis advisor, Professor Dr. Vatanavongs Ratanavaraha for his suggestions in every step of research procedure and helping me to coordinate for road accident data requirement.

Asst. Prof. Dr. Sajjakaj Jomnonkwao, who gives recommendations about data mining statistics and manuscript revision. Ms. Wanpen Suebsai, Secretary of Transportation Engineering, who helps coordinate various documentaries during the study. Suranaree University of Technology which supports the scholarship of Doctoral degree. I would like to thank Bureau of Highway Safety, Department of Highway, Ministry of Transport who supported data collection of road traffic accidents.

Finally, I would like to express great thanks to my parents who give cultivate with love and well support education until I have continuously achieved success in my life.

Thanapong Champahom

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF TABLES

Table **Page**

LIST OF TABLES (Continued)

Table **Page**

LIST OF FIGURES

Figure **Page**

SYMBOLS AND ABBREVIATIONS

α	=	Statistically significant level
β	=	Structural coefficient
λ	=	Factor loading coefficient
χ^2	=	Chi-square
df	=	Degree of freedom
RMSEA	=	Root mean square of approximation
SRMR	=	Standardized root mean residual
CFI	=	Comparative fit index
TLI	=	Tucker Lewis Index
SEM	=	Structural equation modeling
CFA	=	Confirmatory factor analysis
EFA	=	Exploratory factor analysis
CR	=	Composite reliability
AVE	=	Average variance extracted
NB	=	Negative Binomial Regression Model
ZI	=	Zero-Inflated model
ZINB	=	Zero-Inflated Negative Binomial Regression Model
SZINB	=	Spatial Zero-Inflated Negative Binomial Regression Model
LL	=	Log-Likelihood
MI	=	Measurement Invariance

CHAPER I

INTRODUCTION

1.1 Rationale for the research

1.1.1 Excursion

In 2014, from more than 180 countries worldwide, the number of deceases caused by road accidents was about 1.25 million people in each year. The group of deprived countries broke a record of highest deceases (WHO, 2015). In Thailand, the death rate from road accidents was 36.2 over 100,000 people ranking as the fourteenth of the world (WHO, 2015) and the deceases from crashed car accidents over 100,000 people as the second of the world ranked lower than Libya.

Thailand is in the phase of agricultural, commercial, and industrial expansion. The government has supported road transportation to be more comfortable, faster, and safer. This leads to people's increasing uses of personal cars which are one of the factors causing more road accidents affecting a great deal of loss to both government and private sectors (Office of transport and traffic policy and planing, 2014).

The tendency of the number of accidents reported by royal polices. In 2006, the number of accidents was 110, 685 cases causing 12,691 deceases. It had a tendency to decrease until 2013 and continuously increased until 2016 (DOH, 2014). According to all accidents reported from Royal Police in 2015, the responsibilities of Department of Highways for accidents were 20 percent, the proportion of casualty was 66 percent, the proportion of deceases was 34 percent. Accident cost reported from Royal Police in 2015, approximated 219, 233 million baht while the responsibility of Department of Highways was 42,899 million baht as 20 percent estimated from expenditure classified by the aspect of passengers and drivers' injury including deceases, disabilities, serious injuries, and slight injuries (DOH, 2016). Thus, highway road (In control of DOH) has to provide the ways for reducing rate of severity.

According to the number of accidents on Highways classified by types of crashed cars in 2015, from the total amount of 13,575 cases, 4,041 accidents distinguished as rear-end crashes were "crashes on the road in the same direction" about 30 percent inferior to "accidents outside the street on the straight road". When considering the number of deceases of "crashes on the road in the same direction", the highest number was 484 fatal accident cases and the number of casualty ranking second on the list was 2,430 cases (DOH, 2016). The rear-end crash accident is a type of road accidents which occur most often and it has higher proportion than other types of accidents. The type of rear-end crash occurrence is the aspect that the following vehicle crashing into the back of the leading vehicle. The disputants may be one or more. The aspect of traffic when the accidents often occur is heavy and the vehicles continuously run at high speed and in constricted space. Most of the rear-end crashes are not too grave. The severity of accidents, caused by the crashing vehicles' high speed as well as their size and weight, may lead to the fatality. The causes of rear-end crash accidents include abrupt change in front of other cars and too close tracking behind, immediate change of traffic lanes or overtaking, the readiness and driver's skill in driving. Factors causing rear-end crashed cars were1) ducking, the way small vehicles crash the rear-end of big and huge vehicles, which may cause drivers' fatality2) carrying things protruding out of the rear of vehicles which may cause the rear-end crash accidents if the chauffeur driving behind cannot notice them, 3) broken cars parking and obstructing traffic lanes without signaling co-road users 4) other factors such as driving on right lane at low speed, driving at night time (Ministry of Interior Department of Disaster Prevention and Mitigation, 2014). Lerdworawinich (2000) has studied the ways for reduced risk and severity of rear-end collision on Thai highway. He experimented installation of Tailgating Treatment on the pavement for providing drivers know the distance between their vehicle and leading vehicle. After experiments, He found *Tailgating Treatment* can reduce the risk of rear-end crashes. Pawinee Iamtrakul et. al. (2008) Iamtrakul (2008) have studied risk factors, causing the rear-end crash accidents in Phra Nakhon Si Ayutthaya Province, analyzed by building questionnaire and collecting data from the case studies of the factual rear-end crash accidents which were classified into serious and unserious cases.

1.1.2 Factor affecting to rear-end crash size

Chen et al. (2015) studied levels of driver injuries resulting from rear-end collisions, Contribution factors included driver behavior factors (e.g., age, gender), vehicle factors (e.g., vehicle type); road physical features (e.g., road function, pavement); and environmental factors (e.g., light conditions, weather conditions) (Chen et al., 2016) .Das and Abdel-Aty (2011) studied frequency of rear-end collisions and levels of injuries on main roads in urban cities. In the injury levels model, they found that high vehicle speeds resulted in greater severity of injuries. For road surfaces with a high friction coefficient, traffic islands could decrease severity of injuries. Sullivan and Flannagan (2003) studied fatalities resulting from rear-end collisions by comparing crashes that occurred both in lighted and unlighted conditions, finding that collisions that occurred without light had two times more fatalities than those in lighted conditions (Abdel Aty & Abdelwahab, 2004). Qi et al. (2013) studied injury levels in rear-end collisions at work zones; finding that nighttime rear-end collisions increased the level of injuries. (Piccinini et al., 2017); Wiacek et al. (2015) found that rear-end collisions by heavy vehicles increased chances of fatalities. Mohamed et al. (2017) found that rural roads and violation of determined speed limits resulted in more severe rear-end collisions.

1.1.3 Quasi-Induced Exposure Method

Quasi-Induced Exposure Methods (Carr B.R., 1970) have been widely used in the field of traffic accident research. The principle of these methods is to predict the at-fault driver based on the accident report (Chandraratna & Stamatiadis, 2009; Taha & Vinayak, 2013) by supposing that the distribution of not-at-fault drivers closely represents the distribution of exposure to accident hazards (X. Yan & Radwan, 2006; X. Yan et al., 2005).

1.1.4 Differential between urban and rural roads

Several differences can be noted when considering the severity of rear-end crashes between urban and rural areas from various perspectives. For example, the number of intersections results in a decrease in a front car‘s speed upon reaching signalized and unsignalized intersections. This scenario increases the chance of rear-end crashes, but severity may differ from that on rural roads (Islam, 2016). Chatterjee and Davis (2016) aimed to prevent shock waves from forming on freeways where rear-end collisions tend to occur. The speed of urban vehicles is typically low due to the relatively dense traffic, whereas rear crashes occurring in rural zones or on roads that connect the districts and provinces may be more severe because most vehicles on rural roads use high speed (David&Santosh, 2015). In terms of vehicle types that potentially access roads on rural and urban zones as well. For example, large trucks are allotted a limited time for road accessibility in urban areas. Vehicle types as classified by size also affect the severity of rear crashes in these areas. In term of attitude of driver, Zabihi et al. (2019) studied seat belt usage among adult drivers on urban and rural roads.

1.1.5 Modeling the hierarchical structure of road crash data

This section describes the application of the logistic model to the predictive analysis of fatal rear-end crashes caused by the effect of relevant variables. An additional concept for the selection of variables to be incorporated into the model pertains to road accidents classified into more than one level. In other words, explanation variables that affect injury levels should have hierarchical structures. For instance, based on researchers' viewpoints, accident cases should be assigned personal factors that affect the first injury level along with the second level of the physical features of the road where the accident occurred: straight or curved roads, intersection characteristics, (Dupont et al., 2013) or area characteristics such as sub-districts, districts, provinces, etc.

1.1.6 Study of rear-end crash frequency model

A study of Xuedong Yan and Radwan (2009) who studied rear-end collisions with trucks' presence. Meng and Weng (2011) have examined risks of rear-end collisions in work zones, Apart from risks of accidents in work zones, studies of rear-end collisions at crossroads were also conducted (Chu et al., 2015; Cunto &Saccomanno, 2009; Shahi et al., 2009; Wang et al., 2003). Meng and Qu (2012) compared crossroads with and without countdown traffic lights (Ni&Li, 2014). Burdett et al. (2016) analyzed rear-end crash at roundabout approach. Wan et al.

(2013) studied rear-end and lane-changing collisions through car-following behavior, Studies of other types of rear-end collisions included effectiveness of low-speed autonomous emergency braking in rear-end collisions (Fildes et al., 2015) and the proportion of low-speed leading cars affecting rear-end collisions (Nishimura et al., 2015).

1.2 Purpose of the research

- To study the factors affecting the rear-end crash severity
- To study the characteristics of rear-end crash on the intersections
- To study the factors potentially affecting the chance of becoming at fault-driver in the rear- end accidents and the complicated relationships of the independent variables resulting in the likelihood of fatality caused by rear-end collisions
- To study the characteristics affecting the accident severity when the accident areas are different (Urban areas and rural areas)
- To study the factors affecting the difference of rear-end injury levels (Accident and road levels)
- To select an advanced statistical model suitable for predicting the frequency of about-to-happen rear-end accidents and find appropriate ways to reduce the number of rear-end accidents

1.3 Scope of the research

- Study only the rear-end accidents occurring under the responsibility of Department of Highways
- Study the rear-end crash severity
- Study Rear-end crash on intersections
- Study spatial rear-end crash
- Study the appropriate statistical model for predicting the accident number

1.4 Research questions

- What factors affect the severity of rear-end accidents occurring on the highways?
- According to the data of rear-end collisions on the intersections, what are factors causing the rear-end crashes?
- What are the driver characteristics or environmental factors resulting in at-fault-driving in rear-end crash accident?
- For the severity of the rear-end accidents occurring on urban roads and rural roads, are there any differences? Which factor causes the rear end crash to be more or less severe? and is there any difference between the characteristics of both models?
- When considering the two factors including accident level factor and the road level factor, which factor causes higher injury severity level?
- Which count data model is the most appropriate for predicting the number of rear-end collisions occurring on the road segment?

1.5 Contribution of the research

- The model results can be taken to apply to be guidelines for reducing the rear-end crash severity.
- The discovery of factors, affecting the rear-end crashes on the intersection leads to the ways to reduce the accident number at the intersection
- The concepts can be taken to conduct a specific campaign for risk groups causing rear-end accidents, and reduce the death chance caused by them.
- The policy is created to reduce the rear-end accident severity for urban and rural roads.
- The policies are established and road physical characteristics are improved to reduce rear-end severity both at personal level and spatial levels.
- The newly suitable models are acquired for predicting the rear-end accident number.

1.6 Organization of the research

This research has studied the rear-end crash accidents occurring on Thai highways by analyzing the accident data in a hierarchical structure of which the first level is the study of accident level and the road level or spatial level. The components totally comprise 8 chapters including the following details;

Chapter I: Research principles and rationale. This part mentions the background, the importance of each research section, research objectives, scope of research, research questions, and contribution to this research.

Chapter II: The analysis of factors that affect the driver injury level caused by rear-end accidents on the highways by using the structural equation model to find various factors in the structural model with the concept that the rear-end crash severity can be a group factor, considered as Latent variables.

Chapter III: for the chance analysis of rear-end rashes on intersection, this part uses Decision tree for analysis because of the large database

Chapter IV: the decision tree model is applied to find the complicated relationship of the independent variables resulting in the chances of being at-fault driver in the rear end crashes. In addition, the analysis was conducted to find the factors affecting the likelihood of rear-end crash fatality

Chapter V: the rear-end crash severity was analyzed using Structural equation modeling by comparing the different types of factors affecting the crash severity between urban and rural roads.

Chapter VI: The analysis of factors affecting injury levels by using Hierarchical structure or Multi-level modeling Which is a logit model in order to get the probability resulting from various factors affecting the fatal rear-end accidents by considering the accident details as the first level analysis and spatial data as the second level.

Chapter VII: Appropriate statistical methods for predicting the accident number are studied by establishing the model consisting of Poison regression, Negative binomial regression, Zero-Inflated Negative binomial regression and Spatial Zero-Inflated Negative binomial regression (SZINB). SZINB is the application based on random effect to find the relationship of rear-end crashes within the same area.

Chapter VIII: A summary of the analysis of all 6 studies (sections 2 - 7)

1.7 References

Abdel Aty, M., and Abdelwahab, H. (2004). Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. **Accident Analysis & Prevention**, 36(3): 447-456.

Burdett, B., Alsghan, I., Chiu, L.-H., Bill, A. R., and Noyce, D. A. (2016). Analysis of Rear-End Collisions at Roundabout Approaches. **Transportation Research Record: Journal of the Transportation Research Board**, 2585(1): 29-38.

Carr B.R. (1970). *A Statistical Analysis of Rural Ontario Traffic Accidents Using Induced Exposure Data*. Paper presented at the Proc. Sysposium in the Use of Statistical Methods in the Analysis of Road Accidents, ODCE, Patis, France.

Chandraratna, S., and Stamatiadis, N. (2009). Quasi-induced exposure method: evaluation of not-at-fault assumption. **Accident Analysis & Prevention**, 41(2): 308-313.

Chatterjee, I., and Davis, G. A. (2016). Analysis of Rear-End Events on Congested Freeways by Using Video-Recorded Shock Waves. **Transportation Research Record: Journal of the Transportation Research Board**, 2583(1): 110-118.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., and Guan, H. (2015). A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. **Accident Analysis & Prevention**, 80: 76-88.

Chen, C., Zhang, G., Yang, J., Milton, J. C., and Alcantara, A. D. (2016). An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naive Bayes (DTNB) hybrid classifier. **Accident Analysis & Prevention**, 90: 95-107.

Chu, A., Bhat, C., R. , Pendyala, R. M., and Garikapati, V. (2015). A Comprehensive mixed logit analysis od crash type conditional on a crash event. *Civil, Architectural, and Environmental Engineering.* The University of Texas at Austin. Retrieved from http://hdl.handle.net/2152/34140

Cunto, F., and Saccomanno, F. F. (2009). Simulated safety performance of rear-end and angled vehicle interactions at isolated intersections. **Canadian Journal of Civil Engineering**, 36(11): 1794-1803.

Das, A., and Abdel-Aty, M. A. (2011). A combined frequency–severity approach for the analysis of rear-end crashes on urban arterials. **Safety Science**, 49(8-9): 1156-1163.

David, B., and Santosh, T. (2015). *Investigation of Key Crash Types: Rear-end Crashes in Urban and Rural Environments* Retrieved from Austroads Ltd.:

DOH. (2014). *Thailand Traffic Accident on National Highways in* 2013. Retrieved from Bureau of Highway Safety Department of Highways Ministry of Transport,:

DOH. (2016). *Thailand Traffic Accident on National Highways in* 2015. Retrieved from Bureau of Highway Safety Department of Highways Ministry of Transport,:

Dupont, E., Papadimitriou, E., Martensen, H., and Yannis, G. (2013). Multilevel analysis in road safety research. **Accident Analysis and Prevention**, 60: 402-411.

Fildes, B., Keall, M., Bos, N., Lie, A., Page, Y., Pastor, C., . . . Tingvall, C. (2015). Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes. **Accident Analysis & Prevention**, 81: 24-29.

Iamtrakul, P. (2008). *Analysis of factors affecting on rear-end crashed accidents*. Retrieved from http://roadsafetythai.org/uploads/userfiles/ACC_51016.pdf

Islam, M. T. (2016). Investigating the Speed and Rear-End Collision Relationship at Urban Signalized Intersections. **Transportation Research Record: Journal of the Transportation Research Board**, 2601(1): 10-16.

Lerdworawinich, P. (2000). *Report No. Rd* 281- *Studying of risky reducing of Rear-end accident on Thai highway* Retrieved from Deportment of highway, Ministry of Thailand.: http://km.doh.go.th/doh/preview.php?tableName= TB_KNOWLEDGE&tableId=17161&fieldName_show=KNOWLEDGE_NAME&id_text=KNOWLEDGE_ID

Meng, Q., and Qu, X. (2012). Estimation of rear-end vehicle crash frequencies in urban road tunnels. **Accident Analysis & Prevention**, 48: 254-263.

Meng, Q., and Weng, J. (2011). Evaluation of rear-end crash risk at work zone using work zone traffic data. **Accident Analysis & Prevention**, 43(4): 1291-1300.

Ministry of Interior Department of Disaster Prevention and Mitigation. (2014). *Driving to reduce Rear-end crashes*. Retrieved from http://www.sadaocity .go.th /service/cate/29

Mohamed, S. A., Mohamed, K., and Husain, A. A.-H. (2017). *Investigating Factors Affecting the Occurrence and Severity of Rear-End Crashes*. Paper presented at the Transportation Research Procedia - WCTR 2016 Shanghai.

Ni, Y., and Li, K. (2014). Estimating Rear-End Accident Probabilities at Signalized Intersections: A Comparison Study of Intersections With and Without Green Signal Countdown Devices. **Traffic Injury Prevention**, 15(6): 583-590.

Nishimura, N., Simms, C. K., and Wood, D. P. (2015). Impact characteristics of a vehicle population in low speed front to rear collisions. **Accident Analysis & Prevention**, 79: 1-12.

Office of transport and traffic policy and planing. (2014). *Analyzes road accidents (*2557 *B.E.)*. Retrieved from www.otp.go.th/uploads/files/1471501633-dgyrs-3z11i.pdf

Pawinee Iamtrakul et. al. (2008). *Studying factor afficting to risky rear-end crash*. Retrieved from http://roadsafetythai.org/uploads/userfiles/ACC_51016.pdf

Piccinini, G. B., Engström, J., Bärgman, J., and Wang, X. (2017). Factors contributing to commercial vehicle rear-end conflicts in China: A study using on-board event data recorders. **Journal of Safety Research**, 62: 143-153.

Qi, Y., Srinivasan, R., Teng, H., and Baker, R. (2013). Analysis of the Frequency and Severity of Rear-End Crashes in Work Zones. **Traffic Injury Prevention**, 14(1): 61-72.

Shahi, J., Namaki Araghi, B., and Mehdizadeh, A. (2009). Estimation of rear-end potential conflict using Time-to-collision model. **Ingeniería y Desarrollo**(25): 67-79.

Sullivan, J. M., and Flannagan, M. J. (2003). *Risk of fatal rear-end collisions: is there more to it than attention?* Paper presented at the Proceeding of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Taha, H. R., and Vinayak, D. (2013). *Modelling Likelihood of At-Fault and Not-At-Fault Carshare Users*. Paper presented at the International Conference Road Safety and Simulation, Rome, Italy.

Wan, H., Liu, Z., and Wang, H. (2013). Rear-end and Lane-changing Collisions Avoidance Theoretical Models of the Multi-lane Freeway in Disaster Weather. **Procedia - Social and Behavioral Sciences**, 96: 70-81.

Wang, Y., Ieda, H., and Mannering, F. (2003). Estimating rear-end accident probabilities at signalized intersections: occurrence-mechanism approach. **Journal of Transportation engineering**, 129(4): 377-384.

WHO. (2015). *Global status report on road safety* 2015. Retrieved from http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/

Wiacek, C., Bean, J., and Sharma, D. (2015). *Real-world analysis of fatal rear-end crashes*. Paper presented at the 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration. https://trid.trb.org/view/1365430

Yan, X., and Radwan, E. (2006). Analyses of rear-end crashes based on classification tree models. **Traffic Injury Prevention**, 7(3): 276-282.

Yan, X., and Radwan, E. (2009). Analysis of truck-involved rear-end crashes using multinomial logistic regression. **Advances in Transportation Studies an international Journal**, 17.

Yan, X., Radwan, E., and Abdel-Aty, M. (2005). Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. **Accident Analysis & Prevention**, 37(6): 983-995.

Zabihi, F., Davoodi, S. R., and Nordfjaern, T. (2019). The role of perceived risk, reasons for non-seat belt use and demographic characteristics for seat belt use on urban and rural roads. **International Journal of Injury Control and Safety Promotion**, 26(4): 431-441.

CHAPTER II

ANALYSIS OF FACTORS AFFECTING REAR-END CRASH SEVERITYUSING STRUCTURAL EQUATION MODELING

2.1 Abstract

Road accidents regularly cause a high number of fatalities. Thailand's road accident fatality rate of 36.2 of every 100,000 people ranks the second highest in the world. Surprisingly, same-direction collisions comprise the highest proportion of crashes leading to fatality. To determine how best to minimize the number of fatalities and injuries, this research uses structural equation modeling (SEM) to examine factors affecting rear-end collisions' severity. According to SEM results, the driver factor had the greatest effect on collision severity, followed respectively by road and environmental factors. After assessing relevant factors, this study suggested that stakeholder organizations should play an important role in road design and maintenance and in driver training. The study also discussed driving and road policies in Thailand and other developed countries.

2.2 Introduction

In 2014, road accidents caused about 1.25 million deaths in over 180 countries worldwide, and developing countries broke a record for the highest number of deaths

(WHO, 2015). Thailand ranked second in the world, behind Libya, in the number of deaths caused by car crashes per 100,000 people.

At this writing, Thailand is undergoing agricultural, commercial, and industrial expansion, with the Thai government supporting improvements to make road transportation faster, safer, and more comfortable. Expansion has led to increased use of personal vehicles, which is one factor causing both road accidents and significant losses to the government and the private sector (Office of transport and traffic policy and planing, 2014).

Of 13,575 highway crash in 2015, as classified by types of crashed cars, 4,041 were distinguished as rear-end collisions, that is, "crashes on the road in the same direction," about30 percent less than "accidents off the street on a straight road." When considering the number of deaths due to "crashes on the road in the same direction," the highest number was 484 cases, with the number of casualties ranking second at 2,430 cases (Department of Highway Thailand, 2016). Area - end collision, in which a following vehicle crashes into the back of a leading vehicle, is the most frequent type of road accident. There may be one or more vehicles involved because when these accidents occur, traffic levels are often heavy, and vehicles move at high speeds in constricted spaces. Even so, most rear-end collisions are not serious. Their severity, as affected by speed and vehicles size and weight, affects resulting fatalities. Causes of rear-end collisions include abrupt changes in front of other cars following too closely behind, rapid changing between traffic lanes or passing, and drivers' awareness and skills. Lerdworawinich (2000) has studied ways of reducing risks and severity of rear-end collisions on Thai highways. He has experimented with installation of a tailgating treatment on roads to help drivers increase their awareness of the distance between their vehicle and the vehicle in front of them. Lerdworawinich (2000) found that the tailgating treatment can reduce the risk of rear-end collisions. Iamtrakul (2008) Studied risk factors causing rear-end collisions in Phra Nakhon Si Ayutthaya Province, using and analyzing questionnaires and collecting data from case studies of rear-end collisions classified into serious and non-serious cases.

In Thailand, aside from two studies on rear-end collisions, no studies have used historical statistics to build a model for analyzing factors affecting the number and severity of injuries. Because these factors cannot be directly measured, the severity of injuries was divided into three levels :minor injury, serious injury, and fatality. The structural equation model's (SEM's) ability to determine the relationship between latent variables that cannot be directly measured, such as "severity of accidents" is the "structural model (path analysis)," and latent variables measured by observed variables are "measurement models."

SEM has been applied to analyze a variety of accidents in other countries. The model was not designed, however, to find and predict factors on Thai highways affecting rear-end crash severity as indicated by numbers of deaths and serious and slight injuries. When these factors are determined, they can be used for road engineering design and driver training. According to model results in this research, a variety of variables have never been studied in any other research, for example, crash types, traffic quantity, truck percentage, and personal factors such as alcohol use, safety equipment use, and so on. This study contributes by using the model's results to propose policy that can reduce rear-end crash severity.

2.3 Literature review

This research follows global research trends that attend to road accidents as the most frequently occurring type of transportation accident. The rear-end collision is the type most frequently studied, with many studies having discussed factors that affect the probability of rear-end collisions, such as driver age. The study analysis is based on rear-end collisions' pre-crash conditions, which consider leading vehicles' speed. Types of rear-end collisions include "stopped in road," "decelerating speed," and "normal speed" (Ma&Yan, 2014). Comparisons between teen and adult rear-end collisions have also been undertaken (Seacrist et al., 2016). Rear-end crash potential has been assessed in roads' work zone merging areas (Weng et al., 2014). Joon-Ki et al. (2007) established a model for predicting the possibility of rear-end collisions on freeways (Pande&Abdel-Aty, 2008). Liang et al. (2010) Studied multi-agent and driver behavior in rear-end collision notices. Among four warning factors, they included driver repository (e.g., vehicle type), rear-end collision cases, an environment model, and a driving behavior model. These factors resemble those in a study by Xuedong Yan and Radwan (2009) who studied rear-end collisions with trucks' presence. Meng and Weng (2011) have examined risks of rear-end collisions in work zones, Apart from risks of accidents in work zones, studies of rear-end collisions at crossroads were also conducted (Chu et al., 2015; Cunto&Saccomanno, 2009; Shahi et al., 2009; Wang et al., 2003). Meng and Qu (2012) compared crossroads with and without countdown traffic lights (Ni&Li, 2014). Burdett et al. (2016) analyzed rear-end crash at roundabout approach. Wan et al. (2013) studied rear-end and lane-changing collisions through car-following behavior, Studies of other types of rear-end collisions included effectiveness of low-speed autonomous emergency braking in rear-end collisions (Fildes et al., 2015) and the proportion of low-speed leading cars affecting rear-end collisions (Nishimura et al., 2015).

Chen et al. (2015) studied levels of driver injuries resulting from rear-end collisions, Contribution factors included driver behavior factors (e.g., age, gender), vehicle factors (e.g., vehicle type); road physical features (e.g., road function, pavement); and environmental factors (e.g., light conditions, weather conditions) (Chen, Zhang, Yang, et al., 2016). Das and Abdel-Aty (2011) studied frequency of rear-end collisions and levels of injuries on main roads in urban cities. In the injury levels model, they found that high vehicle speeds resulted in greater severity of injuries. For road surfaces with a high friction coefficient, traffic islands could decrease severity of injuries. Sullivan and Flannagan (2003) studied fatalities resulting from rear-end collisions by comparing crashes that occurred both in lighted and unlighted conditions, finding that collisions that occurred without light had two times more fatalities than those in lighted conditions (Abdel Aty & Abdelwahab, 2004). Qi et al. (2013) studied injury levels in rear-end collisions at work zones, finding that nighttime rear-end collisions increased the level of injuries. (Piccinini et al., 2017); Wiacek et al. (2015) found that rear-end collisions by heavy vehicles increased chances of fatalities. Mohamed et al. (2017) found that rural roads and violation of determined speed limits resulted in more severe rear-end collisions.

Variables found and used in previous studies are illustrated in Table 2.1. New variables in this research consisted of two groups (Table 2.2) as follows:

Group1. Using SEM, as used in previous research collecting all crash types to study accident severity, this study focused only on rear-end collisions. Added variables were crash types, safety equipment use, and large truck proportion.

(Hamdar & Schorr, 2013; Hassan & Al-Faleh, 2013; Kim et al., 2011; Lee et al., 2008; Schorr & Hamdar, 2014)

Group2.The study of only rear-end collisions, especially injury severity levels they caused (e.g., (Georgi et al., 2009; Yuan et al., 2017)) found that no research has investigated rear-end crash severity by measuring it as a latent variable. New variables in this research included road maintenance, consideration of rear-end crash types affecting severity, and other variables, including rear-end crashes on straight roads with drivers' sight distance affected, higher speed on main roads than on parallel roads, sudden stops in intersections, and traffic quantity potentially affecting driving speed that reduced crash severity.

2.4 Discussing variables

For measuring rear-end collisions on Thai highways, the following indicators are used. Indicators of rear-end crash severity can be measured by injury at three levels: (i) number of deaths, referring to casualties who die on the road or in the hospital; (ii) serious injuries, meaning an injury that cannot heal in less than 3 weeks; (iii) slight injuries, meaning an injury that can heal in less than 3 weeks. For considering the effect of contributing factors for all injury levels, rear-end crash severity is set as a latent variable.

1) Driver factor indicators are as follow: (i) vehicle types and truck sizes that might increase collisions' numbers and injury levels and truck sizes related to speeding; (ii) drivers' ages divided into three groups (26–35, 36–45, and 46–55 years) (Ma & Yan, 2014) affecting drivers' healing, with younger drivers healing more quickly than older drivers; and (iii) driver genders when female drivers have lower

perception time than male drivers. Other factors included safety equipment use, exceeding speed limits, and order of vehicle involvement (Lee et al., 2008).

2) The road factor is divided into three categories: (i) divided highways with directions separated by a median to reduce accidents and make drivers pay more attention; (ii) work zone safety signs to make drivers reduce their vehicles' speed; (iii) road surfaces, for example, the variety of asphalt and concrete that could affect vehicles' speed (Das & Abdel-Aty, 2011). Other variables included rear-end crashes on straight roads, but with driver sight distance affected, higher speed on main roads than on parallel roads, sudden stops in intersection areas (Dong et al., 2016; Islam, 2016), and traffic quantity potentially affecting speeds that reduced accident severity.

3) Environmental factors were divided into three categories: (i) lighting conditions on road, which could affect the number of accidents (Qi et al., 2013); (ii) accident time, with drivers often increasing their speed in daylight because of the clear vision; (iii) weather affecting driving speed, and (iv) road surface conditions that might affect braking distance (X. Yan&Radwan, 2006).

4) In a rear-end collision case involving two vehicles, vehicular speed could be the important factor. Indicators of rear-end collision are as follow: (i) leading vehicle speed when struck from behind by other vehicles even though the leading vehicle is traveling at normal speed; and (ii) stopped vehicles hit from behind by other vehicles (Ma&Yan, 2014).

Table 2.1 Variables codes and descriptions

Code	Description	Value
Driver factors		
V1	Large vehicle size involvement (6 wheeled truck and larger)	1 =Yes, 0 =other
V2	Gender of driver	1 =Male, 0 =Female
V3	Age of driver from 26–35 Years	1 =Yes, 0 =other
V4	Age of driver from 36–45 Years	1 =Yes, 0 =other
V5	Age of driver from 46–55 Years	1 =Yes, 0 =other
V6	Driver used safety equipment (seat belt, helmet)	1 =Yes, 0 =other
V7	Drunk driver involved	1 =Yes, 0 =other
V8	Exceeding the speed limit	1 =Yes, 0 =other
V9	Order of vehicle involvement	Counts
Road factors		
V10	Per cent trucks	Continuous
V11	Traffic direction separated by road median (barrier, etc.)	1 =Yes, 0 =other
V12	The road was not being repaired	1 =Yes, 0 =other
V13	The road was asphalt or concrete pavement	1 =Yes, 0 =other
V14	Road horizontal alignment	1 =Straight, 0 =Curve
V15	Road graded	1 =slope, 0 =other
V16	Rear-end collision happened in interior lane	1 =Yes, 0 =other
V17	Rear-end collision happened at intersection	1 =Yes, 0 =other
V18	Log of AATD	Continuous
V19	Number of lanes	0 =Rather than 4 lanes, 1 = other

Table 2.1 Variables codes and descriptions (Continued)

Code	Description	Value
Environmental factors		
V20	Collision happened at night in low-light conditions	1 =Yes, 0 =other
V21	Visualization of drivers as accident	1 =Clean, 0 =other
V22	Time of collision	1 =Day, 0 =Night
V23	Status of road surface	1 =Wet, 0 = Dry
Rear-end factors		
V24	Leading vehicle was using normal and stable speed	1 =Yes, 0 =other
V25	Leading vehicle has stopped	1 =Yes, 0 =other
Crash size severity factors		
V26	Numbers of fatalities	Counts
V27	Number of persons seriously injured	Counts
V28	Number of persons slightly injured	Counts

Table 2.2 Gaps of Literatures

Work/variables	V1	V2	V3-V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25
This Study	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(Lee et al., 2008)*	✓	✓	✓	-	-	-	-	-	-	✓	✓	✓	✓	✓	-	-	-	✓	✓	✓	✓	✓	✓
(Kim et al., 2011)*	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
(Hamdar&Schorr, 2013)*	-	✓	✓	-	-	-	-	-	-	-	✓	✓	✓	✓	-	-	-	✓	✓	✓	-	-	-
(Hassan&Al-Faleh, 2013)*	✓	-	✓	-	✓	-	-	✓	-	-	-	✓	✓	✓	-	-	-	-	✓	✓	-	-	-
(Schorr&Hamdar, 2014)*	-	✓	✓	-	-	✓	-	✓	-	✓	-	✓	-	-	-	✓	-	-	-	✓	-	-	-
(Ma&Yan, 2014)**	-	✓	✓	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	✓
(Chen et al., 2015)**	✓	✓	✓	✓	✓	✓	-	-	-	-	✓	✓	✓	-	-	-	-	✓	✓	✓	✓	-	-
(Chen, Zhang, Yang, et al., 2016)**	✓	✓	✓	✓	✓	✓	✓	-	-	-	✓	✓	✓	-	-	-	-	✓	✓	✓	✓	-	-
(Das&Abdel-Aty, 2011)**	-	-	-	-	-	✓	-	-	✓	-	✓	✓	-	-	✓	✓	-	✓	✓	✓	✓	-	-

Table 2.2 Gaps of Literatures (Continued)

Work/variables	V1	V2	V3-V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25
(Xuedong Yan&Radwan, 2009)**	✓	✓	✓	-	✓	-	-	-	✓	-	-	-	-	-	-	-	-	✓	✓	✓	-	-	-
(Andreas Georgi et al., 2009)**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	✓	✓	-	-
(Abdel Aty&Abdelwahab, 2004)**	✓	✓	✓	✓	✓	✓	-	-	-	-	-	✓	✓	-	✓	-	-	✓	✓	✓	✓	-	-
(Qi et al., 2013)**	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	✓	✓	✓	✓	-
(Christopher et al., 2014)**	✓	-	-	-	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
(Mohamed et al., 2017)**	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	-	-	✓	-	✓	✓	✓	✓	✓	-	-
(Yuan et al., 2017)**	✓	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	✓	✓	✓	✓	-	-

Remark * :denotes literature in SEM analyses of accident size. ** denotes studies in analysis of rear-end crash severity.

2.5 Method

2.5.1 Data collection

This study's data collection included gathering original data and collecting data about rear-end collisions from the DOH: For this research, data for analysis of highways in Thailand were drawn from2011 to 2015 (B.E.2554–2558). Data were originally surveyed by area permanent officers who collected details of highways accidents: date, highway data, accident characteristics, crash type, number of accidents, and injury severity levels. Data were subsequently collected in the Highway Accident Information Management System (HAIMS). Consequently, data selected for this study involved only rear-end collisions with consideration of crash types, and data of drivers in accidents were used to establish the model. Selection of rear-end crash data produced 1,902 cases and 4,134 accident cars and drivers.

2.5.2 Analysis methods

Exploratory Factor Analysis (EFA) was developed in the early 20th century by Karl Pearson and Charles Spearman. The aims of EFA are to indicate variable that are unobserved or cannot be estimated directly, and to reduce the number of observed variables. The EFA describes the covariance among many variables in terms of a few unobserved variables (Washington et al., 2011). Factor analysis is calculated by expressing the $X_i^{'}$ whichin alinear function form, such that,

$$\begin{aligned}
X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \mathcal{E}_1 \\
X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \cdots + \ell_{2m}F_m + \mathcal{E}_2 \\
&\vdots \quad \ddots \quad \vdots \\
X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \mathcal{E}_p
\end{aligned} \tag{2.1}$$

In a matrix notation, the factor analysis model will become:

$$(\mathrm{X}-\mu)_{p\times 1} = \mathrm{L}_{\mathrm{p\times m}} F_{m\times 1} + \mathcal{E}_{p\times 1} \tag{2.2}$$

Where F's are factors or variables, and ℓ's are the factor loadings. The $\mathcal{E}$ is associated only with $X_i^{'}$ and, the p are random errors and m factor loading are unobserved or latent variables. The factor rotation method used determines the loading factor. If the loading factor is close to one, this means variable X_i is largely influenced by F_i (Washington et al., 2011).

The results of EFA are ℓ's (loading factor) from equation (1), shown in Table 2.3, consisted of 5 components beginning from the consideration of the variables indicating *rear-end crash size* which was found in the second component with a loading factor of *fatality*, *serious injury* and *slight injury* at 0.369, 0.168 and 0.302 respectively. In first component was call driver factor, including *V1 – V9* with a loading factor of –0.323 *to* 0.881. In third component, it was called road factor consisted of ten variables including V10-V19 with loading factors of –0.744 *to* 0.774, respectively. Regarding the forth component, it was called environmental factor which consisted of four variables with loading factor –0.869 *to* 0.981. The rear-end collision factors was in the fifth component and consisted of V24 – V25 with loading factor of 0.935 and –0.633, respectively.

Data were used to create a correlation matrix examine to what extent a mutual relationship exists between observed variables. Then SEM was run using the MPlus 7.2 program.

Table 2.3 Loading factor of EFA

	Components				
Variable	1	2	3	4	5
V1	**0.337**	0.173	0.303	0.207	-0.226
V2	**-0.207**	-0.129	-0.190	-0.176	0.128
V3	**0.503**				
V4	**0.881**				
V5	**0.769**				
V6	**0.220**		-0.169	-0.113	
V7	**-0.323**		-0.110	0.210	
V8	**0.394**		0.139		0.266
V9	**0.174**		0.110		
V10	0.277		**0.525**		
V11	-0.114		**0.725**		
V12			**0.102**		
V13	-0.168		**0.467**		-0.106
V14	0.194		**-0.744**		
V15	-0.195		**0.920**		
V16	-0.109		**0.607**		0.126
V17	-0.167		**-0.207**	-0.178	
V18			**0.774**		
V19			**0.641**		
V20	-0.271			**0.981**	
V21			0.104	**0.273**	
V22	-0.112			**-0.869**	
V23				**0.432**	
V24	0.262		0.283	0.209	**0.935**
V25				0.132	**-0.633**

Table 2.3 Loading factor of EFA (Continued)

	Components				
V26		**0.369**		0.159	-0.138
V27		**0.168**			
V28		**0.302**			

Note: Calling: component 1 = Driver factor, component 2 = crash size, component 3 = Road factors, component 4 = Environmental factor and component 5 = Crash type factor. The rotation = 'Varimax'.

2.5.3 Structural equation modeling (SEM)

SEM requires specification of the relationship between observed variables and latent variables. SEMs rely on information contained in the variance-covariance matrix, but latent variables' measurement must distinguish between fixed and free parameters. Fixed parameters are set to a reference variable, which is the base of estimation and comparison with the free parameter, for the structural model is a relationship between independent latent variables and dependent latent variables that have similar linear regression loading factors.

The SEM estimation parameter is similar to that of other statistical models. SEMs are used to evaluate theories or hypotheses using empirical data, which are contained in a $P \ x \ P$ variance-covariance matrix **S**, an unstructured estimator of the population variance-covariance matrix Σ (Washington et al., 2011). $\sum(\theta)$ is a variance-covariance matrix which turns from a generated model-implied and uses an estimated parameter vector θ. A dependent variable (exogenous variable) in SEM is a variable that has a one-way arrow pointing to it. The set of dependent variables is collected into a vector η, For independent variables (endogenous variables) are collected in the vector ξ. The relationship between them is the following:

$$\eta = \beta\eta + \gamma\xi + \varepsilon \quad (2.3)$$

Where β is the estimated vector of coefficients that contains regression coefficients for the dependent variable and γ for the independent variable. ε is the vector of regression error terms. The estimator in SEM depends on the distribution assumption of variables and the scale of a variable. This study's scale variables are only discrete data not abnormally distributed. Lee et al. (2008) Suggested that weighted least squares (WLS) methods estimate rather than assume the multivariate normality of variables.

For model goodness-of-fit Measure (GOF), the first part was basic GOF consisting of Chi-square statistic (χ^2) that presented the difference of covariance matrices among empirical data. Degree of freedom (DF) is the amount of mathematical information available to estimate model parameters. The root mean squared error of approximation (RMSEA) was a fairly correct calculation and showed more accurate statistical examples of χ^2. The value of RMSEA must be less than 0.05 (Hair et al., 2010; Kline, 2015; Shi et al., 2011). The Tucker-Lewis Index (TLI) and the comparative fit index (CFI) illustrate the proportion of difference of χ^2. TLI and CFI varied in that it is actually a comparison of the nor med chi-square values for the null and specified model. The value of TLI and CFI ranges from0 and 1and appropriate values must be greater than 0.90 (Hamdar & Schorr, 2013; Hassan & Abdel-Aty, 2013; Yu, 2002). To assess GOF, the error of WLS prediction must be considered. Appropriate values of weighted root mean square residual (WRMR) suggested by Yu (2002) must be less than 1.

2.6 Results and Discussion

2.6.1 Descriptive data

The overall view of data, as shown in Table 2.4, shows the group of variables, the names of variables, the types of variable explanations and percentage of categories, and the mean of slight injuries, serious injuries, and fatalities (dependent variables or endogenous variables). There wer 25 independent (four groups) variables. Disguise variables of injuries consisted of the number of fatalities, serious injuries, and slight injuries.

The highest mean for fatalities was found to have been caused by the driver factor, with (V7) drunk drivers involved in the most fatalities, a mean of 0.69 (1.19. Rear-end collisions with large trucks (V1) showed a mean of 0.58 (85.56%), and drivers aged 36–45 years old (V4) at 0.35 (37.65. (% Considering drivers' gender (V2), women had greater risk of fatalities than men, with a mean of 0.36 (17.66. (%Road factors revealed that non-sloped roads (V15) had the highest mean of fatalities at 0.99 (96.54%), followed by curved roads at a mean of 0.71 (93.53%).

Environmental factors showed rear-end collisions with normal visibility conditions (V21) at 0.4 (92.33%); road with wet surface (V23) at 0.37 (6.94%); accidents occurring during the day (V22) at 0.29 (67%); and accidents occurring without light at night (V20) at 0.58 (10.09%).

Regarding types of rear-end collisions, in which crashes were divided into type of car movement before the crash, the maximum mean of fatalities was with a parked car in front (V25) at 0.4 (5.49%), followed by a leading car slowing down (V24) at 0.34 (65.61%).

For factors affecting serious injuries, road and environmental factors had the highest means at 0.54 and 0.47, respectively, followed by driver factors at a mean of 0.45. Collision factors affecting severity of injuries were in last place with a mean value of 0.44. The highest mean of slight injuries was due to driver factors.

Table 2.4 Descriptive Statistics

Descriptive Statistics					Average (person)		
Group	Code	Categories	Frequency	Percentage	Slight injury	Serious injury	Fatality
Driver Factors	V1	1	3,536	85.56	1.63	0.47	0.58
		0	597	14.44	1.29	0.43	0.30
	V2	1	3,403	82.34	1.20	0.39	0.23
		0	730	17.66	1.37	0.45	0.36
	V3	1	1,055	25.53	1.45	0.41	0.31
		0	3,078	74.47	1.31	0.45	0.35
	V4	1	1,556	37.65	1.35	0.39	0.35
		0	2,577	62.35	1.33	0.46	0.33
	V5	1	707	17.11	1.32	0.44	0.33
		0	3,426	82.89	1.35	0.44	0.36
	V6	1	1,525	36.90	1.34	0.49	0.21
		0	2,608	63.10	1.35	0.41	0.41
	V7	1	49	1.19	0.73	0.73	0.69
		0	4,084	98.81	1.35	0.43	0.33
	V8	1	2,928	70.84	1.38	0.41	0.32
		0	1,205	29.16	1.24	0.50	0.38
	V9	1	1,901	46.00	1.33	0.43	0.33

Table 2.4 Descriptive Statistics (Continued)

Descriptive Statistics					Average (person)		
Group	Code	Categories	Frequency	Percentage	Slight injury	Serious injury	Fatality
Driver Factors		2	1,899	45.95	1.29	0.44	0.33
		3	245	5.93	1.61	0.44	0.35
		4	51	1.23	1.19	0.33	0.53
		5	17	0.41	1.94	0.47	0.54
		6	10	0.24	1.74	0.20	0.20
		7	5	0.12	2.80	0.40	0.20
		8	3	0.07	2.00	0.67	-
		9	2	0.05	4.00	-	-
Road Factors	V11	1	2,638	63.83	1.34	0.37	0.33
		0	1,495	36.17	1.35	0.55	0.35
	V12	1	4,027	97.44	1.35	0.44	0.34
		0	106	2.56	1.18	0.46	0.27
	V13	1	3,646	88.22	1.36	0.45	0.35
		0	487	11.78	1.20	0.35	0.26
	V14	1	265	6.41	1.29	0.38	0.31
		0	3,868	93.59	2.07	1.28	0.71
	V15	1	143	3.46	2.08	1.96	0.31
		0	3,990	96.54	1.32	0.38	0.99
	V16	1	399	9.65	1.51	0.31	0.20
		0	3,734	90.35	1.32	0.45	0.35
	V17	1	753	18.22	1.47	0.43	0.27
		0	3,380	81.78	1.31	0.44	0.35
	V19	1	896	21.68	1.33	0.27	0.22
		0	3,237	78.32	1.34	0.48	0.37

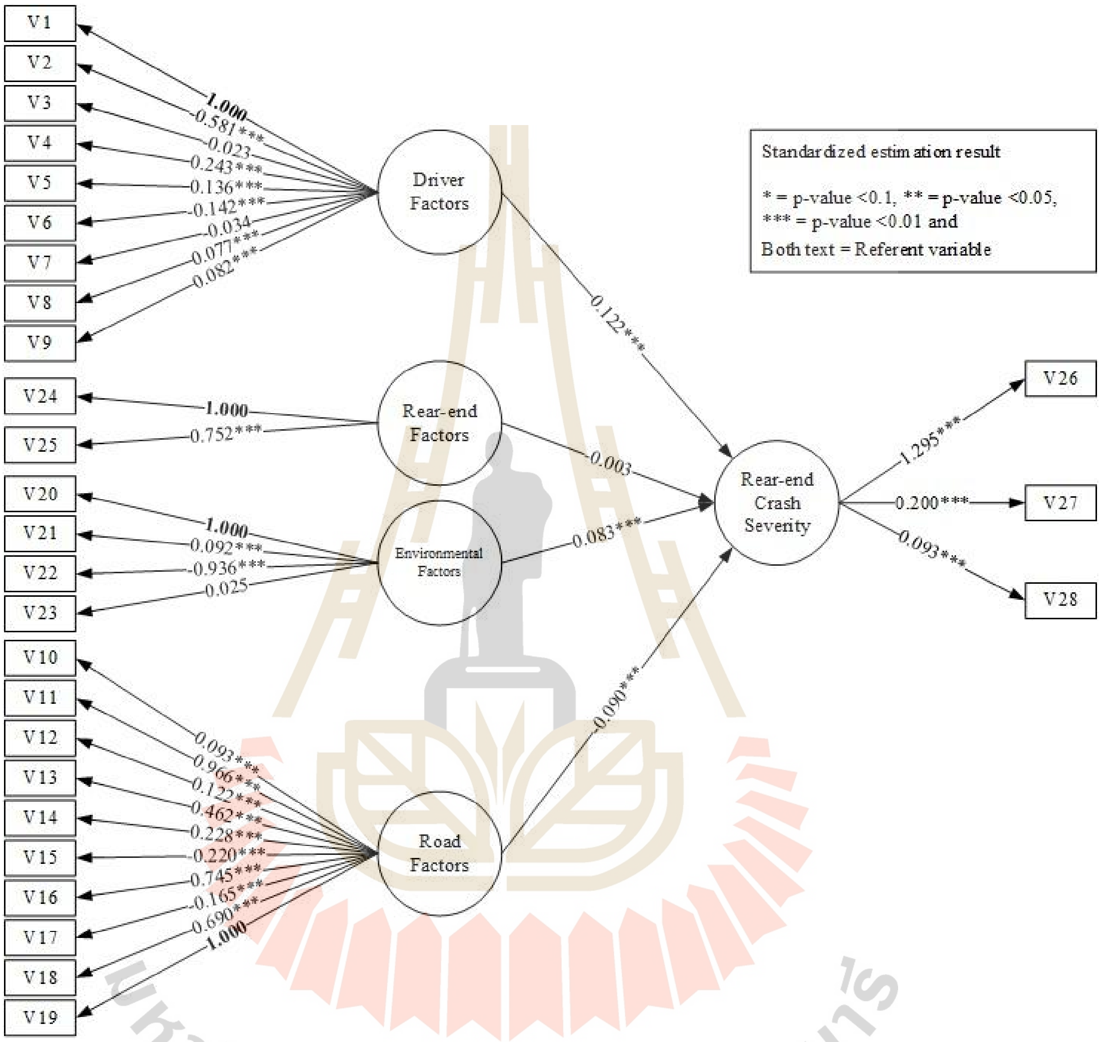
Table 2.4 Descriptive Statistics (Continued)

Descriptive Statistics					Average (person)		
Group	Code	Categories	Frequency	Percentage	Slight injury	Serious injury	Fatality
Environmental Factors	V20	1	417	10.09	1.09	0.40	0.58
		0	3,716	89.91	1.37	0.44	0.31
	V21	1	3,816	92.33	1.44	0.60	0.40
		0	317	7.67	1.33	0.42	0.33
	V22	1	2,769	67.00	1.39	0.46	0.29
		0	1,364	33.00	1.25	0.38	0.42
	V23	1	287	6.94	1.48	0.64	0.37
		0	3,846	93.06	1.33	0.42	0.33
Rear-end crash type Factors	V24	1	2,693	65.16	1.46	0.44	0.34
		0	1,440	34.84	1.13	0.42	0.33
	V25	1	227	5.49	0.98	0.31	0.40
		0	3,906	94.51	1.36	0.44	0.33

Remark: average of percentage truck(V10) =16.7, average of Log AADT (V18) = 10.42.

2.6.2 Results of SEM and Discussion

In analysis of data on rear-end highway collisions, acquired from the DOH, to determine factors affecting levels of driver and passenger injuries, determined factors were classified into four groups of latent variables including individual, road, environmental, and collision factors. The model was compared with empirical data by considering model fit information values as shown in the note to Figure 2.1**,** with achi-square statistic value = 1232.160, df = 302 (p-value <0.000), RMSEA = 0.027, CFI=0.928, TLI = 0.910 and WRMSR = 1.880. Although WRMSR value was greater than the cutoff value, it could be accepted (Baggio et al., 2013; Machado et al., 2016 ; Schnabel et al., 2015). Comparison of this model's goodness of fit with cutoff criteria of other research showed it within acceptance criteria; thus, it can be used to interpret research results.



Note: Model fit information: Chi-square value (χ^2) = 1232.160, degree of freedom (df) = 302(p-value = 0.000), Root mean square error of approximation (RMSEA) = 0.027, CFI = 0.928, TLI = 0.910; Weighted root mean square residual (WRMSR) = 1.880.

Figure 2.1 SEM result model

Consideration of the measurement model of rear-end crash severity using three variables including the number of fatalities (reference variable), the number of serious injuries, and slight injuries, found that the number of fatalities from each accident could evidently indicate severity levels of injuries (β = 1.295, S.E. = 0.004) followed by the number of serious injuries (β = 0.2, S.E. = 0.008) and the number of minor injuries (β = 0.099, S.E. = 0.017). Additionally, operational definitions of injury levels were differently distinguished in Thailand and North America. In Thailand, injuries were classified into three levels including death, serious injury, and slight injury, as distributed by levels of hospital treatment. Injury levels in North America were individually divided into the Abbreviated Injury Scale (AIS) by sorting according to body different parts: head, face, neck, thorax, abdomen, spine, upper extremities, lower extremities, and external. Each injury level is assigned an AIS score on an ordinal scale ranging from 1 (minor injury, probability of death =0%) to 6 (maximum injury, probability of death = 100%) (Stevenson et al., 2001). After some consideration, researchers decided that the AIS system's criteria of injury score and duration of treatment in the hospital could not be directly compared. In addition, treatment systems differ to some extent between the two countries.

The structural model revealed that among the four independent latent variables, rear-end collisions' severity was significantly and respectively affected by three factors: driver, road, and environmental. While collision types did not significantly affect severity of injuries, the driver factor most affected injuries' severity (β = 0.122, S.E. = 0.013). In consideration of the measurement model for the *driver factor* providing the reference variable, *large vehicle(V1),* which is interpreted as the presence of trucks, and the accident would affect the increase of injury severities in accord with studies conducted by Piccinini et al. (2017); Qi et al. (2013); Wiacek et al. (2015). This violent effect may originate from the massive size of trucks causing strike force resulting in more severe injuries. Driver gender was the second variable most affecting levels of injuries, with women receiving more serious injuries possibly because female drivers are hurt more easily than male drivers, conforming to Chen et al. (2015); Mafi et al. (2018) findings that male drivers tended to suffer lower levels of injuries. Another cause may be women‘s longer stop-car decision time compared to men‘s (Warshawsky-Livne&Shinar, 2002). For the age factor, drivers were compared by age ranges, including 26–35, 36–45, and 46–55.Drivers 36–55 years old were in more severe accidents, a finding similar to Lee et al. (2008), which found that the drivers 40–50 years old affected increasing severity of injuries. Along with safety equipment nonuse, drivers’ injury levels increase, following studies of Chen et al. (Chen et al., 2015; Chen, Zhang, Yang, et al., 2016). Other significant factors in causing greater accident severity were the sequential order of involved vehicles and driving over the speed limit.

When considering the *road factors* significantly affecting the levels of injuries (β = −0.09, S.E. = 0.013), overall, every indicator attained statistical significance. The variable with the highest loading factor (reference variable) was the number of traffic lanes (V19). More than four lanes lessened rear-end crashes’ severity. This is relevant to a study finding that more traffic lanes potentially decreased fatalities because more lanes caused drivers to be more careful (Chen, Zhang, Yang, et al., 2016; Mohamed et al., 2017). The divided road variable was determined to compare (V11) rear-end collisions on roads with and without traffic islands. Roads without traffic islands affected levels of injury severity, in accordance with Das and Abdel-Aty (2011) study. According to the main road variable (V16), rear-end collisions occurring on main roads resulted in higher severity. This is relevant to the study of Khorashadi et al. (2005) who found that innermost lanes potentially increased injury levels, probably resulting from higher speed on main roads than on parallel roads. Huang, Chin, and Haque (2008), followed by log AADT (V18), found that higher traffic quantity resulted in decreased rear-end crashes, consistent with a study discovering that increased AADT decreased safety (Abdel-Aty & Haleem, 2011; Schorr & Hamdar, 2014; Stylianou & Dimitriou, 2018). In road surface types (V13), surfaces other than asphalt increased serious injuries in accordance withLee et al. (2008) finding that concrete surfaces affected increasing severity of injuries. For normal roads or work zones (V12), which also affected serious injuries, collisions were caused by drivers exceeding speeds for roads being repaired or maintained (Mohamed et al., 2017). Another variable indicated significantly in the measurement model was collisions at intersections. A leading vehicle's need to brake increased the risk of crash by a following vehicle (Das & Abdel-Aty, 2011). Additionally, a road's grade or slope (V15) created greater severity in rear-end collisions, following Chen, Zhang, Yang, et al. (2016). Lower percentages of trucks (V10) affected greater rear-end crash severity. In Thailand, the trucks usually used the arterial roads, there are many traffic lanes. This related to the results of V19, if the number of lanes increased it will be small of rear-end crash size.

As for the environmental factor and significant effects on levels of injuries (β = 0.083, S.E. = 0.009), the measurement model determined darkness (V20) as a reference variable. If an accident occurred at night with no available light, injury levels increased, confirming much research on low visibility leading to more serious injuries (Chen, Zhang, Huang, et al., 2016; Chen et al., 2015; Sullivan & Flannagan, 2003; Xuedong Yan & Radwan, 2009). Due to reduced traffic at night, drivers who drove at high speed could not stop their cars and crashed in to leading cars at low speeds. This was the cause of serious injuries conforming to the variable that compared nighttime and daytime crashes (V22) - nighttime crashes caused greater severity of injury than daytime crashes (Chen et al., 2015). For the driver visibility factor, the condition of visibility including clear skies, without dust, fog, or smoke to hinder vision, affected greater severity of injuries (Abdel Aty & Abdelwahab, 2004). The road surface variable was not significant in this measurement model (Chen et al., 2015).

2.7 Conclusion

This research studied factors affecting rear-end crash severity on Thailand's highways, as indicated by numbers of fatalities and serious and slight injuries as analyzed with SEM. From analysis of data obtained from the DOH, these research results can assist organizations involved in law enforcement, including inspectors' offices and organizations involved in road design and maintenance, for instance, the Department of Highways or the Department of Rural Roads, in reducing grear-end crash severity.

The first group of factors increasing rear-end crashes' severity the most is the driver factor: trucks involved, female drivers, drivers from 36–55 years old (at which ages Thai drivers often drive at high speed), not using safety equipment, rear-end crashes caused by driving over the speed limit, the high number of traffic violators in Thailand, and the sequential order of involvement in rear-end crashes. Thus, involved organizations should implement policy to reduce injury severity in rear-end collisions by establishing “Truck Only Lanes” (Chrysler, 2016) that can reduce conflicts between trucks and other drivers. Another policy for female drivers’ safety is using the “two dots” or “tailgating” indicator, now available only on Thai motorways, to warn drivers about leaving space behind lead vehicles. This installation would benefit both males and females, of course (Hutchinson, 2008).

The second group of factors concerns roads, which affect rear-end crash severity due to the number of traffic lanes, traffic islands, main roads, road surface types, intersections, road steepness, road bends, and roads in maintenance. Policy from this variable group involves Road Safety Audits, especially, four- or fewer than four-lane roads and roads without traffic islands that decrease rear-end crash severity.

The last group of factors affecting rear-end accident severity is environmental. Indicators causing rear-end crash severity are roads without light at night and clear visibility, which seems to encourage speeding. For potential policy, light installations in risky areas, for instance, truck-parking areas, should be considered. Another potentially useful policy measure is effective speed-limit enforcement. Technology might assist here, with installation of speed-censoring cameras.

Applications of this research in other developed countries might involve differences among the three main factors of driver, road, and environment. The road factor can be directly applied, for instance, by performing Road Safety Audits. The environmental factor can be instantly applied, for instance, light improvement to reduce rear-end collision severity, and “Truck Only Lanes” can be considered for immediate installation. However, some conditions, for example, AADT, and truck percentage, may differ from those in Thailand. As for speed limits and safety equipment use, compulsory enforcement was potentially more successful in more highly developed countries.

This study found factors affecting rear-end collision severity and introduced guidelines for its reduction. However, the study contains model limitations due to unanalyzed passenger characteristics. Those variables potentially result in increasing severity of rear-end collisions, which might result from data collection limitations, that is, not including passenger characteristics: the number in each vehicle, their use of safety equipment, and their gender. Thus, these factors are proposed for additional, future study.

2.8 Acknowledgements

The authors would like to thank the Bureau of Highway Safety, Department of highway, Ministry of transport for providing the dataset in the analysis. And we would like to thank Enago (www.enago.com) for the English language review.

2.9 Reference

Abdel-Aty, M., and Haleem, K. (2011). Analyzing angle crashes at unsignalized intersections using machine learning techniques. **Accident Analysis & Prevention**, 43(1): 461-470.

Abdel Aty, M., and Abdelwahab, H. (2004). Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. **Accident Analysis & Prevention**, 36(3): 447-456.

Andreas Georgi, Marc Zimmermann, Thomas, Lich Lisa Blank, Nils Kickler, and Reiner Marchthaler. (2009). New approach of accident benefit analysis for rear end collision avoidance and mitigation systems.

Baggio, S., Studer, J., Deline, S., Mohler-Kuo, M., Daeppen, J. B., and Gmel, G. (2013). Factor structure of early smoking experiences and associations with smoking behavior: valence or sensitivity model? **International journal of environmental research and public health**, 10(12): 6305-6318.

Burdett, B., Alsghan, I., Chiu, L.-H., Bill, A. R., and Noyce, D. A. (2016). Analysis of Rear-End Collisions at Roundabout Approaches. **Transportation Research Record: Journal of the Transportation Research Board**, 2585(1): 29-38.

Chen, C., Zhang, G., Huang, H., Wang, J., and Tarefder, R. A. (2016). Examining driver injury severity outcomes in rural non-interstate roadway crashes using a hierarchical ordered logit model. **Accident Analysis & Prevention**, 96: 79-87.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., and Guan, H. (2015). A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. **Accident Analysis & Prevention**, 80: 76-88.

Chen, C., Zhang, G., Yang, J., Milton, J. C., and Alcantara, A. D. (2016). An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naive Bayes (DTNB) hybrid classifier. **Accident Analysis & Prevention**, 90: 95-107.

Christopher, W., James, B., and Dinesh, S. (2014). *Real-world analysis of fatal rear-end crashes*. Retrieved from National Highway Traffic Safety Administration:

Chrysler, S. T. (2016). *Preferential lane use for heavy trucks*. Retrieved from https://static.tti.tamu.edu/tti.tamu.edu/documents/PRC-15-39-F.pdf

Chu, A., Bhat, C., R. , Pendyala, R. M., and Garikapati, V. (2015). A Comprehensive mixed logit analysis od crash type conditional on a crash event. *Civil, Architectural, and Environmental Engineering.* The University of Texas at Austin. Retrieved from http://hdl.handle.net/2152/34140

Cunto, F., and Saccomanno, F. F. (2009). Simulated safety performance of rear-end and angled vehicle interactions at isolated intersections. **Canadian Journal of Civil Engineering**, 36(11): 1794-1803.

Das, A., and Abdel-Aty, M. A. (2011). A combined frequency–severity approach for the analysis of rear-end crashes on urban arterials. **Safety Science**, 49(8-9): 1156-1163.

Department of Highway Thailand. (2016). Thailand Traffic Accident on National Highways in 2015. Bureau of Highway Safety Department of Highways Ministry of Transport. Retrieved from http://bhs.doh.go.th/download/accident

Dong, C., Clarke, D. B., Nambisan, S. S., and Huang, B. (2016). Analyzing injury crashes using random-parameter bivariate regression models. **Transportmetrica A: Transport Science**, 12(9): 794-810.

Fildes, B., Keall, M., Bos, N., Lie, A., Page, Y., Pastor, C., Tingvall, C. (2015). Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes. **Accident Analysis & Prevention**, 81: 24-29.

Georgi, A., Zimmermann, M., Lich, T., Blank, L., Kickler, N., and Marchthaler, R. (2009). *New approach of accident benefit analysis for rear end collision* *avoidance and mitigation systems*. Paper presented at the 21st International Technical Conference on the Enhanced Safety of Vehicles.

Hair, J. F., Anderson, R. E., Tatham, R. L., and William, C. (2010). Multivariate Data Analysis. In: Pearson. New Jersey.

Hamdar, S. H., and Schorr, J. (2013). Interrupted versus uninterrupted flow: a safety propensity index for driver behavior. **Accident Analysis & Prevention**, 55: 22-33.

Hassan, H. M., and Abdel-Aty, M. A. (2013). Exploring the safety implications of young drivers' behavior, attitudes and perceptions. **Accident Analysis & Prevention**, 50: 361-370.

Hassan, H. M., and Al-Faleh, H. (2013). Exploring the risk factors associated with the size and severity of roadway crashes in Riyadh. **Journal of Safety Research**, 47: 67-74.

Hutchinson, P. (2008). Tailgating. Safety Centre for Automotive Safety Research, University of Adelaide. Australia, Research Report# CASR046. Retrieved from http://casr.adelaide.edu.au/publications/list/?id=638

Iamtrakul, P. (2008). Analysis of factors affecting on rear-end crashed accidents. Retrieved from http://roadsafetythai.org/uploads/userfiles/ACC_51016.pdf

Islam, M. T. (2016). Investigating the Speed and Rear-End Collision Relationship at Urban Signalized Intersections. **Transportation Research Record: Journal of the Transportation Research Board**, 2601(1): 10-16.

Joon-Ki, K., Yinhai, W., and Gudmundur, F. U. (2007). Modeling the Probability of Freeway Rear-End Crash Occurrence. **Journal of transport engineering**, 113(1).

Khorashadi, A., Niemeier, D., Shankar, V., and Mannering, F. (2005). Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. **Accident Analysis & Prevention**, 37(5): 910-921.

Kim, K., Pant, P., and Yamashita, E. (2011). Measuring Influence of Accessibility on Accident Severity with Structural Equation Modeling. **Transportation Research Record: Journal of the Transportation Research Board**, 2236: 1-10.

Kline, R. B. (2015). Principles and practice of structural equation modeling: Guilford publications.

Lee, J. Y., Chung, J. H., and Son, B. (2008). Analysis of traffic accident size for Korean highway using structural equation models. **Accident Analysis & Prevention**, 40(6): 1955-1963.

Lerdworawinich, P. (2000). Report No. Rd 281- Studying of risky reducing of Rear-end accident on Thai highway Bereau of Road research and Developement, Deportment of highway, Ministry of Thailand. Retrieved from http://km.doh.go.th/doh/preview.php?tableName=TB_KNOWLEDGE&tableId=17161&fieldName_show=KNOWLEDGE_NAME&id_text=KNOWLEDGE_ID

Liang, J., Chen, L., Cheng, X.-y., and Chen, X.-b. (2010). Multi-agent and driving behavior based rear-end collision alarm modeling and simulating. **Simulation Modelling Practice and Theory**, 18(8): 1092-1103.

Ma, L., and Yan, X. (2014). Examining the nonparametric effect of drivers' age in rear-end accidents through an additive logistic regression model. **Accident Analysis & Prevention**, 67: 129-136.

Machado, L. A., Telles, R. W., Costa-Silva, L., and Barreto, S. M. (2016). Psychometric properties of Multidimensional Health Locus of Control - A and General Self-Efficacy Scale in civil servants: ELSA-Brasil Musculoskeletal Study (ELSA-Brasil MSK). **Brazilian Journal of Physical Therapy**, 20(5): 451-460.

Mafi, S., Abdelrazig, Y., and Doczy, R. (2018). Analysis of Gap Acceptance Behavior for Unprotected Right and Left Turning Maneuvers at Signalized Intersections using Data Mining Methods: A Driving Simulation Approach. **Transportation Research Record: Journal of the Transportation Research Board**, 2672(38): 160-170.

Meng, Q., and Qu, X. (2012). Estimation of rear-end vehicle crash frequencies in urban road tunnels. **Accident Analysis & Prevention**, 48: 254-263.

Meng, Q., and Weng, J. (2011). Evaluation of rear-end crash risk at work zone using work zone traffic data. **Accident Analysis & Prevention**, 43(4): 1291-1300.

Mohamed, S. A., Mohamed, K., and Husain, A. A.-H. (2017). *Investigating Factors Affecting the Occurrence and Severity of Rear-End Crashes*. Paper presented at the Transportation Research Procedia - WCTR 2016 Shanghai.

Ni, Y., and Li, K. (2014). Estimating Rear-End Accident Probabilities at Signalized Intersections: A Comparison Study of Intersections With and Without Green Signal Countdown Devices. **Traffic Injury Prevention**, 15(6): 583-590.

Nishimura, N., Simms, C. K., and Wood, D. P. (2015). Impact characteristics of a vehicle population in low speed front to rear collisions. **Accident Analysis & Prevention**, 79: 1-12.

Office of transport and traffic policy and planing. (2014). Analyzes road accidents (2557 B.E.). Retrieved from www.otp.go.th/uploads/files/1471501633-dgyrs-3z11i.pdf

Pande, A., and Abdel-Aty, M. (2008). A Computing Approach Using Probabilistic Neural Networks for Instantaneous Appraisal of Rear-End Crash Risk. **Computer-Aided Civil and Infrastructure Engineering**, 23(7): 549–559.

Piccinini, G. B., Engström, J., Bärgman, J., and Wang, X. (2017). Factors contributing to commercial vehicle rear-end conflicts in China: A study using on-board event data recorders. **Journal of safety research**, 62: 143-153.

Qi, Y., Srinivasan, R., Teng, H., and Baker, R. (2013). Analysis of the Frequency and Severity of Rear-End Crashes in Work Zones. **Traffic Injury Prevention**, 14(1): 61-72.

Schnabel, D. B. L., Kelava, A., van de Vijver, F. J. R., and Seifert, L. (2015). Examining psychometric properties, measurement invariance, and construct validity of a short version of the Test to Measure Intercultural Competence (TMIC-S) in Germany and Brazil. **International Journal of Intercultural Relations**, 49: 137-155.

Schorr, J. P., and Hamdar, S. H. (2014). Safety propensity index for signalized and unsignalized intersections: exploration and assessment. **Accident Analysis & Prevention**, 71: 93-105.

Seacrist, T., Belwadi, A., Prabahar, A., Chamberlain, S., Megariotis, J., and Loeb, H. (2016). Comparison of crash rates and rear-end striking crashes among novice teens and experienced adults using the SHRP2 Naturalistic Driving Study. **Traffic Injury Prevention**, 17(sup1): 48-52.

Shahi, J., Namaki Araghi, B., and Mehdizadeh, A. (2009). Estimation of rear-end potential conflict using Time-to-collision model. **Ingeniería y Desarrollo**(25): 67-79.

Shi, J., Bai, Y., Tao, L., and Atchley, P. (2011). A model of Beijing drivers' scrambling behaviors. **Accident Analysis & Prevention**, 43(4): 1540-1546.

Stevenson, M., Segui-Gomez, M., Lescohier, I., Di Scala, C., and McDonald-Smith, G. (2001). An overview of the injury severity score and the new injury severity score. **Injury Prevention**, 7(1): 10-13.

Stylianou, K., and Dimitriou, L. (2018). Analysis of Rear-End Conflicts in Urban Networks using Bayesian Networks. **Transportation Research Record: Journal of the Transportation Research Board**, 2672(38): 302-312.

Sullivan, J. M., and Flannagan, M. J. (2003). *Risk of fatal rear-end collisions: is there more to it than attention?* Paper presented at the Proceeding of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Wan, H., Liu, Z., and Wang, H. (2013). Rear-end and Lane-changing Collisions Avoidance Theoretical Models of the Multi-lane Freeway in Disaster Weather. **Procedia - Social and Behavioral Sciences**, 96: 70-81.

Wang, Y., Ieda, H., and Mannering, F. (2003). Estimating rear-end accident probabilities at signalized intersections: occurrence-mechanism approach. **Journal of Transportation engineering**, 129(4): 377-384.

Warshawsky-Livne, L., and Shinar, D. (2002). Effects of uncertainty, transmission type, driver age and gender on brake reaction and movement time. **Journal of Safety Research**, 33(1): 117-128.

Washington, S. P., Karlaftis, M. G., and Mannering, F. (2011). Statistical and Econometric Methods for Transportation Data Analysis (2 ed.).

Weng, J., Meng, Q., and Yan, X. (2014). Analysis of work zone rear-end crash risk for different vehicle-following patterns. **Accident Analysis & Prevention**, 72: 449-457.

WHO. (2015). *Global status report on road safety 2015*. Retrieved from http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/

Wiacek, C., Bean, J., and Sharma, D. (2015). *Real-world analysis of fatal rear-end crashes*. Paper presented at the 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration. https://trid.trb.org/view/1365430

Yan, X., and Radwan, E. (2006). Analyses of rear-end crashes based on classification tree models. **Traffic Injury Prevention**, 7(3): 276-282.

Yan, X., and Radwan, E. (2009). Analysis of truck-involved rear-end crashes using multinomial logistic regression. **Advances in Transportation Studies an international Journal**, 17.

Yu, C.-Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. University of California, Los Angeles Los Angeles. Retrieved from https://www.statmodel.com/download/Yudissertation.pdf

Yuan, Q., Lu, M., Theofilatos, A., and Li, Y. B. (2017). Investigation on occupant injury severity in rear-end crashes involving trucks as the front vehicle in Beijing area, China. **Chinese journal of traumatology**, 20(1): 20-26.

CHAPTER III

ANALYSIS OF REAR-END CRASH AT INTERSECTION ON THAI HIGHWAY: DECISION TREE APPROACH

3.1 Abstract

The rear-end crash on highways tends to continuously increase. This is in accordance with many researches pointing out that the collisions are likely to highly occur at the intersection as vehicles have to reduce the velocity to approach into the intersection. Therefore, this research focuses on seeking for the ways decreasing the number of rear-end collisions at the intersection through the guidelines for relevant organizations to improve such as physical features of roads as well as the promotion methods. The accident cases occurring on highways from 2011 to 2015 were the data used to be analyzed by classification and regression tree (CRT). The target variables were rear-end crashes at the intersection/and those outside the intersection. The results from tree model found that the significant variables to be further recommended were: average traffic volume, road surface type and lighting condition factor. In addition, this research has provided the guidelines for reducing the number of crashes at the intersection. It also guides the study of rear-end crashes at the intersections in the future.

3.2 Introduction

In Thailand, It was found that rear-end crashes on Thai highways tend to incessantly increase (Department of Highway, 2016, 2017, 2018). When considering the statistics distributed by the crash type, it was found that the number of crashes from one direction occurred as high as the second largest number. The crash statistics in Figure 3.1 shows only accident cases collected by the officers of the Department of Highways but the actual rear-end crashes on the highways are further (Iamtrakul, 2008). That is, minor rear-end crashes were not recorded in these statistics. (Iamtrakul, 2008). This is in accordance with Chanbunditayanun (2017) identifying that the rear-end crashes occur most frequently in Thailand. Therefore, the need to reduce the number of rear-end collisions is very imperative. Finding of this study would provide such guidelines must be considered based on the road physical characteristics where the accidents always occur. As the rear-end collision, the back car crashes against the rear of the front car, is caused at the point where the speed of the front vehicle is reduced. This is relevant to previous research which often found that the point of rear-end crash occurrence is the tunnel area (Meng & Qu, 2012), work zone, (Weng et al., 2014; Weng et al., 2015) intersections (Yan & Radwan, 2006; Yan et al., 2005; Yinhai Wang et al., 2002). When simultaneously considering with characteristics of Thai highway, the highest risk of rear-end crashes was identified at the intersections where there are many traffic lights, especially in the urban area where the rear-end crashes are highly potential to occur. (Iamtrakul, 2008).

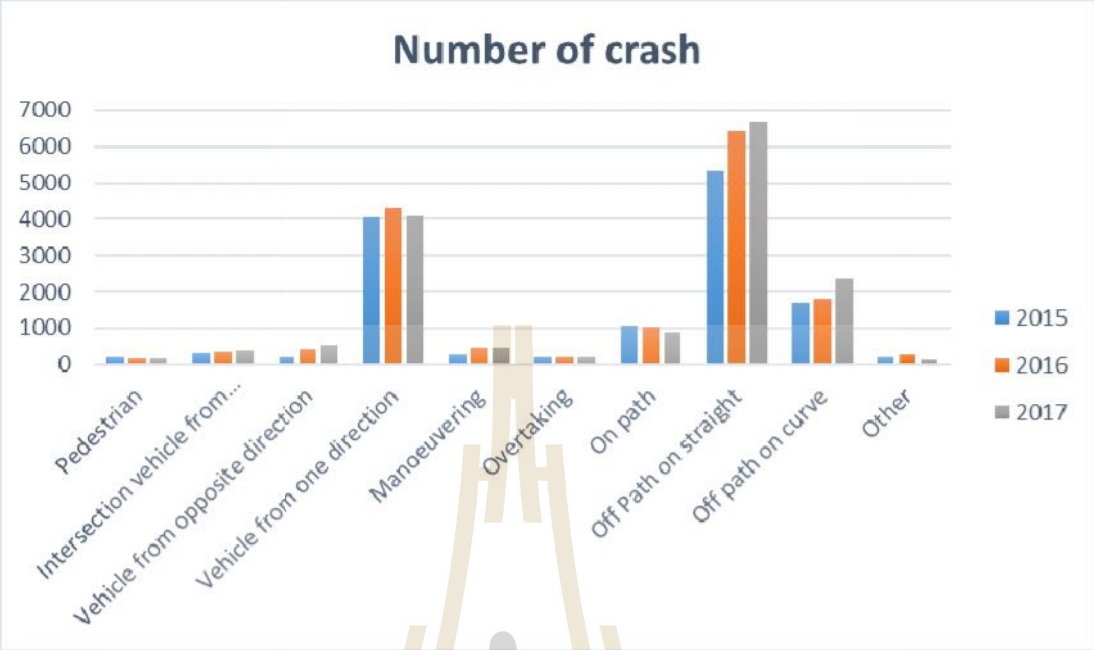


Figure 3.1 Crash on Thai highway by crash type

Factors potentially causing rear-end crashes at the intersection depend on driver's individual differences including breaking distance (considered from the decision period), sight distance etc. consisting of gender, age, alcohol use (Anvari et al., 2017; X. Li et al., 2016; Nikiforos & John, 1997), environment (time period, weather condition), roads including road surface conditions, physical characteristics (Z. Li et al., 2014; Mendez & Izquierdo, 2010), vehicle type which affects the aspect of parking along the road, and breaking distance resulting from load and vehicle weight (Abdel Aty & Abdelwahab, 2004; Harb et al., 2007; Nikiforos, 2008), road physical characteristics before the intersection such as the number of traffic lanes affecting road characteristics afore the intersection, such as the number of lanes affecting driver's behavior as well as speed used for driving. (Kim et al., 2016; X. Li et al., 2016; Meng & Qu, 2012). Other factors that possibly contribute to the notion promoting traffic collision avoidance at the intersection are rear- end crash types that are classified by the movement of the front vehicle before the rear-end crash (Ma&Yan, 2014). Visibility factor, which affects the speed the driver uses, and visual range before deciding to make a brake to enter into the intersection. (Chen et al., 2015)

The statistical theories used for analysis are based on the dependent variable data characteristics, which categorical data consisting of two variables are including the rear- end crashes occurring at the intersection, and those occurring outside the intersection. The method widely used for this variable type is the analysis of whether the parameter is estimated or not. (If it is a parameter type, the relationship between the independent variables and the dependent variables is considered. If not, the data will be sorted to see the data proportion, called data mining). It was found that binary logistic regression is an alternative for parametric analysis. Yan and Radwan (2006) indicated its limitation of the difficulty of investigating the relationship between the two variables. Consequently, the suitable alternative is nonparametric analysis (that is, Decision tree of Classification tree; DT) which is arranged through an algorithm to perceive the data proportion according to the determined dependent variables. (Agouti et al., 2017). So if an appropriate amount of data is available, the characteristics of complex independent variables could be analyzed. (Yan & Radwan, 2006; Zheng et al., 2016). The characteristics of a decision tree is a structure that includes a root node, branches, and leaf nodes (Muhammad et al., 2017). Yan and Radwan (2006) have used DT for the rear-end collision data analysis which is divided into two models: the first model is to analyze which accident characteristics would be classified as a rear-end collision, and the second is to analyze what rider's characteristics would be likely to become at-fault-driver. The area of study is Florida, USA. This study found that rear-end crashes are over-presented at signalized intersections due to higher speed-limits; therefore, the recommendation is to reduce the speed-limits to 40 mph which efficiently contribute to lower rear-end crash rate.

However, in Thailand there has never been study of rear-end crashes occurring at intersections of highways nationwide, especially, the application of decision tree model for rear-end crashes reduction. Therefore, this study aims to fulfill the direction of reducing the number of rear-end crashes at the intersection by proposing the policies to relevant agencies whose role in driver training or organization or organization or road design and maintenances such as Department of Highways, and Department of Land Transport for implementation across the country.

3.3 Highway crash reporting

The data used in this study were accident cases occurring on highways obtained from Department of Highways (DOH) from 2011 to 2015, the characteristics of data considerably collected included date, road segment, physical characteristics at the traffic accident scene (e. g straight, curved, median, intersection), environment (e.g, rain, lighting conditions, time of accident), presumed causal information (such as excessive speeding) and the number of accident injuries (including fatalities, serious injuries, and slight injuries). The data provided by this department may not cover every accident occurrence. In case that a minor collision and the victims could agree with each other, that accident was not recorded.

After selecting the rear-end crash types which comply with the movement of the front vehicle before collision. (Ma & Yan, 2014), they were classified into 3 main types including 1) going straight (the front car using the normal speed) 2) decelerating speed (the car is slowing down, such as turning a car or U-Turn using), and 3) Stopped (the front car is parking on the roadside/ road edge or is parking at traffic lights). After screening, there were 2,115 cases which were used for analysis. As this analysis took the vehicle data into account to add driver and vehicle factors into the model, the data set comprised 5,445 vehicles involved in rear-end crashes. The traffic volume was obtained from the report of the DOH, 2017. It also matched the road segment where the accidents occurred.

According to Descriptive statistics shown in Table 3.1, for vehicle type factor (Veh_Type), it was found that medium small cars (personal cars, pick-up trucks) are the vehicles having the highest accident rate (18.9%), followed by large trucks (18%). For the rear-end crash types, decelerating speed crash is the highest proportion (35.5%). For traffic volume, it was found that the rear-end crashes at the intersection are average 20,605 per day, less than those outside the intersection with the average truck proportion 15.49%.

3.4 Method

3.4.1 Variable setting

For variable management to enter into the model according to Table 3.1, there were two types of variables including 1) categorical variable: The values of the independent variables were divided according to the characteristics of the variables in numerical data, for example, Gender (0 = male, 1 = female), Vehicle type (1 = a small car such as a motorcycle, 2 = medium-sized vehicles such cars and pick-up trucks, 3 = large vehicles or over six- wheeled trucks), Crash Type (1=Going straight, 2=Decelerating, 3=Stopped), and 2) Continuous such as Traffic volume (AADT), Per Ctruck which is the proportion of trucks.

3.4.2 Classification tree and Building Model

The elements of DT model consist of 3 main parts which included decision nodes, branches, and leaf nodes. Within DT structure of each decision node, the variables were displayed and each branch exposed one variable value based on decision rules. In addition, leaf node was the expected value of the target variables (Song & Lu, 2015). This study used a Decision tree model (Decision Tree or Classification Tree; DT) for the rear end crash analysis starting from the target variable (dependent variables) by setting value 1 for the rear-end crashes occurring at the intersection, and value 0 for the rear-end crash not occurring at the intersection.

This study used SPSS Program and chose classification and regression tree (CRT) by which influence variables were analyzed. As this research aims to find the relationship between the target variables and other variables, each independent (predictor) variable will be shown in the form of ranks according to its importance to the model (IBM, 2012). Previously, many researches have used CRT to analyze accident data. (Kashani & Mohaymany, 2011; Pakgohar et al., 2011; Pande et al., 2010). CRT functions to seek for maximizing within-node homogenous. The extent to which a node does not represent a homogenous subset of cases is an indication of impurity. (IBM, 2012).

Table 3.1 Variables' characteristics and descriptive statistics

		Intersection					
		Non-intersection			Intersection		
		Count	Row N %	Mean	Count	Row N %	Mean
SpeedExc	Other wise	1044	76.1%		327	23.9%	
	Yes	2694	84.6%		489	15.4%	
Veh_Type	Small	1259	81.1%		294	18.9%	
	Middle	1923	82.0%		421	18.0%	
	Large	556	84.6%		101	15.4%	
Driver	non-at-fault	1733	81.9%		382	18.1%	
	at-fault	2005	82.2%		434	17.8%	
CrashType	Stopped	2628	88.9%		327	11.1%	
	Decelerating	871	64.5%		479	35.5%	
	Go Straight	239	96.0%		10	4.0%	
Gender	Female	2967	82.3%		636	17.7%	
	Male	615	79.6%		158	20.4%	
Slight_In				1.3			1.5
Serious_In				0.5			0.4
Veh_Involve				2.3			2.2
Normal	Other wise	94	83.2%		19	16.8%	
	Yes	3644	82.1%		797	17.9%	
Slope	Other wise	3593	81.7%		804	18.3%	
	Yes	145	92.4%		12	7.6%	
env_light	Day	2478	81.5%		564	18.5%	
	Night without light	414	90.0%		46	10.0%	
	Night with light	846	80.4%		206	19.6%	
Weather	Otherwise	3431	81.4%		782	18.6%	
	Clean	307	90.0%		34	10.0%	
No.ofLane	4 and less	2867	80.5%		695	19.5%	
	Otherwise	871	87.8%		121	12.2%	
SafetyEqui	Other wise	2319	80.6%		557	19.4%	
	Use	1419	84.6%		259	15.4%	
Alcohol	Other wise	3695	82.1%		805	17.9%	
	Yes	43	79.6%		11	20.4%	
PerCTruck				16.97			15.49
AADT60				36,397			20,605
env_surfaces	Dry	3446	81.4%		789	18.6%	
	Wet	292	91.5%		27	8.5%	
Concrete	Other wise	3320	82.9%		683	17.1%	
	Yes	418	75.9%		133	24.1%	

Note: SpeedExc = Exceeding speed limit; Veh_type = vehicle type; Slight_In = slight injury; Serious_In = serious injury; Normal = Veh_involve = number of vehicle involvement; env_light = lighting condition; SafetyEqui = using safety equipment; PerCTruck = Percentage of trucks; env_surfaces= road surface condition.

The following thing for consideration was the choice of algorithms in splitting. In SPSS, there are two types of CRT including Gini and Twoing, but Gini splits are widely used. For Gini's principle, splits are found that maximize the homogeneity of child nodes with respect to the value of the dependent variables. It is based on squared probabilities of membership for each category of the dependent variable (L.-Y. Chang&Chien, 2013; IBM, 2012; Kashani & Mohaymany, 2011). For further details, readers are offered to read supplementary articles (L.-Y. Chang & Chien, 2013; IBM, 2012; Kashani & Mohaymany, 2011). Regarding the consideration on realism of CRT, it was conducted by using unit misclassification costs which are the proportion of observed and predicted data comparisons (Khan et al., 2015).

For determining optimal tree model, validation equaled cross validation 70:30 (Yan & Radwan, 2006). The samples were divided into two sets called training. The large data set which was 70% of the total number of rear-end crashes was used for being main interpretation, and testing (30%) was used to measure the data consistency whether it was in the same direction or not. For avoidance of over-fitting model, maximum was determined as: tree depth=5 nodes, minimum cases in parent node = 150, and minimum cases in child node=75 (Khan et al., 2015).

3.5 Results and Discussion

The misclassification costs of the classification and regression tree (CRT) indicating predictive accuracy are shown in Table 2. It was found that the overall percent correct is 68.4% which is considered acceptable. (L. Y. Chang & Chen, 2005; Khan et al., 2015).

Table 3.2 Misclassification costs

Observed	Predicted		
	Non-intersection	Intersection	Percent Correct
Non-intersection	2564	1174	68.6%
Intersection	264	552	67.6%
Overall Percentage	62.1%	37.9%	68.4%

According to the results from tree model as shown in Figure 3.2, it was found that the number of nodes = 13, terminal = 7, and Depth = 5 starting from the root node which is intersection where the rear-end collisions occur at 17.9%. The most significant variable is crash type which is decelerating speed (35.5%). This makes sense because at the intersection where rear-end crashes occur, most cars often have to slow down to approach the intersection (Wiacek et al., 2015), followed by traffic volume which is less than 153,624 vehicles per day. The number of this traffic volume is often obtained on the main roads or inter-city highways which sometimes cross into downtown with plenty of intersections.

The secondly consequential variable is road pavement of which concrete type is more likely to cause rear-end crashes than other types (19.4%). This result is not consistent with the studies of Flask et al. (2014); Zhan et al. (2015) indicating that the concrete road surface could reduce the number of accidents due to its greater friction coefficient. When considering the coefficients of adhesion for different pavement surfaces in the study of Wang et al. (2012), it found that in the dry condition the peak friction coefficient of concrete pavement and asphalt pavement were equaled (0.7-0.8). So friction coefficient may actually not affect. Moreover, according to the parent node was the crash type which is stopped in a road, and going straight. It was not related to breaking factors.

The subsequent variable is ‘envi light’ which found that the rear-end crashes much potentially occur at daytime and nighttime with illuminated light or having lighting poles (10.9%) as the clear visibility making drivers drive the cars with speed so great that the drivers of back car cannot manage to stop them when arriving at the intersection. This causes high chances of rear-end crashes at the intersections.

Leaf node is traffic volume for adequate vision. Traffic volume which is in the range of 8,949 -29,218 vehicles per day has high chances to cause rear-end crashes at the intersection. The highways having the mentioned traffic volume are often major highways connecting between provinces and districts. This result is consistent with the study of L. Y. Chang and Chen (2005) that ADT 20,000 vehicles per day resulted in Highway accident frequency. Yan et al. (2005) have described the relationship between ADT and rear-end crashes at the intersection that if ADT increases every 2,000 vehicles per day, the chances of rear-end crashes will increase by 12%. The reason is that the decrease in headway of vehicles certainly affects the likelihood of rear-end collisions.

The measurement of importance variables of CART which is a prediction variable X in relation of the final tree. It was defined as the weighted sum across all splits in the tree of the improvement. In other word it was measurement of attributable to each variable in its role as a surrogate to the primary split (Banerjee et al., 2008). The highest score is 100 which means that variable was the best performing variable. The other significant variables but not shown in tree model due to their small proportions as shown in Fig.3 (Variable importance), consists of using excessive speed, truck volume, age of driver, and the using of safety equipment, etc. These variable might effect to the occurrence of rear-end crash on intersection.

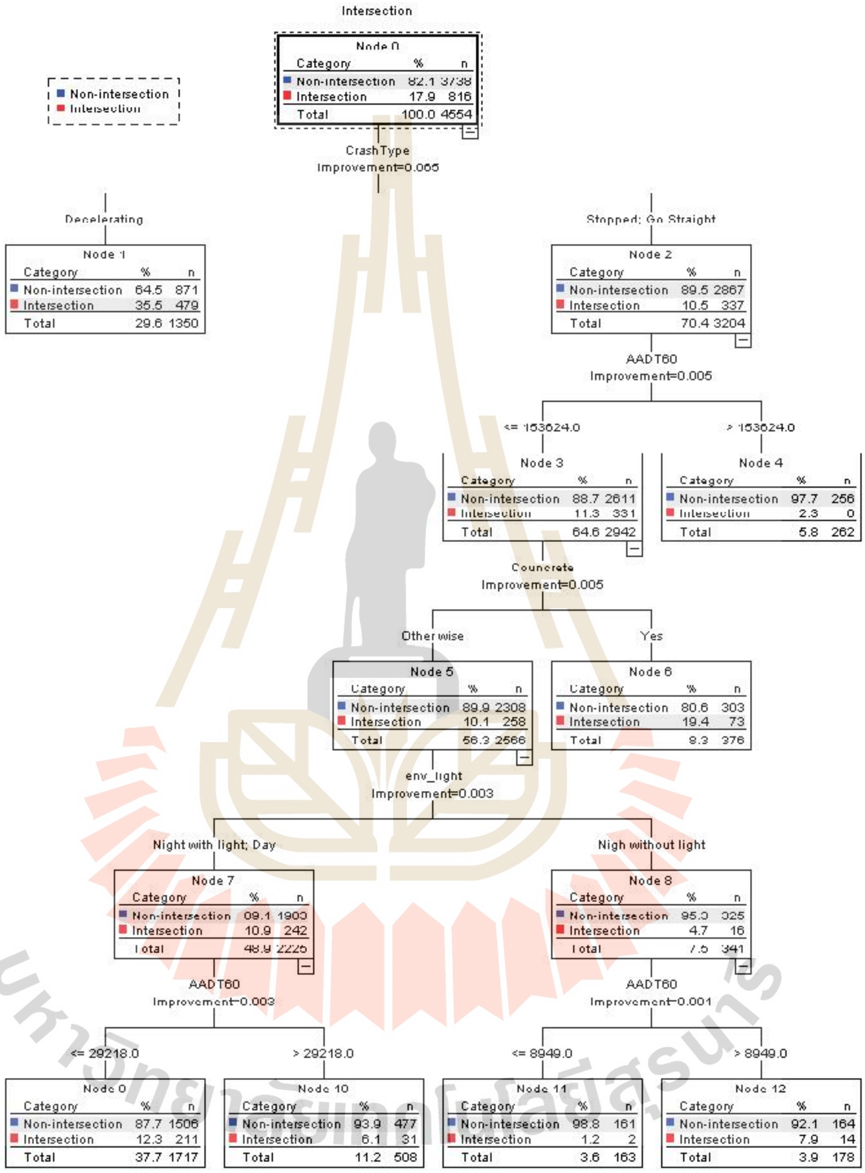


Figure 3.2 Tree model

3.6 Conclusion and implementation

This study aims to find a way to reduce the number of rear-end crashes at the intersection, by analyzing the factors likely to cause them by using the Classification and regression tree due to its ability to analyze the relationship between complicated variables. In addition, the misclassification costs of the model are considered acceptable.

The most significant variable of rear-end crash at intersections model is speed deceleration. The secondly significant variable is concrete pavement road which increase the likelihood of rear-end crashes. Lastly, rear-end crash is likely occur at nighttime with present of lighting poles.

Related agencies such as Department of Highways, Department of Land Transport (driver training license) potentially apply the study results to reduce the number of rear-end crashes at the intersection. The mostly important factor was rear-end crash type, which was decelerating speed. It can be emphatically concluded that during the rear-end crash, the front vehicle driver is decelerating the speed to enter the intersection. Highway authorities may consider warning signs recommending the drivers to slow down before reaching the intersection. According to the light condition factors, it found that the rear-end crash on day or night time with light. The driver training should be emphasized that might be ‘don’t use high speed when entering the intersection’.

This is in line with the results which acquire ADT in the range of 9,000 - 30,000 vehicles per day. This policy should be urgently considered. Additionally, Official departments of highway should review about number of rear-end crash at intersection of road with 9,000 – 30,000 vehicle per day and concrete pavement. Then, they could consider improving some intersections with high rear-end crash rate.

In addition, speeding before entering the intersection will cause rear-end crashes in the case of pavement type, because there has been no evident research that concrete road surface results in higher rear-end collisions at the intersection. In the future, this issue can be taken to further investigation.

Regarding to the other countries where is developing country, this result could be applied to reduce the rear-end crash at intersection.

The limitation of this research is that the number of variables resulting from the tree model is small; the trend of those variable's effects cannot be indicated. In the future, the parametric analysis method, such as binary logistic regression, can be used for the mentioned trend analysis. Additionally, other variables to be potentially analyzed are specified their importance as shown in Fig. 3.3.

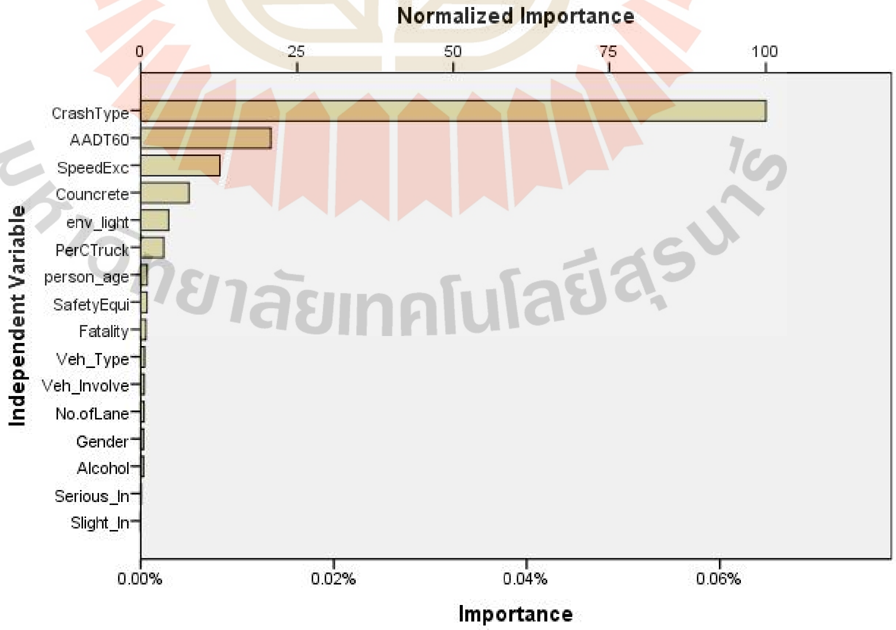


Figure 3.3 Variable importance

3.7 Reference

Abdel Aty, M., and Abdelwahab, H. (2004). Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. **Accident Analysis & Prevention**, 36(3): 447-456.

Agouti, T., Gharnati, F., and Ait-Mlouk, A. (2017). Application of Big Data Analysis with Decision Treefor Road Accident. **Indian Journal of Science and Technology**, 10(29): 1-10.

Anvari, M. B., Tavakoli Kashani, A., and Rabieyan, R. (2017). Identifying the Most Important Factors in the At-Fault Probability of Motorcyclists by Data Mining, Based on Classification Tree Models. **International Journal of Civil Engineering**, 15(4): 653-662.

Banerjee, A. K., Arora, N., and Murty, U. S. N. (2008). Classification and Regression Tree (CART) Analysis for Deriving Variable Importance of Parameters Influencing Average Flexibility of CaMK Kinase Family.

Chanbunditayanun, W. (2017). The major factor of accident in Thailand. Retrieved from http://www.js100.com/th/site/post_share/view/43883

Chang, L.-Y., and Chien, J.-T. (2013). Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. **Safety Science**, 51(1): 17-22.

Chang, L. Y., and Chen, W. C. (2005). Data mining of tree-based models to analyze freeway accident frequency. **Journal of Safety Research**, 36(4): 365-375.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., and Guan, H. (2015). A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. **Accident Analysis & Prevention**, 80: 76-88.

Department of Highway. (2016). Thailand Traffic Accident on National Highways in 2015. Bureau of Highway Safety, Department of Highways. Retrieved from http://bhs.doh.go.th/files/accident/58/report_accident2558.pdf

Department of Highway. (2017). Thailand Traffic Accident on National Highways in 2016. Bureau of Highway Safety Department of Highways Ministry of Transport, . Retrieved from http://bhs.doh.go.th/files/accident/59/report_accident2559.pdf

Department of Highway. (2018). Thailand Traffic Accident on National Highways in 2017. Bureau of Highway Safety Department of Highways Ministry of Transport, . Retrieved from http://bhs.doh.go.th/files/accident/60/report_accident2560.pdf

Flask, T., Schneider, W. H., and Lord, D. (2014). A segment level analysis of multi-vehicle motorcycle crashes in Ohio using Bayesian multi-level mixed effects models. **Safety Science**, 66: 47-53.

Harb, R., Radwan, E., Yan, X., and Abdel-Aty, M. (2007). Light truck vehicles (LTVs) contribution to rear-end collisions. **Accident Analysis & Prevention**, 39(5): 1026-1036.

Iamtrakul, P. (2008). *Analysis of factors affecting on rear-end crashed accidents*. Retrieved from http://roadsafetythai.org/uploads/userfiles/ACC_51016.pdf

IBM. (2012). IBM SPSS Decision Trees 21. Retrieved from www.sussex.ac.uk/its/pdfs/SPSS_Decision_Trees_21.pdf

Kashani, A. T., and Mohaymany, A. S. (2011). Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. **Safety Science**, 49(10): 1314-1320.

Khan, G., Bill, A. R., and Noyce, D. A. (2015). Exploring the feasibility of classification trees versus ordinal discrete choice models for analyzing crash severity. **Transportation Research Part C: Emerging Technologies**, 50: 86-96.

Kim, S., Song, T.-J., Rouphail, N. M., Aghdashi, S., Amaro, A., and Gonçalves, G. (2016). Exploring the association of rear-end crash propensity and micro-scale driver behavior. **Safety Science**, 89: 45-54.

Li, X., Yan, X., Wu, J., Radwan, E., and Zhang, Y. (2016). A rear-end collision risk assessment model based on drivers' collision avoidance process under influences of cell phone use and gender-A driving simulator based study. **Accident Analysis & Prevention**, 97: 1-18.

Li, Z., Wang, W., Chen, R., and Liu, P. (2014). Conditional inference tree-based analysis of hazardous traffic conditions for rear-end and sideswipe collisions with implications for control strategies on freeways. **IET Intelligent Transport Systems**, 8(6): 509-518.

Ma, L., and Yan, X. (2014). Examining the nonparametric effect of drivers' age in rear-end accidents through an additive logistic regression model. **Accident Analysis & Prevention**, 67: 129-136.

Mendez, A. G., and Izquierdo, F. A. (2010). Quasi-induced exposure: the choice of exposure metrics. **Accident Analysis & Prevention**, 42(2): 582-588.

Meng, Q., and Qu, X. (2012). Estimation of rear-end vehicle crash frequencies in urban road tunnels. **Accident Analysis & Prevention**, 48: 254-263.

Muhammad, L. J., Salisu, S., Yakubu, A., Malgwi, Y. M., Abdullahi, E. T., Mohammed, I. A., and Muhammad, N. A. a. (2017). Using Decision Tree Data Mining Algorithm to Predict Causes of Road Traffic Accidents, its Prone Locations and Time along Kano –Wudil Highway. **International Journal of Database Theory and Application**, 10(1): 197-206.

Nikiforos, S. (2008). Quasi-Induced Exporsure: Issue and Validation: Nova Science Publishers, Inc.

Nikiforos, S., and John, A. D. (1997). Quasi-Induced Exposure: Methodology and Insight. **Accident Analysis & Prevention**, 29(1): 37-52.

Pakgohar, A., Tabrizi, R. S., Khalili, M., and Esmaeili, A. (2011). The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. **Procedia Computer Science**, 3: 764-769.

Pande, A., Abdel-Aty, M., and Das, A. (2010). A classification tree based modeling approach for segment related crashes on multilane highways. **Journal of Safety Research**, 41(5): 391-397.

Song, Y. Y., and Lu, Y. (2015). Decision tree methods: applications for classification and prediction. **Shanghai Archives of Psychiatry**, 27(2): 130-135.

Wang, Y.-g., Chen, K.-m., and Hu, L.-w. (2012). Killer Tailgating: Recommendation of Traveling Intervals between Consecutive Motor Vehicles for Rear-end Collision Avoidance. **Arabian Journal for Science and Engineering**, 37(3): 619-630.

Weng, J., Meng, Q., and Yan, X. (2014). Analysis of work zone rear-end crash risk for different vehicle-following patterns. **Accident Analysis & Prevention**, 72: 449-457.

Weng, J., Xue, S., Yang, Y., Yan, X., and Qu, X. (2015). In-depth analysis of drivers' merging behavior and rear-end crash risks in work zone merging areas. **Accident Analysis & Prevention**, 77: 51-61.

Wiacek, C., Bean, J., and Sharma, D. (2015). Real-world analysis of fatal rear-end crashes. *24th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration.* Retrieved from https://www-esv.nhtsa.dot.gov/proceedings/24/files/24ESV-000270.PDF

Yan, X., and Radwan, E. (2006). Analyses of rear-end crashes based on classification tree models. **Traffic Injury Prevention**, 7(3): 276-282.

Yan, X., Radwan, E., and Abdel-Aty, M. (2005). Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. **Accident Analysis & Prevention**, 37(6): 983-995.

Yinhai Wang, Hitoshi Ieda, and Fred Mannering. (2002). *Estimating Rear-End Accident Probabilities at Signalized Intersections: An Occurrence-Mechanism Approach*. Retrieved from

Zhan, X., Aziz, H. M. A., and Ukkusuri, S. V. (2015). An efficient parallel sampling technique for Multivariate Poisson-Lognormal model: Analysis with two crash count datasets. **Analytic Methods in Accident Research**, 8: 45-60.

Zheng, Z., Lu, P., and Tolliver, D. (2016). Decision Tree Approach to Accident Prediction for Highway–Rail Grade Crossings. **Journal of the Transportation Research Board**, 2545(1): 115-122.

CHAPTER IV

ANALYSIS OF REAR-END CRASH ON THAI HIGHWAY: DECISION TREE APPROACH

4.1 Abstract

Objective: Among crash types on Thai highways, rear-end crashes have been found to cause the largest number of fatalities. This study aims to find ways to decrease rear-end crashes and fatal rear-end crashes. Methods: Classification and regression tree (CART) was used to analyze the complicated relationship of variables of big data. The analysis was conducted by creating two models: 1) a model which indicates the causes of rear-end crashes by applying Quasi-Induced Exposure to at-fault driver characteristics; 2) a determined model which studies fatal crashes. Results: Predictor variables in the model of at-fault and not-at-fault drivers found that driver age is most significant, followed by number of lanes and median opening area. For the mode of fatality, the use of safety equipment was found to be most importance. Conclusion: The model results can be used to develop guidelines for public awareness programs for motorists and to propose policy changes to the Department of Highway in order to reduce the severity of rear-end crashes. Moreover, this paper discusses the variables that may result in both the perspective of rear-end crash number and the fatality rate of rear-end crashes as strategies in future research.

4.2 Introduction

Crash trends on Thailand highways are continuously on the increase (Department of Highway, 2016, 2017, 2018). Crash type statistics reveal that rear-end collision is the second most common type of collision. However, the highest number of fatalities occur as a result of rear-end collisions (Figure 4.1). Therefore, finding strategies to decrease the number and severity of rear-end crashes is urgently needed.

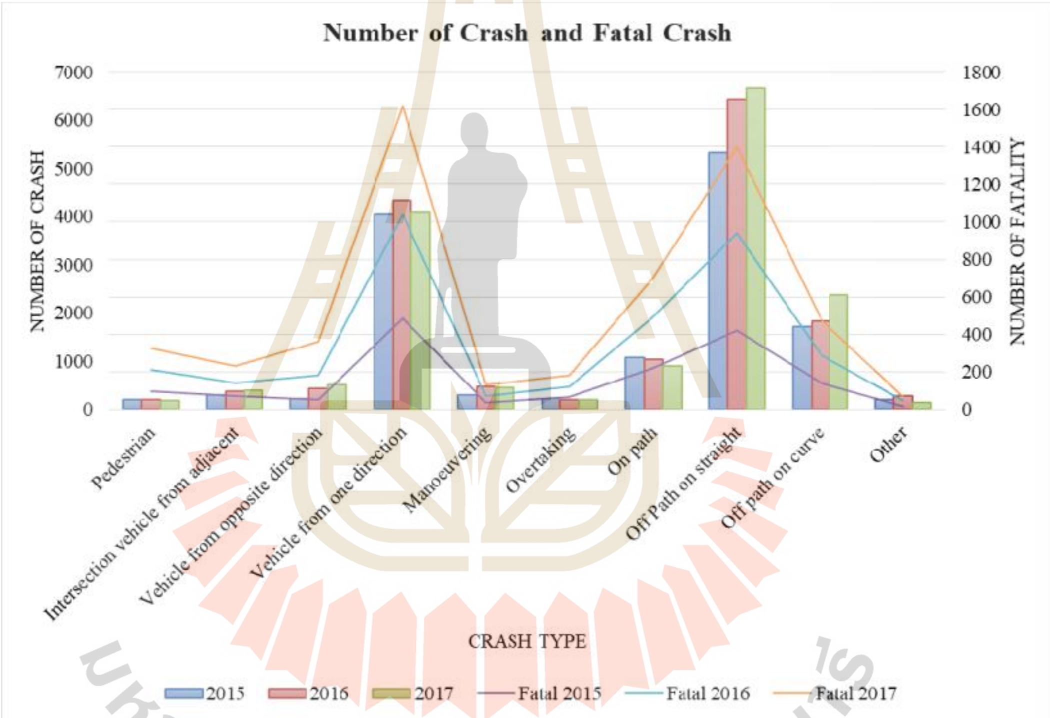


Figure 4.1 Crash on Thai highways by crash type

There are two important issues in the study of rear-end crashes at intersections. First, a study of the causes of rear-end crashes, focusing on at-fault and not-at-fault drivers, has found that most crashes are caused by drivers not leaving enough space between their own car and cars in front (Department of Disaster

Prevention and Mitigation, 2014). Therefore, the cause of rear-end collisions is the car behind (L. Ma&Yan, 2014). This study focuses on the driver characteristics of the at-fault driver, that is, the driver of the car behind that crashes into the car in front, by applying Quasi-Induced Exposure Methods (Carr B.R., 1970). These methods have been widely used in the field of traffic accident research. The principle of these methods is to predict the at-fault driver based on the accident report (Chandraratna&Stamatiadis, 2009; Taha&Vinayak, 2013) by supposing that the distribution of not-at-fault drivers closely represents the distribution of exposure to accident hazards (X. Yan&Radwan, 2006; X. Yan et al., 2005). Second, this research explores the relevance of the high fatality rate caused by rear-end crashes. Fatal crashes must be considered from a characteristic study of rear-end crashes by focusing on ways to reduce fatalities. Sullivan and Flannagan (2003) studied fatal crash risks and found that darkness is the risk factor causing the greatest number of fatalities, due to the invisibility of vehicles parked along the roadside. Wiacek et al. (2015) found that the greater the difference in velocity of the struck car and the striking car the higher the number of rear-end crash fatalities. However, if a truck is involved in a rear-end crash, the chances of fatality are further increased.

There are numerous factors affecting the causes of rear-end crashes. These include driver characteristics that affect driving decision making, such as driver characteristics (gender, age, alcohol use) (Anvari et al., 2017; X. Li et al., 2016; Nikiforos&John, 1997), environment (time, weather conditions), roads (surface condition, physical characteristics) (Z. Li et al., 2014; Mendez&Izquierdo, 2010), vehicle type (Nikiforos, 2008), and number of traffic lanes (Kim et al., 2016; X. Li et al., 2016; Meng&Qu, 2012).

Previous research has found that factors causing death in rear-end crashes included driver characteristics affecting braking, such as gender, age, and alcohol or substance abuse (Chen et al., 2015). Use of a seatbelt has been found to be another important contributing factor to rear-end fatalities (Chen et al., 2016). Vehicle type is an important factor in all accident types (Weng et al., 2014), but especially in rear-end crashes. If the types of vehicle involved in a crash are very different, the chances of severity are higher (Xuedong Yan&Radwan, 2009). Speed limit factors also affect the severity of the crash (Wiacek et al., 2015). Other important characteristics of fatal crashes are physical road characteristics and visibility (Chen et al., 2015).

A statistical analysis of rear-end crashes involves independent variables, such as weather conditions, vehicle type, seat belt use, and dependent variables, such as at-fault driver and not-at-fault driver, and fatal and non-fatal rear-end crash. The distribution analysis method has been widely used to generalize whether an estimated parameter exists. If there is an estimated parameter, the relationship between independent and dependent variables is considered. If there is no estimated parameter, data are investigated proportionally. X. Yan and Radwan (2006) have stated that there are limitations to the use of parametric analysis (binary logistic regression) due to the difficulty in using it to investigate the relationship between two variables. Thus, an appropriate alternative is non-parametric analysis, or Decision tree or classification tree (DT). This is an algorithmic arrangement to perceive proportions of data according to determined dependent variables (also known as data mining) (Agouti et al., 2017). Thus, appropriate data can be used to analyze complex independent data (X. Yan & Radwan, 2006; Zheng et al., 2016). A decision tree is a structure that includes a root node, branches, and leaf nodes (Muhammad et al., 2017). X. Yan and

Radwan (2006) have used DT to study rear-end crash data in Florida, USA, by analyzing two models. The first was an analysis of which accidents involved rear-end crashes, and the second was an analysis of driver characteristics of individuals who could potentially become at-fault-drivers.

In choosing a model for this study, other models that can analyze the relationship between independent variables and target or categorical variables were considered. A traditional model using multiple logistic regression which has been widely used (Xuedong Yan & Radwan, 2009). Another common model is the multinomial log it model which theoretically analyzes data using the nested log it model (NLM), which can examine hierarchical dependent variables (Abdel Aty &Abdelwahab, 2004). Odds ratio is used to interpret probability. The advantage of this method is the ability to compare the effects of explanatory variables on dependent variables, especially when independent variables result in statistical significance. However, the limitation of each of these models is their inability to find relationships between explanatory variables. The Decision Tree Model (DT), however, potentially solves this problem. As mentioned earlier, rear-end crashes are the cause of high fatalities. Therefore, the presentation of this model simultaneously identifies relationships between independent variables, which may allow for the application of findings to policy development. For example, an examination of whether the different ages of drivers in different traffic lanes affects the role of the driver (at-fault / not-at-fault) in a crash can influence the development of effective policy. Research by Khan et al. (2015), which compared DT and ordinal discrete choice model, confirmed that DT can help to address issues of multicollinearity and variable redundancy.

Among studies that have analyzed rear-end crashes (Table 4.1), most have analyzed crash frequency, followed by crash severity (fatal/non-fatal). One study has analyzed both crash frequency and severity outcome (Das & Abdel-Aty, 2011). However, the crash data used in that study came from a country with different roads, conditions, driver behaviors to Thailand, leading to the development of a very different model. No concentrated road crash study of highways in Thailand has been conducted which applies the DT model to the reduction of the number of rear-end crash fatalities and fatal rear-end crashes. This research will discuss model consistency with the number of fatalities, by comparing the two with previous studies as a guideline for conducting future research.

Table 4.1 Comparison with others studies in analysis of rear-end crash field

Studies/Model	At-fault / not-at-fault	Fatal Injury	Comparison of two models	Raised issues
This study	✓	✓	✓	Case accidents in Thailand; Comparison of the result of two model.
Yan, et al. [10]	✓			Rear-end crash at signalized intersections.
Yan and Radwan [9]	✓			Model#1: Rear-end vs Non-Rear-end, Model#2: At-fault / not-at-fault.
Chandraratna and Stamatiadis [7]	✓			Evaluation of not-at-fault assumption.
Meng and Qu [20]	✓			At Urban road tunnels.
Ma and Yan [5]	✓			Focused on age of driver.
Weng, et al. [23]	✓			On work zone.
Chen, et al. [21]		✓		Hybrid approach.
Chen, et al. [22]		✓		Hybrid classifier.
Sullivan and Flannagan [11]		✓		Lighting conditions.
Joon-Ki, et al. [40]	✓			Probability of Freeway Rear-End Crash Occurrence.
Das and Abdel-Aty [30]	✓	✓	✓	Genetic programing approach, Rear-end crash in Florida.

Note: At-fault or not-at-fault driver are assumed related to the rear-end crash frequency

4.3 Highway Crash Reporting

This study used Department of Highway (DOH) road accident data from 2011 to 2015. These data included dates, road segments, physical characteristics of accident scenes (e.g. straight road, curved road, work zone, median, intersection), environmental conditions (e.g. rain, lighting conditions, time of accident), cause-and-effect data (e.g. driving over the speed limit) and injury data (including fatalities, serious injuries and minor injuries). The information provided by the DOH may not cover all accidents. In cases of minor collisions, where victims came to an agreement, accidents were not recorded.

Rear-end type collisions were selected from these data, and divided into three main types according to the movement of the front car prior to the collision (L. Ma & Yan, 2014). These three are 1) going straight, with the front car traveling at normal speed, 2) decelerating speed, with the front car decelerating, such as when turning the car or executing a u-turn, and 3) stopping, with the front car parked on the roadside or on the hard shoulder or stopped at traffic lights. After screening, there were 2,096 cases of rear-end collision. As vehicle data had to be considered in this analysis, driver and vehicle factors were added to the model. The dataset comprised 5,445 vehicles involved in accidents.

Table 4.2 Categorical Variables' characteristics and descriptive statistics

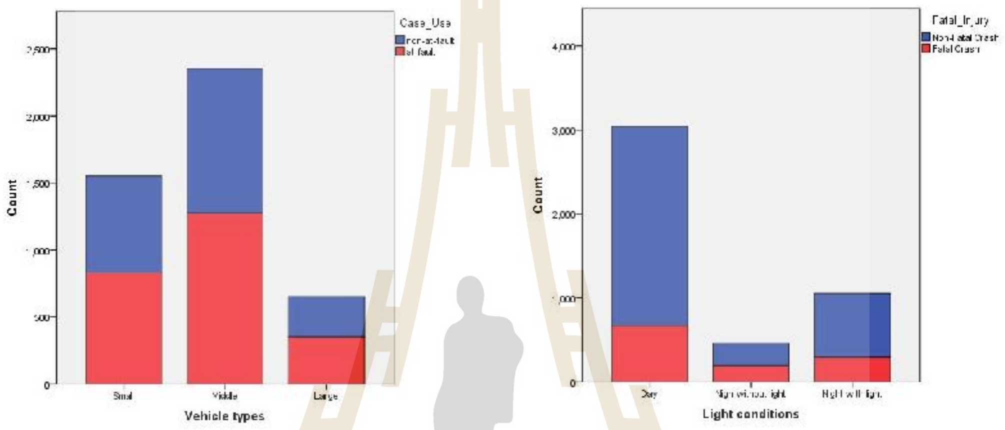
		Driver Exposure				Fatal Injury			
		not-at-fault		at-fault		Fatal Crash		Non-Fatal	
		Count	%	Count	%	Count	%	Count	%
Veh_Type	Small	718	15.8%	833	18.3%	387	8.5%	1164	25.6%
	Middle	1078	23.7%	1274	28.0%	533	11.7%	1819	39.9%
	Large	300	6.6%	351	7.7%	236	5.2%	415	9.1%
CrashType	Stopped	1344	29.5%	1611	35.4%	721	15.8%	2234	49.1%
	Decelerating	645	14.2%	705	15.5%	347	7.6%	1003	22.0%
	Go Straight	107	2.3%	142	3.1%	88	1.9%	161	3.5%
Gender	Female	1731	38.0%	2037	44.7%	1000	22.0%	2768	60.8%
	Male	365	8.0%	421	9.2%	156	3.4%	630	13.8%
Main_Road	Other wise	1895	41.6%	2193	48.2%	1070	23.5%	3018	66.3%
	Yes	201	4.4%	265	5.8%	86	1.9%	380	8.3%
Entran_Exit	Other wise	2084	45.8%	2448	53.8%	1154	25.3%	3378	74.2%
	Yes	12	0.3%	10	0.2%	2	0.0%	20	0.4%
Non-Repairing road	Other wise	54	1.2%	59	1.3%	28	0.6%	85	1.9%
	Yes	2042	44.8%	2399	52.7%	1128	24.8%	3313	72.7%
road_lane	2	710	15.6%	792	17.4%	385	8.5%	1117	24.5%
	3	10	0.2%	11	0.2%	4	0.1%	17	0.4%
	4	926	20.3%	1083	23.8%	573	12.6%	1436	31.5%
	5	9	0.2%	8	0.2%	5	0.1%	12	0.3%
	6	155	3.4%	202	4.4%	63	1.4%	294	6.5%
	7	0	0.0%	2	0.0%	0	0.0%	2	0.0%
	8	160	3.5%	205	4.5%	92	2.0%	273	6.0%
	9	3	0.1%	4	0.1%	5	0.1%	2	0.0%
	10	33	0.7%	59	1.3%	21	0.5%	71	1.6%
	12	60	1.3%	65	1.4%	8	0.2%	117	2.6%
	14	30	0.7%	27	0.6%	0	0.0%	57	1.3%
road_isle	No median	779	17.1%	861	18.9%	399	8.8%	1241	27.3%
	Flush	115	2.5%	133	2.9%	96	2.1%	152	3.3%
	Raised	452	9.9%	561	12.3%	298	6.5%	715	15.7%
	Depressed	513	11.3%	612	13.4%	264	5.8%	861	18.9%
	Barrier	237	5.2%	291	6.4%	99	2.2%	429	9.4%
Asphalt	Other wise	242	5.3%	309	6.8%	124	2.7%	427	9.4%
	Yes	1854	40.7%	2149	47.2%	1032	22.7%	2971	65.2%
Straight	Other wise	125	2.7%	172	3.8%	96	2.1%	201	4.4%
	Yes	1971	43.3%	2286	50.2%	1060	23.3%	3197	70.2%
Slope	Other wise	2039	44.8%	2358	51.8%	1090	23.9%	3307	72.6%
	Yes	57	1.3%	100	2.2%	66	1.4%	91	2.0%
Intersection	Non-intersection	1725	37.9%	2013	44.2%	985	21.6%	2753	60.5%
	Intersection	371	8.1%	445	9.8%	171	3.8%	645	14.2%
Median_opening	Other wise	1853	40.7%	2129	46.8%	985	21.6%	2997	65.8%
	Yes	243	5.3%	329	7.2%	171	3.8%	401	8.8%
env_surfaces	Dry	1951	42.8%	2284	50.2%	1073	23.6%	3162	69.4%
	Wet	145	3.2%	174	3.8%	83	1.8%	236	5.2%
Weather	Otherwise	1947	42.8%	2266	49.8%	1063	23.3%	3150	69.2%
	Clean	149	3.3%	192	4.2%	93	2.0%	248	5.4%
env_light	Day	1413	31.0%	1629	35.8%	669	14.7%	2373	52.1%
	Nigh without light	217	4.8%	243	5.3%	194	4.3%	266	5.8%
	Night with light	466	10.2%	586	12.9%	293	6.4%	759	16.7%
SafetyEqui	Other wise	1312	28.8%	1585	34.8%	856	18.8%	2041	44.8%
	Use	784	17.2%	873	19.2%	300	6.6%	1357	29.8%
Alcohol	Other wise	2070	45.5%	2433	53.4%	1131	24.8%	3372	74.0%
	Yes	26	0.6%	25	0.5%	25	0.5%	26	0.6%
SpeedExc	Other wise	651	14.3%	734	16.1%	387	8.5%	998	21.9%
	Yes	1445	31.7%	1724	37.9%	769	16.9%	2400	52.7%

Table 4.3 Descriptive statistics of continuous variables

	Driver Exposure							
	not-at-fault				at-fault			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.
Person_age	38.6	14.0	9.0	86.0	38.0	13.9	11.0	85.0
Veh_Involve	2.2	0.5	2.0	8.0	2.4	0.8	2.0	8.0
AADT60	33,318	53,089	183	270,050	33,779	52,363	163	270,050
PerCTruck	16.4	10.6	0.0	68.0	17.0	10.8	0.0	68.0
	Fatal Injury							
	Fatal Crash				Non-Fatal Crash			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.
Person_age	39.8	14.3	12.0	86.0	37.8	13.8	9.0	85.0
Veh_Involve	2.3	0.7	2.0	7.0	2.3	0.7	2.0	8.0
AADT60	26,583	39,286	483	270,050	35,943	56,342	163	270,050
PerCTruck	18.2	11.9	1.1	68.0	16.2	10.2	0.0	59.4

Note: Person_age = age of deriver; Veh_Involve = number of vehicle involvement; AADT60 = Average annual traffic volume on 2017; PerCTruck = Average truck volume (%).

Descriptive statistics, shown in Table 4.2, define the dependent variables: 1) The at-fault driver is the driver of the striking car, while the not-at-fault driver is the driver of the struck vehicle, 2) Fatal rear-end crash refers to a collision with at least one fatality either at the accident scene or at the hospital, while non-fatal rear-end crash denotes rear-end crash without fatality. For all 22 independent variables of the two models, they exhibited with the values of the two dependent variables. Data description was displayed to help illustrate the overall picture created by the data (L. Ma et al., 2016; Lu Ma et al., 2015). After cleaning the data for driver exposure, there were 2,458 at-fault drivers and 2,096 not-at-fault drivers. With regard to crash fatalities, 1,156 vehicles were involved in fatal rear-end crashes, and 3,396 vehicles were involved in non-fatal rear-end crashes. According to vehicle type (Veh_Type), medium cars, such as private cars and pickup trucks had a 28.0% chance of being at-fault. According to fatal rear-end collisions, larger vehicles were the cause of 11.7% of collisions, and small vehicles, such as motorcycles, were the cause of 8.5% of accidents (Figure 4.2 (a)). Light condition (env_light) was the dominant environmental factor affecting fatalities, with 42.2% of rear-end collision fatalities occurring at night in the absence of light, 27.9% occurring at night with light, and 22% occurring in the daytime (Figure 4.2 (b)).



(a) (b)

Figure 4.2 Relationship between dependent variables and some independent variables

The distribution of continuous variables is shown in Table 4.3. With regard to driver age distribution, there was little difference between the ages of at-fault drivers and not-at-fault drivers. The average age of at-fault drivers was 38.04 years and of not-at-fault driver was 38.58 years. The mean value of trucks involved in fatal rear-end crashes was 18.2% and in non-fatal rear-end crashes was 16.2%. The DT model was then used for further analysis. Predictions could then be presented as logical if-then conditions at the terminal node. Thus, data did not require normal distribution. In other words, the relationships between independent and dependent variables was not obligatory for the existence of linear relationships (Akanbi et al., 2015).

Relationships between the independent variables are shown in the pair wise coefficient correlation model (Table 4.4). Two highly correlated pairs were found: 1) Road surface factor (env surface) correlated with weather condition (r = 0.840). This was particularly evident in cases where there were unusual conditions, such as rain resulting in a wet road surface; 2) Factor of the number of traffic lanes and median type (r = 0.621). This relationship is rational, as roads in Thailand typically have four or more traffic lanes and median types usually include a depressed median and barrier. Some pairs exhibited no relationship, such as driver age and road slope, or driver gender and road surface type.

4.4 Methods

4.4.1 Variable setting

The dependent variables were determined as categorical values, such as fatal = 1, non-fatal = 0. According to independent variables, there were two variable types: 1) categorical variables, the values of which were divided according to variable characteristics in numeric form, for example, gender (0 = male, 1 = female), vehicle types (1 = small vehicle, i.e., motorcycle, 2 = medium vehicle, i.e., car, pickup truck, 3 = large vehicle, i.e., six-wheel truck), crash types (1 = going straight, 2 = decelerating, 3 = stopped), and 2) continuous variables, such as number of lanes (2,3,4, ...). 'Per C Truck' was the proportion of trucks traffic volume.

Table 4.4 Correlation among independent variables

	1.	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**	**9.**	**10**	**11.**	**12.**	**13.**	**14.**	**15.**	**16.**	**17.**	**18.**	**19.**	**20.**	**21.**	**22.**
1.Vehicle types	1.00	**-0.05**	**0.11**	**-0.20**	**0.07**	0.02	-0.01	**0.14**	**0.16**	-0.03	**-0.05**	**0.05**	-0.03	0.02	**0.07**	**0.08**	**0.06**	-0.02	-0.02	**0.09**	**0.15**	**0.03**
2.CrashType		1.00	0.01	0.02	**-0.09**	0.00	0.01	**-0.10**	**-0.10**	0.00	**0.04**	**-0.07**	**0.16**	**0.16**	**-0.10**	**-0.08**	-0.02	**-0.05**	0.00	**-0.16**	**-0.10**	**-0.23**
3.person_age			1.00	**-0.13**	-0.01	-0.02	-0.01	0.02	0.03	-0.01	0.02	0.00	0.01	0.02	0.00	0.01	-0.01	0.00	0.02	0.01	0.00	**-0.04**
4.Gender				1.00	-0.02	-0.01	-0.01	**-0.05**	-0.04	0.00	0.02	-0.02	0.03	-0.03	-0.02	-0.02	-0.08	0.04	-0.01	**-0.04**	**-0.04**	0.00
5.Main_Road					1.00	-0.02	-0.01	**0.40**	**0.23**	**-0.17**	**0.06**	**-0.04**	**-0.05**	**-0.04**	0.01	0.01	0.08	0.04	-0.01	**0.35**	**0.07**	**0.06**
6.Entran_Exit						1.00	0.01	**0.13**	**0.05**	-0.01	0.02	-0.01	0.02	-0.03	-0.02	-0.02	0.00	0.00	-0.01	**0.13**	0.02	0.03
7.Normal							1.00	**0.06**	**0.05**	-0.03	0.02	0.00	0.00	0.00	0.03	0.03	0.00	-0.02	0.02	**0.05**	0.01	-0.03
8.road_lane								1.00	**0.62**	**-0.22**	**0.11**	**-0.09**	**-0.12**	**0.05**	-0.01	-0.01	**0.12**	**0.06**	**-0.05**	**0.60**	**0.17**	**0.09**
9.road_isle									1.00	**-0.13**	**0.12**	**-0.08**	**-0.12**	**0.16**	0.02	0.03	**0.09**	**0.09**	**-0.04**	**0.48**	**0.26**	**0.06**
10.Asphalt										1.00	**-0.05**	**0.04**	**-0.06**	**-0.08**	**0.06**	0.03	**-0.10**	**0.07**	0.03	**-0.10**	**-0.04**	**-0.06**
11.Straight											1.00	**-0.39**	0.03	0.03	**-0.11**	**-0.08**	0.03	**0.05**	0.01	**0.07**	**-0.07**	0.01
12.Slope												1.00	**-0.05**	**-0.04**	0.05	0.04	-0.02	**-0.07**	0.00	**-0.06**	**0.09**	**-0.04**
13.Intersection													1.00	**0.08**	**-0.07**	**-0.06**	0.00	**-0.06**	0.03	**-0.11**	**-0.05**	**-0.09**
14.Median_open														1.00	-0.02	-0.03	-0.01	-0.01	0.00	**-0.05**	0.03	**-0.11**
15.env_surfaces															1.00	**0.84**	-0.02	0.00	-0.01	-0.03	**0.04**	**0.07**
16.Weather																1.00	0.02	0.00	0.01	-0.03	**0.04**	**0.06**
17.env_light																	1.00	**-0.05**	0.02	**0.10**	**0.06**	**0.04**
18.SafetyEqui																		1.00	-0.01	**0.08**	-0.01	0.00
19.Alcohol																			1.00	**-0.04**	**-0.04**	**-0.08**
20.AADT60																				1.00	0.03	**0.12**
21.PerCTruck																					1.00	0.03
22.SpeedExc																						1.00

Note: Bold number represented correlation, which is significant at the 0.05 level (2-tailed).

4.4.2 Classification tree and building model

This study used a decision tree or classification tree (DT) model for rear-end crash data analysis, which started by determining target variables (dependent variables). Two models were constructed. Model#1 analyzed at-fault/not-at-fault drivers. In order to consider this variable, the driver factor was only selected for the first and second vehicles, as the first vehicle was clearly identifiable as accident-prone. Therefore, 4,192 vehicles (2,096 rear-end crashes) were analyzed in the model. Model#2 was an analysis of factors resulting in fatal and non-fatal rear-end collisions. Therefore, data included the two or more vehicles involved in a rear-end crash. Out of a total of 4,554 vehicles, 2,096 were involved in those crashes.

The DT model consists of three components. These are decision node, branches, and leaf nodes. Within the DT structure, each decision node displays the variable, and each branch displays one variable value based on decision rules, while leaf nodes exhibit the expected values of target variables (Song & Lu, 2015).

SPSS was used to conduct the analysis. In order to create the DT, the full dataset was first split according to root node, which was the proportion of values in the target variable. This was then split into a number of smaller subsets. Several SPSS types can be used to carry out splitting and growing, including CHAID, CART and QUEST. Each of these types has advantages and disadvantages. This study chose CRT for two reasons. First, CRT is capable of analyzing binary node splitting, which is suitable for the interpretation of accident data analysis results (X. Yan & Radwan, 2006). Second, CRT can potentially analyze influence variables. This research sought to find the relationship between target variables and other variables expressed in form of the rank of each independent (predictor) variable according to its importance to the model (IBM, 2012). A great deal of previous research has used CRT to analyze accident data (Kashani & Mohaymany, 2011; Pakgohar et al., 2011; Pande et al., 2010), as CRT functions to emphatically focus on maximizing within-node homogeneity. The extent to which a node does not represent a homogenous subset of cases is an indication of impurity (IBM, 2012).

Choosing the correct splitting algorithm is also important. SPSS CRT offers two types of splitting, Gini and Twoing. Gini splits, which are widely used, function to maximize the homogeneity of child nodes with respect to the values of the dependent variables. Gini is based on squared probabilities of membership for each category of the dependent variable (Chang & Chien, 2013; IBM, 2012; Kashani & Mohaymany, 2011). For CART acceptance, splitting was achieved by using unit misclassification costs. This is the proportion of observed and predicted data comparisons (Khan et al., 2015).

In order to determine the optimal tree model, ten-fold cross-validation was undertaken, which is one of several cross-validation techniques to select for appropriate tree size. To avoid over-fitting the model, the maximum tree depth was five nodes, minimum cases in the parent node were 150, and minimum cases in child node were 75 (Khan et al., 2015).

4.5 Results and Discussion

According to the results from the CART of the two models, when considering misclassification costs for predictive accuracy (Table 4.5), Model#1 had overall correctness of 52.9% and Model#2 of 65.1%. Despite these low values, as confirmed by Kashani and Mohaymany (2011); Khan et al. (2015), they can be accepted and interpreted.

Table 4.5 Misclassification costs

Model#1			
Observed	Predicted		
	Not-at-fault	At-fault	Percent Correct
Not-at-fault	1654	442	78.9%
At-fault	1538	558	26.6%
Overall	76.2%	23.8%	52.9%
Model#2			
	Predicted		
Observed	Non-Fatal Crash	Fatal Crash	Percent Correct
Non-Fatal Crash	2332	1066	68.6%
Fatal Crash	522	634	54.8%
Overall	62.7%	37.3%	65.1%

4.5.1 Model#1

Model#1 (Figure 4.3) found six major variables related to the target variables. The most significant variable is driver's age (*person_age*). Drivers aged less than 21 years were at-fault drivers in 57.3% of accidents. This may be because younger drivers are less careful. Chandraratna and Stamatiadis (2009); L. Ma and Yan (2014) found that young drivers are more likely to be at fault than middle-aged drivers. Those aged over 21 years were at-fault only 48.9%. The significant variable was road lane, which can be interpreted that if a driver aged more than 21 years drives on a road with 10 or more traffic lanes (considering at only 10 lanes as there is no frequency of seven lanes), the chances of being at-fault drivers are 61.7%. This is consistent with research by Pande et al. (2010). This may be because roads with many lanes provide greater opportunities for speeding and vehicles are often parked on the roadside. Some less observant older drivers may be at fault for rear-end collisions. For accidents occurring at the median (*median_opening*), where the median is on a road with fewer than 10 traffic lanes, drivers older than 21 years were more likely to be at-fault. Due to the characteristics of median openings, front car drivers are more likely to reduce car speed in order to turn or execute a u-turn. If the car behind is too close, the chances of a rear-end collision are high. Dividing drivers into less and more than 25 years is a variable that has not previously been investigated. This research found that drivers in these two age ranges potentially consist of not-at fault drivers. When considering drivers aged over 25 years together with median type, there are more at-fault drivers when driving on unoccupied streets with a raised or flush median, with a greater chance of being at-fault than drivers on roads with barriers or depressed medians. The causes of these results were raise median, no median, and painted median. In Thailand, most of these median types are used on roads with low traffic flow, such as in residential areas or urban streets. Therefore, when driving too close, there is a chance of rear-end collision. This is consistent with research conducted by Joon-Ki et al. (2007); MRJ Baldock et al. (2005), who concluded that spacing on low-speed roads is a major cause of rear-end collisions. However, a study by Das and Abdel-Aty (2011) indicated that median type had no effect on the frequency of rear-end collisions.

Overall policy and public relations, therefore, should promote the reduction of rear-end collisions in the following ways: driver training should place special emphasis on drivers under 21 years of age, focusing on driving at the legal speed limit, and maintaining an appropriate distance from the vehicle in front. For drivers aged 21 years and older, it is important to pay special attention to roads with more than 10 lanes, and to take greater care of median openings on roads with fewer lanes. In other words, drivers should observe whether the car in front is executing a u-turn. Drivers aged 25 years or older should take special care on roads with no median, with a raised median, or with a depressed median, and they should maintain a greater distance from the car in front.

4.5.2 Model#2

The results of Model#2 (Figure 4.4) reveal 14 variables essential to fatal/non-fatal crashes. The most significant variable was safety equipment (*SafertEqui*), such as seatbelts or helmets. Those who did not use safety equipment were a 29.5% risk of dying in a rear-end collision. This is consistent with other research which has found that the use of safety equipment can reduce accident severity (Chen et al., 2015; MRJ Baldock et al., 2005). The next most significant variable was visibility, with a rear-end crash at night with no light having a 49% risk of fatality. This result supports findings by Chen et al. (2015); Sullivan and Flannagan (2003). Low light driving leads to rear-end crashes against cars parked along roadsides. In addition, a lower quantity of night-time traffic leads to drivers driving at higher speeds, which, in turn, causes a greater number of fatalities due to high velocity while crashing. In the case of sufficient light (in the daytime and at night with light), the variable of roads with a minimum of 2-8 traffic lanes on which a large number of trucks are parked, the chances of rear-end crashes are high. Moreover, the second variable, vehicle type (*Veh_type*), shows that large cars and trucks with six wheels or more result in 39.7% of deaths. This is relevant to the findings of Chang and Chien (2013); Chen et al. (2015), who found that the chances of fatality while decelerating and going straight were 53.1% (60/113 of crash accidents). Large vehicles which hit small vehicles on 2-4 lane roads have a high chance of fatality due to the vehicle body size factor (Xuedong Yan&Radwan, 2009). With regard to other crash types, stopped crash type has a 33% chance of fatalityy (80/240 crash accidents). In other words, rear crashes, occurring when the front cars are stopping, have a high fatality rate. With regard to medium and small vehicles, the chances of fatal crashes are high when the driver is aged more than 36 years (31.4%).

For drivers who use seatbelts, the second variable of raised and flush median led to a higher chance of fatality than other median types as these two types exist in areas of low-speed driving. If drivers violate the rules, the chances of rear-end collisions will be very high. For example, roads with a flush median type usually have no auxiliary lane to separate turning cars. Therefore, if a speeding car comes from behind, the resulting rear-end crash will be severe. This is consistent with the second variable, median opening, where there is a 48.8% probability of death. For other median types, two to four traffic lanes had 16.2% fatal rear-end crashes. With regard to leaf node, *envi_light* was found to be in accordance with Chen et al. (2016), who found that collisions occurring at night with both light and no light have a greater chance of fatality chances than collisions occurring during the daytime.

Policy recommendations to reduce fatalities from rear-end collisions are as follows: promoting awareness of seatbelt use by focusing on the driving license test, and increasing the strictness of law enforcement. For light conditions affecting visibility, drivers must be made aware of the danger of driving on roads with no lights, especially at night. Relevant authorities should consider increasing light installation on roads where the risk of rear-end collision is high. With regard to vehicle type, truckers must increase their awareness of parking their vehicles on roads with a high risk of rear-end collision, such as where there are no parking lanes and no light. In other words, the relevant departments, such as the DOH, should consider setting up illuminated roadside rest stops for trucks.

4.5.3 Discussion of the two models

Considering the overall picture of the two models, similar variables result in frequent rear-end collisions and fatalities. The first variable is the small number of lanes (2-4 traffic lanes), which is common in Thailand. The results of the models differed. Model#1 found fewer at-fault drivers in cases of a small number of traffic lanes, while model#2, found a high chance of fatalities. Future research should analyze this issue with regard to how different traffic lanes affect the frequency and severity of rear-end collisions. Another variable which was significant in both models was median type. Barrier and depressed median types result in a small number of rear-end collisions, and a low fatal crash rate. Therefore, when subordinate units of the DOHs make road improvements, these two median types should be considered. With regard to median opening point, both models found that rear-end collisions occurring at the median opening had a high incidence of at-fault, and caused high proportion of fatal crashes as the front vehicle decelerated or executed a u-turn. In these conditions, there is a high probability for the occurrence of a rear-end collision. In the case of fatal crashes, if the following vehicle has not seen the turning signal, a serious rear-end collisions will occur.

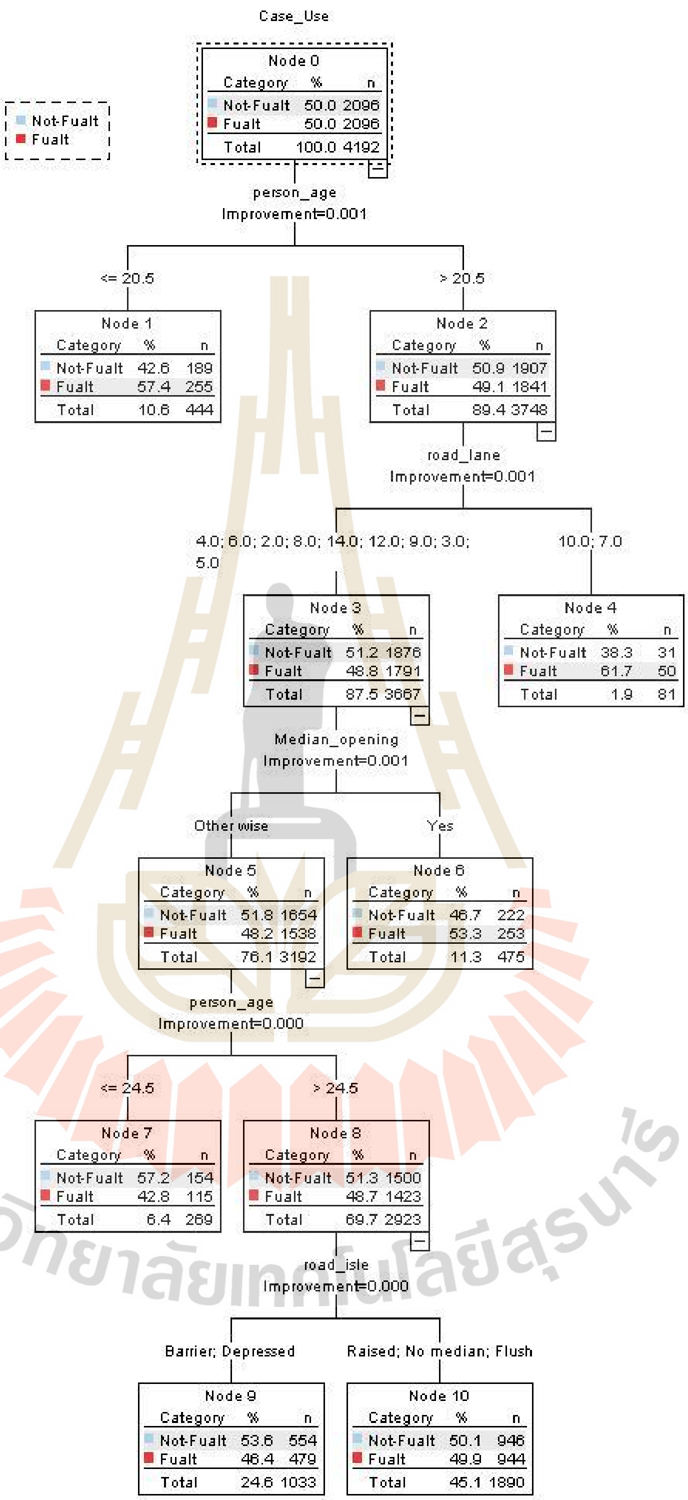


Figure 4.3 Tree Model#1

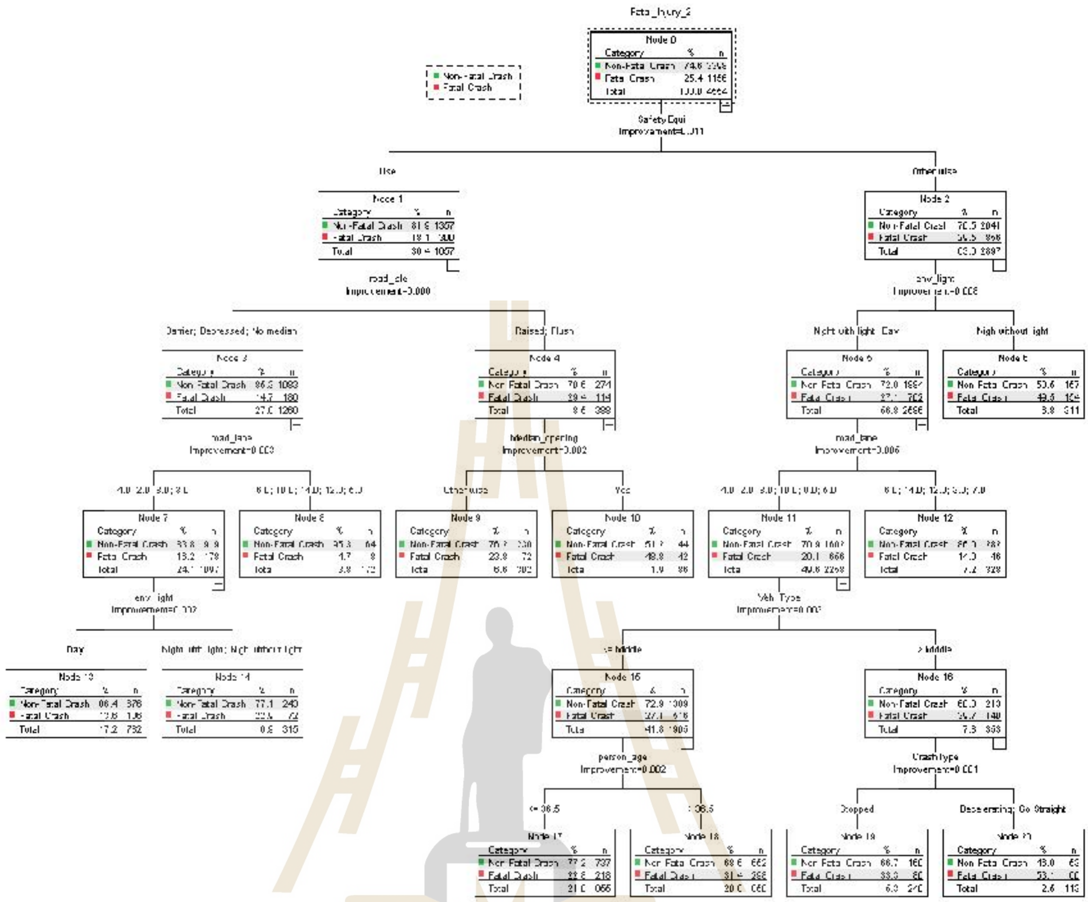


Figure 4.4 Tree Model#2

4.6 Conclusion

This research sought to explore two issues related to rear-end crashes. First, to find the factors which increase the number of rear-end collisions. This was achieved by focusing on the driver and environmental characteristics that cause rear-end collisions. Second, to find the factors causing fatal rear-end collisions. Using highway rear-end collision data from 2011 to 2015, non-parametric analysis was conducted on the significance of other variables which affect target variables, using an overview of factors, including drivers, the driving environment, and physical road characteristics. The model results were found to be able to predict rear-end collisions and fatalities with acceptable accuracy. The factors can contribute to a reduction in the number of at-fault drivers, and a reduction in the fatality rate of rear-end collisions.

The factors acquired from this analysis can be used to develop transportation office and rural road office policy and public relations practices, in order to reduce the number and severity of rear-end collisions.

It is recommended that future research parametric and non-parametric analysis to compare these factors in order to better understand the factors affecting crashes. In addition, a further investigation of lane numbers, median type, and median opening affecting the number of rear-end collisions, and fatal crashes, is called for, as these three variables were imperative for both models.

4.7 Acknowledgments

We would like to thank Enago (www.enago.com) for the English language review.

4.8 Reference

Abdel Aty, M., and Abdelwahab, H. (2004). Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. **Accident Analysis & Prevention**, 36(3): 447-456.

Agouti, T., Gharnati, F., and Ait-Mlouk, A. (2017). Application of Big Data Analysis with Decision Treefor Road Accident. **Indian Journal of Science and Technology**, 10(29): 1-10.

Akanbi, O. A., Amiri, I. S., and Fazeldehkordi, E. (2015). Chapter 3 - Research Methodology. In O. A. Akanbi, I. S. Amiri, & E. Fazeldehkordi (Eds.), *A*

Machine-Learning Approach to Phishing Detection and Defense (pp. 35-43). Boston: Syngress.

Anvari, M. B., Tavakoli Kashani, A., and Rabieyan, R. (2017). Identifying the Most Important Factors in the At-Fault Probability of Motorcyclists by Data Mining, Based on Classification Tree Models. **International Journal of Civil Engineering**, 15(4): 653-662.

Carr B.R. (1970). *A Statistical Analysis of Rural Ontario Traffic Accidents Using Induced Exposure Data*. Paper presented at the Proc. Sysposium in the Use of Statistical Methods in the Analysis of Road Accidents, ODCE, Patis, France.

Chandraratna, S., and Stamatiadis, N. (2009). Quasi-induced exposure method: evaluation of not-at-fault assumption. **Accident Analysis & Prevention**, 41(2): 308-313.

Chang, L.-Y., and Chien, J.-T. (2013). Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. **Safety Science**, 51(1): 17-22.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., and Guan, H. (2015). A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. **Accident Analysis & Prevention**, 80: 76-88.

Chen, C., Zhang, G., Yang, J., Milton, J. C., and Alcantara, A. D. (2016). An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naive Bayes (DTNB) hybrid classifier. **Accident Analysis & Prevention**, 90: 95-107.

Das, A., and Abdel-Aty, M. A. (2011). A combined frequency–severity approach for the analysis of rear-end crashes on urban arterials. **Safety Science**, 49(8): 1156-1163.

Department of Disaster Prevention and Mitigation. (2014). Redueing-Stopping Disaster: Prevention of rear-end accident. Available from: http://www.disaster.go.th/en/index.php Date Accessed: 10-Jan-2019

Department of Highway. (2016). Thailand Traffic Accident on National Highways in 2015. Bureau of Highway Safety, Department of Highways. Retrieved from http://bhs.doh.go.th/files/accident/58/report_accident2558.pdf

Department of Highway. (2017). Thailand Traffic Accident on National Highways in 2016. Bureau of Highway Safety Department of Highways Ministry of Transport, . Retrieved from http://bhs.doh.go.th/files/accident/59/report_accident2559.pdf

Department of Highway. (2018). Thailand Traffic Accident on National Highways in 2017. Bureau of Highway Safety Department of Highways Ministry of Transport, . Retrieved from http://bhs.doh.go.th/files/accident/60/report_accident2560.pdf

IBM. (2012). IBM SPSS Decision Trees 21. Retrieved from www.sussex.ac.uk/its/pdfs/SPSS_Decision_Trees_21.pdf

Joon-Ki, K., Yinhai, W., and Gudmundur, F. U. (2007). Modeling the Probability of Freeway Rear-End Crash Occurrence. **Journal of transport engineering**, 113(1): 11-19.

Kashani, A. T., and Mohaymany, A. S. (2011). Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. **Safety Science**, 49(10): 1314-1320.

Khan, G., Bill, A. R., and Noyce, D. A. (2015). Exploring the feasibility of classification trees versus ordinal discrete choice models for analyzing crash severity. **Transportation Research Part C: Emerging Technologies**, 50: 86-96.

Kim, S., Song, T.-J., Rouphail, N. M., Aghdashi, S., Amaro, A., and Gonçalves, G. (2016). Exploring the association of rear-end crash propensity and micro-scale driver behavior. **Safety Science**, 89: 45-54.

Li, X., Yan, X., Wu, J., Radwan, E., and Zhang, Y. (2016). A rear-end collision risk assessment model based on drivers' collision avoidance process under influences of cell phone use and gender-A driving simulator based study. **Accident Analysis & Prevention**, 97: 1-18.

Li, Z., Wang, W., Chen, R., and Liu, P. (2014). Conditional inference tree-based analysis of hazardous traffic conditions for rear-end and sideswipe collisions with implications for control strategies on freeways. **IET Intelligent Transport Systems**, 8(6): 509-518.

Ma, L., Wang, G., Yan, X., and Weng, J. (2016). A hybrid finite mixture model for exploring heterogeneous ordering patterns of driver injury severity. **Accident Analysis & Prevention**, 89: 62-73.

Ma, L., and Yan, X. (2014). Examining the nonparametric effect of drivers' age in rear-end accidents through an additive logistic regression model. **Accident Analysis & Prevention**, 67: 129-136.

Ma, L., Yan, X., and Weng, J. (2015). Modeling traffic crash rates of road segments through a lognormal hurdle framework with flexible scale parameter. **Journal of Advanced Transportation**, 49(8): 928-940.

Mendez, A. G., and Izquierdo, F. A. (2010). Quasi-induced exposure: the choice of exposure metrics. **Accident Analysis & Prevention**, 42(2): 582-588.

Meng, Q., and Qu, X. (2012). Estimation of rear-end vehicle crash frequencies in urban road tunnels. **Accident Analysis & Prevention**, 48: 254-263.

MRJ Baldock, AD Long, VL Lindsay, and AJ McLean. (2005). Rear end crashes. Centre for Automotive Safety Research, The University of Adelaide. Retrieved from http://casr.adelaide.edu.au/reports

Muhammad, L. J., Salisu, S., Yakubu, A., Malgwi, Y. M., Abdullahi, E. T., Mohammed, I. A., and Muhammad, N. A. a. (2017). Using Decision Tree Data Mining Algorithm to Predict Causes of Road Traffic Accidents, its Prone Locations and Time along Kano –Wudil Highway. **International Journal of Database Theory and Application**, 10(1): 197-206.

Nikiforos, S. (2008). Quasi-Induced Exporsure: Issue and Validation: Nova Science Publishers, Inc.

Nikiforos, S., and John, A. D. (1997). Quasi-Induced Exposure: Methodology and Insight. **Accident Analysis & Prevention**, 29(1): 37-52.

Pakgohar, A., Tabrizi, R. S., Khalili, M., and Esmaeili, A. (2011). The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. **Procedia Computer Science**, 3: 764-769.

Pande, A., Abdel-Aty, M., and Das, A. (2010). A classification tree based modeling approach for segment related crashes on multilane highways. **Journal of Safety Research**, 41(5): 391-397.

Song, Y. Y., and Lu, Y. (2015). Decision tree methods: applications for classification and prediction. **Shanghai Archives of Psychiatry**, 27(2): 130-135.

Sullivan, J. M., and Flannagan, M. J. (2003). *Risk of fatal rear-end collisions: is there more to it than attention?* Paper presented at the Proceeding of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. http://drivingassessment.uiowa.edu/DA2003/pdf/55_Sullivanformat.pdf

Taha, H. R., and Vinayak, D. (2013). *Modelling Likelihood of At-Fault and Not-At-Fault Carshare Users*. Paper presented at the International Conference Road Safety and Simulation, Rome, Italy.

Weng, J., Meng, Q., and Yan, X. (2014). Analysis of work zone rear-end crash risk for different vehicle-following patterns. **Accident Analysis & Prevention**, 72: 449-457.

Wiacek, C., Bean, J., and Sharma, D. (2015). Real-world analysis of fatal rear-end crashes. *24th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration.* Retrieved from https://www-esv.nhtsa.dot.gov/proceedings/24/files/24ESV-000270.PDF

Yan, X., and Radwan, E. (2006). Analyses of rear-end crashes based on classification tree models. **Traffic Injury Prevention**, 7(3): 276-282.

Yan, X., and Radwan, E. (2009). Analysis of truck-involved rear-end crashes using multinomial logistic regression. *Advances in Transportation Studies an* *international Journal.* Retrieved from http://www.atsinternationaljournal.com/index.php/2009-issues/xvii-april-2009/433-analysis-of-truck-involved-rear-end-crashes-using-multinomial-logistic-regression

Yan, X., Radwan, E., and Abdel-Aty, M. (2005). Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. **Accident Analysis & Prevention**, 37(6): 983-995.

Zheng, Z., Lu, P., and Tolliver, D. (2016). Decision Tree Approach to Accident Prediction for Highway–Rail Grade Crossings. **Journal of the Transportation Research Board**, 2545(1): 115-122.

CHAPTER V

REAR-END CRASH MODELS: A COMPARISON BETWEEN URBAN AND RURAL ZONES

5.1 Abstract

This study compares models that demonstrate the severity of rear-end collisions between the said areas. Severity was categorized into three levels based on the extent of injuries, namely, fatal, serious, and minor injuries. Occurrences of rear-end collisions in Thailand were classified according to the rural and urban zones of the municipalities in various districts. Afterward, factor analysis was used to reduce the number of variables for analysis of latent variables. Then, the integrated model was built using a structural equation modeling of both zones to verify consistency and build individual models. Lastly, the indicators of the latent variables of the rear-end collision models for urban and rural areas were compared using the measurement invariance method. Results showed that the models differed with regard to crash type and vehicle involvement factors. The findings will be beneficial for decision makers who are responsible for implementing traffic schemes or reducing fatalities from accidents.

5.2 Introduction

Rear-end collision is considered an important incident due to the frequent recurrence of crashes. In several countries, the number of deaths from rear-end crashes are numerous in comparison with other types of collisions.

Several differences can be noted when considering the severity of rear-end crashes between urban and rural areas from various perspectives. For example, the number of intersections results in a decrease in a front car's speed upon reaching signalized and unsignalized intersections. This scenario increases the chance of rear-end crashes, but severity may differ from that on rural roads (M. T. Islam, 2016). Chatterjee and Davis (2016) aimed to prevent shock waves from forming on freeways where rear-end collisions tend to occur. The speed of urban vehicles is typically low due to the relatively dense traffic, whereas rear crashes occurring in rural zones or on roads that connect the districts and provinces may be more severe because most vehicles on rural roads use high speed (David & Santosh, 2015). In terms of vehicle types that potentially access roads on rural and urban zones as well. For example, large trucks are allotted a limited time for road accessibility in urban areas. Vehicle types as classified by size also affect the severity of rear crashes in these areas. In term of attitude of driver, Zabihi et al. (2019) studied seat belt usage among adult drivers on urban and rural roads.

The studies on rear-end collisions in urban and rural zones are few. Chen et al. (2015) investigated the severity of driver injury based on road function factors, such as urban, rural interstate, and rural non-interstate roads. The result indicated that urban roads significantly influenced no level of injury. David and Santosh (2015) explored the factors that affect the severity of rear crashes in urban and rural zones. In terms of road environmental factor, the authors' key finding is that rear crashes in rural areas were less frequent. The speed of cars in rural zones leads to severe rear-end crashes. In addition old drivers (> 65 years) sustained high levels of injurybut had a slight effect on the incidence of rear crashes. According to gender, female drivers tend to obtain severe injury, especially, in high-speed environments, which increases the severity of rear-end crashes.

As previously mentioned, many dimensions in research place an emphasis on the "injury level" of drivers. For example, Chen et al. (2015) investigated various factors that influence injury levels in rear-end crashes in the United States using data on New Mexico from 2010 to 2011. Injury levels were divided into three, namely, property damage only, injury, and fatality, which is similar to that of Chen et al. (Chen et al., 2016), death risks (Sullivan & Flannagan, 2003), and severity of rear crash incidents on urban arterials (Das & Abdel-Aty, 2011). However, the number of victims involved in said rear crashes has not been studied. The other dimension of analysis was identifying the "rear-end crash size," which is the integration of the number of injured people across levels. The present study considers latent variables and employs structural equation modelling to analyze the relationship between factors. Previous research has identified crash severity in terms of latent variables. For example, Lee et al. (2008) examined crash size on expressways using accident size as a dependent variable, which was indicated by the number of deaths, injuries, cars, and number of vehicles damaged. Hassan and Al-Faleh (2013) explored the relationship between exploratory factors associated with crash size indicated by the number of cars, injuries, and damages to the government and private sectors. Schorr and Hamdar (2014) compared the severity between signalized and unsignalized intersections by defining indicators from four aspects, namely, severity, number of cars, number of injuries, and number of deaths. Hamdar and Schorr (2013) Compared roads between interrupted and uninterrupted flows and used factors, such as injury count, fatality count, severity, and traffic violation of aggressive maneuver as dependent variables. Kim et al. (2011) studied the severity of accidents as indicated by the extent of damage to vehicle, type of injury, and involved vehicle or unit.

However, research that provides a comparative study of crash severity using latent variables, especially that of rear-end crashes between urban and rural zones. Therefore, the present study aims to fill this research gap. Knowledge of different characteristics that influence the size of crashes occurring on urban and rural zones can be applied for improving the physical characteristics of specific roads to reduce the severity of rear-end crashes.

The study poses the following research question: “What are the differences between the indicators of latent factors influencing accident severity caused by rear-end crashes in urban and rural zones?” Subsequently, the following issues will be discussed:

Issue 1: What are the differences between crash size indicators for rear-end crashes on urban and rural roads?

Issue 2: Do differences exist between the indicators of road factors of rear-end crashes between urban and rural zones? How?

Issue 3: Do differences exist between the indicators of environmental factors of rear-end crashes between urban and rural zones? How?

Issue 4: Do differences exist between the indicators of vehicles of rear-end collisions between urban and rural zones? How?

Issue 5: Do differences exist between the indicators of crash types between urban and rural zones? How?

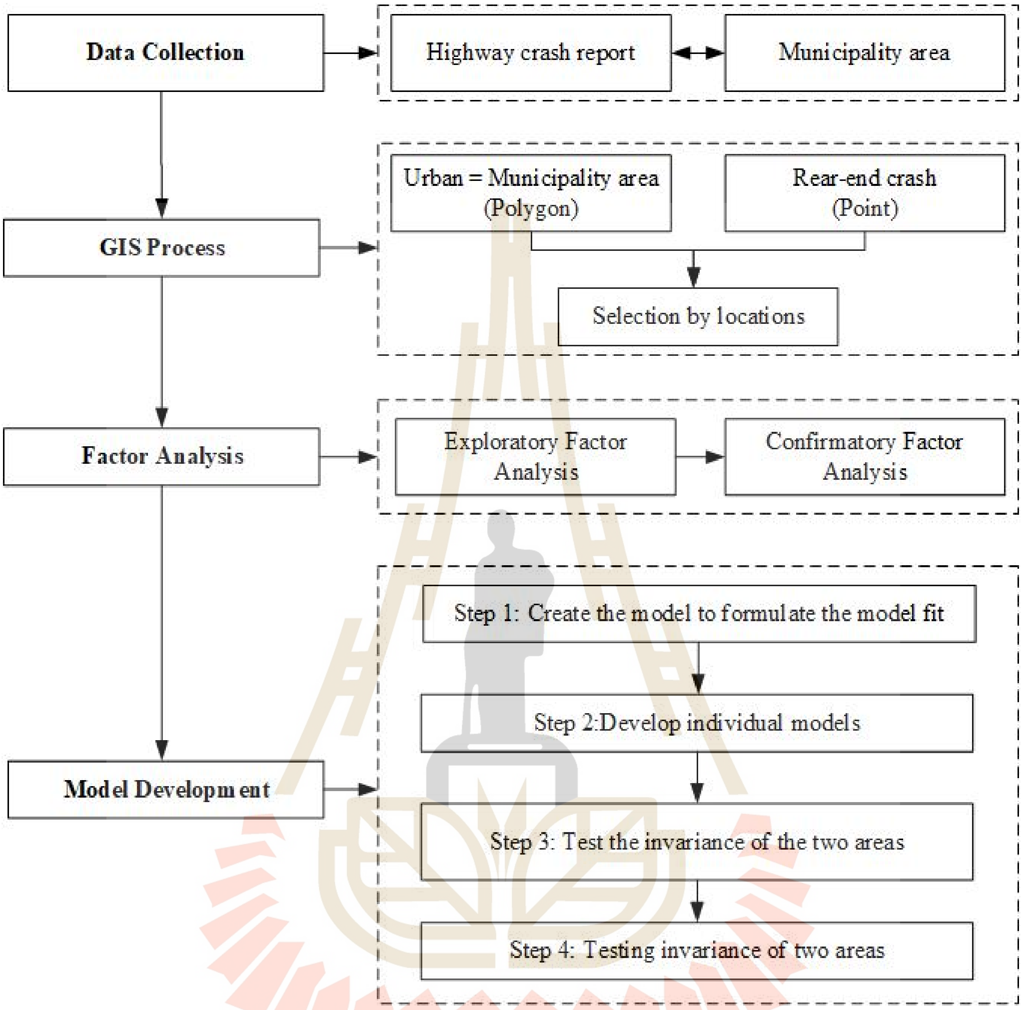
Issue 6: What are the differences between the overall structure of both models?

5.3 Method

This section discusses the procedures and methods used to achieve the research objectives. Figure 5.1 illustrates the procedure, which consists of four main steps, namely, (1 data collection, 2) Geometric Information System (GIS) process, 3) factor analysis, and 4) model development. The figure describes the process in detail.

5.3.1 Data collection

Data consist of two parts. The first is highway crash reports on cases of road accidents for 2011–2017 occurring on highways across the country. During accidents, a highway officer investigates and records each case. The record contains information about the scene of the accident (specific name of the road and location in kilometer), date, physical characteristics of the road (i.e., median type and intersection), environmental factors (i.e., weather, time, and lighting), vehicle type, crash type, and UTM. Data were collected from the Highway Accident Information Management System of the Department of Highways in Thailand. After filtering, a total of 11,976 rear-end crash cases were obtained. Data on traffic volume were retrieved from the Traffic Information Movement System. The two sets of data were subsequently matched with highway number and phase of kilometre range. The second part is 2) GIS (shape file) data on the WGS84 datum of UTM zone 47 that shows the administrative area, which is a municipality and characterized as polygon data. The reason for dividing the zones per municipality is due to a clearly divided area of crowded buildings, which differentiates traffic flow or physical road characteristics, vehicle type, number of users, and connected roads.



Note: Rectangles in solid lines denote the main processes. Rectangles in dashed lines pertain to sub-processes.

Figure 5.1 Framework of the study

To specify the actual zones of the occurrences of rear crashes in urban or rural zones, the GIS program was used by running a selection according to location and determining the shape files of rear crashes in the municipality. Results showed a total of 3,303 and 8,664 cases of rear-end crashes in urban and rural zones, respectively. After dividing the crash sites into two areas, the descriptive statistics are presented in Table 5.1.

Table 5.1 Descriptive Statistics (Urban)

Variables	Description	Value	Urban				
			Count	%	Mean (person)		
					Slight injury	Serious injury	Fata
Main_Road	Crashing on main roads (including parallel lanes and exits)	1 = Yes	372	12.27	0.50	0.14	0.09
		0 = Other	2,661	87.73	0.67	0.19	0.17
Raised	Raised median	1 = Yes	884	29.15	0.69	0.19	0.17
		0 = Other	2,149	70.85	0.63	0.18	0.16
Depressed	Depressed median	1 = Yes	535	17.64	0.78	0.23	0.20
		0 = Other	2,498	82.36	0.62	0.17	0.15
Barrier	Roads divided by barriers	1 = Yes	655	21.60	0.43	0.07	0.07
		0 = Other	2,378	78.40	0.71	0.22	0.19
Asphalt	Roads made of asphalt and/or concrete	1 = Yes	2,418	79.72	0.68	0.19	0.18
		0 = Other	615	20.28	0.52	0.16	0.11
Straight	Crashing on straight roads	1 = Yes	2,907	95.85	0.64	0.18	0.15
		0 = Other	126	4.15	0.80	0.29	0.37
Slope	Degree for slope more than 3%	1 = Yes	36	1.19	0.83	0.50	0.42
		0 = Other	2,997	98.81	0.65	0.18	0.16
Intersection	Crashing on intersection	1 = Yes	497	16.39	0.71	0.23	0.15
		0 = Other	2,536	83.61	0.64	0.18	0.16
Med_Open	Crashing on opening median point	1 = Yes	297	9.79	0.74	0.23	0.22
		0 = Other	2,736	90.21	0.64	0.18	0.16
Weather	Visualization of drivers	1 = Other	162	5.34	0.70	0.2	0.25
		0 = Clean	2,871	94.66	0.65	0.18	0.16
Day	Time of accident	1 = Day	1,884	62.12	0.67	0.16	0.12
		0 = Other	1,149	37.88	0.62	0.22	0.23
Night_NoLight	Night-time crashes on roads without lighting poles	1 = Yes	170	5.61	0.89	0.39	0.39
		0 = Other	2,863	94.39	0.63	0.17	0.15
Decelerating	Leading vehicle was decelerating speed	1 = Yes	710	23.41	0.69	0.23	0.19
		0 = Other	2,323	76.59	0.63	0.17	0.15
Stopped	Leading vehicle has stopped	1 = Yes	127	4.19	0.50	0.24	0.25
		0 = Other	2,906	95.81	0.65	0.18	0.16
Motorbikes	Small-sized vehicles (i.e., motorcycles, three-wheeled vehicles)	1 = Yes	639	21.07	0.66	0.27	0.23
		0 = Other	2,394	78.93	0.64	0.16	0.14
Trucks	Large-sized vehicles (trucks with six wheels or more)	1 = Yes	301	9.92	0.71	0.24	0.28
		0 = Other	2,732	90.08	0.64	0.18	0.15

Table 5.1 Descriptive Statistics (Rural) (Continued)

Variables	Description	Value	Rural				
			Count	%	Mean (person)		
					Slight injury	Serious injury	Fatal
Main_ Road	Crashing on main roads (including parallel lanes and exits)	1 = Yes	1,310	15.12	0.41	0.10	0.05
		0 = Other	7,354	84.88	0.77	0.27	0.23
Raised	Raised median	1 = Yes	1,530	17.66	0.71	0.22	0.20
		0 = Other	7,134	82.34	0.71	0.24	0.20
Depressed	Depressed median	1 = Yes	2,310	26.66	0.79	0.26	0.21
		0 = Other	6,354	73.34	0.69	0.23	0.20
Barrier	Roads divided by barriers	1 = Yes	2,000	23.08	0.41	0.08	0.05
		0 = Other	6,664	76.92	0.80	0.29	0.24
Asphalt	Roads made of asphalt and/or concrete	1 = Yes	7,757	89.53	0.72	0.25	0.21
		0 = Other	907	10.47	0.67	0.20	0.14
Straight	Crashing on straight roads	1 = Yes	8,194	94.58	0.69	0.23	0.19
		0 = Other	470	5.42	1.19	0.43	0.32
Slope	Degree for slope more than 3%	1 = Yes	199	2.30	1.49	0.62	0.49
		0 = Other	8,465	97.70	0.7	0.23	0.19
Intersection	Crashing on intersection	1 = Yes	956	11.03	0.89	0.30	0.17
		0 = Other	7,708	88.97	0.69	0.23	0.20
Med_Open	Crashing on opening median point	1 = Yes	725	8.37	0.93	0.31	0.27
		0 = Other	7,939	91.63	0.69	0.23	0.19
Weather	Visualization of drivers	1 = Other	564	6.51	0.97	0.31	0.25
		0 = Clean	8,100	93.49	0.70	0.24	0.19
Day	Time of accident	1 = Day	5,710	65.90	0.72	0.23	0.15
		0 = Other	2,954	34.10	0.71	0.26	0.30
Night_No Light	Night-time crashes on roads without lighting poles	1 = Yes	796	9.19	0.88	0.37	0.47
		0 = Other	7,868	90.81	0.70	0.23	0.17
Deceleratin g	Leading vehicle was decelerating speed	1 = Yes	1,719	19.84	0.82	0.31	0.24
		0 = Other	6,945	80.16	0.69	0.22	0.19
Stopped	Leading vehicle has stopped	1 = Yes	411	4.74	0.65	0.31	0.36
		0 = Other	8,253	95.26	0.72	0.24	0.19
Motorbikes	Small-sized vehicles (i.e., motorcycles, three-wheeled vehicles)	1 = Yes	1,505	17.37	0.67	0.30	0.32
		0 = Other	7,159	82.63	0.72	0.23	0.17
Trucks	Large-sized vehicles (trucks with six wheels or more)	1 = Yes	908	10.48	0.84	0.33	0.34
		0 = Other	7,756	89.52	0.70	0.23	0.18

Given the exploratory variables associated with the average number of deaths, we found that the top two variables were road characteristics, with a degree for slope more than 3%, and rear crashes at night (i.e., without light). Crashes in urban zones occurred in curves and intersections, whereas accidents in rural areas were caused by crashing against parked cars and trucks.

For the proportion of a small number of samples, crashes that occurred on slopes reached only 1.19% and 2.3%in urban and rural zones, respectively. The reason behind the finding is the small number of outstanding physical road characteristics in Thailand, especially when separated into urban and rural zones. However, similar research has been conducted and analyzed using a small sample size of approximately 2% (Hassan & Al-Faleh, 2013; Lee et al., 2008). The continuous variable in the urban and rural zones is the average traffic volumes of 24,101 and 21,590 vehicles per day with the proportions of trucks at 16.48% and 17.06%, respectively.

5.3.2 Factor analysis

As the rear crash model based on latent variables has not been previously studied, a dimension reduction was required using exploratory factor analysis (EFA) as the research tool. Observable variables were collected, which will indicate latent variables as structural equation modelling method was used. Thus, EFA was used to select the observable variables for indicating each latent variable. EFA was developed in the early 20th century by Pearson and Spearman. Its purpose is to identify variables that cannot be directly observed or measured and reduce the number of observable variables. EFA was used to describe the covariance value of all variables in the form of unobserved variables (Washington et al., 2011).

Table 5.2 provides the results of EFA, were analyzed using the unweighted least square method, which is ideal for categorical data in SPSS (Basto&Pereira, 2012). In terms of goodness of fit, a Kaiser-Meyer - Olkin (KMO) value of 0.585 and significant of the Chi-square test at p-value < 0.000 were considered acceptance criteria (Cerny & Kaiser, 1977).

Table 5.2 EFA Result

	Factor			
	1	2	3	4
Main_Road	**.466**	−.019	.061	.031
Raised	**−.309**	−.214	.108	.141
Depressed	**.965**	−.136	.261	.012
Barrier	**.782**	−.159	−.065	−.071
Asphalt	**−.111**	.001	−.082	−.056
Straight	**.512**	−.021	.206	.167
Slope	**−.458**	.031	−.209	−.110
Intersection	**−.266**	−.143	.178	.067
Med_Open	**.211**	−.044	−.183	.090
LN_AADT	**.730**	.148	.118	.031
Per_Truck	**.256**	.059	.026	.009
Weather	−.083	.063	**−.075**	−.069
Day	−.017	−.096	**.463**	−.312
Night_NoLight	−.166	.158	**−.574**	.256
Motorbike	−.078	−.028	.027	**−.057**
Truck	.021	.145	.000	**−.113**
Stopped	0.050	**−.156**	−.036	.001
Decelerating	−.179	**−.346**	.243	.073

Note: Goodness of fit: Chi-square (87) = 4208.458, p-value < 0.000, and KMO measure of sampling adequacy = 0.585

Given the high values of factor loadings, findings show that theycan be reduced to four factors. These values were named factor 1 (i.e., Main_Road, Raised, Depressed, Barrier, Asphalt, and Straight, Slope, Intersection, and Median_Open). This group of variables is considered road factors. The variables Stopped and Decelerating constitute factor 2, which was named crash type. Factor 3 or environmental factor included Weather, Day, and Night_No Light. Factor 4 consisted of Motorbike and Truck and named vehicle involvement.

5.3.3 Model development

The individual models were developed by categorizing the locations of rear-end crash scenes before analyzing data by invariance analysis.

Step 1: Create separate models for both zones in which all variables are taken into consideration and adjust such models to be reasonably suitable (based on EFA results) and create fit indices. For modelling, all independent variables are discrete and have count variables outcomes; thus, using an estimator as weighted least squares (WLS) or WLSMV in Mplus (Lee et al., 2008; Yu, 2002) is necessary. Parameter estimation in SEM was carried out by predicting the population covariance matrix of an observe variable: (i.e., Σ that can be specified in the terms of an unknown parameter:(θ)), which consists of B, Γ, Φ, and Ψ matrices. The components of the covariance matrix in the model is $\Sigma = \Sigma(\theta)$ .Therefore, the parameters of θ can be predicted by minimizing the discrepancies between the sample and population covariance matrices, which results in $\Sigma(\theta)$. The components in SEM consist of two models, namely, measurement model (pertains to the relationship between indicators) and structural model (denotes the relationship between latent variables). For further reference, see Hair Jr et al. (2010).

To verify whether the relationship between latent variables (i.e., whether or not the measurement and structural models are adequate for SEM analysis) and to complete this step, the values indicating the model parameters were examined, such as comparative fit index (CFI)> 0.95 and Root mean square error of approximation (RMSEA)< 005. (Mulaik & Millsap, 2000). The developed models relevant to the empirical data were realized according to the index of item–objective congruence (IOC)of models, where each parameter line, reasonableness of size, and parameter line direction were considered using the values of goodness-of-fit statistics.

Results show that the standard criteria were acceptable, as shown in Table 5.3 (Hair Jr et al., 2010; Kenny, 2016; Yu, 2002).

Table 5.3 Goodness-of-fit indices

Measure	Definition	Fit indices
Chi-square statistic (χ^2)	-	Expected significant p-values[a]
Degree of freedom (df)	-	
p-Value	-	
χ^2/df	-	Value < 3[b]
CFI	$\mathrm{CFI} = 1 - \frac{\max[(\chi^2_{H_0} - df_{H_0}), 0]}{\max[(\chi^2_{H_0} - df_{H_0}), (\chi^2_b - df_b), 0]}$	>0.92[a]
TLI	$\mathrm{TLI} = \frac{\chi^2_b/df_b - \chi^2_{H_0}/df_{H_0}}{(\chi^2_b/df_b) - 1}$	>0.92[a]
WRMR	$WRMR = \sqrt{\frac{2NF(\hat{\theta})}{e}}$	Value <1.00[c]
RMSEA	$RMSEA = \sqrt{\max\left[\left(\frac{2F(\hat{\theta})}{d} - \frac{1}{N}\right), 0\right]}$	Values <0.07 with CFI values of 0.92 or higher*

Note:

References:

[a](Hair Jr et al., 2010).

[b](Washington et al., 2011).

[c](Yu, 2002).

Tucker–Lewis index (TLI) and CFI, where df_b and df_{H_0} are the degrees of freedom for the baseline and hypothesized (under H0) models, respectively.

Weighted root mean square residual, where e denotes the number of sample statistics, and $F(\hat{\boldsymbol{\theta}}) = F_{WLS}(\theta) = \min\left[\left(\frac{1}{2}\right)s - \sigma(\theta)\right]' W^{-1}[s - \sigma(\theta)]$, which is the minimum of the WLS fitting function.

RMSEA, where d denotes the degrees of freedom of the model, and $F(\hat{\boldsymbol{\theta}})$ is the minimum of fitting function $F(\theta)$.

Step 2: Create two individual SEM models of rear-end crashes for urban and rural zone. Refer to the structures of the models buildings of the 1 step.

Step 3:Multi-group SEM analysis, which is also known as measurement invariance analysis, is an application for the comparison of similarly structured models. This test aims to determine the similarity of two groups of samples (invariance measurement or equivalence) using similar models (Hair Jr et al., 2010). This process is also known as cross-validation. For example, there are comparison for number of constructs, factor loading of indicator, mean, covariance, etc. The statistics used for comparison are the differences of Chi-square value and degrees of freedom. Consequently, both values are considered to test the levels of significance and indicate whether or not the various parameters of the two models are different (Hair Jr et al., 2010; Jomnonkwao et al., 2015; Nambulee et al., 2019).

The first step of for the measurement invariance (MI) of categorical variable is to check the structure of the model (configural invariance).This test pertains to the factor structure within each group or each latent variable separately (Hortensius, 2012). It consists of two sub-models as follows. (1)A model that will determine that all parameter values are independent in both groups (free across groups). However, the threshold value must be set to 1, and all other parameters must be set to zero. This model is called configural equivalence or base model. (2)The next step is building a model that specifies that all parameter values are equal except for a threshold value of 1, whereas all other parameters are set to zero. This model is called full equivalence or full model. Lastly, both models are employed for the Chi-square difference test.

5.4 Results

Step 1: Considering the consistency between the models and empirical data with the goodness of fit values, as shown in Table 6,we found that nearly all variables meet the criteria according to Table 5.3, except for WRMR, which is 2.512.This finding is due to the sufficiently high correlation between categories. However, the result remains acceptable, which is in agreement with Hanson and Kim (2007), where WRMR values at less than 3 were considered acceptable. Moreover, a similar research does not reflect WRMR values as fit indices because it is not a well-studied fit statistic and has failed to behave as well (Hassan&Al-Faleh, 2013; Schorr& Hamdar, 2014).

Table 5.4 SEM results of Step 1

Model/Variables	**Step 1**			
	Estimate	**S.E.**	**Est./S.E.**	**p-Value**
ACCZ				
Serious	0.371	0.01	38.947	<0.000
Fatal	0.515	0.015	35.179	<0.000
Slight	0.325	0.009	37.763	<0.000
ROAD				
Main_Road	1.12	0.05	22.238	<0.000
Raised	−0.161	0.016	−10.287	<0.000
Depressed	0.004	0.027	0.162	0.871
Barrier	1.461	0.062	23.636	<0.000
Asphalt	−0.238	0.018	−13.07	<0.000
Straight	0.271	0.017	15.72	<0.000
Slope	−0.232	0.021	−10.868	<0.000
Intersection	−0.322	0.016	−19.934	<0.000
Med_Open	−0.177	0.015	−11.505	<0.000
LN_AADT	0.722	0.028	25.422	<0.000
Per_Truck	0.037	0.007	5.267	<0.000
ENVI				
Weather	1	0.00	999	999.000
Day	−0.168	0.019	−8.748	<0.000
Night_NoLight	3.285	0.384	8.566	<0.000
VEH_INV				
Motorbike	1	0.00	999	999.000
Truck	0.08	0.023	3.49	<0.000
CRASH_T				
Decelerating	2.465	0.016	150.053	<0.000
Stopped	−0.188	0.002	−88.077	<0.000
ACCZ ON				
ROAD	−0.361	0.018	−20.092	<0.000
ENVI	0.058	0.008	7.242	<0.000
VEH_INV	0.158	0.009	18.429	<0.000
CRASH_T	−0.012	0.005	−2.331	0.02

Note: ACCZ = crash size, ENVI = environmental factor, VEH_INV = vehicle involvement, CRASH_T = crash type, S.E. = standard error, Est. = estimate

We then considered the rationale, as shown in Table 5.4, and found that all variables are consistent with reality. For example, we observed that the weight of the fatal variable has the highest loading value in terms of crash size. We infer that this variable is the main indicator of the measurement followed by serious and slight injuries.

Step 2: This step is a result of running the model by separating the two sample groups, as shown in Table 5.5 and Figure 5.2 and Figure 5.3. In summary, both trends are similar. In other words, the highest loading factor belongs to fatality followed by serious and slight injuries for crash size (ACCZ). In the urban zone models, slight injury has a negative value because this variable negatively correlates with certain independent variables. However, this result is possible due to the weight given by the fatal variable. A slight difference is observed in the structural model, that is, the maximum weight variable is ROAD followed by ENVI and VEH_INV, respectively, but in the model of rear crashes in rural zone, CRASH_T variable significantly affects ACCZ in negative direction.

Step 3: Table 5.6 provides the analysis results from invariance analysis, which first considers the goodness of fit of the configural equivalence model (i.e., the model enables the parameters to be independent from one another) and full equivalence model (i.e., the model forces the parameters of both sample groups to be equivalent). Result ofthe difference test reveals that the value of delta Chi-square (5) 30 =.182 ($p < 0.000$). This finding indicates that the measure of the latent variables of rear-end crash models in urban and rural zones is significantly different.

Table 5.5 SEM results of Step 2

Model/Variables	Step 2 (Urban Zone)				Step 2 (Rural Zone)			
	Estimate	**S.E.**	**Est./S.E.**	**p-Value**	**Estimate**	**S.E.**	**Est./S.E.**	**p-Value**
Measurement model								
ACCZ								
Serious	1.766	0.006	281.515	<0.000	0.347	0.006	58.158	<0.000
Fatal	2.035	0.171	11.93	<0.000	0.526	0.008	61.855	<0.000
Slight	0.028	0.004	−7.447	<0.000	0.292	0.015	19.075	<0.000
ROAD								
Main_Road	0.715	0.043	16.64	<0.000	1.074	0.02	54.77	<0.000
Raised	−0.527	0.049	−10.805	<0.000	−0.224	0.018	−12.168	<0.000
Depressed	−0.073	0.048	−1.529	0.126	−0.022	0.018	−1.234	0.217
Barrier	1.169	0.04	29.078	<0.000	1.266	0.018	72.186	<0.000
Asphalt	−0.128	0.058	−2.199	0.028	−0.093	0.017	−5.334	<0.000
Straight	0.249	0.044	5.694	<0.000	0.27	0.018	14.773	<0.000
Slope	−0.252	0.062	−4.048	<0.000	−0.226	0.024	−9.347	<0.000
Intersection	−0.414	0.026	−15.636	<0.000	−0.373	0.016	−23.261	<0.000
Med_Open	−0.155	0.033	−4.63	<0.000	−0.336	0.018	−18.721	<0.000
LN_AADT	0.651	0.021	31.48	<0.000	0.619	0.008	82.15	<0.000
Per_Truck	0.191	0.02	9.629	<0.000	−0.005	0.008	−0.633	0.527
ENVI								
Weather	0.192	0.047	4.057	<0.000	0.117	0.028	4.249	<0.000
Day	−0.923	0.072	−12.876	<0.000	−0.821	0.036	−22.828	<0.000
Night_NoLight	0.913	0.071	12.806	<0.000	1.072	0.046	23.073	<0.000
VEH_INV								
Motorbike	1	0	999	999.000	1	0	999	999.000
Truck	1.418	0.234	6.051	<0.000	0.121	0.027	4.469	<0.000
CRASH_T								
Decelerating	2.362	0.027	87.697	<0.000	2.508	0.02	123.082	<0.000
Stopped	−0.202	0.005	−43.138	<0.000	−0.182	0.002	−76.698	<0.000
Structural model								
ACCZ								
ROAD	−0.073	0.007	−10.27	<0.000	−0.453	0.011	−40.906	<0.000
ENVI	0.073	0.008	9.332	<0.000	0.258	0.014	18.292	<0.000
VEH_INV	0.043	0.004	9.636	<0.000	0.119	0.011	11.22	<0.000
CRASH_T	0.003	0.002	1.321	0.187	−0.021	0.006	−3.707	<0.000

Note: ACCZ = crash size, ENVI = environmental factor, VEH_INV = vehicle involvement, CRASH_T = crash type, S.E. = standard error, Est. = estimate

Table 5.6 Measurement of invariance results and goodness of fit

	χ^2	*df*	*p-Value*	χ^2/df	RMSEA	CFI	TLI	WRMR	Delta χ^2	Delta *df*	p-Value
Step 1: Total samples	2634	163	<0.000	16.159	0.036	0.953	0.939	2.512			
Step 2: Single model											
Urban Model	561.044	157	<0.000	3.574	0.029	0.942	0.923	1.620			
Rural Model	3101.094	172	<0.000	18.030	0.044	0.935	0.921	2.802			
Step 3: Invariance analysis											
Configural equivalence (base model)	3253.643	339	<0.000	9.687	0.039	0.936	0.920				
Full equivalence (full model)	3283.825	344	<0.000	9.458	0.038	0.936	0.922				
Difference test									30.182	5	<0.000

Note: RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker–Lewis index, WRMR = weighted root mean square residual

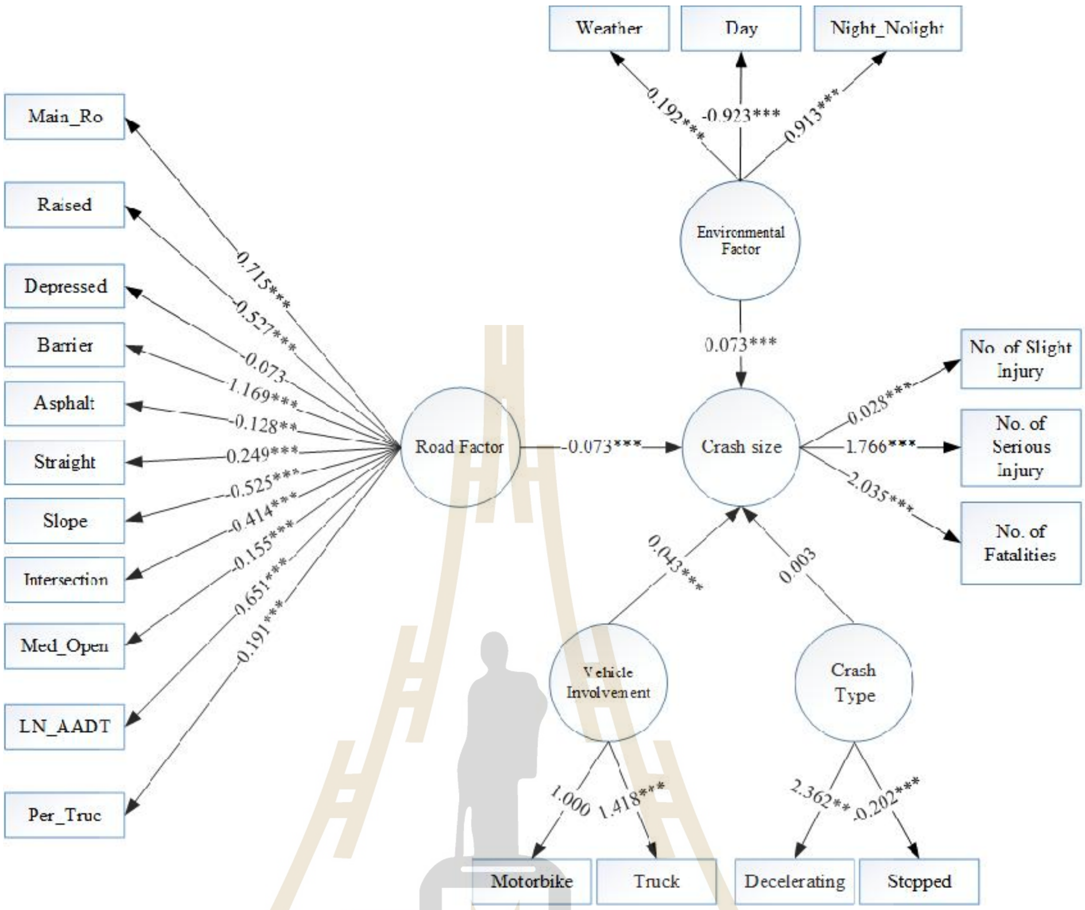


Figure 5.2 Model for urban areas

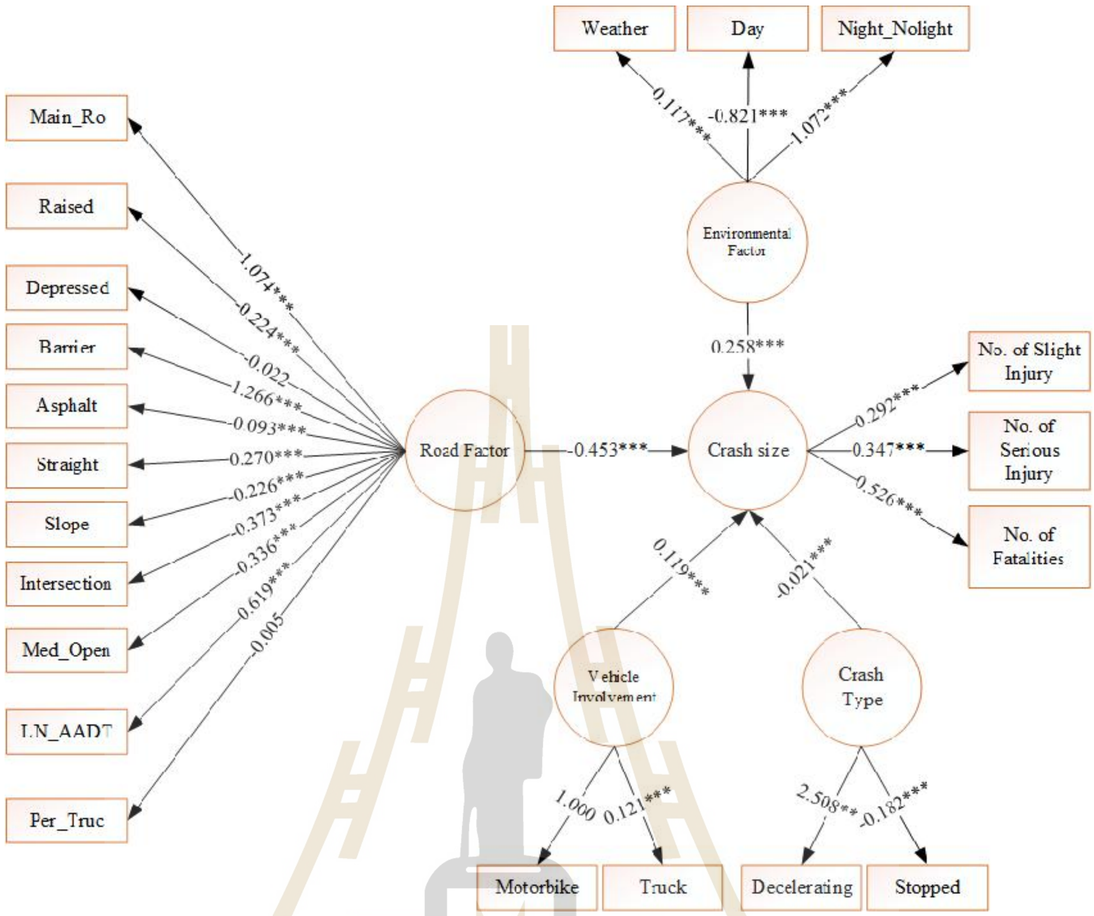


Figure 5.3 Model for rural areas

5.5 Discussion

Measurement Invariance (MI) is conducted to test for the differences in the indicators of each latent variable and consecutively compare rear-end crash models between urban and rural zones. The study addresses the research question with the following result: The latent variables of the rear crash models in urban and rural zones are significantly different. This finding has been confirmed by similar previous works.

Yan et al. (2005) Investigated the road types that influence rear crashes in intersections using the logistic regression model. Analysis indicated thatthe location of crashes or roads when classified as urban and rural areas significantly affect the severity of rear crashes. That is, urban roads increase the chances of rear crashes compared with rural roads 20%. Khorashadi et al. (2005) highlighted that accident occurrences in urban and rural areas resulted in different levels of injury. Lord et al. (2005) differentiated between crash flow density and crash flow V/C using the differences in road characteristics in urban and rural areas, which is similar to that of Stylianou and Dimitriou (2018). Kmet and Macarthur (2006) explored the injury levels of children and young drivers from road accidents between two zones. The authors found that accidents occurring on rural roads were more likely to lead to mortality than those on urban roads. This finding is possibly due to the shortage of road safety features, such as traffic control devices, graded curves, lighting, and divided traffic streams. Li et al. (2008) studied medical service utilization for traffic and compared fatalities between accidents in urban and rural zones. The authors also concluded that providing good medical services or promoting the use of helmets and seat belts can decrease fatality rates in rural zones. Czech et al. (2010) compared the cost of alcohol-related traffic crashes in rural and urban zones. The results indicated that alcohol-related traffic crashes in rural areas were 15. times higher than those in urban zones, which resulted in seven to eight times higher accident costs. Peek-Asa et al. (2010) Considered the factors influencing teenage driver-involved crashes and found that in terms of number of accidents, those that occurred in rural areas were less than those in urban zones. However, in terms of injury levels, accidents in rural zones had increased chances of leading deaths and serious injury than those in urban zones. According to crash type, single-vehiclecollision at night typically occurred in rural areas. Factors affecting severity included traffic quantity, vehicle type, and proportion of heavy trucks, among others. In addition, the impact of the industry could reduce severity by promoting campaigns to abstain from changing traffic lanes while driving at high speeds. Other crash types include agriculture equipment crashes (Harland et al., 2014) and pedestrian collisions (S. Islam & Jones, 2014). Studies from the perspectives of driving attitude and behavior of young drivers resulted in urban and rural area, found that driver behavior differs in terms of speed, vehicle selection, and public transport facility. (Eiksund, 2009). In the following sections, we will discuss detailed issues.

Issue 1: The similar indicators of crash size in both models mainly focus on the number of fatalities, which is reasonable because crash size is based on the weighting of the highest level of injury followed by the number of serious injuries and number of slight injuries. This issue is similar to that presented by Schorr and Hamdar (2014), which indicated that the higher the level of severity, the higher the severity index.

Issue 2: Several differences were observed for the indicators of road factors. However, the loading factors in the top three ranks were similar. (1) Barrier type influenced slight injury. This finding is in agreement with Zou and Tarko (2018), that is, barrier type can reduce severity. (2) Rear-end crashes on main traffic lanes (Main_Road) influenced slight severity; research on this factor is relatively few. However, given the rationale, commonly, differences in speed on traffic lanes are relatively less compared with roads with parallel lanes. Moreover, the greater the number of connected roads, the higher the chances of collision (David & Santosh, 2015). (3) High traffic volumes resulted in smaller crash sizes. This finding is relevant to research that aims to explain the possibility that the increase in AADT will lead to a decrease in speed. This situation potentially caused the crash severity (Haleem

&Abdel-Aty, 2010; Haleem et al., 2019; Haleem & Gan, 2013). In terms of the difference in both models, Per Truck was a significant indicator of rear crashes in urban zones but not in rural zones.

Issue 3: Regarding the indicators of environmental factor, results indicate that both models follow the same direction. That is, the first rank increases the severity of accidents at night with no lighting. Several studies (Chen et al., 2015; Chen et al., 2016; Reeves et al., 2019) supported this issue. The previous literature indicated that sight distance was short during crashes followed by night crashes, which increased severity and poor visualization and thus led to increased injury. The results of the two variables are consistent with those of Lee et al. (2008).

Issue 4: Distinctive differences were noted for vehicle indicator. The model posits that motorbike is a reference variable. In the rear-end crash model for urban area, the present study found that trucks increased the severity of crashes. This finding is in agreement with David and Santosh (2015). If motorcycles are involved in rear-end crashes in rural zones, then crash size becomes more severe. On rural roads in Thailand, motorcycles use high speed in general due to the low-density traffic volume. Truck drivers frequently stop at road shoulders to sleep. However, motorcycles at high speed can collide with the truck's rear end, which leads to the death of motorcycle drivers.

Issue 5: The indicators of rear crash type showed that both models were identical. In other words, the variable that resulted in increased crash severity is the one that identifies whether or not the leading vehicle is decelerating. However, for the rear crash model in urban zones, this latent variable did not influence crash size. In terms of urban zones, the study found that collision with front cars that are parked increased the severity of rear crashes due to different speeds while crashing. Rear-end crashes resulted in fatal and serious injuries in high-speed zones (David & Santosh, 2015).

Issue 6: The structural model shows the relationship between latent variables. It considers the weighting of the loading factors that influence crash size by comparing rear crashes between urban and rural zones. The identified distinct differences were as follows. (1) The orders of weight in the urban model that influence crash size are equal but are in opposite directions (i.e., −0.073 for road factor and 0.073 for environmental factor). Road factor has the highest weight in rural zones followed by environmental factor (−0.453 and 0.258, respectively). (2) For rear-end crashes in urban zones, crash type did not significantly influence crash size; however, this variable significantly affected crash size in rural zones due to the speed of leading car during crashing. This factor evidently influenced crash size. For clarity, road conditions in rural zone led to the differences in speed of rear crashes. This factor clearly influenced crash size (David & Santosh, 2015). Moreover, the problem of parked trucks resulted in rear crash occurrences, which involved cars driving at high speeds without any precaution. These factors are the possible causes of rear crashes and increased chances of deaths, especially in the case of motorcycles.

5.6 Conclusion

This study compared the severity of rear crashes in terms of latent variables, namely, number of fatality, serious injury, and slight injury. Furthermore, SEM was applied to determine the indicators of crash size. Difference between urban and rural zones are characterized by various factors, such as speed factor of different vehicles and vehicle type. In addition, studies that compared these factors between the two areas are relatively few. Therefore, the results of the present study can be considered for further study. Additionally, relevant parties, such as highway maintenance agencies, can use the results to enhance the design ofroad schemes.

Significant differences were denoted based on the results using the MI of two models that tested the indicators of each latent variable, such as road factors, environmental factors, vehicle involvement, and crash type. For the overall individual model, the study found that both models had a similar structure. The only difference between them is the extent to which crash type influenced crash size but only for rural zones. Discussion of the detailed issues of the latent variables also pointed to several differences. In the case of a truck-involved crashes in urban zones, there was more severity than the motorcycle-involved crash. While casein rural zone, there was more severity in a motorcycle-involved. To reduce the severity of rear-end crashes in both areas, the first factor pertains to the median, which is a barrier type, followed by the occurrence of rear crashes on main channels. Lastly, the variable Night_NoLight was found as the main indicator of crash size severity in both areas.

5.7 Limitation and future research

This study comes with certain limitations. The first is the data of crashes occurring on highways in Thailand are used, such that potential drivers should consider the relevant physical road characteristics, crashes occurrences and vehicle type for appropriate application. The second is that the driver factor has been omitted to specifically focus on physical road characteristics and environment factors that influence rear crash severity. However, driver factors, such as gender, age, occupation, and experience, are undeniably significant factors.Future studies should integrate these potential factors into the model for analysis.

5.8 Acknowledgment

The authors would like to thank Enago (www.enago.com) for the English language review.

5.9 Reference

Basto, M., and Pereira, J. M. (2012). An SPSS R-menu for ordinal factor analysis. **Journal of Statistical Software**, 46(4).

Cerny, B. A., and Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. **Multivariate behavioral research**, 12(1): 43-47.

Chatterjee, I., and Davis, G. A. (2016). Analysis of Rear-End Events on Congested Freeways by Using Video-Recorded Shock Waves. **Transportation Research Record: Journal of the Transportation Research Board**, 2583(1): 110-118.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., and Guan, H. (2015). A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. **Accident Analysis & Prevention**, 80: 76-88.

Chen, C., Zhang, G., Yang, J., Milton, J. C., and Alcantara, A. D. (2016). An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naive Bayes (DTNB) hybrid classifier. **Accident Analysis & Prevention**, 90: 95-107.

Czech, S., Shakeshaft, A. P., Byrnes, J. M., and Doran, C. M. (2010). Comparing the cost of alcohol-related traffic crashes in rural and urban environments. **Accident Analysis & Prevention**, 42(4): 1195-1198.

Das, A., and Abdel-Aty, M. A. (2011). A combined frequency–severity approach for the analysis of rear-end crashes on urban arterials. **Safety Science**, 49(8-9): 1156-1163.

David, B., and Santosh, T. (2015). *Investigation of Key Crash Types: Rear-end Crashes in Urban and Rural Environments* Retrieved from Austroads Ltd.:

Eiksund, S. (2009). A geographical perspective on driving attitudes and behaviour among young adults in urban and rural Norway.

Hair Jr, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2010). Multivariate Data Analysis: Global Edition, 7/E. New York.

Haleem, K., and Abdel-Aty, M. (2010). Examining traffic crash injury severity at unsignalized intersections. **Journal of Safety Research**, 41(4): 347-357.

Haleem, K., Azam, S., Manepalli, U., and Mays, M. (2019). Identifying and comparing the injury severity risk factors on rural freeways in different states in the United States. **International Journal of Injury Control and Safety Promotion**, 26(4): 343-353.

Haleem, K., and Gan, A. (2013). Effect of driver's age and side of impact on crash severity along urban freeways: a mixed logit approach. **Journal of Safety Research**, 46: 67-76.

Hamdar, S. H., and Schorr, J. (2013). Interrupted versus uninterrupted flow: a safety propensity index for driver behavior. **Accident Analysis & Prevention**, 55: 22-33.

Hanson, T. L., and Kim, J.-O. (2007). *Measuring Resilience and Youth Development: The Psychometric Properties of the Healthy Kids Survey. Issues & Answers. REL 2007-No. 34*. In *Regional Educational Laboratory West*. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/west/pdf/REL_2007034.pdf

Harland, K. K., Greenan, M., and Ramirez, M. (2014). Not just a rural occurrence: differences in agricultural equipment crash characteristics by rural-urban crash site and proximity to town. **Accident Analysis & Prevention**, 70: 8-13.

Hassan, H. M., and Al-Faleh, H. (2013). Exploring the risk factors associated with the size and severity of roadway crashes in Riyadh. **Journal of Safety Research**, 47: 67-74.

Hortensius, L. (2012). Project for Introduction to Multivariate Statistics: Measurement Invariance. Available from: https://pdfs.semanticscholar.org/6d69/f889d6acd9a58706e8e230e4e43ab4ae3bda.pdf Date Accessed: Jul 1, 2019

Islam, M. T. (2016). Investigating the Speed and Rear-End Collision Relationship at Urban Signalized Intersections. **Transportation Research Record: Journal of the Transportation Research Board**, 2601(1): 10-16.

Islam, S., and Jones, S. L. (2014). Pedestrian at-fault crashes on rural and urban roadways in Alabama. **Accident Analysis & Prevention**, 72: 267-276.

Jomnonkwao, S., Ratanavaraha, V., Khampirat, B., Meeyai, S., and Watthanaklang, D. (2015). Factors influencing customer loyalty to educational tour buses and measurement invariance across urban and rural zones. **Transportmetrica A: Transport Science**, 11(8): 659-685.

Kenny, D. A. (2016). Measuring model fit. Available from: http://davidakenny.net/cm/fit.htm Date Accessed: June-2017

Khorashadi, A., Niemeier, D., Shankar, V., and Mannering, F. (2005). Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. **Accident Analysis & Prevention**, 37(5): 910-921.

Kim, K., Pant, P., and Yamashita, E. (2011). Measuring Influence of Accessibility on Accident Severity with Structural Equation Modeling. **Transportation Research Record: Journal of the Transportation Research Board**, 2236: 1-10.

Kmet, L., and Macarthur, C. (2006). Urban-rural differences in motor vehicle crash fatality and hospitalization rates among children and youth. **Accident Analysis & Prevention**, 38(1): 122-127.

Lee, J. Y., Chung, J. H., and Son, B. (2008). Analysis of traffic accident size for Korean highway using structural equation models. **Accident Analysis & Prevention**, 40(6): 1955-1963.

Li, M. D., Doong, J. L., Chang, K. K., Lu, T. H., and Jeng, M. C. (2008). Differences in urban and rural accident characteristics and medical service utilization for traffic fatalities in less-motorized societies. **Journal of Safety Research**, 39(6): 623-630.

Lord, D., Manar, A., and Vizioli, A. (2005). Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. **Accident Analysis & Prevention**, 37(1): 185-199.

Mulaik, S. A., and Millsap, R. E. (2000). Doing the Four-Step Right. **Structural Equation Modeling: A Multidisciplinary Journal**, 7(1): 36-73.

Nambulee, W., Jomnonkwao, S., Siridhara, S., and Ratanavaraha, V. (2019). Modelling of seat belt use intention for intercity buses based on health belief model. **Transportmetrica A: Transport Science**, 15(2): 944-962.

Peek-Asa, C., Britton, C., Young, T., Pawlovich, M., and Falb, S. (2010). Teenage driver crash incidence and factors influencing crash injury by rurality. **Journal of Safety Research**, 41(6): 487-492.

Reeves, K., Chandan, J. S., and Bandyopadhyay, S. (2019). Using statistical modelling to analyze risk factors for severe and fatal road traffic accidents. **International Journal of Injury Control and Safety Promotion**, 26(4): 364-371.

Schorr, J. P., and Hamdar, S. H. (2014). Safety propensity index for signalized and unsignalized intersections: exploration and assessment. **Accident Analysis & Prevention**, 71: 93-105.

Stylianou, K., and Dimitriou, L. (2018). Analysis of Rear-End Conflicts in Urban Networks using Bayesian Networks. **Transportation Research Record: Journal of the Transportation Research Board**, 2672(38): 302-312.

Sullivan, J. M., and Flannagan, M. J. (2003). *Risk of fatal rear-end collisions: is there more to it than attention?* Paper presented at the Proceeding of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Washington, S. P., Karlaftis, M. G., and Mannering, F. (2011). Statistical and Econometric Methods for Transportation Data Analysis (2 ed.). 90 B/W Illustrations: Chapman and Hall/CRC.

Yan, X., Radwan, E., and Abdel-Aty, M. (2005). Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. **Accident Analysis & Prevention**, 37(6): 983-995.

Yu, C.-Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes (Vol. 30): University of California, Los Angeles Los Angeles.

Zabihi, F., Davoodi, S. R., and Nordfjaern, T. (2019). The role of perceived risk, reasons for non-seat belt use and demographic characteristics for seat belt use on urban and rural roads. **International Journal of Injury Control and Safety Promotion**, 26(4): 431-441.

Zou, Y., and Tarko, A. P. (2018). Barrier-relevant crash modification factors and average costs of crashes on arterial roads in Indiana. **Accident Analysis & Prevention**, 111: 71-85.

CHAPTER VI

THE APPLICATION OF HIERARCHICAL LOGISTIC MODELS TO COMPARE URBAN AND RURAL ROADWAY MODELING OF FATAL REAR-END VEHICULAR CRASHES

6.1 Abstract

A rear-end crash is a widely studied type of road accident. The road area at the crash scene is a factor that significantly affects the likelihood of fatalities from rear-end collisions. These road areas may be classified as urban or rural and evince obvious differences such as speed limits, number of intersections, vehicle types, etc. However, no study comparing rear-end crashes occurring in urban and rural areas has yet been conducted. Therefore, the present investigation focused on the comparison of diverse factors affecting the likelihood of rear-end crash fatalities in the two types of roadways. Additionally, hierarchical logistic models grounded in a spatial basis concept were applied by determining varying parameter estimations with regard to road segments. Additionally, the study compared coefficients with multilevel correlation model and those without multilevel correlation. Four models were established as a result. The data used for the study pertained to rear-end crashes occurring on Thai highways between 2011 and 2015. The results of the data analysis revealed that the model parameters for both urban and rural areas are in the same direction with the larger number of significant parameter values present in the rural

rear-end crash model. The significant variables in both the urban and rural road segment models are the number of traffic lanes, seat belt use by drivers, and the time of the incident. To conclude, the present study is useful because it provides another perspective of rear-end crashes to encourage policy makers to apply decisions that favor rules that assure safety.

6.2 Introduction

The classification of fatal accidents by their crash type yielded the result that the highest number of fatalities occur from rear-end crashes, as illustrated in Figure 6.1. In terms of statistical consideration, the numbers of such accidents rose continuously between 2014 and 2016 (DOH, 2015, 2016, 2017). Therefore, the injury levels or severity of the rear-end accidents must first be ascertained and emphasized to reduce the number of rear-end crashes.

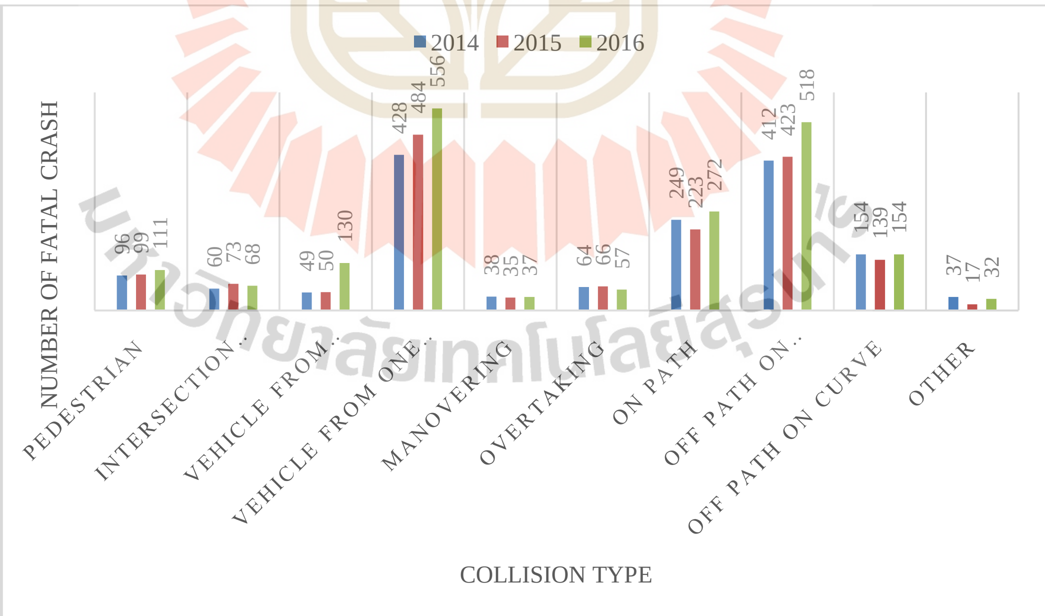


Figure 6.1 Trend of fatal road traffic accidents classification by collision type (DOH, 2015a, 2016, 2017)

Rear-end crashes occur frequently and are considered to cause high fatalities. Numerous studies have been conducted on rear-end crashes and have found the factors that cause loss of life. For example, Sullivan and Flannagan (2003) studied the relationship between deaths from rear-end crashes and time periods and found the fatality risks of rear-end crashes were two times higher when they occurred in darkness than at daytime. This outcome is consistent with the study conducted by Yan and Radwan (2009). Chen et al. (2015) found that the involvement of a truck was the primary potential cause of the high severity of injuries from rear-end crashes (Wiacek et al., 2015). This element was followed by driving under the influence of alcohol, and the third aspect on the list was rear-end crashes occurring at night in conditions of darkness. In addition, Chen, Zhang, Yang, et al. (2016) also discovered that visibility and road-slope were features predicting fatal driver injuries.

Some studies have examined road factors related to rear-end injury levels. Among them, Shawky et al. (2016) have found that rural roads cause the most driver fatalities, followed by the number of traffic lanes. This study found a high risk of fatality caused by rear-end crashes when the number of traffic lanes was less than 4. Similarly, Chen et al. (2015) found that rear-end crashes that happened in urban areas caused fewer deaths than those that took place in rural areas. Research on other factors causing fatalities from such accidents included results such as curved roads resulted in fewer rear-end crashes, or that the involvement of a motorcycle in the accident resulted in high fatality rates from rear-end crashes.

The differing physical characteristics of urban and rural roads possibly result in the dissimilar severity of injuries found in rear-end accidents in cities vis-à-vis villages. Chen, Zhang, Liu, et al. (2016) discovered that collisions on curved urban roads presented the possibility of a high risk of injuries due to high speed driving. The truck volume is also another critical reason. The dissimilar levels of severity between urban and rural roads were also attributed to the differing nighttime illumination in these areas. The results of an investigation by Uddin and Huynh (2017) found that increased traffic volume was likely to reduce the injury level, especially in rural areas.

The road area characteristics for both urban and rural areas are summarized in Table 6.1. Previous research initiatives on the severity of injuries compared models classified according to these characteristics. These investigations found that before 2016, researchers rarely compared the two roadway types, focusing instead on either the urban or rural model (Chen, Zhang, Huang, et al., 2016; Das & Abdel-Aty, 2011; Khorashadi et al., 2005). After 2016, however, some clear comparisons (Islam&Brown, 2017; Uddin & Huynh, 2017; Wu et al., 2016) between the urban and rural locations were undertaken. Nonetheless, none of these extant investigations has attempted a vigorous comparison of rear-end crashes in urban and rural locales. The present study fulfills this research gap.

In terms of methodology, most studies have applied the mixed logit model. The mixed effect multinomial logit model is an analysis tool that allows the parameters of exogenous variable values to vary to accident cases by specifying the function of unobserved heterogeneity into the linear function (Washington et al., 2011)

Table 6.1 Summary study injury severity analysis base on roadways

Works	Focus on Rear-end crash	Road location			Raised issues	Method
		Urban	Rural	Comparison		
This study	✓	✓	✓	✓	Comparison fata rear-end crash based on road location (Urban and Rural)	Hierarchical Binary Logistic model
Uddin and Huynh (2017)	-	✓	✓	✓	Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways	Mixed logit model
Islam and Brown (2017); Wu et al. (2016)	-	✓	✓	✓	A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways	Mixed logit model
Wu et al. (2016)	-	✓	✓	✓	Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways	Mixed logit model and Nested logit model
Chen, Zhang, Liu, et al. (2016)	-	-	✓	-	Driver injury severity outcome analysis in rural interstate highway crashes	a two-level Bayesian logistic regression
Chen, Zhang, Huang, et al. (2016)	-	-	✓	-	Examining driver injury severity outcomes in rural non-interstate roadway crashes	Hierarchical ordered logit model
Das and Abdel-Aty (2011)	✓	✓	-	-	A combined frequency–severity approach for the analysis of rear-end crashes	Genetic Programming
Khorashadi et al. (2005)	-	-	✓	-	Differences in rural and urban driver-injury severities in accidents involving large-trucks	Multinomial logit

Research projects on road safety often consider accident data on hierarchically ordered levels. It has been observed that accidents do not occur at single planes. Rather, they happen at multiple levels: the driver may represent the first level, while roads characteristics divided by road range or area of each province may be considered at a higher plane (Dupont et al., 2013). The multilevel or hierarchical structure analysis offers the advantage of clarity in the perception of the features of different accident occurrences. For example, an accident analysis may be divided into two levels: the driver factor level (such as gender, age, drunk driving, etc.) and vehicle type (such as personal cars, buses, trucks, etc.) form the first level, while the type of road is held as the second level. The analysis results can then display the factors that most influence accidents or cause the most critical injuries. For level 2, the roads may be divided into segments or according to the controlled phases of each highway. The differences between the road conditions selected for analysis may include features such as road type (main road, secondary road, etc.), police vigilance, and traffic volume passing through the particular stretch of the road. Inevitably, this examination will yield discrete results for accidents and accident severity as mentioned above. By analyzing two distinct planes, the different policies may be observed both at the organization level of road maintenance and of drivers who use the road types (driver characteristics). The results of such a model could be employed both for the determination of guidelines for a spatial policy and for a personal level campaign to reduce rear-end accidents (Park et al., 2017).

6.3 Multilevel analysis for road safety research

This section describes the application of the logistic model to the predictive analysis of fatal rear-end crashes caused by the effect of relevant variables. An additional concept for the selection of variables to be incorporated into the model pertains to road accidents classified into more than one level. In other words, explanation variables that affect injury levels should have hierarchical structures. For instance, based on researchers’ viewpoints, accident cases should be assigned personal factors that affect the first injury level along with the second level of the physical features of the road where the accident occurred: straight or curved roads, intersection characteristics, (Dupont et al., 2013) or area characteristics such as sub-districts, districts, provinces, etc.

The application of multilevel analyses for road safety studies may be categorized according to the characteristics of data analysis management, which comprise 3 types:

1) Multilevel modeling of aggregate accident data. This type of analysis examines spatial data distribution, dividing data into hierarchical structures according to the nature of the area (spatial data). The division of the area depends on the way the research questions are framed. For example, the Adanu et al. (2017) study determined the second level factor to be the area attributes classified according to the postal code of each location. The aggregate data analysis encompassed the risks of accidents occurring on each road or in each area. The area classification yielded a large amount of accident data, enabling the researchers to conduct a multilevel modeling of the aggregate data analysis (Dupont et al., 2013).

2) Multilevel modeling of disaggregate accident data. This kind of analysis focuses on separate accident cases. Most multilevel disaggregate data are analyzed for injury level patterns and can be organized as hierarchical data: for example, road characteristics, regional characteristics, and so on. Analyses based on the multilevel modeling of disaggregate accident data focus on accidents that involve a small number of vehicles (Dupont et al., 2013). The solution is to specify the estimated variance of the random effect of the second level data. For example, Kim et al. (2007) investigation analyzed the intersection characteristics (level two). The variables selected at this level were traffic signal lights, the angles of the intersections angle,

road segment types, etc. These were tested to determine the type of crashes that occurred more frequently: rear or side collisions.

3) Multilevel modeling of behavioral and attitudinal data. This sort of analysis investigates the attitudes of drivers through aspects such as speed, drunk driving, and seatbelt use. This method is grounded in the concept that the questioned drivers vary according to road-sites. To cite an example, Vanlaar (2005) examined drivers who drank alcohol before driving and determined variables according to diverse to road characteristics such as the traffic flow, the estimated density of parking cars etc.

The present study determines the factors affecting the gravity of the injury at the second level to be the road characteristics (divided by controlled road segment) because the researchers believe that the features of each controlled road segment depend on the supervision of each highway district divided by the boundary of the province or by the district in the case of the large province. The multilevel analysis provides the benefit of the ability to interpret the results of the factors affecting injury levels both at the stage of the occurrence of the accident at the individual level and at the plane of different road characteristics.

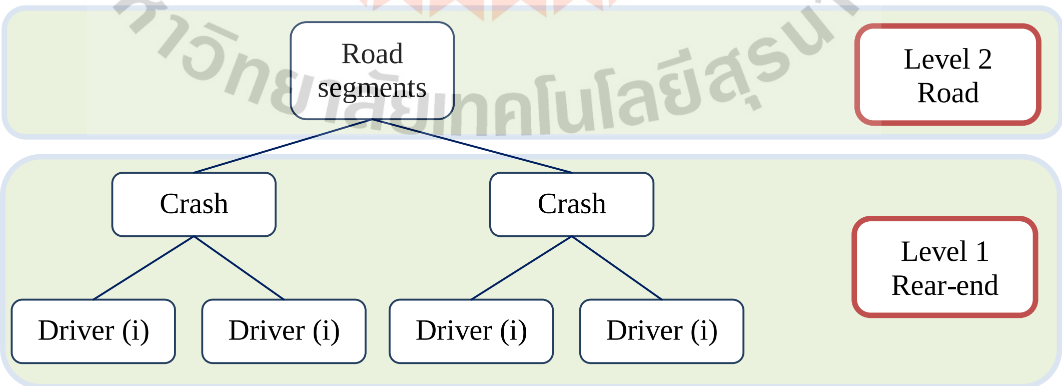


Figure 6.2 Hierarchical structure of rear-end crash

Logistics models applying multilevel concepts begin by determining the driver factor (driver *i*) in the road group *j*. Each accident involves the driver (driver *i* th) on road j (road *j* th) (Figure 6.2) and determines the two severity levels including

$$Y_{ij} = \begin{cases} 1; \text{ in case of fatal crash} \\ 0; \text{ in case of non} - \text{fatal crash} \end{cases}$$

where $Y_{ij}|\mathrm{p_{ij}} \sim Bernouilli(p_{ij}), p_{ij} = \Pr(Y_{ij} = 1)$ is the probability of driver *i* from road *j* being fatally injured in a crash emanating from the relationship of the estimated parameters of explanation variables such as sex, age, which can be calculated from

$$\mathrm{log}it(P_{ij}) = \eta_{ij} = \beta_{0j} + \beta_{1j}X_{ij} \tag{6.1}$$

where η_{ij} represents the log odds of driver i on road j which has fatal rear-end crashes. β_{0j} is the constant or the average value of log odds1 which has fatal rear-end crashes on road j only. X_{ij} is an explanatory variable at the individual level for predicting likelihood odds of fatality *j* and β_{1j} are parameter values that indicate the relationship (slope) between the driver level variables and the log odds likelihood that possibly causes the fatalities from rear-end collisions. Equation (6.1) assumes that accidents on each road result in different degrees of severity. Thus, the equations are different from those of common logistic models. Each road has a constant value (β_{0j}) but the slopes (β_{1j}) are different.

Considering a multilevel model as the explanation of the variation of the regression coefficient and variables at the second level from the road characteristics, these qualities are both supposed to be constant and the slope values are the effect at level 1 (Equation (6.1)) which can be varied according to the explanation variables of each road (Z_j) such as length, annual average daily traffic (AADT), etc. This equation will be:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + \mu_{0j} \tag{6.2}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + \mu_{ij} \tag{6.3}$$

where γ_{00} indicates the log odds log odds likelihood of fatal rear-end crashes on the road, Z_j is an explanation variable for each road j, and γ_{01} is the slope or the relationship between the predictor variables. Z_j, $\mu_{0j,}$ and μ_{ij} are prediction errors of information at the road level and demonstrate the unique effect of each road j. γ_{10} is the mean value of the predictor effect from level 1 and γ_{11} is the slope showing the relationship between the variables at road level, and the fatality rate on each road j.

The multilevel model is created by substituting β_{0j}andβ_{1j}into Equation (6.1) to get

$$\eta_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{11}X_{ij}Z_j + \gamma_{01}Z_j + \mu_{ij}X_{ij} + \mu_{0j} \tag{6.4}$$

Equation (*4*) can be explained by asserting that the effect of the variable at the driver level is predicted by the fixed effect of each road. For example, the value obtained through Equation (2) is not substituted in Equation (1). Therefore, to reduce the errors of Equation (*1)*, the effects of factors at the driver level are first allowed to vary from the characteristics of each road. This equation subsequently demonstrates a random intercept and random slope. For models that only exhibit a random intercept the reference to the driver odds log i from road j will yield the fatality risk.

γ_{00} represents the calculation of the log odds of fatality of common drivers on general roads. In other words, the different road characteristics indifferently affect the likelihood of death. $\gamma_{10}X_{ij}$ represents the effect value from the driver level and $\gamma_{01}Z_j$ indicates the effect values of each road.

To obtain additional effects in a multilevel model consisting of $\gamma_{11}X_{ij}Z_j$, the parameters of the driver level are believed to affect the road level and $\mu_{ij}X_{ij}$ are variables indicating the randomization of both the constant and slope values. Marginal effects are the effect that one-unit increase of an explanatory variable has on fatality probabilities on rear-end collision.

Test results must be obtained from models without parameter estimation (unconditional model) before a multilevel model analysis is conducted. This task is accomplished by considering the proportion of variance of dependent variables (outcome) within the group (accidents happening on the same road) and between groups (each road). This intra-class correlation coefficient (ICC) is potentially calculated by assuming that the dependent variables have logistic distribution. For the driver level errors of which the variance value is $\pi^2/3$ (Bryk & Raudenbush, 1992), ICC values can be calculated as follows:

$$\rho = \frac{\sigma_{\mu_0}^2}{\sigma_{\mu_0}^2 + \pi^2/3} \tag{6.5}$$

Where$\sigma_{\mu_0}^2$ is the variance of the dependent variables between each road (level two). ICC values should be nonzero for them to be suitable for multilevel modeling. If the ICC value is zero, there is no variation between the data pertaining to accidents occurring on each road. Previous research has evinced that ICC values should be higher than 0.09 or 9%. At the second level, the variables included the intersection (Kim et al., 2007), for which the variance at accident level was 16% (Andrew P. Jones & Stig H. Jørgensen, 2003), and the injury levels in accident cases, for which the ICC value was measured as 28.9% (Huang et al., 2008)

The following step is taken to assess mode suitability by comparing between the intercept-only model with all parameters set to zero and the convergence model with parameter vector βThe values used for comparison are the log-likelihood of both models, called Pseudo R-Squared or McFadden ρ^2

$$\rho^2 = 1 - \frac{LL(\beta)}{LL(0)} \tag{6.6}$$

where $LL(\beta)$ is the log-likelihood value of the model with parameter estimation and $LL(0)$ is the log-likelihood of the model without parameter estimation or with parameter estimation that only takes a constant value. Ifρ^2 has a value close to 1, the model's predictive accuracy is close to the actual data (Washington et al., 2011).

In this study, the R program package that included: glm: (*Fitting Generalized Linear Models)* was used to analyze for coefficients without the multilevel modeling technique. Additionally, glmer: (*Fitting Generalized Linear Mixed-Effects Models*) was used to analyze for coefficients with multilevel modeling technique.

6.4 Developing the models

6.4.1 Data collection

The data consist of 3 parts: 1) the highway crash report data including road accidents occurring nationwide on Thai highways between 2011 and 2015. Each case was surveyed and recorded by highway district officials who are regularly stationed at various provinces throughout the country. The data consist of information relating to accident locations (road/kilometer post), date of accident, physical road characteristics (road median type, intersection, etc.), environmental information (weather, time, lighting conditions), driver evidence (car type, driver age, gender, seat belt use, and alcohol or drugs consumption) and the Universal Transverse Mercator (UTM) coordinate, etc. All data were collected from the Department of Highway Accident Information Management System (HAIMS). Next, the data were screened for rear-end accidents and incomplete information was removed. Finally, 2,096 cases involving 4,554 drivers remained for analysis 2) Traffic volume data were drawn from the Traffic Information Movement System. The two data-sets were matched with highway numbers and kilometer posts, and 3) According to the classification of urban or rural areas, the data were obtained on the geographical information system (GIS) in the form of a shape file) on the datum WGS84, UTM zone 47 range system, showing the administrative controls of a municipality as polygon data due to the clear division of building densities resulting in the differences in traffic or physical road characteristics, car types, number of road users, and connecting roads.

To specify the location of cases in either urban or rural areas, the GIS program was used by commanding selection by location and by determining rear-end shape files comprising 953 drivers in urban areas and 3,061 drivers in rural areas. The number of victims in urban areas was rather small because the department of highways officers did not survey the data if the accidents were not serious if minor injuries were sustained in accidents occurring in urban areas, and if the victims could settle the matter successfully with each other. Consequently, such instances were not available in the system.

The description data presented in Table 6.2 shows the 24 independent variables divided by road areas (urban and rural) and classified according to the dependent variable, a fatal rear-end crash with 1 or more fatality. After initial consideration, it was found that when accidents occur, the highest proportion of the victims of fatal rear-end accidents include a drunk driver. This ratio remains high in both urban (5.25%) and rural (19.65%) locations, followed by roads that are regularly used. In terms of road factors, the notable proportion of fatal rear-end crashes were found to be caused by painted road medians with values of 5.09% in urban areas, and 19.57% in rural areas.

6.4.2 Model developments

The model development process shown in Figure 6.3 included 3 essential steps including 1) analyzing the correlation test for the relationship between the independent and dependent variables, 2) considering data consistency through the creation of hierarchical structural models, and 3) creating a hierarchical model comprising 2 sub-models incorporating the random intercept model, and the random parameters model.

Conducting a correlation test can help to obtain an overview of the relationships between the independent variables. It can also assist in the selection of the variables to assimilate into the model by distinguishing independent and dependent variables between fatal and non-fatal rear-end crashes. Three variables were not included into the models based on the selection criteria of significant values at p-value>0.05: “normal,” “env_surface,” and “weather.”

Table 6.2 Data Description

Variable name	Description	Code	Description	Road Area							
				Urban				Rural			
				Non-Fatal Crash		Fatal Crash		Non-Fatal Crash		Fatal Crash	
				Freq.	%	Freq.	%	Freq.	%	Freq.	%
VehType	Type of vehicle	1	Small Vehicle	244	5.36%	84	1.84%	919	20.18%	306	6.72%
		2	Middle Vehicle	385	8.45%	113	2.48%	1,434	31.49%	412	9.05%
		3	Large Vehicle	80	1.76%	47	1.03%	336	7.38%	194	4.26%
Gender	Gender of driver	0	Male	550	12.57%	202	4.62%	2,094	47.85%	757	17.30%
		1	Female	140	3.20%	30	.69%	485	11.08%	118	2.70%
Main Road	Crash location divided by traffic lane	0	Outer traffic lane	640	14.05%	226	4.96%	2,378	52.22%	844	18.53%
		1	Inner traffic lane	69	1.52%	18	.40%	311	6.83%	68	1.49%
Normal	The road was not being repaired	0	Other	9	.20%	7	.15%	76	1.67%	21	.46%
		1	Yes	700	15.37%	237	5.20%	2,613	57.38%	891	19.57%
Divided road	Road was divided by median island	0	No	264	5.80%	103	2.26%	977	21.45%	296	6.50%
		1	Yes	445	9.77%	141	3.10%	1,712	37.59%	616	13.53%
Flush	Road was divided by flush median	0	No	676	14.84%	232	5.09%	2,570	56.43%	828	18.18%
		1	Yes	33	.72%	12	.26%	119	2.61%	84	1.84%
Raised	Road was divided by raised median	0	No	567	12.45%	179	3.93%	2,116	46.46%	679	14.91%
		1	Yes	142	3.12%	65	1.43%	573	12.58%	233	5.12%
Depressed	Road was divided by depressed median	0	No	532	11.68%	199	4.37%	2,005	44.03%	693	15.22%
		1	Yes	177	3.89%	45	.99%	684	15.02%	219	4.81%
Barrier	Road was divided by barrier	0	No	616	13.53%	225	4.94%	2,353	51.67%	832	18.27%
		1	Yes	93	2.04%	19	.42%	336	7.38%	80	1.76%
Concrete	The road was concrete pavement	0	No	623	13.68%	217	4.77%	2,348	51.56%	815	17.90%
		1	Yes	86	1.89%	27	.59%	341	7.49%	97	2.13%
Straight	Road horizontal alignment	0	Other	39	.86%	29	.64%	162	3.56%	67	1.47%
		1	Straight	670	14.71%	215	4.72%	2,527	55.49%	845	18.56%
Slope	Road graded	0	Other	696	15.28%	234	5.14%	2,611	57.33%	856	18.80%
		1	Slope	13	.29%	10	.22%	78	1.71%	56	1.23%
Intersection	Rear-end collision happened near intersection (<100 m)	0	Other	585	12.85%	196	4.30%	2,168	47.61%	789	17.33%
		1	Yes	124	2.72%	48	1.05%	521	11.44%	123	2.70%
Median opening	Rear-end collision happened at opening median point	0	Other	633	13.90%	211	4.63%	2,364	51.91%	774	17.00%
		1	Yes	76	1.67%	33	.72%	325	7.14%	138	3.03%
env_surfaces	Status of road surface	0	Dry	656	14.40%	226	4.96%	2,506	55.03%	847	18.60%
		1	Wet	53	1.16%	18	.40%	183	4.02%	65	1.43%
Weather	Visualization of drivers as accident	0	Other	654	14.36%	218	4.79%	2,496	54.81%	845	18.56%
		1	Clean	55	1.21%	26	.57%	193	4.24%	67	1.47%

Table 6.2 Description Data (Continued)

Variable name	Description	Code	Description	Road area							
				Urban				Rural			
				Non-Fatal Crash		Fatal Crash		Non-Fatal Crash		Fatal Crash	
				Freq.	%	Freq.	%	Freq.	%	Freq.	%
Day	Time of crash	0	Night	201	4.41%	89	1.95%	824	18.09%	398	8.74%
		1	Day	508	11.16%	155	3.40%	1,865	40.95%	514	11.29%
Night_NoLight	Collision happened at night in low-light conditions	0	Other	670	14.71%	209	4.59%	2,462	54.06%	753	16.53%
		1	Yes	39	.86%	35	.77%	227	4.98%	159	3.49%
Night_light	Collision happened at night in high-light conditions	0	Other	547	12.01%	190	4.17%	2,092	45.94%	673	14.78%
		1	Yes	162	3.56%	54	1.19%	597	13.11%	239	5.25%
safety_equip	Driver used safety equipment	0	No	422	9.27%	182	4.00%	1,591	34.94%	681	14.95%
		1	Yes	287	6.30%	62	1.36%	1,098	24.11%	231	5.07%
alcohol	Driver was drunk	0	No	702	15.42%	239	5.25%	2,664	58.50%	895	19.65%
		1	Yes	7	.15%	5	.11%	25	.55%	17	.37%
Speed_Exceed	Cause of crash was exceeding speed	0	No	181	3.97%	86	1.89%	805	17.68%	299	6.57%
		1	Yes	528	11.59%	158	3.47%	1,884	41.37%	613	13.46%
four_lane	Road was 4 lanes or more	0	No	537	11.79%	217	4.77%	2,059	45.21%	749	16.45%
		1	Yes	172	3.78%	27	.59%	630	13.83%	163	3.58%
Crash Type	Moving of leading vehicle before rear-end crash	1	Going Straight	458	10.06%	152	3.34%	1,776	39.00%	569	12.49%
		2	Decelerating speed	212	4.66%	71	1.56%	791	17.37%	276	6.06%
		3	Stopped on traffic lane	39	.86%	21	.46%	122	2.68%	67	1.47%

Note: Freq. = Frequency.

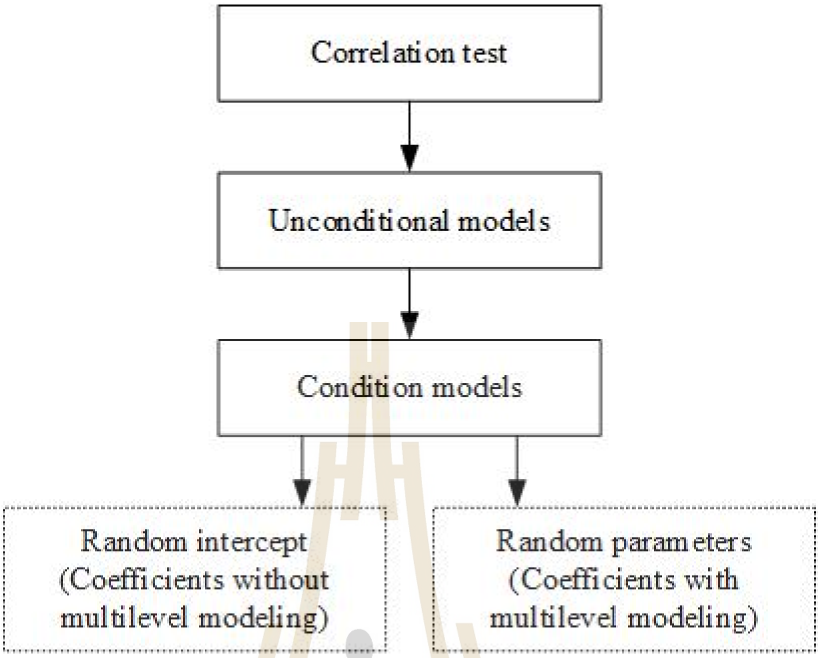


Figure 6.3 Developing Model

With regard to the random effect of unconditional models, the results depicted in Table 6.3 evidence that the urban roadway model was generated with random effect with an ICC value of 0.09, and the value of 0.093 was computed for the rural roadway. It may thus be interpreted that the fatal rear-end crash variance on the same road accounted for approximately 9% of the total variance. Both models met the acceptance criteria for the creation of multilevel models (Kim et al., 2007).

All the independent variables included in the construction of the models were divided into two standards: 1) the random intercept prototype (coefficients without multilevel modeling) in which the random effects values of independent variables and the physical characteristics of the road were not set according to road data, and 2) the random parameters exemplar (coefficients with multilevel modeling) in which the random slopes of the physical characteristics variables (such as vehicle types, road median types, intersections, slopes, etc.) were set according to the road segments where the accidents occurred.

Table 6.3 Results of unconditional model

Effect	Urban road		Rural road	
	Estimate	std.	Estimate	std.
Fixed Effect				
Intercept	-1.066	0.0742	-1.0813	0.0383
Random effect				
Intercept	0.3086	0.0233	0.3295	0.0151
ICC	0.086		0.091	
- 2log L	782		3255.2	

Note: N of Urban = 953 and N of Rural = 3601

6.5 Results and discussion

The results of the analysis of the 4 constructed models included coefficients without multilevel modeling technique (NMLM) and coefficients with multilevel modeling technique (MLM) and were divided into 2 representative areas, urban and rural, as displayed in Table 6.4.

6.5.1 Urban models

Overall, 922 drivers on 331 roads were included in the models of fatal and non-fatal rear-end crashes occurring on urban roadways. There were. In the NMLM model, the value of ρ^2 was 0.085, which met the acceptance criteria (Kockelman & Kweon, 2002) but in comparison, the MLM value of $\rho^2$was 0.1498, indicating that the prediction capability achieved through MLM was much better than via NMLM. Random effect values were estimated at 4.454 (S.E. = 0.746) evincing that fatal rear-end crashes varied according to road segments. This result significantly affected the models.

The comparison between the estimate parameter values of NMLM and MLM revealed their tendency to move along the same direction. In other words, in the case of a positive estimated direction, the movement would be positive for both models with only a slight difference in the significance level. The first group pertained to the consistent variables of the two models, “safety_eqiup_1 = 1” which had a marginal effect of –0.0939. This outcome suggested that the use of driver safety equipment could reduce fatal rear-end crashes by 9.39%. This result is precisely reasonable and is in congruence with numerous previous studies (Abdel-Aty, 2003; Hassan & Meguid, 2017; Shawky et al., 2016; Wiacek et al., 2015). The next variable in this group was driver age with marginal effect of 0.0013. This finding demonstrated the increasing likelihood of fatality when older drivers were involved in rear-end accidents. This result opposes the findings of Chen et al. (2015); Chen, Zhang, Yang, et al. (2016) for whom the driver age did not affect the injury level. The studies conducted by Abdel-Aty (2003); Xiao et al. (2019) also obtained equivalent results and were aligned to the two previously mentioned studies. The third variable was the number of traffic lanes. This study discovered that roads with four or more traffic lanes reduced the likelihood of fatal rear-end crashes, perhaps because roads with two traffic lane have narrow boundary lanes. This inadequate space may result in more severe rear-end crashes when the car driver wants to turn immediate right and must break suddenly because of the differing speeds of the two cars involved in the accident. This finding is congruent with the results obtained by a study conducted by Flask et al. (2014), which found that roads with 4 or more traffic lanes activated the reduction of the risk of fatality. However, this result is not consistent with the outcomes of investigations undertaken by Yuan et al. (2017), which reported that the number of traffic lanes increased the injury levels of rear-end crashes in which a truck was involved. In addition, Hyodoa and Todorokia (2018) found that this variable did not significantly affect injury levels.

For the second variable group, “day = 1” was not significant in NMLM, but significant in MLN (p <0.1) with a marginal effect of –0.0462. It may be inferred that fatality risks could be reduced when the rear-end crash happened in the daytime. This result is understandable because a large number of vehicles ply on urban roads during the daytime and also because clear vision is a factor for the prevention of accidents. In a daytime scenario, a driver may be able stop the car in time before an accident occurs or the presence of a large number of vehicles on urban thoroughfares may cause drivers not to operate vehicles at vastly different speeds (ITARDA, 2011; Sullivan & Flannagan, 2003; Yuan et al., 2017)

For the third variable group, “lnAADT” (marginal effect = –0.0399) was significant (p-value <0.1) in NMLM, but not significant in MLM. Perhaps this variable was not significant in MLM because the traffic volume was not accorded enough weighted values to exert a significant effect when road factors were allowed to vary according to the road segments. Observed mostly in the rear-end crash frequency model, these variables often report a significantly positive effect (Chen, Zhang, Yang, et al., 2016; Das & Abdel-Aty, 2011).

Table 6.4 Estimate results model of urban and rural roadways (Urban roadway)

Variables	**Urban**					
	NMLM			**MLN**		
	Est.	Std.	Marg.	Est.	Std.	Marg.
Fixed effects:						
(Intercept)	2.819	4.100		1.966	4.496	
VehType=2	−0.726	0.447	-	−0.672	0.527	-
VehType=3	0.187	0.577	-	0.601	0.679	-
Gender=1	−0.518	0.547	-	−0.528	0.672	-
safety_equip_1=1	−1.343**	0.630	−0.0751	−1.224*	0.739	−0.0393
alcohol_1=1	0.332	1.418	-	0.379	2.131	-
person_age	0.032	0.013	0.0017	0.036**	0.016	0.0013
Day=1	−0.696	0.499	-	−1.171*	0.664	−0.0462
Night_NoLight=1	1.185	0.954	-	1.064	1.108	-
CrashType=2	−0.224	0.550	-	−0.289	0.788	-
CrashType=3	1.046	0.886	-	1.172	1.067	-
env_surfaces=1	−0.861	0.865	-	−0.685	1.242	-
Main_Road=1	0.302	1.228	-	0.294	2.003	-
Devided_Median=1	0.229	1.516	-	0.177	2.457	-
four_lanes=1	−1.756**	0.779	−0.0992	−2.345*	1.391	−0.0342
Flush=1	−0.180	1.880	-	0.439	2.817	-
Riased=1	0.848	1.473	-	0.909	2.354	-
Depressed=1	0.432	1.452	-	1.014	2.340	-
Intersection=1	−0.699	0.585	-	−0.930	1.096	-
Median_opening=1	0.318	0.663	-	−0.433	1.662	-
PerCTruck	0.043	0.042	-	0.045	0.041	-
Straight=1	−0.820	0.989	-	−1.216	1.064	-
Councrete=1	0.619	0.785	-	−0.641	1.091	-
lnAADT	−0.750*	0.443	−0.0399	−0.629	0.490	-
Slope=1	−1.001	2.710	-	−0.824	3.111	-
Random effects:						
[illegible] mean)				4.454**	0.746	
LL(0)	−391.0					
LL(NMLM)	−357.7					
LL(MLM)				−332.4		
McFadden [illegible]	0.085			0.1498		

Note: ** $p<0.05$, * $p<0.1$,
LL(0): *log-likelihood value at convergence*
NMLM*: Coefficients without multilevel modeling technique,*
MLM*: Coefficients with multilevel modeling technique,* ***Est.****: Estimation,* ***Std.****:Standard error,*
Marg.: *Marginal effect*
Urban roadway models: Number of observation: 922, groups: Route_Con2, 331 segments.
Rural roadway models: Number of observation: 3454, groups: Route_Con2, 715 segment.

Table 6.4 Estimate results model of urban and rural roadways (Rural roadway) (Continued)

Variables	**Rural**					
	NMLM			**MLN**		
	Est.	Std.	Marg.	Est.	Std.	Marg.
Fixed effects:						
(Intercept)	0.410	1.530		0.806	2.007	
VehType=2	−0.258*	0.141	−0.0248	−0.122	0.169	-
VehType=3	0.457**	0.188	0.0478	0.600**	0.229	0.0522
Gender=1	−0.190	0.179	-	−0.260	0.216	-
safety_equip_1=1	−0.737**	0.176	−0.0713	−0.923**	0.213	−0.0710
alcohol_1=1	0.472	0.575	-	−0.238	0.810	
person_age	0.015**	0.004	0.0014	0.013**	0.005	0.0010
Day=1	−0.782**	0.168	−0.0790	−0.924**	0.220	−0.0764
Night_NoLight=1	0.317	0.259	-	0.343	0.333	-
CrashType=2	0.310*	0.178	0.0303	0.033	0.266	-
CrashType=3	0.952**	0.282	0.0998	1.198**	0.353	0.1111
env_surfaces=1	−0.483**	0.262	-0.0448	−0.686*	0.353	−0.0493
Main_Road=1	−0.884**	0.342	−0.0797	−1.208**	0.452	−0.0817
Devided_Median=1	−0.106	0.460	-	1.380	0.903	-
four_lanes=1	−0.059	0.237	-	−1.332*	0.749	−0.0608
Flush=1	1.225**	0.537	0.1340	0.171	1.016	-
Riased=1	0.470	0.452	-	−0.866	1.053	-
Depressed=1	0.178	0.449	-	−0.642	0.971	-
Intersection=1	−0.789**	0.213	−0.0724	−1.938**	0.704	−0.0690
Median_opening=1	−0.289	0.221	-	−0.680	0.714	-
PerCTruck	0.041**	0.015	0.0040	0.046**	0.018	0.0037
Straight=1	0.117	0.305	-	0.318	0.449	-
Councrete=1	−0.027	0.322	-	−0.749	0.474	-
lnAADT	−0.314*	0.167	−0.0305	−0.470**	0.218	−0.0372
Slope=1	0.443	0.385	-	0.923	5.022	-
Random effects:						
[illegible]mean)				1.350**	0.411	
LL(0)	−1621.7					
LL(NMLM)	−1486.8					
LL(MLM)				−1312.3		
McFadden [illegible]	0.0832			0.1907		

6.5.2 Rural models

The overview of the models of rear-end accidents on rural roadways encompassed 3,454 drivers on 715 road segments. The value of ρ^2 in the NMLM model equaled 0.0832 which met acceptance criteria. However, when compared with the value of ρ^2 of MLM which was 0.1907, the prediction ability of MLM was demonstrated to be superior to that of NMLM. The random effect was estimated to be

1.354 (SE = 0.411) and indicated that fatal rear-end crashes varying according to road segments significantly affected the models with regard to rural roads.

The acquired overall picture of the rural model was alike the urban model, which evinced the tendency estimation in the same direction. In terms of the first variable group, vehicle type represented the primary significant variable for both NMLM and MLM. Large vehicles were found to be more likely to cause more fatal rear-end crashes. When considering the MLM, it was found that the estimated value was higher (marginal effect = 0.0478 and 0.0522). Thus, this variable is quite reasonable because the vehicle size could affect the force of the impact. In considering the physical characteristics of rural roads on which most vehicles are driven at high speed along with the vehicle size, the increase in fatal rear-end crashes (Abdel-Aty, 2003; Wiacek et al., 2015; Yan & Radwan, 2009; Zeng et al., 2016) becomes even more plausible. The second most important variable is seatbelt usage. In terms of driver age, it was found that older drivers caused a slightly higher fatality risk from rear-end crashes (marginal effect = 0.001). The results obtained from both NMLM and MLM models were very similar. This result is consistent with Zheng et al. (2018) study. With regard to the timing of the occurrence of the rear-end crash as a variable, the result was obvious: there were higher fatality risks when rear-end crashes occurred at night. In term of tendency estimation, MLM was found to present a clearer parameter value. For variables with rather high parameter estimates, the rear-end crash was type 3, or an accident with a parked lead car. This finding depicts that in comparisons between rear-end crash types, rural road rear-end accidents with a parked lead car were 3.3 times more likely to cause fatalities than when the lead car was driving at a normal speed (marginal effect = 0.1111). This result is consistent with the rear-end crash study conducted by Beck and Tripathi (2015) who anticipated that a high degree of difference in the speeds of the front and rear vehicles in an accident augmented the risk of fatality (Misener et al., 2000). The analysis of road surface variables revealed that dry road surfaces increased fatality risk (marginal effect = –0.0493, p-value <0.1). This outcome is aligned with the findings of Kim et al. (2007)'s study. They reported that most drivers tended to drive carefully in wet road conditions by decelerating. Therefore, when a rear-end crash occurred, there was a decreased risk of fatality in comparison to dry road conditions in which divers could accelerate to the extent they desired (Chen, Zhang, Huang, et al., 2016; Chen, Zhang, Yang, et al., 2016).

The accident location variable compared rear-end crashes occurring on main roads to those that happened on parallel pathways. Rear-end accidents on the main roads reported lower fatality risks (marginal effect = –0.0817), perhaps because of superior access control into the interior traffic lanes, which causes only a slight difference in vehicle speed. When a rear-end crash occurs in such conditions, the accidents are not very grave. Conversely, the parallel road often allows the parked car scenario. This result is consistent with Khorashadi et al. (2005). The variable of accident location near an intersection caused a noticeably elevated result in the MLM (marginal effect = –0.069). When considering the effect, it was found that accidents occurring at intersections cause less fatal rear-end crashes, perhaps because of the presence of clear signs on intersections on Thai highways. This signage makes drivers accessing intersections reduce their speed (Li et al., 2019). In terms of the involvement of trucks, these were found to cause a greater proportion of fatal rear-end crashes. The results posted by Kidando et al. (2019) indicating that increasing truck proportions caused lower fatality are deemed irrelevant to the present study; however, Yan and Radwan (2009) reported that rear-end crashes in which trucks were involvement caused more severe injuries. With regard to the AADT, an increase in this variable was found to lower the likelihood of fatality (marginal effect = –0.0372). This outcome is consistent with the studies conducted by Haghighi et al. (2018); Kidando et al. (2019). Both these research endeavors found that an increase in traffic quantity resulted in decreased vehicular speed.

The number of traffic lanes was a variable that was found to be significant in MLM but not in NMLM. The MLM analysis results for this study revealed that rear-end collisions that happened on roads with 4 or more traffic lanes were less likely to cause fatal rear-end crashes than those that occurred on roads with only two lanes (marginal effect = –0.0608, p-value <0.1). The number of traffic lanes influenced crash severities in varied ways since the parameters of the models were very different. A study undertaken by Flask et al. (2014) analyzed collisions with mixed effect models and also reported that increasing traffic lanes reduces accident severity.

Three variables were significant in the NMLM model, but not significant in the MLM: 1) rear-end crashes that occurred when the front car was decelerating caused fewer fatal rear-end crashes than those in which the lead car was moving at normal speed (marginal effect = 0.0303, p-value <0.1); 2) medium cars, private vehicles, and pick-up vans caused fewer fatality risks than motorcycles, a reasonable finding because of the presence of superior safety equipment such as airbags, seat belts within the vehicle body, equipment that can soundly reduce the injury severity (Abdel-Aty, 2003); and 3) painted road medians increased the risks of fatal rear-end crashes (marginal effect = 0.1340, p-value <0.05) because, in Thailand, these features are used on roads with relatively small traffic volumes and are associated with stretches on which most cars move at speed. This finding is consistent with the outcomes of a study accomplished by Tarko et al. (2008), which demonstrated that flush medians influenced the severity of injuries from accidents. It may be contended all the three variables mentioned above do not vary according to road segments: the rear-end crash severity attributed to these factors are similar regardless of the road type on which they occur.

6.5.3 Comparison of coefficients urban and rural roadway models

Some variables were significant with multilevel correlation coefficients in terms of the estimated parameter in both urban and rural models. It was found that rear-end crashes on rural roads tended to be more severe than those that occurred on urban roads (Kidando et al., 2019; Li et al., 2018). Seat belt use was the most important and obvious variable because of its relatively high estimate value to potentially reduce the risks of fatality. This result is understandable because using seatbelts reduces the chances of death regardless of whether accidents occur on urban or rural roads. The second most vital variable for both types of roads was the driver age, a factor that was discovered to move in the same direction for all the models. The involvement of an older driver in a rear-end crash probably increases the risk of death (urban: marginal effect = 0.0013, rural: marginal effect = 0.0010). Both the MLM and NMLM comparisons evinced the significant effect of driver age on the risk of fatality and this danger did not vary by road segments. In relation to environmental factors, nighttime collisions were found to move in the same direction for both types of roadways: the risk of fatalities increased for nighttime collisions. In terms of the significance level of this variable, both the NMLM and NMLM analyses of rural roads yielded significant results (urban: marginal effect = –0.0462; rural: marginal effect = –0.0764). With regard to urban roads, this variable was significant only when for the random effect model.

The variables selected for the urban road model were not significant while many significant variables were discovered for the rural model. However, the overall image represented the same direction of the estimate values:

1) Large vehicle type: It can be interpreted from the obtained results that a rear-end crash involving a truck on urban roads will not significantly result in fatalities. This outcome is deemed reasonable because of the speed limits applied on urban roads. This finding is also aligned to the results reported by Khorashadi et al. (2005)'s study: road accidents caused by speed violations are significant influential only on rural road models.

2) Crash type with a parked lead car: The speed of vehicles plying on urban roads is restricted by heavy traffic conditions and a large number of intersections. When a rear-end crash occurs, the speed of the involved vehicles is not vastly different (Beck&Tripathi, 2015). Khorashadi et al. (2005) found unequivocally that the variable identified as a parked lead car (at a standstill on a roadway) had much higher estimate coefficients in the rural road models than in the urban road models.

3) Dry or wet road surface characteristics: Vehicles operating on urban roadways exhibit comparable speeds regardless of wet or dry road conditions. Thus, effects are not significantly different for accidents occurring in either circumstance.

On rural roads, however, vehicles tend to accelerate on dry roads (Chen, Zhang, Liu, et al., 2016), causing more chances of fatalities in the case of an accident.

4) Crash locations (main roads or parallel lanes): Despite their ability to accommodate high vehicular speeds, main traffic lanes include a large number of intersections and access roads in urban areas, which obstruct their operating at full speed (Greibe, 2003).

5) Location of accident (near or further from intersections): In spite of a greater number of intersections, the severity of rear-end crash injuries were not found to be at significance level for urban areas (Greibe, 2003).

6) Truck volume specifying whether or not a truck is involved in an accident: This result was not significant in the urban area model.

7) Traffic volume: This factor yielded the interesting outcome that when it was specified that the likelihood of fatality did not vary according to the road type or segment, the variable achieved the significance level (p-value <0.1).

6.6 Conclusion

A rear-end crash is a primordial order crash type and is deemed important because of the continuously increasing number of fatalities attributed to such accidents. The physical road characteristics resulting from land use (divided into urban roads and rural roads) evidently influenced the severity of crash injuries in distinct manners. However, no study has yet compared the differences in the gravity of fatal and non-fatal rear-end crashes. This identified research gap required to be filled to benefit agencies engaged in road design, bodies supervising security policies, and scholars looking to apply these concepts to develop appropriate models.

This study found that the unconditional model can be analyzed via a layered structure and applied a hierarchical logistic regression model that determined the estimate parameters to vary according to road segments. It also compared modeling with multilevel coefficient correlation (MLM) to casting without the use of multilevel coefficient correlation (NMLM). The 4 models were consequently established through the utilization of data pertaining to rear-end accidents occurring on Thai highways in a given period. The results of a comparison study of model suitability effected by comparing the ρ^2 values found that the analyses achieved through MLM potentially yielded results superior to NMLM.

An overview of the results obtained from the models of urban and rural roads, the urban road models were discovered to yield fewer numbers of significant parameters than the rural road models. The number of traffic lanes commanded the highest estimate value, followed by seat belt use by the driver, and the time of the accidents. Therefore, policy makers can apply the results of this study to both rural and urban areas. For example, they could initiate a public campaign favoring seat belt use, and caution the populace to be more careful when driving at night or in the dark.

The risks of fatal rear-end accidents were found to increase on rural roads. Significant variables in the direction of escalation were identified as roads with less than 4 lanes, the lack of adequate intersections to reduce speed and prevent rear-end crashes, rear-end crashes on interior lanes, rear-end crashes with a parked lead car, rear-end crashes at night, the lack of seat belt use, acceleration on dry road surface conditions, rear-end crashes with large truck involvement, low-traffic roads, a large number of trucks, and the presence of older drivers.

This study presents the limitations of investigating a relatively small number of rear-end crashes, especially in urban areas. This difficulty emanated from the fact that small accident cases were neither surveyed nor recorded in the system. Nonetheless, the results of this study have revealed interesting variables in the form of crash types. The findings obtained from the constructed models evidenced that rear-end crashes with a parked lead car were most likely to cause fatalities. Accordingly, future studies should compare the severity of injuries vis-à-vis types of accidents and should also develop models that incorporate injury levels as dependent variables to obtain comprehensive information that may be applied to ensure higher road safety across Thailand.

6.7 Acknowledgment

The authors would like to thank Enago (www.enago.com) for the English language review.

6.8 Reference

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

Adanu, E. K., Smith, R., Powell, L., and Jones, S. (2017). Multilevel analysis of the role of human factors in regional disparities in crash outcomes. **Accident Analysis and Prevention**, 109: 10-17.

Andrew P. Jones, and Stig H. Jørgensen. (2003). The use of multilevel models for the prediction ofroad accident outcomes. **Accident Analysis and Prevention**, 35: 59–69.

Beck, D., and Tripathi, S. (2015). Investigation of Key Crash Types: Rear-end Crashes in Urban and Rural Environments Sydney NSW 2000 Australia: Austroads Ltd.

Bryk, A. S., and Raudenbush, S. W. (1992). Hierarchical linear models : applications and data analysis methods: Sage Publications, Newbury Park.

Chen, C., Zhang, G., Huang, H., Wang, J., and Tarefder, R. A. (2016). Examining driver injury severity outcomes in rural non-interstate roadway crashes using a hierarchical ordered logit model. **Accident Analysis and Prevention**, 96: 79-87.

Chen, C., Zhang, G., Liu, X. C., Ci, Y., Huang, H., Ma, J., . . . Guan, H. (2016). Driver injury severity outcome analysis in rural interstate highway crashes: a two-level Bayesian logistic regression interpretation. **Accident Analysis and Prevention**, 97: 69-78.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., and Guan, H. (2015). A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. **Accident Analysis and Prevention**, 80: 76-88.

Chen, C., Zhang, G., Yang, J., Milton, J. C., and Alcantara, A. D. (2016). An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naive Bayes (DTNB) hybrid classifier. **Accident Analysis and Prevention**, 90: 95-107.

Das, A., and Abdel-Aty, M. A. (2011). A combined frequency–severity approach for the analysis of rear-end crashes on urban arterials. **Safety Science**, 49(8-9): 1156-1163.

DOH. (2015). Thailand Traffic Accident on National Highways in 2014. Available from: https://bhs.doh.go.th/download/accident Date Accessed: June-2019

DOH. (2016). Thailand Traffic Accident on National Highways in 2015. Available from: https://bhs.doh.go.th/download/accident Date Accessed: June-2019

DOH. (2017). Thailand Traffic Accident on National Highways in 2016. Available from: https://bhs.doh.go.th/download/accident Date Accessed: June-2019

Dupont, E., Papadimitriou, E., Martensen, H., and Yannis, G. (2013). Multilevel analysis in road safety research. **Accident Analysis and Prevention**, 60: 402-411.

Flask, T., Schneider, W. H., and Lord, D. (2014). A segment level analysis of multi-vehicle motorcycle crashes in Ohio using Bayesian multi-level mixed effects models. **Safety Science**, 66: 47-53.

Greibe, P. (2003). Accident prediction models for urban roads. **Accident Analysis and Prevention**, 35: 273–285.

Haghighi, N., Liu, X. C., Zhang, G., and Porter, R. J. (2018). Impact of roadway geometric features on crash severity on rural two-lane highways. **Accident Analysis and Prevention**, 111: 34-42.

Hassan, M. T. Z., and Meguid, S. A. (2017). Effect of seat belt and head restraint on occupant's response during rear-end collision. **International Journal of Mechanics and Materials in Design**.

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

Hyodoa, S., and Todorokia, T. (2018). An Analysis of Risk Factors for Rear-Ender Accident on Urban Expressway Considering Accident Severity. International Symposum of Transport Simulation (ISTS'18) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS'18).

Islam, S., and Brown, J. (2017). A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama. **Accident Analysis and Prevention**, 108: 163-171.

ITARDA. (2011). *Rear-end collisions with a moving bicycle—posing a high risk of death*. Retrieved from Institute for Traffic Accident Research and Data Analysis, Japan:

Khorashadi, A., Niemeier, D., Shankar, V., and Mannering, F. (2005). Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. **Accident Analysis and Prevention**, 37(5): 910-921.

Kidando, E., Moses, R., Ozguven, E. E., and Sando, T. (2019). Incorporating travel time reliability in predicting the likelihood of severe crashes on arterial highways using non-parametric random-effect regression. **Journal of Traffic and Transportation Engineering (English Edition)**.

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

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

Li, Z., Chen, C., Ci, Y., Zhang, G., Wu, Q., Liu, C., and Qian, Z. S. (2018). Examining driver injury severity in intersection-related crashes using cluster analysis and hierarchical Bayesian models. **Accident Analysis and Prevention**, 120: 139-151.

Li, Z., Wu, Q., Ci, Y., Chen, C., Chen, X., and Zhang, G. (2019). Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. **Accident Analysis and Prevention**, 129: 230-240.

Misener, J. A., Tsao, H. S. J., Song, B., and Steinfeld, A. (2000). Emergence of a Cognitive Car-Following Driver Model: Application to Rear-End Crashes with a Stopped Lead Vehicle. **Transportation Research Record**, 1724(1): 29-38.

Park, H. C., Kim, D. K., Kho, S. Y., and Park, P. Y. (2017). Cross-classified multilevel models for severity of commercial motor vehicle crashes considering heterogeneity among companies and regions. **Accident Analysis and Prevention**, 106: 305-314.

Shawky, M. A., Kishta, M., and Al-Harthi, H. A. (2016). Investigating Factors Affecting the Occurrence and Severity of Rear-End Crashes. World Conference on Transport Research; Shanghai.

Sullivan, J. M., and Flannagan, M. J. (2003). *Risk of fatal rear-end collisions: is there more to it than attention?* Paper presented at the Proceeding of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Tarko, A. P., Villwock, N. M., and Blond, N. (2008). Effect of Median Design on Rural Freeway Safety: Flush Medians with Concrete Barriers and Depressed Medians. **Transportation Research Record**, 2060(1): 29-37.

Uddin, M., and Huynh, N. (2017). Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways. **Accident Analysis and Prevention**, 108: 44-55.

Vanlaar, W. (2005). Drink driving in Belgium: results from the third and improved roadside survey. **Accident Analysis and Prevention**, 37(3): 391-397.

Washington, S. P., Karlaftis, M. G., and Mannering, F. (2011). Statistical and Econometric Methods for Transportation Data Analysis (2 ed.). Boca Raton London New York Washington, D.C.

Wiacek, C., Bean, J., and Sharma, D. (2015). Real-world analysis of fatal rear-end crashes. 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration.

Wu, Q., Zhang, G., Zhu, X., Liu, X. C., and Tarefder, R. (2016). Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways. **Accident Analysis and Prevention**, 94: 35-45.

Xiao, D., Xu, X., and Duan, L. (2019). Spatial-Temporal Analysis of Injury Severity with Geographically Weighted Panel Logistic Regression Model. **Journal of Advanced Transportation**, 2019: 1-15.

Yan, X., and Radwan, E. (2009). Analysis of truck-involved rear-end crashes using multinomial logistic regression. **Advances in Transportation Studies an international Journal**, 17.

Yuan, Q., Lu, M., Theofilatos, A., and Li, Y. B. (2017). Investigation on occupant injury severity in rear-end crashes involving trucks as the front vehicle in Beijing area, China. **Chinese Journal of Traumatology**, 20(1): 20-26.

Zeng, Q., Wen, H., and Huang, H. (2016). The interactive effect on injury severity of driver-vehicle units in two-vehicle crashes. **Journal of Safety Research**, 59: 105-111.

Zheng, Z., Lu, P., and Lantz, B. (2018). Commercial truck crash injury severity analysis using gradient boosting data mining model. **Journal of Safety Research**, 65: 115-124.

CHAPTER VII

SPATIAL ZERO-INFLATED NEGATIVE BINOMIAL REGRESSION MODELS: AN APPLICATION TO ESTIMATE REAR-END CRASH FREQUENCIES ON THAI HIGHWAY

7.1 Abstract

Rear-end crash is a type of road traffic accidents that often occur. Currently, the application of advanced statistical models to predict the frequency of the accident number has been increasingly used as it makes the model predictions more accurate. This study focuses on fulfilling the application of statistical models to find the relationship between the explanatory variable and the rear end crash frequency. The data used in the study are rear-end collisions occurring on highways throughout Thailand in the years 2011-2018. The number of rear-end collisions was distributed according to the segments of which road physical characteristics were similar. In this study spatial correlation was applied by varying according to the jurisdiction of the Department of Highways. For model development, there were 4 models starting with Poisson regression model, Negative binomial model, Zero-inflated negative binomial model and spatial zero-inflated negative binomial model (SZINB). When compared with AIC, it was found that SIZNB was the model that suit data most. Regarding random effect results, the effect of the significance which was constant both

significant variables of conditional sate and zero state included Segment length, Number of lanes, and Traffic volume. This study can be a starting point for those who are interested in applying the spatial model in rear-end crash analysis.

7.2 Introduction

Rear-end crash is a crash type that worldwide researchers value. To predict the accident number, they have developed a model which is one method they have been focusing on for a long time. When considering the number of accidents classified by type of collisions on the highways in Thailand (Figure 7.1). The number of rear-end crashes or car crashes coming along the same direction was found the second-highest number of collisions, followed by off path in straight. For the number of deaths, the rear-end crashes caused the highest number of fatalities. In addition, when annually considered, it was found that its trend has been continuously increasing (DOH, 2016, 2017, 2018).

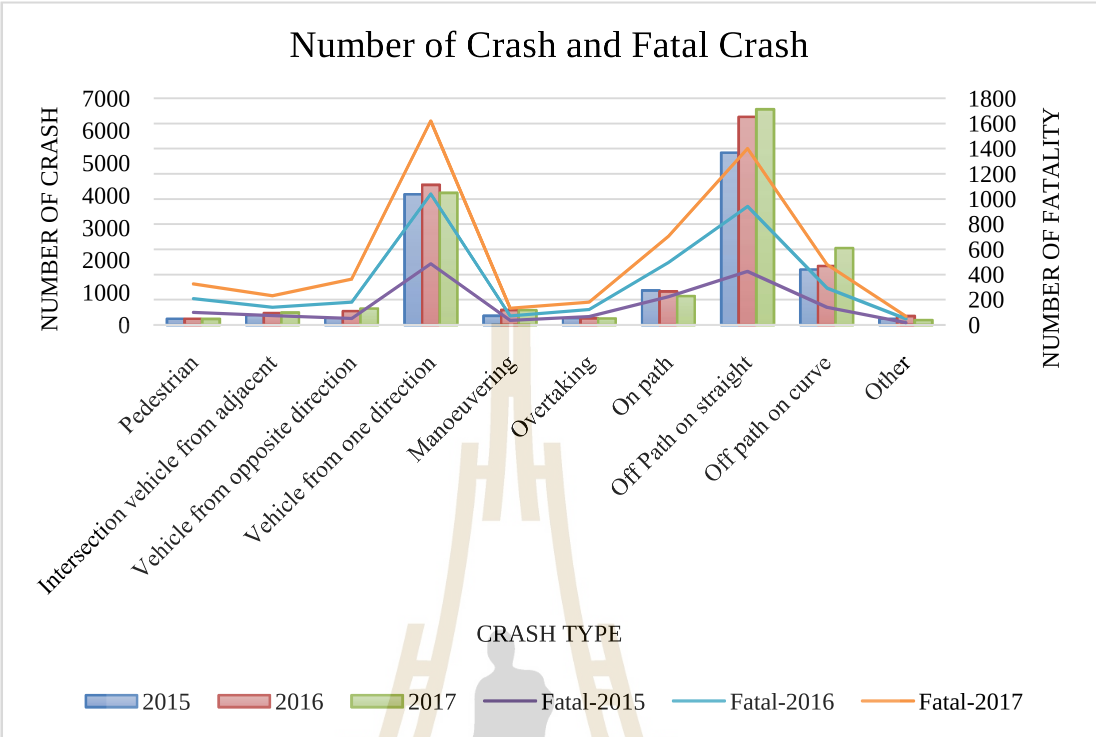


Figure 7.1 Number of crash and fatalities on Thai highway classification by crash type

According to previous studies which focused on finding the relationship between various factors affecting the number of rear-end crashes as shown in Table 7.1, it was found that the different methods were used in many studies including the study of probabilities of rear-end accidents on signalized intersections analyzed by Negative binomial regression (Wang et al., 2002), the rear-end collisions on freeway analyzed by Poisson probability (Joon-Ki et al., 2007) as well as the rear-end collision on the freeway analyzed by Probabilistic Neural network for instantaneous Appraisal of rear-end crash risk (Anurag Pande & Abdel-Aty, 2008) the analysis of rear-end collisions on urban arterials using Genetic programming (Das & Abdel-Aty, 2011), the risk of rear-end collisions at work zone analyzed by Stepwise regression method

(Meng & Weng, 2011) and Truncated count data models (Qi et al., 2013), the risk of rear-end accidents caused by drivers' merging behavior (Weng et al., 2015a), frequency analysis of rear- end accidents in urban road tunnels (Meng & Qu, 2012), the risk of rear-end collisions at signalized intersections with and without green signal countdown devices. (Ni & Li, 2014), the factors such as vehicle by vehicle interactions with road physical characteristics and operational condition on urban roads (Dimitriou et al., 2018). In Thailand, Champahom et al. (2019) have studied the rear-end collisions on the highways throughout the country using Classification and Regression Tree to analyze the relationship between explanatory and target variables. The study consisted of 2 models which wereat-fault vs not at-fault model and fatal vs non-fatal model.

However, there have not been any studies establishing the models to predict the frequency of the rear-end collisions by using spatial model of crash frequency. Actually, spatial model Technique which is multilevel modeling of aggregate accident data is in the current trend of technique used for analysis which distributes and divides the spatial data into hierarchical structures according to their nature. In addition, the division of the area depends on the way of research questions. For example, the studies of Adanu et al. (2017) determined the second level factor to be the area characteristics, classified according to the postal code of each area. For the purpose of aggregate data analysis, most data analyzed the risks of accidents occurring on each road or each area. Due to the area classification, there was a large number of accident data which enabled the researcher to conduct multilevel modeling of aggregate data analysis (Dupont et al., 2013). The study of the accident number on intersections which were grouped into a spatial model in order to identify factors at the area level potentially affecting the number of accidents both observable zonal effects and unobserved heterogeneity which can be done by considering heterogeneous and spatial correlations (Cai et al., 2018). Zone classification for analysis spatial correlation may be considered from traffic analysis zones (TAZs) which have been divided by traffic characteristics, Urbanization Density of the junction etc. (Huang et al., 2019; Osama & Sayed, 2017).

When considering the research that studied the rear-end accident number. It was found that many factors affected the frequency or the probability of rear-end collisions were segment length, traffic volume or travel, the truck proportion, the urban area, the road physical characteristics such as traffic lane number, Innermost traffic lanes, shoulder width, the existence of medians, and median width etc. (Bhowmik et al., 2018; Das & Abdel-Aty, 2011; Joon-Ki et al., 2007; Ma et al., 2017; Mothafer et al., 2017).

Liu et al. (2018) stated that a spatial model was used to find the relationship between explanatory factors with the rear-end collisions frequency (Compared with side swipe and other crashes). However, the study of rear-end crashes were conducted only in urban areas. Therefore, this study has implemented the spatial model to find the relationship between road physical characteristics and the rear end collision frequency. The areas of this study are under the responsibility of Department of Highways throughout the country divided by the provinces which have different spatial characteristics, number of passing vehicles etc. The contribution of this study is to be a starting point for applying spatial models for rear end crash studies.

Table 7.1 Previous work in rear-end crash occurrences modeling

Work	Method	Location	Raised issue
Wang et al. (2002)	negative binomial regression	Tokyo, Japan	Rear-end crash occurrence at signalized intersections
Joon-Ki et al. (2007)	Poisson probability	Tokyo, Japan	Freeway Rear-End Crash Occurrence
Anurag Pande and Abdel- Aty (2008)	Probabilistic Neural Networks	Orlando FL, USA	Road segments were divided into two groups based on the average traffic speeds observed around the crash location prior to the crash occurrence.
Das and Abdel-Aty (2011)	The Genetic Programming (GP)	Florida, USA	Rear-end crashes on urban arterials which analyzed rear-end crash frequency and severity of injuries
Meng and Weng (2011)	Stepwise regression	Singapore	Rear-end crash risk at work zone using work zone
Meng and Qu (2012)	Inverse Gaussian regression	Singapore	Analyze the time to collision (TTC) data collected from rear-end crash on road tunnels
Qi et al. (2013)	Truncated count data models	New York, USA	Frequency and Severity of Rear-End Crashes in Work Zones
Ni and Li (2014)	A microscopic modeling approach	Suzhou, China	Rear-end crash proability at intersections with and without Green Signal Countdown Devices
Weng et al. (2015b)	Mixed probit model	Ang Mo Kio Avenue 3 in Singapore	The relationship of drivers' merging behavior and rear-end crash risks work zone merging
Dimitriou et al. (2018)	Multinomial Logit model	Nicosia, Cyprus	Rear-end crash potential in urban environmental road including factor vehicle-by vehicle interactions, geometric characteristics and operational conditions
Champahom et al. (2019)	Classification and Regression Tree Model	Thailand	Modeling of at-fault/not-at-fault and fatal/non-fatal form rear-end collision on Thai highway. Rear-end crash occurrences was predicted from the at-fault driver in differences environmental factor and road characteristics factors
This study	*Spatial zero-inflated negative binomial regression model*	*Thailand*	*Rear-end crash frequencies modeling which is specifics of road geometry explanatory factors. Spatial correlation is applied to develop models*

7.3 Method

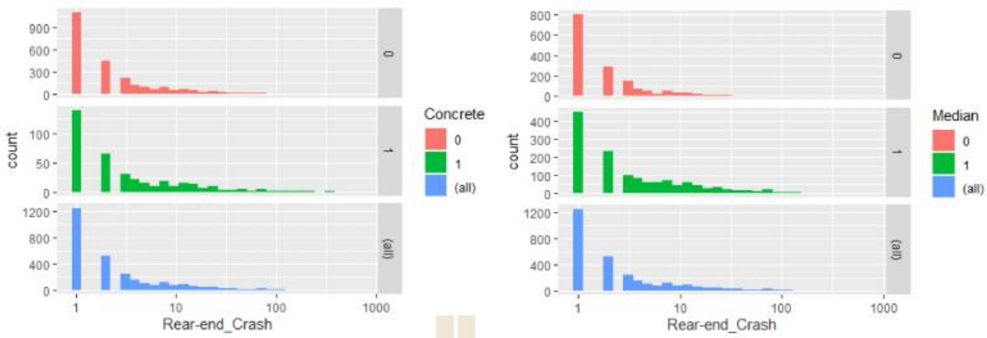
7.3.1 Data collection

The data used in this study consisted of two parts including 1) Collision data which were data collected from the Highway offices located throughout the country. The data were accidents on Thai highways occurring from 2011 to 2018 and the rear-end collisions which were subsequently screened totaled 22,536 cases 2) Road segment data which were also divided by the Department of Highways officials. Segments were divided due to any change of the road characteristics (Agbelie, 2016; Anastasopoulos, 2016; Mothafer et al., 2016) including lane numbers, traffic surface types, lane width, shoulder width, available medians, median width. When compiling national highways, it consisted of 16,939 segments as data description shown in Table 7.2 indicating the average segment length of 3.082 kilometers, the average lane number of 3.186, the average width shoulder of 1.738 meters. The traffic volume logarithm value was 8.881, and the average truck proportion was 16.329%.

Considering the number of rear-end collisions occurring in each segment, it was found that the mean was 1.331 (SD = 11.56). Figure 7.2 shows the distribution of the number of rear-end accidents indicating that most of them are 0, accounting for 82.5%, so statistical models which would be applied have to not only match count data but also suit the data distribution in case that the data contain a lot of 0 (Liu et al., 2018).

Table 7.2 Data description

Variables	Description	Mean	SD	Min	Max
Rear-end crash	Number of rear-end crashes	1.331	11.560	0	679
Length	Length of road segment (Kilometers)	3.082	5.022	0.100	63.165
No_Lane	Number of lanes	3.186	1.778	1	14
Concrete	Pavement type (1 = concrete, 0 = other)	0.099	0.299	0	1
Lane_width	Lane width (Meters)	3.474	0.209	2.500	6
Footpath	Indicate type of shoulder (1 = footpath, 0 = otherwise)	0.048	0.213	0	1
Shoulder_width	Shoulder width(Meters)	1.738	0.876	0	7.200
Median	Divided road (1 = yes, 0 = other)	0.331	0.470	0	1
Median_Width	Median width (Meters)	1.844	3.273	0	15
LogAADT	Log of annual traffic volume	8.881	1.177	4.060	12.734
Percent_Truck	Percentage of heavy truck	16.329	11.771	0	72.507



(a) (b)

Figure 7.2 Histogram of rear-end crash frequency in log10 scale

Note:(a) = divided by pavement type (1 = concrete pavement; 0 = otherwise); (b) = Road segment is divided a median (1= yes; 0=otherwise)

For spatial data shown in Figure 7.3, it can be explained that 18 highways are governed by Department of Highways nationwide and Sub-departments of highways (Sub-DOH) are classified by province. In case of a large province, there will be many Sub-DOHs, for example, in the small figure which presents a large province divided into 3 Sub-DOHs. Additionally, the different line colors represent roads that are under the control of each Sub-DOH. Totally, there are 104 Sub-DOHs in Thailand.

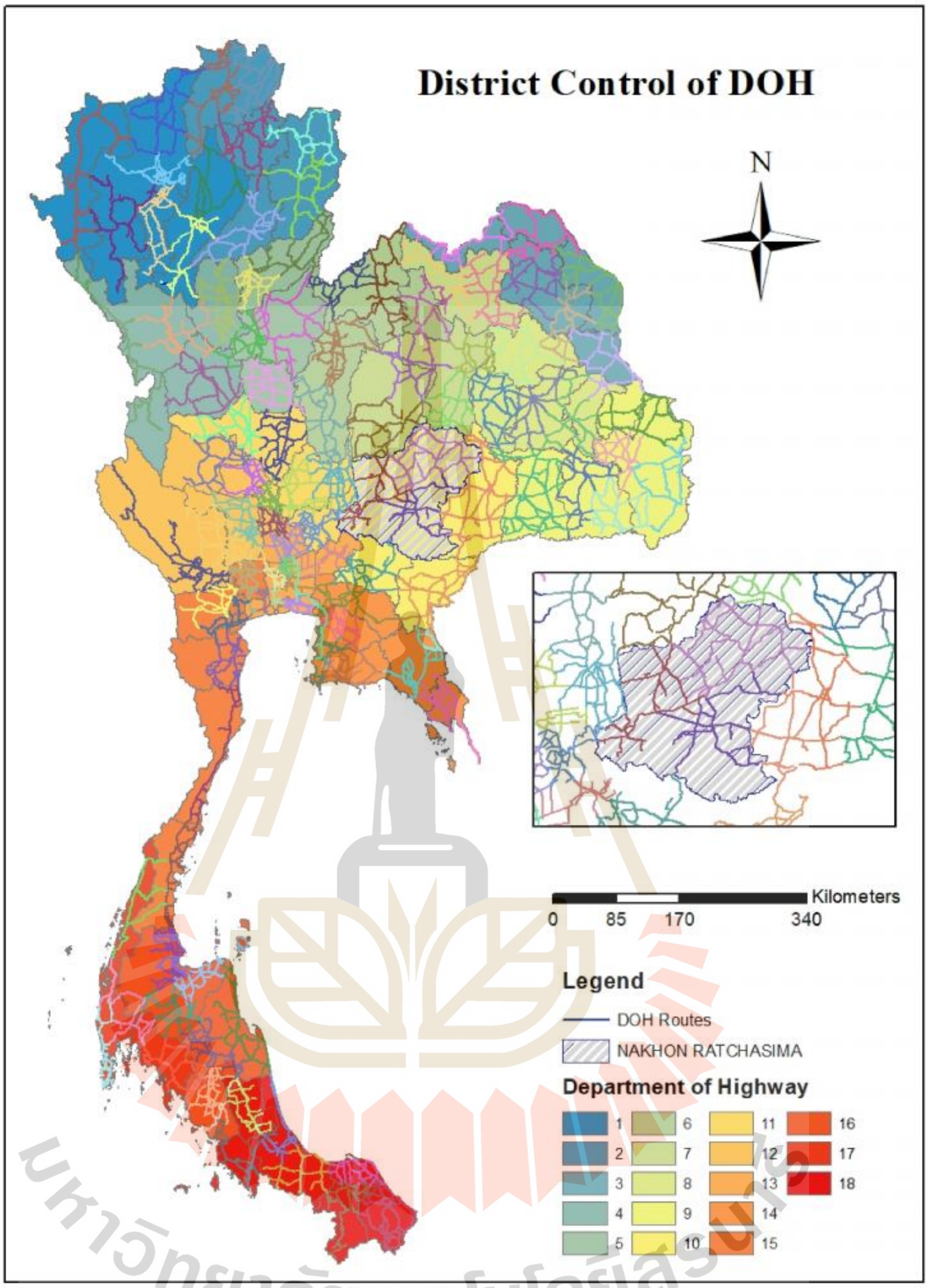


Figure 7.3 District control of sub department of highway

7.3.2 Model development

Count data consist of Nonnegative integer value and is normally found in transportation models such as number of routes drivers change per week, accident number occurring on each road per year, etc. The common error for the count data model is to see the number of occurrences as continuous data. Thus, the application of broad-spectrum regression theory such as Standard least squares regressions clearly inaccurate as the general regression model can predict the values of variables both non-negative and negative. (Washington et al., 2011).

- **Poisson regression model**

The count data model begins with the consideration of the dependent variable distribution referring to, in this case, the number of rear-end collisions which occur on each road. The most distribution of count data is usually Poisson distribution (Caliendo et al., 2019) where the probability of road i will have the number of accident occurrences. yi can be calculated from

$$P(y_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \tag{7.1}$$

When $P(y_i)$ is the probability of the accident number y_i on road i และ λ_i is Poisson parameter for each road. $E[y_i]$ is the number of predicted accidents that will occur on each road where $E[y_i]$ is a prediction of the number of occurrences due to Explanation variables including traffic volume, each road length, road physical characteristics, road surface characteristics, median types, driver visibility conditions, etc. The relationship between the explanation variables and the Poisson parameter is in the type of a log-linear model.

$$\lambda_i = \exp(\beta X_i) \tag{7.2}$$

Where X_i is the vector of the explanation variable and β is the vector of the parameter estimation. The number of occurrences or the number of accidents can be predicted from $E[y_i] = \lambda_i = \exp(\beta X_i)$ and this model is predicted using maximum likelihood methods.

$$L(\beta) = \prod_i \frac{EXP[-\mathrm{EXP}(\beta X_i)][EXP(\beta X_i)]^{y_i}}{y_i!} \tag{7.3}$$

Log of the likelihood function is easier to manage and more suitable for estimation. It can be calculated from

$$L(\beta) = \sum_{i=1}^{n}[-\exp(\beta X_i) + y_i\beta X_i - LN(y_i!)] \tag{7.4}$$

- **Negative binomial regression model (NB)**

The NB model is used when the invariance results from the Poisson model are not appropriate, that is, the mean of the estimation is not equal to the variance. If the expected value is greater than the variance, it is called under dispersed $(E[y_i] > VAR[y_i])$ or over - dispersed $(E[y_i] < VAR[y_i])$. The phenomenon that most often occurs is Over-dispersed. Negative binomial regression model will be used to adjust from the equation (7.2) (Saeed et al., 2019; Washington et al., 2011).

$$\lambda_i = \mathrm{EXP}(\beta X_i + \varepsilon_i) \tag{7.5}$$

Where $\mathrm{EXP}(\varepsilon_i)$ is Gamma - distribution has mean value equal to 1, and the variance into which has been added in order to change the mean value.

$$VAR[y_i] = E[y_i][1 + \alpha E[y_i]] = E[y_i] + \alpha E[y_i]^2 \tag{7.6}$$

For Probability equation of the Poisson model. α is assigned to be 0, which means that the choice between two models (Poisson and Negative binomial) depends on the value of α, which is most often the over-dispersed model. The probability of the number of rear-end collisions occurring on the road segment that is considered to be a negative binomial distribution can be calculated from

$$\mathrm{P}(y_i) = \frac{\Gamma\left(\left(\frac{1}{\alpha}\right)+y_i\right)}{\Gamma\left(\frac{1}{\alpha}\right)y_i!}\left(\frac{1/\alpha}{\left(\frac{1}{\alpha}\right)+\lambda_i}\right)^{1/\alpha}\left(\frac{\lambda_i}{\left(\frac{1}{\alpha}\right)+\lambda_i}\right)^{y_i} \tag{7.7}$$

Where Γ(.) Is the gamma function, for parameter estimation, it can be calculated from

$$\mathrm{P}(\lambda_i) = \prod_i \frac{\Gamma\left(\left(\frac{1}{\alpha}\right)+y_i\right)}{\Gamma\left(\frac{1}{\alpha}\right)y_i!}\left(\frac{1/\alpha}{\left(\frac{1}{\alpha}\right)+\lambda_i}\right)^{1/\alpha}\left(\frac{\lambda_i}{\left(\frac{1}{\alpha}\right)+\lambda_i}\right)^{y_i} \tag{7.8}$$

- **Zero-Inflated Negative binomial regression model**

For predicting the annual number of accidents occurring each year, there may be some roads on which no accidents have taken place. Thus, these 2 different characteristics can be divided into Normal-countand Zero-count. The general model may not be comprehensive for separating the analysis into two parts. Therefore, the most suitable model for a dual-state is the Zero-inflated model. When established on negative binomial model, it is called Zero-inflated Negative binomial:ZINB (Mahmud et al., 2019).

For ZINB similar to the model of the independent event equation $Y = (y_1, y_2 \dots, y_n)$

$$y_i = 0 \ \ with\ proability\ p_i + (1+p_i)\left(\frac{\frac{1}{\alpha}}{\left(\frac{1}{\alpha}\right)+\lambda_i}\right)^{\frac{1}{\alpha}} \tag{7.9}$$

$$y_i = y \ \ with\ proability\ (1+p_i)\left(\frac{\Gamma\left(\left(\frac{1}{\alpha}\right)+y\right)\mu_i^{\frac{1}{\alpha}}(1-\mu_i)^y}{\Gamma\left(\frac{1}{\alpha}\right)y_i!}\right), \qquad y = 1,2,3,\dots$$

Where$\mu_i = (1/\alpha)[1/\alpha + \lambda_i\,]$ is the maximum likelihood method used for parameter estimation again in the ZINB model. For parameter estimation, the maximum likelihood method is used, while the confidence value is determined by the value of likelihood ratio test.

- **Random effects count models**

In some cases, there may be reasons for predicting that there is a mutual correlation between the observed data. This relationship occurs because of spatial data such as accident data occurring in the same area. The effect should be determined not to be observed. In this study, the road is divided into areas according to the supervision of Sub-Department of Highways. Having such a relationship, the model application should generate random effects and Fixed effect (When the unpredictable influence is considered an indicator variable)

For the equation of random influence of numbers, improve from the equation (7.5) to

$$LN(\lambda_{\mathrm{i}}) = \beta \mathrm{X_i} + \varepsilon_i + \eta_i \ \ \text{or}, \lambda_{\mathrm{i}} = \mathrm{EXP}(\beta \mathrm{X_i} + \varepsilon_i)\mathrm{EXP}\ (\eta_i) \tag{7.10}$$

Where λ_i is the expected number of events for the data *i* in group *j* (the regulatory space of Sub-Department of Highway, which is expected to have unobserved heterogeneity. X_i is the vector of explanation variable variable, β is the parameter prediction vector and η_i is the random effect for the data group *j*. Spatial correlation is calculated from the *spatial variation proportion out of the total variation* (*Huang et al.*, 2017*; Osama&Sayed*, 2017) as follows.

$$\text{Spatial correlation} = \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_e^2} \quad (7.11)$$

Where σ_η^2 is the variance obtained from predicting i or is the variance within the same Sub-DOH area. σ_e^2 is the variance obtained from fixed effect estimation or is the variance between the areas of Sub-DOH responsibilities.

The general model is derived from the assumption that η_j is randomly distributed through each group. For example $EXP(\eta_j)$ as a Gamma-distribution with mean value equal to 1 and variance is α which is established from Negative binomial regression model.

7.4 Results

7.4.1 Parameter estimation

Table 7.3 shows the parameter estimation from 4 models. The Poisson regression model (POI) found that there are all statistically significant parameters except "Median width". When considering the prediction of the number of zero accidents it is found that their proportion is 66%. Therefore, it can be considered that the model should be analyzed between the two groups of rear-end collisions including zero and non-zero rear collisions. Dispersion ratio is the consideration of the model results $(E[y_i] < VAR[y_i])$with its value of 34.194; when considered together with Pearson's Chi-square, it is statistically significant. For model development, the elements of Over-dispersion model should be supplemented.

The following model is Negative binomial regression model (NB). An overview of the parameters found that there are 3 variables which are not statistically significant at the confidence level of 95% such as "Concrete", "Lane width" and "Footpath". For Over-dispersion, its value is 0.099 (SD = 0.002) which is considered significant to the model. Therefore, it can be concluded that the model is better than POI.

Zero-inflated negative binomial model (ZINB). The results of the ZINB parameter estimation come out in two states, namely Conditional model and Zero-inflation model. For conditional model, it was found that most factors have statistical significance except "Concrete". Regarding Zero-inflation model, there are many significant variables, including Intercept, No_Lane, Concrete, Shoulder_ width, Median and LogAADT. For examining the Over-dispersion, its value is 0.181 (SD = 0.03) with statistical significance. It can be repeatedly interpreted that NB analysis is more appropriate than POI.

For the model that applies the Spatial Zero-inflated negative binomial model (SZINB), the spatial correlation value equaling 0.355 shows the large proportion of rear end variability within the area. Therefore, spatial analysis is appropriate. For the random parameter (RP) components, it means that the rear-end crash frequency is allowed to vary to the areas within the responsibility of Sub-DOH. Regarding the estimation results showing those of 4 variables, Intercept is 5.556 (SD

= 2.357) which is considered that η_i is not significant zero (Han et al., 2018). Although the remaining 3 variables are not statistically significant, they can also help improve the model to be more efficient and simultaneously reduce the number of significant variables (Osama & Sayed, 2017).

Table 7.3 Estimated parameter results

Explanation variable	POI			NB			ZINB			SZINB		
	Mean	SD	P-value	Mean	SD	P-value	Mean	SD	P-value	Mean	SD	P-value
Random effect												
(Intercept)										5.556	2.357	
LogAADT										0.078	0.279	
Percent_Truck										0.002	0.012	
Distance										0.001	0.028	
Fixed effect												
Conditional model:												
(Intercept)	–7.213	0.170	<0.000	–6.594	0.555	<0.000	–5.018	0.485	<0.000	-6.932	0.743	<0.000
Length	0.091	0.001	<0.000	0.120	0.005	<0.000	0.044	0.005	<0.000	0.064	0.006	<0.000
No_Lane	0.116	0.003	<0.000	0.171	0.019	<0.000	0.121	0.020	<0.000	0.093	0.024	<0.000
Concrete[=1]	–0.128	0.018	<0.000	–0.134	0.095	0.157	–0.070	0.109	0.517	-0.091	0.130	0.484
Lane_width	–0.339	0.045	<0.000	–0.158	0.141	0.263	–0.233	0.117	0.047	0.040	0.175	0.820
Footpath[=1]	0.792	0.019	<0.000	0.237	0.127	0.061	0.344	0.143	0.016	0.286	0.159	0.073
Shoulder_width	–0.081	0.009	<0.000	0.100	0.038	0.008	0.082	0.039	0.035	0.225	0.045	0.000
Median[=1]	0.417	0.022	<0.000	0.561	0.096	<0.000	0.468	0.109	<0.000	0.711	0.121	0.000
Median_Width	0.002	0.002	0.323	0.027	0.013	0.044	0.026	0.015	0.080	0.026	0.016	0.111
LogAADT	0.820	0.008	<0.000	0.606	0.031	<0.000	0.583	0.031	<0.000	0.599	0.052	<0.000
Percent_Truck	–0.011	0.001	<0.000	-0.010	0.002	<0.000	–0.009	0.003	<0.000	-0.001	0.003	0.709
Zero-inflation model:												
(Intercept)							4.186	1.128	<0.000	2.680	1.227	0.029
Length							–1.172	0.094	<0.000	–1.239	0.102	<0.000
No_Lane							–0.130	0.030	<0.000	–0.130	0.031	<0.000
Concrete[=1]							–0.165	0.143	0.247	–0.243	0.149	0.104
Lane_width							–0.113	0.293	0.699	0.151	0.323	0.639
Footbath[=1]							–0.127	0.188	0.500	–0.186	0.196	0.344
Shoulder_width							–0.141	0.065	0.029	–0.056	0.066	0.399
Median[=1]							–0.430	0.151	0.004	–0.224	0.157	0.154
Median_Width							0.010	0.020	0.621	0.014	0.019	0.476
LogAADT							–0.158	0.055	0.004	–0.126	0.058	0.029
Percent_Truck							0.004	0.004	0.328	0.004	0.004	0.330
Predicted Zero	66%											
Dispersion ratio	34.194											
Over-dispersion				0.099	0.002		0.181	0.030				
Spatial correlation										0.355		

7.4.2 Comparison of Models Goodness-of-fit model

An overall picture of the all model suitability started with the concept of creating a model based on the distribution of accident number and got the Poisson regression model (POI). The Negative binomial model (NB) was subsequently developed from POI with the principle of Over-dispersed. The following step was considering the distribution of rear end collisions in each road segment which showed a large number of 0. Thus, the model was developed into Zero-inflated negative binomial regression model (ZINB). Consequently, the application of spatial concept which was the analysis of the relationship of rear-end collisions within the responsibility area of Sub-DOH was the final model Zero-inflated negative binomial regression with spatial correlation model (SZINB). According to Table 7.4, the model accuracy is determined by The Akaike Information Criterion (AIC). It was found that the SZINB value is extremely close to zero. This can therefore interpret that SZINB is the most consistent with empirical data (Fountas & Anastasopoulos, 2018). For ρ^2 of SZINB is 0.090. Despite its relatively small number, it is still acceptable *(*Ma et al., 2017*;* Venkataraman et al., 2013)

Table 7.4 Models goodness-of-fit

Model	Log-Likelihood	AIC	ρ2
POI(m0)	–74797	149596	
POI(m1)	–50753	101527	0.321
NB(m0)	–15203	30410	
NB(m1)	–14164	28345	0.068
ZINB(m0)	–15200	30412	
ZINB(m1)	–13670	27392	0.101
ZINBS(m0)	–14697	29407	
ZINBS(m1)	–13380	26831	0.090

Note: m0 is intercept only model, m1 is convergence models. The Akaike Information Criterion (AIC) is calculated as: AIC = 2[K-LL(m1)]. Where, K is the number of model parameters. McFadden ρ^2 *is calculated as:* $\rho^2 = 1 - LL(m1)/LL(m0)$

7.5 Discussion

The direction of the parameters in all 4 models tend to go in the same direction. However, this section has focused on evaluating results from spatial zero-inflated negative binomial models (SZINB) due to its best goodness-of-fit by taking direction and size of the parameters into consideration.

Regarding the Conditional model, Intercept has value of –6.932 (P <0.000) which means that regardless of other factors, it was found that on each segment road there was no rear-end crash. The variable with the highest parameter value which is more likely to cause rear-end collisions was the road segment with medians. This is consistent with the Baldock et al. (2005) whose study found that most rear-end collisions occurred on roads with raised medians (73.7%). Regarding reasons, it can be considered from the area issues as most highways in Thailand usually built the medians in urban or community areas (Bureau of location and design, 2011). The urban roads tend to have higher rear-end collisions than rural roads on which there are a small number of raised medians. For the result of Parameter estimation of LogAADT which is 0.599 (P<0.000), this result is very reasonable due to the increasing traffic volume which resulted in reducing vehicle headway and giving the opportunity to increase rear end crashes (Das & Abdel-Aty, 2011; Dimitriou et al., 2018; Gaca et al., 2011; Liu et al., 2018; Zavareh et al., 2017). The next variable is the footpath type which found that the availability of a footpath increased the rear- end collisions frequency since a footpath provides a lot of pedestrians leading to the high demand for sidewalks which are most often in urban areas where there is relatively high traffic volume causing the high likelihood of rear-end collisions as well. Many researchers have found that the number of rear-end collisions in urban areas is higher than that in rural areas (Joon-Ki et al., 2007). For the shoulder width, it was found that The width of the shoulder potentially increased the frequency of rear end collisions. This result does not correspond with Joon-Ki et al. (2007) which found that the width of the shoulder which was narrower resulted in increasing rear-end crashes while Das and Abdel-Aty (2011) found that the shoulder width did not affect the number of rear-end collisions. However, when considering the correlation coefficient as shown in Table 7.5, it was found that the shoulder width had a relatively high relationship with LogAADT. This can be said that the wide shoulder width design was built for high traffic volume roads which subsequently and inevitably increased the rear-end crash opportunities. For the number of traffic lanes, it was found that the increasing lane number led to an increase in rear-end collisions. This is in accordance with the study of Venkataraman et al. (2013) who found that road segments with 4 traffic lanes were more likely to cause rear-end collisions than those with 3 lanes. For all types of collisions, Agbelie (2016); Anastasopoulos (2016) discovered that the increasing lane number also increased the crash frequency. The variable specifying the length of the road segment has been found that the road segment length increased, the number of rear-end collisions increased consequently. This result is consistent with many researches (Agbelie, 2016; Caliendo et al., 2019).

For Zero-inflation model: The results were interpreted by considering directions together with significance. If there is a positive direction, increasing the parameter value will make the number of rear-end collisions on the road segment to be 0. For Intercept with a positive direction, it shows that when other variables are not considered, the overall picture of rear-end collisions is 0. The variable with the highest estimation is "length". It was found that if the length of the segment increases, the rear end collision will not be zero. This result resembles the zero sate model of Dong et al. (2014) found that Intercept was positive and length was also positive. For the increasing traffic lane number, the rear-end collision will not be 0. For traffic volume, which has been discovered that when the amount of traffic volume increased, the rear end collisions were not significant zero.

7.6 Conclusion

This study aims to be the starting point for the study of rear-end collisions with advanced statistical models. In order to get a precise model for analyzing, the relationship between road physical characteristics factors and the frequency of rear-end accidents. The data used are rear-end collisions occurring on Thai highways distributed on the road segments with the same characteristics in each such as Number of lanes, Shoulder width etc. For model development, it was established on the concept of a linear relationship between the number of rear-end collisions and the explanatory factors. The process began with the Poisson regression model in which the problem of Over-dispersion was adjusted with the Negative binomial regression model. Since the data on the number of rear-end collisions in each segment contained a lot of 0, the model was developed by Zero-inflated model and finally added with spatial correlation of which the areas were classified in authority of Highway District. The final model is the Spatial Zero-inflated Negative binomial regression model (SZINB). When considering the statistical values to compare with the model. (The Akaike Information Criterion: AIC), it was found that the SZINB model was the most suitable for this data set.

For the results of SZINB random effect, it was found that only constant value was significant. This showed that the relationships between the remaining variables have relationships within the Sub-DOH areas were not strong enough. However, spatial correlation allowance helped make the model more appropriate. In terms of fixed effect, when considering conditional model together with zero-inflated model, there are 3 significant variables, which are segment length, number of lanes and traffic volume. It was found that when these variables increased, they resulted in the increasing number of rear-end collisions. For other variables, which were insignificant in the zero state but significant in the conditional state, were the existence of the medians which increased the chances of rear-end collisions, especially the raised median. In addition, other variables which increased the number of rear-end collisions included Shoulder width, lane number, and the availability of footpath shoulder.

This study has limitations in terms of road physical characteristics of which a few factors potentially analyzed as independents were missing such as road surface roughness, number of junctions per segment, distance of median openings, land-use benefits, and parking permission etc. All of these factors tend to affect rear-end collisions. Future studies may add these factors to the rear-end crash model for potentially new insights. However, this study was the beginning point to those interested in applying the spatial model to analyze the rear-end crashes. Further model development may increase comparisons between urban areas and suburb areas by using multivariate for analysis.

7.7 Reference

Adanu, E. K., Smith, R., Powell, L., and Jones, S. (2017). Multilevel analysis of the role of human factors in regional disparities in crash outcomes. **Accident Analysis and Prevention**, 109: 10-17.

Agbelie, B. R. D. K. (2016). A comparative empirical analysis of statistical models for evaluating highway segment crash frequency. **Journal of Traffic and Transportation Engineering (English Edition)**, 3(4): 374-379.

Anastasopoulos, P. C. (2016). Random parameters multivariate tobit and zero-inflated count data models: Addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. **Analytic Methods in Accident Research**, 11: 17-32.

Baldock, M., Long, A., Lindsay, V., and McLean, A. (2005). *Rear end crashes*. Retrieved from Centre for Automotive Safety Research, The University of Adelaide:

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

Bureau of location and design. (2011). Road median design giudeline. Retrieved from http://www.ldhighway.com/gallary/Road%20Medians%20&%20Road%20Widening.pdf

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., and Wang, X. (2018). Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. **Analytic Methods in Accident Research**, 19: 1-15.

Caliendo, C., De Guglielmo, M. L., and Russo, I. (2019). Analysis of crash frequency in motorway tunnels based on a correlated random-parameters approach. **Tunnelling and Underground Space Technology**, 85: 243-251.

Champahom, T., Jomnonkwao, S., Chatpattananan, V., Karoonsoontawong, A., and Ratanavaraha, V. (2019). Analysis of Rear-End Crash on Thai Highway: Decision Tree Approach. **Journal of Advanced Transportation**, 2019: 1-13.

Das, A., and Abdel-Aty, M. A. (2011). A combined frequency–severity approach for the analysis of rear-end crashes on urban arterials. **Safety Science**, 49(8-9): 1156-1163.

Dimitriou, L., Stylianou, K., and Abdel-Aty, M. A. (2018). Assessing rear-end crash potential in urban locations based on vehicle-by-vehicle interactions, geometric characteristics and operational conditions. **Accident Analysis and Prevention**, 118: 221-235.

DOH. (2016). Thailand Traffic Accident on National Highways in 2015. Available from: https://bhs.doh.go.th/download/accident Date Accessed: June-2019

DOH. (2017). Thailand Traffic Accident on National Highways in 2016. Available from: https://bhs.doh.go.th/download/accident Date Accessed: June-2019

DOH. (2018). Thailand Traffic Accident on National Highways in 2017. Available from: https://bhs.doh.go.th/download/accident Date Accessed: June-2019

Dong, C., Clarke, D. B., Yan, X., Khattak, A., and Huang, B. (2014). Multivariate random-parameters zero-inflated negative binomial regression model: an application to estimate crash frequencies at intersections. **Accident Analysis and Prevention**, 70: 320-329.

Dupont, E., Papadimitriou, E., Martensen, H., and Yannis, G. (2013). Multilevel analysis in road safety research. **Accident Analysis and Prevention**, 60: 402-411.

Fountas, G., and Anastasopoulos, P. C. (2018). Analysis of accident injury-severity outcomes: The zero-inflated hierarchical ordered probit model with correlated disturbances. **Analytic Methods in Accident Research**, 20: 30-45.

Gaca, S., Kieć, M., and Zielinkiewicz, A. (2011). Analysis of headways for assessment of risk rear-end collision. Retrieved from http://onlinepubs.trb.org/onlinepubs/conferences/2011/RSS/1/Gaca,S.pdf

Han, C., Huang, H., Lee, J., and Wang, J. (2018). Investigating varying effect of road-level factors on crash frequency across regions: A Bayesian hierarchical random parameter modeling approach. **Analytic Methods in Accident Research**, 20: 81-91.

Huang, H., Chang, F., Zhou, H., and Lee, J. (2019). Modeling unobserved heterogeneity for zonal crash frequencies: A Bayesian multivariate random-parameters model with mixture components for spatially correlated data. **Analytic Methods in Accident Research**, 24.

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

Joon-Ki, K., Yinhai, W., and Gudmundur, F. U. (2007). Modeling the Probability of Freeway Rear-End Crash Occurrence. **Journal of transport engineering**, 113(1).

Liu, C., Zhao, M., Li, W., and Sharma, A. (2018). Multivariate random parameters zero-inflated negative binomial regression for analyzing urban midblock crashes. **Analytic Methods in Accident Research**, 17: 32-46.

Ma, Z., Zhang, H., Chien, S. I., Wang, J., and Dong, C. (2017). Predicting expressway crash frequency using a random effect negative binomial model: A case study in China. **Accident Analysis and Prevention**, 98: 214-222.

Mahmud, S. M. S., Ferreira, L., Hoque, M. S., and Tavassoli, A. (2019). Micro-level safety risk assessment model for a two-lane heterogeneous traffic environment in a developing country: A comparative crash probability modeling approach. **Journal of Safety Research**, 69: 125-134.

Meng, Q., and Qu, X. (2012). Estimation of rear-end vehicle crash frequencies in urban road tunnels. **Accident Analysis and Prevention**, 48: 254-263.

Meng, Q., and Weng, J. (2011). Evaluation of rear-end crash risk at work zone using work zone traffic data. **Accident Analysis and Prevention**, 43(4): 1291-1300.

Mothafer, G. I. M. A., Yamamoto, T., and Shankar, V. N. (2016). Evaluating crash type covariances and roadway geometric marginal effects using the multivariate Poisson gamma mixture model. **Analytic Methods in Accident Research**, 9: 16-26.

Mothafer, G. I. M. A., Yamamoto, T., and Shankar, V. N. (2017). A negative binomial crash sum model for time invariant heterogeneity in panel crash data: Some insights. **Analytic Methods in Accident Research**, 14: 1-9.

Ni, Y., and Li, K. (2014). Estimating Rear-End Accident Probabilities at Signalized Intersections: A Comparison Study of Intersections With and Without Green Signal Countdown Devices. **Traffic Injury Prevention**, 15(6): 583-590.

Osama, A., and Sayed, T. (2017). Investigating the effect of spatial and mode correlations on active transportation safety modeling. **Analytic Methods in Accident Research**, 16: 60-74.

Pande, A., and Abdel-Aty, M. (2008). A Computing Approach Using Probabilistic Neural Networks for Instantaneous Appraisal of Rear-End Crash Risk. **Computer-Aided Civil and Infrastructure Engineering**, 23: 549–559.

Pande, A., and Abdel- Aty, M. (2008). A Computing Approach Using Probabilistic Neural Networks for Instantaneous Appraisal of Rear-End Crash Risk. **Computer-Aided Civil and Infrastructure Engineering**, 23(7): 549–559.

Qi, Y., Srinivasan, R., Teng, H., and Baker, R. (2013). Analysis of the Frequency and Severity of Rear-End Crashes in Work Zones. **Traffic Injury Prevention**, 14(1): 61-72.

Saeed, T. U., Hall, T., Baroud, H., and Volovski, M. J. (2019). Analyzing road crash frequencies with uncorrelated and correlated random-parameters count models: An empirical assessment of multilane highways. **Analytic Methods in Accident Research**, 23.

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

Wang, Y., Ieda, H., and Mannering, F. (2002). Estimating Rear-End Accident Probabilities at Signalized Intersections: An Occurrence-Mechanism Approach. **Journal of Transportation Engineering**, 129(4).

Washington, S. P., Karlaftis, M. G., and Mannering, F. (2011). Statistical and Econometric Methods for Transportation Data Analysis (2 ed.).

Weng, J., Xue, S., Yang, Y., Yan, X., and Qu, X. (2015a). In-depth analysis of drivers' merging behavior and rear-end crash risks in work zone merging areas. **Accid Anal Prev**, 77: 51-61.

Weng, J., Xue, S., Yang, Y., Yan, X., and Qu, X. (2015b). In-depth analysis of drivers' merging behavior and rear-end crash risks in work zone merging areas. **Accident Analysis & Prevention**, 77: 51-61.

Zavareh, M. F., Mamdoohi, A. R., and Nordfjærn, T. (2017). The effects of indicating rear-end collision risk via variable message signs on traffic behaviour. **Transportation Research Part F: Traffic Psychology and Behaviour**, 46: 524-536.

CHAPTER VIII

CONCLUSION AND RECOMMENATIONS

Road accidents in Thailand have been a problem for a long time. At present, both government sectors and private organizations have been regularly focusing on campaigning and solving problems. When considering the statistics of deaths from accidents on highways, it was found that rear-end collisions is the crash type causing the highest fatality. Therefore, the focus to find a way for reducing the number of road accidents and the death rate caused by rear-end collisions is absolutely necessary. According to the literature review, it was found that the Important dimensions related to rear-end collisions which, if studied, will be able to reduce the number and death rate from rear-end accidents The review results showed that there are 6 important dimensions. This study is therefore divided the studies into 6 dimensions which can be summarized as follows:

8.1 Factors affecting the rear-end crash size

Study 1: A model to analyze factors affecting the rear-end crash size was created. It is measured by the number of injured and deaths in form of latent variables and using the Structural Equation Modeling to analyze the mentioned relationship such as female drivers, drivers aged 35-55 years, high speeding, a truck - involved accident resulting in the driver and passengers' higher level of injuries, asphalt road surface and roads with medians which potentially reduce the severity of injuries.

8.2 Rear-end collision at Intersections

Study 2: Rear-end collisions at intersections are more likely to occur because the front vehicles must reduce the speed to get to the crossroads Therefore, this research aims to find ways to reduce the number of rear-end collisions on the intersection through the guidance for relevant agencies to improve road physical characteristics as well as campaign guidelines. The data used for analysis were accident cases occurring on highways from 2011 to 2015 and were analyzed classification and regression tree (CRT) and specifying the target variables as a collision on an intersection / an external collision outside the intersection. From the tree model, it was found that the important variables to be suggested were rear end crash type, average traffic volume, road surface type, and light factors. In addition to this research which made suggestions for reducing the number of rear-end collisions on intersection area, it also provided guidance for the study of rear-end collisions on crossroads.

8.3 Models of at-fault driver vs not at-fault driver and Fatal crash vs non-fatal crash

Study 3: has applied Quasi-Induced Exposure, which is one way to study the rear-end accidents in order to focus on reducing the number of rear-end accidents and their severity. The result of this objective can be used as a way to train drivers to become more aware of the rear-end accident severity. The study found that the factors causing the driver to become at-fault driver were number of traffic lanes and areas at the median opening etc. For another model that analyzed the relationship between fatal rear –end crash and non-fatal rear-end crash, it was found that the use of safety equipment such as helmets or seat belts mostly reduced the chance of death from rear-end collisions.

8.4 Comparison of rear-end crash size between urban and rural areas

Study 4: Differences between urban and rural roads result from different road physical characteristics as well as traffic flow. The occurrence of rear-end accidents varies accordingly. This objective is to study the differences between rear-end accidents that occur between urban and rural roads when mitigation policies were differently identified. This study applied the Measurement of Invariance to compare the differences between rear collisions in urban and rural areas. The results showed that both two models were different, especially crash type and vehicle involvement factors.

8.5 Factors affecting to fatal rear-end crash: Hierarchical model approach

Study 5: Rear-end crash is a type of road accidents which have been abundantly studied. One factor that quite affects the likelihood of fatalities caused by rear-end collisions is the road area at the accident scene, classified as urban and rural roads, which are obviously different such as speed, number of intersections, car types, etc. However, there has never been any comparison study of rear-end collisions occurring between พurban and rural areas. Therefore, this study has focused on comparing factors that affected the likelihood of fatality in rear-end crash which is different between 2 roadways under the concept of spatial basis. Hierarchical logistic model was applied by determining estimation of parameters to vary according to the road segment, and comparing the models having coefficient with multilevel correlation and coefficient without multilevel correlation. Therefore, there were 4 models. The data used in this study were the rear-end collisions occurring on the Thai highways from 2011 to 2015. The study found that the direction of the parameter values of the model in the rural rear-end collisions model went in the same direction. However, the number of significant parameters in rural rear-end crash are higher. The significant variables in both models were the number of traffic lanes, the driver's seat belt usage, and the accident time. To conclude, this study can help fulfil the rear end knowledge. Additionally, the policy decision makers can apply the results to make decisions on safety policy.

8.6 Rear-end crash frequency models: Spatial zero-inflated negative binomial approach

Study 6: Currently, the application of advanced statistical models to predict the frequency of the accident number has been increasingly used. This can make the model predictions more accurate. This study focuses on fulfilling the application of statistical models to find the relationship between the explanatory variables and the rear-end crash frequency. The data used in the study were the rear-end collisions occurring on Thai highways from 2011-to 2018 and were distributed by segment with the same road physical characteristics. The spatial correlation varies according to the jurisdiction of the Department of Highways. For the development of the model, there are 4 models, starting from Poisson regression model, Negative binomial model, Zero-inflated negative binomial model and spatial zero-inflated negative binomial model (SZINB). The AIC comparison results show that SIZNB had the lowest value. This showed that this model was the most suitable for the data. The effect of the random effect was significant only for the constant values, both in the conditional state and the zero state, which are segment length, number of lanes, and traffic volume. This study can be a starting point for those interested in applying a spatial model for rear-end crash analysis.

8.7 Recommendations

This study focuses on finding factors potentially affecting rear-end collisions, both in terms of severity (meaning fatality risk) and number (referring to the chances causing rear crashes or crash frequency occurring on road segment), which lead to the policy of reducing severity and the number of rear-end collisions. Another contribution is the analysis of rear-end collisions with statistics. Accordingly, the recommendation consists of three issues:

The policy to reduce the rear-end collisions severity. From many models of severity, it was found that Rear-end collisions tended to be severe when going together with a trucks the agencies involved with the driving license should emphatically warn the truck driver not to closely approach the car in front. In addition, other drivers, especially motorcycles sharing roads with large vehicles should be simultaneously warned to have appropriate space and be aware of the road where the trucks are parked. The study results showed that the crash type when a front car was parked was highly severe. In terms of the environmental factors, it was found that the rear-end collisions at night without illumination causing higher fatal rear-end crash.   Therefore, involved departments, such as the Department of Highways, should regularly pay attention to those areas. For spatial rear-end collisions, it has been discovered that the crashes occurring between urban and rural areas were different.

For the policy to reduce the rear-end collision frequency which focuses on the driver factor, the results showed that the chances of a driver being the cause of a rear-end collision often the accident points where the front car has slowed down such as the traffic island opening point and at the junction area. Therefore, the Department of Land Transport should emphasize that drivers must be aware of the warning signs, especially when entering an intersection or median openings.

The advanced models were applied on the basis of distributing the data containing a large number of zero and adding the estimation values of unobserved heterogeneity which varied to the spatial value. This study has confirmed that these two concepts make the model more accurate. Therefore, those interested in creating a model to predict the frequency of accidents potentially apply these two principles.

8.8 References

Champahom, T., Jomnonkwao, S., Chatpattananan, V., Karoonsoontawong, A., and Ratanavaraha, V. (2019). Analysis of Rear-End Crash on Thai Highway: Decision Tree Approach. **Journal of Advanced Transportation**, 2019: 1-13.

APPENDIX A

LIST OF PUBLICATIONS

List of Publications

Champahom T., Jomnonkwao S. and Ratanavaraha V. (14-15 March 2019).Analysis Of Rear-End Crash at Intersection on Thai Highway: Decision Tree Approach. **In 13th South East Asian Technical University Consortium Symposium (SECTUC 2019),** Hanoi, Vietnam.

Champahom T., Jomnonkwao S. and Ratanavaraha V. (2019). Analysis Of Rear-End Crash At Intersection on Thai Highway: Decision Tree Approach. **SEATUC Journal of Science and Engineering (SJSE)**. (In press)

Champahom T., Jomnonkwao S., Chatpattananan V., Karoonsoontawong A. and Ratanavaraha V. Analysis of Rear-end Crash on Thai Highway: Decision Tree Approach. **Journal of Advanced Transportation**, Vol.2019. DOI: https://doi.org/10.1155/2019/2568978.

BIOGRAPHY

Mr. Thanapong Champahom was born on the 30 of April, 1992 at Nongwuaso District, Udontani Province. He started his primary education at Nongwuaso Pittayakom School, secondary education at Udonpattanakarn School. Then, he further studied Bachelor's degree in Transportation Engineering Institute of Engineering at Suranaree University of Technology. After his graduation in 2014, he worked in the position of Transportation Engineer for Bangkok Maine Enterprises Ltd. (BME). As he was the student who had the first rank of scores of bachelor's degree curriculum, he was selected to win the scholarship of achieving outstanding school record to doctoral degree in Transportation Engineering at the same university.

His expertise includes the field of the road safety research and crashed models. During his Master's degree study, he presented one oral presentation entitled "*Analysis of Rear-End Crash at Intersection on Thai Highway: Decision Tree Approach*" at 13th South East Asian Technical University Consortium Symposium at Hanoi, Vietnam. Form 14-15 March 2019 (SECTUC 2019). And published two papers entitled of "*Analysis of Rear-End Crash at Intersection on Thai Highway: Decision Tree Approach*" SEATUC Journal of Science and Engineering (SJSE). And "*Analysis of Rear-end Crash on Thai Highway: Decision Tree Approach*" Journal of Advanced Transportation, Vol.2019. DOI: https://doi.org/10.1155/2019/2568978 (ISI; JIF = 1.983).