Steunpunt Mobiliteit & Openbare Werken
Spoor Verkeersveiligheid

A review of accident prediction models for road intersections

B.B. Nambuusi, T. Brijs, E. Hermans

PROMOTOR ▶ Prof. dr. Geert Wets
ONDERZOEKSLIJN ▶ Infrastructuur en ruimte
ONDERZOEKSGROEP ▶ UHasselt, PHL, VITO, VUB, UGent
RAPPORTNUMMER ▶ RA-MOW-2008-004

WETENSCHAPSPARK 5
B 3590 DIEPENBEEK
T ▶ 011 26 91 12
F ▶ 011 26 91 99
E ▶ info@steunpuntmowverkeersveiligheid.be
I ▶ www.steunpuntmowverkeersveiligheid.be

A review of accident prediction models for road intersections

RA-MOW-2008-004

B.B. Nambuusi, T. Brijs, E. Hermans

Onderzoekslijn Infrastructuur en ruimte

DIEPENBEEK, 2008.
STEUNPUNT MOBILITEIT & OPENBARE WERKEN
SPOOR VERKEERSVEILIGHEID

Documentbeschrijving

Rapportnummer:	RA-MOW-2008-004
Titel:	A review of accident prediction models for road intersections
Auteur(s):	B.B. Nambuusi, T. Brijs, E. Hermans
Promotor:	Prof. dr. Geert Wets
Onderzoekslijn:	Infrastructuur en ruimte
Partner:	Universiteit Hasselt
Aantal pagina's:	69
Projectnummer Steunpunt:	2.1
Projectinhoud:	Cross-sectionele risicoanalyse

Uitgave: Steunpunt Mobiliteit & Openbare Werken – Spoor Verkeersveiligheid, mei 2008.

Steunpunt Mobiliteit & Openbare Werken
Spoor Verkeersveiligheid
Wetenschapspark 5
B 3590 Diepenbeek

T 011 26 91 12
F 011 26 91 99
E info@steunpuntmowverkeersveiligheid.be
I www.steunpuntmowverkeersveiligheid.be

Samenvatting

Het doel van dit rapport is het bestuderen van de literatuur met betrekking tot modellen die het aantal ongevallen op kruispunten voorspellen. Meerdere modellen werden geëvalueerd waaronder meervoudige logistische regressies, meervoudige lineaire regressies, Poisson modellen, negatief binomiaal modellen, random effects modellen en classificatie- en regressietechnieken (CART). De data, methodologie en resultaten van verschillende studies worden beschreven. De richting van het effect van significante verklarende variabelen wordt besproken en aanbevelingen worden gedaan.

Verschillende modellen die het aantal ongevallen voorspellen voor verscheidene kruispunttypes en ongevaltypes werden ontwikkeld in de literatuur. Een apart model fitten voor een bepaald kruispunt- en ongevaltype (indien specifieke data beschikbaar zijn) verdient hierbij de voorkeur aangezien dit tot een betere fit en beschrijving van de data leidt vergeleken met één model (Reurings et al., 2005; Turner en Nicholson, 1998).

Verschillende kruispunttypes in landelijke en stedelijke omgevingen werden beschouwd. In de literatuur kunnen meerdere modelstructuren gevonden worden. De meerderheid van de kruispuntmodellen is van de vorm hieronder gegeven en op basis van dit rapport, raden we het volgende model aan voor het voorspellen van ongevallen op kruispunten:

$\mu_i = \beta_o * Q_{MI}^{\beta_1} * Q_{MA}^{\beta_2} * e^{\sum \beta_j x_{ij}}$ met

μ_i = verwacht aantal ongevallen op kruispunttype *i*

Q_{MA} = aantal voertuigen dat het kruispunt oprijdt via de hoofdweg

Q_{MI} = aantal voertuigen dat het kruispunt oprijdt via de zijweg

x_{ij} = vector van verklarende variabelen, *j*, verschillend van de verkeersstroom op kruispunt *i*

β_o = intercept

β_1, β_2 = effect van verkeersvolume op het verwacht aantal ongevallen (elasticiteit)

β_j = regressiecoefficient die het effect van de j^{de} verklarende variabele weergeeft.

De elasticiteit toont de procentuele verandering in het verwacht aantal ongevallen dat geassocieerd kan worden met een 1% verandering in verkeersvolume. De effecten van risicofactoren die de kans op een ongeval beïnvloeden gegeven een mate een blootstelling worden gemodelleerd als een exponentiële functie. Deze keuze hangt samen met de kenmerken van de Poisson verdeling: kruispuntongevallen zijn gebeurtenissen die zelden voorkomen en bovendien positieve getallen (Reurings et al., 2005).

De keuze van het model hangt af van de aard van de afhankelijke variabele en het doel van het onderzoek. Indien het de bedoeling is conclusies te trekken voor de hele populatie dan zijn modellen gebaseerd op populatiegemiddeldes geschikt (H 2, 3 en 4). Anderzijds moeten onderzoekers geïnteresseerd in locatiespecifieke conclusies opteren voor random effects modellen (H 5). Voor onderzoekers die ongevallen wensen te groeperen op basis van bepaalde criteria is CART een plausibele keuze (H 6).

De variabelen jaarlijks gemiddeld dagelijks verkeer op hoofd- en zijwegen, totaal aantal voertuigen en voetgangers die het kruispunt oversteken, verlichting en timing van verkeerslichten bleken statistisch significant in de meeste modellen. Daarom is het wenselijk deze variabelen op te nemen in een kruispuntenmodel. Overige relevante variabelen worden opgesomd in hoofdstuk 7. In het algemeen had tenminste één verklarende variabele in de categorieën verkeersstroom, verkeerscontrole, geometrie, eigenschappen van de bestuurders, het voertuig en de omgeving en ruimtelijke ordening een significant effect op het gebeuren van een ongeval. Alle categorieën zijn bijgevolg van belang bij het voorspellen van ongevallen op kruispunten.

English summary

The objective of this report is to review accident prediction models for intersections used in literature to identify which variables have a significant effect on accident occurrence so that we can have a starting point for future research. Several models have been reviewed including multiple logistic regression, multiple linear regression, Poisson models, negative binomial models, random effects models and, classification and regression trees (CART) technique. The data, methodology and results of several studies are described. The direction of the effect of several significant explanatory variables is discussed and recommendations are made.

Different APMs for different intersection types and accident types have been developed in the literature. It is recommended that fitting separate models for different intersection types and accident types gives a better fit and description of the data than one model for all intersection types. Provided data on intersection and accident types are available, it is recommended to fit disaggregated models rather than aggregated models (Reurings et al., 2005; Turner and Nicholson, 1998).

Although similar techniques were applied on rural and urban road intersections, in literature a different model structure was used. Nevertheless, the majority of the models discussed on rural and urban road intersections were of the form given below and based on this report, we would prefer a model for intersections to be of the following form:

$\mu_i = \beta_o * Q_{MI}^{\beta_1} * Q_{MA}^{\beta_2} * e^{\sum \beta_j x_{ij}}$ with

μ_i = expected number of accidents at intersection type *i*

Q_{MA} = number of vehicles entering an intersection from the major road

Q_{MI} = number of vehicles entering an intersection from the minor road

x_{ij} = vector of explanatory variables, *j*, other than traffic flow on intersection *i*

β_o = intercept

β_1, β_2 = effect of traffic volume on the expected number of accidents (elasticity)

β_j = regression coefficient representing the effect of the *j*th explanatory variable other than traffic flow

The elasticity shows the percentage change in the expected number of accidents associated with a 1% change in traffic volume. The effects of risk factors that influence the probability of accidents given exposure are modelled as an exponential function. The choice of an exponential form is logical in the view of the characteristics of the Poisson distribution since accident counts are positive and rare events at intersections (Reurings et al., 2005).

However, the choice of the model depends on the nature of the response and the objective of the research. If interest is in making inference on the entire population, population average based models (chapters 2, 3 and 4) are suitable. In contrast, researchers interested in location specific inference would opt for random effects models (chapter 5). Researchers who wish to group accidents based on particular criteria, the CART is a credible choice (chapter 6).

The variables annual average daily traffic (AADT) on major and minor roads, total vehicle counts and pedestrians crossing all arms, lighting and signal timing were statistically significant in most models. Therefore, it is desirable that APMs for intersections include these variables. The other variables are listed in chapter seven of the report. Generally, atleast one explanatory variable in the categories of traffic flow, traffic control, geometry, driver characteristics, vehicle type or features, environmental factors and land use had a significant effect on accident occurrence. Therefore, all categories are essential in predicting accidents at intersections.

Table of contents

1. INTRODUCTION

Accident prediction models (APMs) are used for a variety of purposes; most frequently to estimate the expected accident frequencies from various roadway entities (highways, intersections, interstates, etc) and also to identify geometric, environmental and operational factors that are associated with the occurrence of accidents. It is important to examine the nature of relationships between roadway, environmental and operational factors and, accidents to understand the causal mechanisms involved in accidents on the one hand and to better predict their occurrence on the other hand. APMs are one path of inquiry often used to gain these insights (Reurings et al., 2005).

In this report, focus is on APMs for road intersections. Intersections are a common place for accidents, which may be due to the fact that there are several conflicting movements as well as many different intersection design characteristics. Intersections also tend to experience severe accidents due to the fact that some of the injury crashes such as angle and left turn collisions commonly occur at intersection. Therefore, there is a need to identify the methodologies to assess the effects that geometric, traffic flow, traffic control, environmental and operational characteristics have on the safety of intersections (Abdel-Aty and Keller, 2005). Since different roads meet at an intersection, different types of accidents occur as a result. This calls for separate models to assess factors associated with different accident types and the safety of different intersections types. Examples of types of intersections studied in this report include signalized intersections, stop controlled intersections, intersections with cameras, etc.

Several approaches have been developed to identify elements that affect the safety of road intersections. These include the multiple logistic regression models, multiple linear regression models, Poisson regression models, negative binomial regression models, random effects models and the classification and regression tree (CART) technique. These regression approaches focus on predicting total accidents with fatalities, injuries, etc for assessing the safety effects of various factors. However, effective road safety management requires knowledge of the present safety performance and what it is likely to be in future if proposed actions are taken. In effect, reliable methods for estimating safety performance of an existing or planned roadway are required (Harwood et al., 2000). As one of the main methods, APMs which describe relationships between the number of accidents and factors that are believed to be related to accident occurrence have been developed by several authors to estimate current or future road safety performance.

1.1. Objective of the report

Since APMs provide an estimate of road safety performance, this report aims at reviewing APMs employed in literature to predict the number of accidents on road intersections so that appropriate actions are taken in the future. Moreover, we will focus on the variables considered and significant findings from these models so that one can know which variables had significant effects on accident occurrence.

In this report, the estimates of the coefficients of variables will not be presented as our objective is to identify which variables have a significant effect on accident occurrence so that we can have a starting point for further research (i.e. the development of an appropriate accident prediction model for intersections in Flanders). In addition, in a number of reviewed papers (e.g Harnen et al., 2003, Bauer and Harwood, 2000, Vogt, 1999 and Chin and Quddus, 2003), the authors provided the values of the coefficients for some variables while for others the coefficients were not given. The authors just mentioned whether a variable was significant or not. For reasons of consistency, we decided to only indicate the direction of the significant explanatory variables in terms of road safety in all the studies reviewed in this report.

1.2. Organization of the report

The report is organized according to the different classes of statistical methodologies applied in previous studies. Chapters 2, 3, 4, 5 and 6 of the report describe the data, statistical methodologies and results obtained from these methodologies. First, we shall present multiple logistic regression models in chapter 2, this will be followed by multiple linear regressions in chapter 3, Poisson and negative binomial regression models in chapter 4, random effects models in chapter 5 and we shall conclude with the CART technique in chapter 6. Discussion and conclusions will be presented in chapter 7.

2. MULTIPLE LOGISTIC REGRESSION MODELS

The term multiple refers to many explanatory variables. Explanatory variables refer to characteristics whose effect on the outcome is being investigated. Multiple logistic regression model is a function describing associations between a binary outcome and a set of explanatory variables (Agresti, 2002).

In this chapter we describe a study in which this technique was applied. More specifically, we investigate the accident tendency at signalized intersections for striking and struck vehicles to identify risk factors related to traffic environment, driver characteristics and vehicle types.

2.1 Study 1

The study described in this section aimed at investigating the accident propensity at signalized intersections for striking and struck vehicles and identified the significant risk factors related to the traffic environment, the driver characteristics and the vehicle types. Both rural and urban signalized intersections were considered in the study. The study was done in the state of Florida and utilized the 2001 Florida traffic accident data obtained from the Florida Department of Highway Safety and Motor Vehicles (DHSMV). In this investigation, Yan et al. (2005) used the Quasi-induced exposure concept in section 2.1.2.1 (Stamatiadis and Deacon, 1995, 1997) and multiple logistic regression technique (section 2.1.2.2).

The DHSMV data comprise a relational database that contains seven files. Each file handles a specific feature of a traffic accident. The data in these files can be linked to obtain the required information. In this study, three files were used and included the event file (containing the characteristics and environment of the accident), the drivers' file (containing the drivers' characteristics) and the vehicles' file (information about the vehicles' characteristics and vehicles' actions in the traffic accident).

Vehicle accidents occurring at signalized intersections were categorized into two groups: rear-end accidents and non-rear-end accidents. Rear-end accidents include information of struck drivers/vehicles and striking drivers/vehicles at fault, as well as the corresponding road environment conditions. Non-rear-end accidents include information on drivers/vehicles that had proper driving action but were involved in accidents, as well as their corresponding road environment conditions. There were 7666 two-vehicle rear-end accidents and 15,734 non rear-end accidents involved by not-at-fault drivers. The response was coded 1 for rear-end and 0 for non rear-end accidents. Three models were fitted and they included: road environment factors, striking drivers/vehicle characteristics and struck drivers/vehicle characteristics.

2.1.1 Explanatory variables

Data were collected on road environmental factors, driver characteristics and vehicle types. Also, data on annual average daily traffic were collected. The explanatory variables in each of these categories are given in Table 1 below. In addition, Table 1 provides the direction of the effects of explanatory variables on the accident risk. However, only results from the road environment factors and striking drivers/vehicle characteristics models are presented. The results from the struck drivers/vehicle characteristics model are not presented in Table 1 but they are explained in section 2.1.3.3.

Table 1: Explanatory variables in each category

Category of explanatory variables	Variable levels	Effect	Causes more rear end accidents
1. Road environment factors			
Number of lanes	Others vs. 2-lane	/	
	6-lane vs. 2-lane	/	
	4-lane vs. 2-lane	-	2-lane
Divided/undivided highway	Undivided vs. divided	-	Divided
Accident time	Night vs. daytime	-	Day time
Road surface condition	Wet vs. dry	+	Wet
	Slippery vs. dry	+	Slippery
Urban/rural	urban vs. rural	+	Urban
Highway character	curveupgrade/downgrade vs. straight-level	+	curveupgrade/ downgrade
	curve-level vs. straight-level	+	curve-level
	straightup/downgrade vs. straight-level	+	straightup/ downgrade
Speed limit	55 mph vs. 25 mph	+	55 mph
	50 mph vs. 25 mph	+	50 mph
	45 mph vs. 25 mph	+	45 mph
	40 mph vs. 25 mph	+	40 mph
	35 mph vs. 25 mph	+	35 mph
	30 mph vs. 25 mph	+	30 mph
2. Striking drivers - Driver characteristics			
Alcohol/drug use	Alcohol under influence vs. No	+	Alcohol under influence
	Drug under influence vs. No	+	Drug under influence
	Alcohol-Drug under influence vs. No	+	Alcohol-Drug under influence
	Had been drinking vs. No	+	Had been drinking alcohol
Age	26–35 vs. <26	-	<26
	36–45 vs. <26	-	<26
	46–55 vs. <26	-	<26
	56–65 vs. <26	-	<26
	66–75 vs. <26	-	<26
	>75 vs. <26	/	
Residence	live in the local county vs. other county	-	Other county
	elsewhere in the state of Florida vs. other county	-	Other county
	other state vs. other county	-	Other county
Gender	Male vs. female	+	Male drivers
Vehicle type			
Accident vehicles	passenger van vs. passenger car	+	passenger van
	pickup/light truck vs. passenger car	+	pickup/light truck
	large size vehicle vs. passenger car	+	large size vehicle
AADT	Average daily traffic flow	+	Higher traffic volume

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

2.1.2 Model structure

This comprises the quasi-induced exposure and multiple logistic regression techniques described in sections 2.1.2.1 and 2.1.2.2 respectively.

2.1.2.1 Quasi-induced exposure technique

In the Quasi-induced exposure analysis, the relative accident involvement ratio (RAIR) is used as the measure of accident causing propensity. The plot of RAIR versus each explanatory variable indicates whether a variable has a positive or negative effect on the occurrence of accidents.

Three types of relative accident involvement ratios were calculated to test the main effects of driver, vehicle and environment factors related to rear-end accidents. RAIRs were calculated using the formula (Stamatiadis and Deacon, 1995, 1997) as follows:

$RAIR_i = (D_{1i} / \sum D_{1i}) / (D_{2i} / \sum D_{2i})$ or $RAIR_i = (V_{1i} / \sum V_{1i}) / (V_{2i} / \sum V_{2i})$ or

$RAIR_i = (E_{1i} / \sum E_{1i}) / (E_{2i} / \sum E_{2i})$ with

$RAIR_i$ = relative accident involvement ratio for type *i* drivers or vehicles or environments

D_{1i} = number of striking drivers of type *i* in rear-end accidents

D_{2i} = number of not-at-fault drivers of type *i* in non-rear-end accidents

V_{1i} = number of striking vehicles of type *i* in rear-end accidents

V_{2i} = number of not-at-fault vehicles of type *i* in non-rear-end accidents

E_{1i} = number of rear-end accidents involving environment type *i*

E_{2i} = number of non-rear-end accidents involving environment type *i*

Furthermore, to test the interaction between type *i* drivers/vehicles/ environments and type *j* drivers/vehicles/environments, the RAIR can be defined as below:

$RAIR_{i,j} = (N1_{i,j} / \sum\sum N1_{i,j}) / (N2_{i,j} / \sum\sum N2_{i,j})$ with

$RAIR_{ij}$ =relative accident involvement ratio for types *i* and *j* drivers/ vehicles/environments

$N1_{ij}$ = number of rear-end drivers, vehicles, environments of types *i* and *j* in rear-end accidents

$N2_{ij}$ = number of not-at-fault drivers, vehicles, environments of types *i* and *j* in non-rear-end accidents

2.1.2.2 Multiple logistic regression technique

Since the dependent variable *Y* (accident classification) can only take on two values: *Y* = 1 for rear-end accidents and *Y* = 0 for non-rear-end accidents, the probability that a rear-end accident will occur is modelled by a multiple logistic regression as below:

$In[\pi(x)/(1-\pi(x))] = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$ with

$\pi(x)$ = conditional probability of a rear-end accident given values of explanatory variables, x, are observed

x_j = vector of explanatory variables describing the j^{th} characteristic of drivers, vehicle type or the environment

β_j = model coefficients which determine odds ratios involved in the rear-end accident when the j^{th} characteristic of drivers, vehicle type or the environment is observed

Further, separate models were fitted for striking and struck vehicles in accidents.

2.1.3 Results

Results obtained from applying the quasi-induced exposure and multiple logistic regression techniques are explained in this section. First, results on road environment will be presented, next results on striking drivers/vehicle characteristics and we close with results on struck drivers/vehicle characteristics. However, we start with a remark on number of lanes and AADT. The number of lanes and AADT were positively correlated. This is because roadways with larger daily traffic have more lanes to handle heavy traffic. As a result, AADT was dropped in this analysis and the number of lanes was used to fit the model. However, a univariate logistic regression model was fitted for AADT and higher traffic volumes were significantly associated with an increase in rear-end accidents.

2.1.3.1 Road environment factors

Seven road environment factors including number of lanes, divided/undivided highway, accident time, road surface condition, highway character, urban/rural and speed limit showed significant associations with the risk of rear-end accidents.

The risk of rear-end accidents happening on 4-lane highways was lower than on 2-lane ones. However, there was no significant difference between 6-lane highways and 2-lane highways and, 6-lane highways and other highways. The risk for undivided highways was lower than that of divided highways.

The rear-end accident risk for the night condition was lower than that for the daytime condition. This may be due to the higher traffic volume during daytime than during the night. The morning and afternoon peaks may affect driving attitude and contribute to accident occurrence. The rear-end accident risk on wet and slippery surfaces was higher compared to a dry road surface. The wet and slippery surfaces reduce drivers' braking ability and as a result accidents occur.

The accident risk in urban area was higher than rural area. This is because urban areas are more dominated by signalized intersections and have complex mobility patterns.

Accident risk on straight up / downgrade (S-U/D), curve-level (C-L) and curveup grade / downgrade (C-U/D) was higher than straight-level (S-L) highway characters. If both curve and grade were present at the same time, the rear-end risk was twice as high as that for normal straight highways. This is because motorists may not account for vehicle mass and momentum which will require a longer stopping time while approaching the intersection on a downhill grade. Further, if the intersection is located on a horizontal curve, there is possible sight obstruction on the inside of curves, such as cut slopes, walls, buildings, bridge piers and longitudinal barriers. These obstructions limit stop sight distance and result in rear-end accidents.

As the speed limit at intersections increased, the risks of getting involved in rear-end accidents increased. At signalized intersections with higher speed limits, drivers are more likely to fall into the dilemma zone (where they can neither execute the intersection crossing nor execute the stopping maneuver safely and comfortably at the onset of yellow). This condition may result in rear-end accidents due to relatively higher operation speeds. On the other hand, at the lower operation speed, the following driver can easily change lane or brake to avoid striking the lead vehicle.

For all findings, similar trends were exhibited by plots of RAIR vs. independent variables except for the comparison between 6- and 2-lane highways. Plots of RAIRs showed that 6 lane highways had higher risk while results from the logistic model did not show a significant difference.

2.1.3.2 Striking driver / vehicle characteristics

Vehicle type and four factors related to driver characteristics including drivers' age, alcohol/drug use, drivers' residence and gender showed significant associations with the risk of rear-end accidents.

Results from the model and RAIR indicated that drivers under influence of alcohol, drugs, or alcohol and drugs and those that had been drinking legal levels of alcohol had higher accident involvement risk than non-drinking drivers. These results were expected since alcohol reduces alertness, interferes with judgment and impairs vision. In addition, most drugs that affect the central nervous system may have the potential to impair driving ability.

Results from the model revealed that there was no significant difference between the oldest (>75 years) and youngest drivers (<26 years). However, results from the comparison of the oldest and youngest drivers using RAIR revealed a higher accident risk for drivers >75 years than drivers <26 years. The remaining age categories of drivers showed lower accident risks than the youngest group. Also, the plot of RAIRs showed a decreasing trend of drivers involved in rear-end accidents with an increase in age until 56–65 years. The oldest group (>75 years) had the highest risk, possibly because of age-related deterioration of their physical and cognitive abilities. The younger group has a larger accident propensity, presumably because of risk-taking and attitudinal factors.

Findings from the model and RAIR plots showed that as the degree of drivers' familiarity with the driving environment decreases, the more they were likely to be involved in rear-end accidents. The local drivers benefit from their driving experiences with the familiar traffic environments so as to avoid the adverse traffic conditions.

Male drivers had a higher risk of accident involvement than female drivers. The analysis is consistent with the study of Ulleberg (2001) in which gender was significant in predicting the involvement in accidents. Further, these findings are in line with what Storie (1977) found: men were involved more often than women in accidents caused by speeding and driving under the influence of alcohol while women were more frequently involved in accidents caused by judgment errors.

The RAIRs for passenger vans, pickup/light trucks and large size vehicles are higher than passenger cars. Similarly, the logistic regression model revealed that accident risks increase with vehicle size. The reason for this result could be that trucks strike other vehicles in the rear much more often than they are struck by other vehicles. Further, truck drivers sit up much higher than passenger vehicle drivers and can see much further down the road, but they may have difficultly responding to the brake light of the leading car with a small headway hence resulting in accidents (Yan et al., 2005).

2.1.3.3 Struck drivers / vehicle characteristics

Based on the results of logistic regression, four factors including vehicle type, driver age, driver residence and gender were found to be associated with rear-end accidents.

Compared to drivers younger than 26 years and older than 75 years, middle age drivers had higher chances on the conditional probability of being struck. The probable reason is that for middle age drivers, better driving experience and physical performance may be helpful to detect a potential conflict earlier than younger and older drivers, but their emergent stop could contribute to their struck role in a rear-end accident.

Female divers were more likely to be in the struck role compared to the male drivers, presumably because the female drivers are more likely to brake when faced with critical traffic events (for example, a yellow signal) than male drivers.

Light trucks were more likely to be in the struck role as compared to the passenger cars while the risk of large size vehicles and vans was lower than passenger cars. These results are consistent with Abdel-Aty and Abdelwahab (2004) who explained that light truck vehicles obscure drivers' visibility of other passenger cars. They may prevent drivers in cars behind them from being aware of traffic situation ahead, therefore more susceptible to collide with the light trucks in case of a sudden application of breaks. Moreover, the results indicated that the risk of large size vehicles was lower than passenger cars. This is because large size vehicles are slowly and noticeable in the traffic stream, their deceleration rate is lower than automobiles and aggressive drivers dislike following a large size vehicle. So, the large size vehicles are less likely to be in the struck role.

Like striking drivers, struck drivers are more likely to be involved in the accidents if they are unfamiliar with the driving environment. The non-local drivers have more difficulties to find their destination and possibly have abnormal behaviors at intersections, such as improper lane changing or sudden stop for right/left-turn, which may contribute to rear-end accidents. Similar results were obtained from the Quasi-induced exposure and multiple logistic regression techniques.

2.1.4 Conclusion and Recommendation

The results showed that seven road environment factors (number of lanes, divided/undivided highway, accident time, road surface condition, highway character, urban/rural and speed limit), five factors related to striking role (vehicle type, driver age, alcohol/drug use, driver residence and gender) and four factors related to struck role (vehicle type, driver age, driver residence and gender) were significantly associated with the risk of rear-end accidents.

The analysis showed that the risk of rear-end accidents for 2-lane highways was higher than 4-lane. The rear-end accidents were more likely to happen at divided highways than undivided highways. The risk of accidents during the night was lower than daytime; compared to a dry road surface, the wet and slippery road surfaces significantly contributed to rear-end accidents; furthermore, as the highway character becomes more complex, rear-end accidents were more likely to occur. When horizontal curve and grade were present at the same time, the rear-end risk could be twice as high as that for normal straight highways. Moreover, the analysis showed a clear trend that as the speed limit increases, the risk of the rear-end accidents increases, especially when the speed limit is higher than 40 mph. Finally, the results confirmed that higher traffic volumes contribute to a higher rear-end risk.

Related to the unfavorable road environment factors, Yan et al. (2005) suggest to consider appropriate engineering countermeasures to reduce the rear-end crash rate. From the perspective of the intersection design and operation, improvement of geometrics will contribute to reducing reaction and stopping times, eliminating motorist confusion and improving visibility of traffic control devices. It is suggested to avoid designing intersections located on a horizontal curve and vertical curve where possible. For existing intersections with a curve or up/down grade, adequate sight distance not only to the signal head but also to the other approaches should be considered so that drivers going-through the intersection can detect potential conflicting vehicles in time.

In addition, motorist information countermeasures such as advanced warning signs installed upstream at the intersection will provide in advance information to the driver about the signal ahead. This will reduce sudden-stop behaviors because of insufficient reaction time due to a signal change in the dilemma zone.

In areas where drivers frequently drive in rainy weather condition, the drainage design should ensure that the rainwater is removed from the intersection surface in time. If intersections with a 50 or 55 mph speed limit have been identified to have higher rear-end crash rates, reducing the speed limit to 40 or 45 mph will efficiently contribute to the lower accident rate.

Corresponding to the higher risk driver populations, younger drivers have less driving experience and tend to drive in conditions that increase their risk. An education program to emphasize the rear-end risk at signalized intersections is strongly suggested for the younger group. According to a previous study (Eby and Molnar, 1998), some of these young drivers may engage in risky driving behaviors because they are risk taking and others engage in this behavior because they are risk ignorant. Those who are risk ignorant may benefit from the education countermeasure and become safer drivers by having a better understanding of the risk inherent in their driving behaviors.

For the older drivers, their higher rear-end risk may result from deteriorating physical conditions, decreasing judgment ability and vision problem. It is necessary to make a further analysis of the criteria of issuing driving license related to driver age and health condition.

Lastly, drivers who had been drinking under legal alcohol use level were more involved in a rear-end accident than non-drinking drivers. This strongly suggests that the threshold of legal alcohol use level on the road should be reduced. If the data of illegal blood alcohol concentration (BAC) were available in traffic accident database, further quantitative analysis of relationship between BAC and accident risk (not limited in rear-end accident) is strongly suggested.

2.2 Summary of multiple logistic regression models

In this section, multiple logistic regression models which capture associations between binary accident outcomes and explanatory variables are summarized. In the study described in the report, the Quasi-induced exposure analysis and multiple logistic regression were used to evaluate the effect of variables on the accident causing propensity. One disadvantage of the Quasi-induced exposure is that it treats each variable independently. This approach is right if one is interested in the effect of only one variable without controlling for other variables. However, if we wish to study the effect of one variable while controlling for other variables, the multiple logistic regression technique is suitable. Apart from the multiple logistic regression technique, several other methods are used to model binary outcomes and an example of such methods is the Probit model. The Probit and the logistic models share some properties, for example, the range of their functions is between 0 and 1. This makes it suitable for their use as probability models representing individual risk. Despite the existence of other methods, the logistic regression has the advantage that its regression effects can be interpreted using odds ratios. Therefore, the logistic regression model is preferred to other methods (Agresti, 2002).

From the results above, it seems that traffic control, traffic flow, vehicle type, geometric and driver characteristics have an effect on accident occurrence at road intersections. We would like to remark that the goodness of fit of the models was not mentioned in the study. However, we assume that the results are valid and provide us with insight into the effect of variables on accidents. Goodness of fit statistics that can be used for this type of model are the deviance and Pearson’s Chi-square since the explanatory variables studied are discrete. The deviance statistic compares the likelihood ratio statistic for a given model versus that of a saturated model. The Pearson Chi-square is used to look at the goodness of fit of a given model versus the saturated model. The values of these statistics are compared with a Chi-square with the degrees of freedom being the difference in the degrees of freedom of a given model and a saturated model. If the values of these statistics are larger than the Chi-square value, then there is evidence of lack of fit (Agresti, 2002).

Although one study in which the multiple logistic regression technique was applied has been described, the technique is widely used (e.g Wang et al., 2007; Al Ghamdi, 2002) utilized the logistic regression model to investigate the influence of explanatory variables on accident occurrence at intersections.

The multiple logistic regression technique described in this chapter is used to analyze only accident binary outcomes. However, there are studies in which accident outcomes are continuous. In such cases, multiple linear regression analysis which describes relationships between continuous outcomes and explanatory variables are more credible. Thus, in chapter 3 several studies in which the accident outcome was continuous and multiple linear regression analysis was applied are described.

3. MULTIPLE LINEAR REGRESSION MODELS

Multiple refers to many explanatory variables. Explanatory variables are characteristics whose effect on the outcome is being assessed. Multiple linear regression is a statistical methodology describing relationships between a continuous outcome and a set of explanatory variables (Kutner et al., 2005).

In this chapter, two studies in which this technique was used are presented. In the first study, the analysis of accident data from four-arm (STOP-controlled and signalized) and three-arm/STOP controlled intersections is presented and in the second study, we investigate relationships between roundabout geometry and accident rates.

3.1 Study 1

Bauer and Harwood (2002) developed APMs for urban at-grade intersections in California. The data used involved collision types from 1990-1992. Three types of urban intersections were discussed, namely four-arm/STOP controlled, three-arm/STOP controlled and four-arm/signalized intersections. The response variable was the number of accidents at the intersections. Multiple linear regression was used to analyze data from four-arm (STOP-controlled and signalized) intersections while Poisson and negative binomial regression models were used for three-arm/STOP controlled intersections. Therefore, analysis of three-arm/STOP controlled intersections will be presented in chapter 4 on Poisson and negative binomial regression models.

3.1.1 Explanatory variables

Data were collected on geometric characteristics, traffic control and traffic volume. Based on engineering judgement, a decision on which explanatory variables should be used in the model was made. However it should be noted that selecting variables based on subjective judgement may lead to biased results. Applying statistical model selection procedures would yield a better model. Depending on the objective of the model, procedures such as adjusted R-square, Mallow's Cp or the prediction sum of squares (PRESS) could be used (Kutner et al., 2005). Not only are these procedures used for variable selection, they are also used to assess the quality of the model. For example, if the purpose is to determine the explanatory power of the variables in the model, the adjusted R-square criterion is used, if it is about how well the fitted values can predict the observed responses, the prediction sum of squares (PRESS) is a good measure, etc. (Kutner et al., 2005). Table 2 below presents the variables collected on each type of intersection. In addition, the directions of the effects of the variables on the number of accidents are provided.

Table 2: Explanatory variables in each category

Category of explanatory variables	**Variable levels**	**Effect**	**Causes more accidents**
1. Four-arm / STOP controlled intersections			
AADT on major and minor roads	continuous	+	Increased AADT
Design speed of major roads	continuous	/	
Outside shoulder width on major roads	continuous	/	
Average lane width on major road	continuous	-	Narrow width
Terrain		/	
Functional class of the major road	Major collector vs. Principal arterial	/	
	Minor arterial vs. Principal arterial	/	
Lighting	yes =0, no = 1	-	Lighting
Major road left-turn channelization	Discrete	/	
Major road right-turn channelization	Discrete	/	
Major road left-turn prohibition	Left turn prohibited vs. Left turns permitted	-	Left turns permitted
Number of lanes on major road	3 vs. ≥ 6	+	3
	4 or 5 vs. ≥ 6	/	
Minor road left-turn channelization	Discrete	/	
Minor road right-turn channelization	No free right turns vs. free right turns	-	Free right turns
Presence of median on major road	yes =0, no = 1	/	
Access control on the major road	None vs. Partial	-	Partial access control on major roads
2. Four-arm / signalized intersections			
AADT on major and minor roads	continuous	+	Increased AADT
Design speed of major roads	continuous	+	Increased speed
Outside shoulder width on major roads	continuous	/	
Average lane width on major road	continuous	-	Narrow width
Terrain		/	
Functional class of the major road	Major collector vs. Principal arterial	/	
	Minor arterial vs. Principal arterial	/	
Major road left-turn channelization	Discrete	/	
Major road right-turn channelization	No free right turns vs. free right turns	-	Free right turns
Number of lanes on minor road	≤ 3 vs. ≥ 4	-	≥ 4
Number of lanes on major road	≤ 3 vs. ≥ 6	-	≥ 6
	4 or 5 vs. ≥ 6	-	≥ 6
Minor road left-turn channelization	Discrete	/	
Minor road right-turn channelization	Discrete	/	
Presence of median on major road	Discrete	/	
Access control on the major road	None vs. Partial	-	Partial access control on the major roads
Signal timing	Fully actuated vs. Semi actuated	+	Fully actuated
	Pretimed vs. Semi actuated	/	
Signal phase	Multiphase vs. Two-phase	-	Two-signal phase
Presence of a Minor road signal mast arm	Yes =1, no= 0	/	

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

3.1.2 Model structure

A log-linear regression model was fitted to the four-arm intersections (STOP-controlled and signalized). In this case, the logarithm of the number of accidents is supposed to be normally distributed and the parameters are estimated by the least square method. Separate models were fitted for the total number of accidents and, fatal and injury accidents and all models were of the form below:

$\mu_i = e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\ldots+\beta_q x_{iq}}$, with

μ_i = mean of the number of accidents at intersection type *i*

x_{ij} = vector of explanatory variables collected from the j^{th} characteristic at the intersection *i*

β_j = parameters describing the effect of explanatory variables*, j,* on the mean number of accidents at intersection *i*

x_{i1} = log of traffic flow in AADT on major roads at intersection *i*

x_{i2} = log of traffic flow in AADT on minor roads at intersection *i*

3.1.3 Results

Only results of four-arm/signalized and four-arm/STOP controlled intersections are presented. The analysis for three arm stop controlled intersections will be presented in chapter four since the Poisson and negative binomial regression models were used for its analysis. The results for four-arm/signalized intersections are described first and those for four-arm/STOP controlled intersections are explained after.

3.1.3.1 Four-arm / signalized intersections

Findings from the four-arm/signalized models for all accidents showed a positive relationship between increased AADT on major and minor road and, signal timing on the total number of accidents. Fully actuated signal timing increased accident numbers compared to semi actuated signal timing. On the other hand, increased average lane width on the major road, multiphase signal phasing compared to two phase signal phasing, the absence of free right turn channelization on the major road, the absence of access control on major road, the existence of three or fewer than three lanes on minor road and having three or fewer than three lanes and four or five lanes compared to six or more than six lanes on the major road had a negative effect on the number of accidents.

The four-arm/signalized models for fatal and injury accidents disclosed a positive relationship between AADT on major and minor road and design speed on major road. Fully actuated signal timing increased accident numbers compared to semi actuated signal timing. On the other hand, multiphase signal phasing compared to two phase signal phasing, the absence of access control on major road, the existence of three or fewer than three lanes on minor road and four or five lanes compared to six or more than six lanes on the major road had a negative effect on the number of fatal and injury accidents.

3.1.3.2 Four-arm /STOP controlled intersections

Results from four-arm/STOP controlled models for the total number of accidents indicated that increased AADT on the major and minor road and number of lanes on major road had a positive effect on the total number of accidents. Having three or fewer than three lanes on the major road compared to six or more than six lanes increased the number of accidents. On the other hand, an increase in the average lane width on major road, presence of major road left-turn prohibition, absence of access control on major road, absence of minor-road right-turn channelization and absence of lighting reduced the total number of accidents.

The four-arm/STOP controlled intersections model for fatal and injury accidents included an extra variable of outside shoulder width on the major road but excluded lighting. The effects of variables common to the fatal and injury accident model for four-arm/STOP controlled intersections as well as a model for total accidents were similar. The extra variable of outside shoulder width on the major road in the fatal and injury accidents' model reduced accidents.

In this study, we would like to mention that the multiple linear regression model used is not the best approach as the response variable was a count and takes on only nonnegative numbers. This method may predict negative numbers of accidents. Further, the information on how good the fitted model is, was not provided. For the goodness of fit, the approaches described in section 3.1.1 of this study can be used depending on the aim of the model. On the other hand, if the analysis was to be redone, we recommend the use of methods for analysing count data, for example, the Poisson and negative binomial models. All the same, the results provide us with an insight of which variables have an effect on the number of accidents despite the shortfalls of the methodology used for the analysis.

3.1.3 Conclusion and recommendation

It appears from this study that increased AADT on major and minor roads, fewer number of lanes on the major road, higher design speeds of major roads and fully actuated signal timing increased the number of accidents. Therefore, it is suggested to reduce the design speed of major roads and to control the AADT on major and minor roads, for example by channelization.

3.2 Study 2

Arndt and Troutbeck (1998) applied multiple linear regression analysis to determine how much variation in the number of accidents was explained by roundabout geometry variables and accident rates in rural and urban areas of Queensland and Australia. A total of 492 accidents and 100 roundabouts on rural and urban arterial roads were studied. Data on geometric design, traffic volume, traffic control and accidents were collected for a five year period (1986 to 1990). Three types of accidents including single vehicle, approaching rear-end collisions and entering/circulating collisions were considered and a separate model was fitted for each type.

3.2.1 Explanatory variables, model structure and results

The variables and model fitted to the single vehicle type of accidents are described first, followed by a model fitted to approaching rear-end collisions and finally a model fitted to entering/circulating collisions.

3.2.1.1 Single vehicle accident

The single vehicle model includes traffic flow, length of driver path on geometric element, 85th percentile speed on previous geometric element and radius of geometric element as explanatory variables. The following equation was fitted:

$$E(\mu_{\sin gle}) = 3.63 * 10^{-14} * Q * L * (S + \Delta S)^2 \left((S + \Delta S)^3 / (R^{1.5} + 47.4)\right)$$ with

$E(\mu_{\sin gle})$ = expected annual single-vehicle accident frequency

Q = annual average daily traffic (AADT) in the direction considered

L = length of vehicle path on the geometric element (m)

S = 85th percentile speed on the geometric element (km/h)
ΔS = decrease in S at the start of the geometric element (km/h)
R = radius of the geometric element

Explanatory variables in this model explained 18.1% of the observed variance in the expected annual single-vehicle accident frequency.

3.2.1.2 Approaching rear-end collisions

The approach rear-end collision model includes approaching and circulating flows, and relative speed between entering and circulating vehicles as explanatory variables. The model fitted to the data was of the form below:

$E(\mu_{rearend}) = 0.62 * 10^{-11} * Q_a * Q_c^{0.5} * S_a^2$ with

$E(\mu_{rearend})$ = expected annual rear-end accident frequency
Q_a = AADT on the approach (one-way traffic)
Q_c = AADT on the circulating carriageway adjacent to the approach (one-way traffic)
S_a = 85th percentile speed on the approach curve (km/h)

Explanatory variables in this model explained 30.6% of the variations in the expected annual approaching rear-end accident frequency.

3.2.1.3 Entering / circulating collisions

The entering/circulating collision model includes approaching and circulating flows and relative speed between entering and circulating vehicles as explanatory variables. Relationships between accident rates and these explanatory variables were studied using the following equation:

$E(\mu_{entering}) = 3.45 * 10^{-12} * Q_a * \sum (Q_{ci} * S_{ri}^2)$ with

$E(\mu_{entering})$ = expected annual accident frequency involving entering and circulating vehicles
Q_a = AADT on the approach (one-way traffic)
Q_{ci} = AADT for each of the *i* movements on the approach entrance curve and on the circulating carriageway adjacent to the approach
S_{ri} = 85th percentile speed for each of the *i* movements on the approach entrance curve and on the circulating carriageway adjacent to the approach (km/h)

The variables in this model explained 10.7% of the variance in the expected annual accident frequency involving entering and circulating vehicles.

3.2.2 Conclusion and recommendation

The results from this study enable the design and construction of roundabouts at a similar cost to minimize accidents. In addition these results may be applied to all intersection and roadway types to minimize accidents (Arndt and Troutbeck, 1998).

The adjusted R-square values of the above models imply a rather low quality as the variables in the models explain a small proportion in the variation of the response. It was also noted that there was no selection procedure used to enter the variables in the models. This might be one of the possible reasons for the low quality of the models.

Using model selection procedures helps to choose a good model depending on the purpose of the research. For example, if the purpose is to determine the explanatory power of the variables in the model, the adjusted R-square criterion is used, if it is about how well the fitted values can predict the observed responses, the prediction sum of squares (PRESS) is a good measure, etc. (Kutner et al., 2005). However, considering the objective of this report, the variables are related to the number of accidents and have a significant effect. Therefore, the results are retained taking the low quality of the model into account.

3.3 Summary of multiple linear regression analysis

In this section, the multiple linear regression technique is summarized. This technique is appropriate to describe relationships between continuous accident outcomes and explanatory variables. It appears from the previous studies that traffic flow, traffic control and geometric characteristics play a big role in the accident risk at intersections.

Although multiple linear regression models have been applied in the previous studies, the response variables were the numbers of accidents. We would like to comment that the multiple linear regression modelling is not appropriate for count data since counts are positive numbers yet the response variable in multiple linear regression analysis is assumed to follow a normal distribution which covers all numbers on a real interval. This method has limitations of predicting negative numbers of accidents which is not logical. This undesirable statistical property limits multiple linear regression models to describe adequately the random, discrete and nonnegative accident events (Chin and Quddus, 2003). For such reasons, there is need to utilize techniques which can sufficiently describe the specific characteristics of accident events. Such techniques include Poisson regression, negative binomial regression and others as described in the subsequent chapters.

4. POISSON AND NEGATIVE BINOMIAL MODELS

Since accident occurrences are unavoidably discrete and more likely random events, the family of Poisson regression models appears to be more suitable than multiple linear regression models. The Poisson regression model is used when discrete response variables have counts as possible outcomes, for example, the number of accidents. However, Poisson models have potential problems; one constraint is that the mean must equal to the variance. If this assumption is not valid, that is, the accident data are significantly overdispersed (the variance is much greater than the mean), the standard errors usually estimated by the maximum likelihood method, will be biased and the test statistics derived from the model will be incorrect. This results in incorrect estimation of the likelihood of accident occurrence (Chin and Quddus, 2003).

To solve the problem of overdispersion, the negative binomial distribution has been employed instead of the Poisson. To establish the negative binomial regression model, an overdispersion parameter is introduced into the relationship of the mean and the variance. By relaxing the condition of mean equal to variance, the negative binomial regression models have more desirable properties than Poisson models to describe the relationship between accident occurrence and road characteristics (Chin and Quddus, 2003). Hence, in the successive sections Poisson and negative binomial regression models are presented as a more credible alternative to multiple linear regression analysis. The majority of the authors presented results from only the negative binomial models although the Poisson model was also fitted.

There are intersections at which zero accidents are recorded on a number of occasions. When there is a zero accident record over a period of time, it may indicate either that the intersection is nearly safe, or that the zero record is a chance occurrence or accidents are not reported. Since the standard Poison and negative binomial models do not help to identify accident contributory factors in this case, it becomes necessary to model the two states. Moreover if the two states are modelled as a single state, the estimated models may be biased as there may be an overrepresentation of zero accidents. Hence the presence of excess zeros in the accident count data may be mistakenly regarded as the presence of overdispersion in the data set, which arises because of an incorrectly specified model. To handle count data with excess zeros, the zero-inflated negative binomial or Poisson models are employed (Kumara and Chin, 2003). This helps to distinguish safe intersections with little probability of accident occurrence from those with zero accident record due to chance or when accidents are not reported. Only the zero-inflated negative binomial model by Kumara and Chin (2003) is presented in section 4.8.

In this chapter, we will discuss eight studies in which Poisson and negative binomial models were applied. All these studies investigated relationships between the number of accidents and traffic volume, traffic control, geometric, land use and environmental characteristics.

4.1 Study 1

Salifu (2004) developed negative binomial regression models to study the relationships between the number of accidents and, traffic flow, traffic control and geometric characteristics of urban intersections in Ghana. The study comprised of 91 intersections of which 57 were three-arm and 34 were four-arm unsignalized urban junctions. A total of 354 accidents and 238 accidents were recorded at three-arm and four-arm junctions respectively for the period 1996 to 1998 inclusive. These accidents account for more than 60% of intersection accidents in Ghana. Accident types covered include property-damage as well as person-involved collisions. The data used in this study were retrieved from the national accident database at the Building and Road Research Institute. The database is compiled from police files using a standard accident report form, which contains information on about 90 variables relating to the time, place, circumstances, the parties involved, etc.

4.1.1 Explanatory variables

Data on traffic flow, traffic control and geometric design features were collected for each junction. Table 3 below provides the variables in each category and the direction of their effects on accident occurrence.

Table 3: Explanatory variables in each category

Category of explanatory variables	Variable levels	Effect	Causes more accidents
1. Four-arm intersections			
Traffic flow			
Vehicle counts & pedestrians crossing all arms (AADT)	continuous	+	Increased vehicles & pedestrians
Proportion of heavy good vehicles as a percentage of the total traffic flow	Proportion	+	Increased proportion of heavy good vehicles
Traffic control			
Spot speeds of vehicles approaching the intersection along the major road	continuous	+	High spot speed
Geometric design features			
Junction layout		/	
Type of major and minor roads		/	
Type and width of lanes		/	
Types of control		/	
Road markings		/	
Street lighting	present =0, absent = 1	+	Absence of street lights
Status of crossing facilities		/	
Left turn lane on major road	present =0, absent = 1	+	Absence of left-turn lanes
Width of the minor roads at neck of Intersections	continuous	+	Wider width at neck of intersections
2. Three-arm intersections			
Traffic flow			
AADT on major and minor roads	continuous	+	Increased AADT on major and minor roads
Traffic control			
Spot speeds of vehicles approaching the intersection along the major road	continuous	+	High spot speed
Traffic control of level 2 (yield on minor road)	present =1, absent = 0	-	Absence of traffic control level 2
Traffic control of level 3 (No control on minor road)	present =1, absent = 0	-	Absence of traffic control level 3
Geometric design features			
Average width of median on major roads	continuous	-	Narrower width of median

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

4.1.2 Model structure

The general form of the fitted models was as follows:

$$E(\mu_i) = \beta_o * AADT_{\min or}^{\beta_1} * AADT_{major}^{\beta_2} * e^{\sum \beta_j x_{ij}}$$ with

$E(\mu_i)$	= expected number of accidents at the i^{th} intersection
$AADT_{\min or}$	= traffic flow on minor roads
$AADT_{major}$	= traffic flow on major roads
β_o	= intercept to be estimated
β_1	= effect of traffic flow on minor roads on the expected number of accidents
β_2	= effect of traffic flow on major roads on the expected number of accidents
x_{ij}	= vector of explanatory variables, *j*, other than traffic flow on intersection *i*
β_j	= regression coefficient representing the effect of the j^{th} explanatory variable other than traffic flow

The above model is transformed using a log-link function as follows:

$$\ln(\mu_i) = \ln(\beta_o) + \beta_1 \ln AADT_{\min or} + \beta_2 \ln AADT_{major} + \sum \beta_j x_{ij}$$

4.1.3 Results

The findings obtained from the analysis of four-arm and three-arm intersections are explained in the following sections. First, findings from four-arm intersections are described, followed by results from three-arm intersections. The Freeman-Tukey test was used to assess the adequacy of the models (Kulmala, 1995). For the model of four-arm intersections, the value of Freeman-Tukey was 0.91. This means the variables in the model explained about 91% of the variation in the number of accidents. For the model of three-arm intersections, the value of Freeman-Tukey was 0.50. These results show that the variables in the model explained about half of the variation in the number of accidents.

4.1.3.1 Four-arm intersections

The model for four-arm intersections revealed a positive relationship between the number of accidents and the variables crossing flow products (AADT), absence of left turn lane on major road, proportion of heavy good vehicles as a percentage of the total traffic flow, average width of the minor road at the neck of the junction, absence of streetlights and standard deviation of average spot speeds on major approaches.

4.1.3.2 Three-arm intersections

Results from three-arm intersections showed that the variables major and minor road daily traffic and standard deviation of average spot speeds on major approaches had a positive effect on the number of accidents while the variables average width of median on major roads, traffic control of level 2 (Yield) on minor road and traffic control of level 3 (No control) on minor road had a negative effect on the number of accidents. These findings are in line with those of Harnen et al. (2003).

4.1.3 Conclusion and recommendation

The results showed that increased vehicle counts and pedestrians crossing all arms, high speeds of vehicles approaching the intersection along the major road, the absence of street lighting, the absence of left-turn lanes and wider width of the minor roads at the neck of intersections increased the number of accidents. From these results, we recommend the provision of street lights, narrowing the width of the minor roads at the neck of intersections, reducing speeds of vehicles approaching the intersection along the major road and regulating traffic flow on all intersecting arms.

4.2 Study 2

Harnen et al. (2003) fitted Poisson and negative binomial regression models to investigate the relationships between motorcycle crashes on urban unsignalized intersections and several explanatory variables in Malaysia. In Malaysia, half of the registered motor vehicles are motorcycles and more than 60% of all casualties are motorcycle casualties. Therefore, traffic safety for motorcyclists is important in Malaysia. A total of 53 urban unsignalized intersections were considered in Selangor, Malaysia. Data were collected over the period 1997 – 2000.

Only intersections which had marginal land use changes, had not undergone major modifications and had an equal number of lanes on the major and minor-road approaches were selected. In addition, intersections were required to have a history of personal injury accidents. Due to this, some bias might have been introduced into the model. To minimise this problem, sampling techniques are suggested as an alternative. This would reduce bias in a sense that intersections are chosen randomly and entirely by chance. Each intersection will have the same probability of being included in the study regardless of its history (Yates et al., 2008).

4.2.1 Explanatory variables

Data were collected on traffic flow, traffic control and intersection geometry. Table 4 below presents the variables in each category and the direction of their effects on the number of accidents.

Table 4: Explanatory variables in each category

Category of explanatory variables	Variable labels	Effect	Causes more accidents
Traffic flow			
Non-motorcycle on major roads	QNMm	+	Increased non-motorcycle traffic on major roads
Non-motorcycle on minor roads	QNMn	+	Increased non-motorcycle traffic on minor roads
Motorcycle flows on major roads	QMm	+	Increased motorcycle traffic on major roads
Motorcycle flows on minor roads	QMn	+	Increased non-motorcycle traffic on minor roads
Total pedestrian flow on major & minor roads	QPED	*	
Traffic control			
Approach speed on major and minor road	SPEED	+	High approach speed
Intersection geometry			
Average Lane width on major roads	LWm	-	Narrow width
Average Lane width on minor roads	LWn	-	Narrow width
No. of lanes on major road	LNm	-	Few lanes on major road
No. of lanes on minor road	Lnn	-	Few lanes on minor road
Average shoulder width on major & minor roads	SHDW	-	Narrow width
No. of intersection arms	NL	/	
Land use			
commercial or non-commercial	LU	*	

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

4.2.2 Model structure

The general form of the fitted model was:

$$E(\mu_i) = \beta_o * QNMm^{\beta_1} * QNMn^{\beta_2} * QMm^{\beta_3} * QMn^{\beta_4} * QPED^{\beta_5} *$$
$$e^{\beta_6 * SPEED + \beta_7 * LWm + \beta_8 * LWn + \beta_9 * LNm + \beta_{10} * NL + \beta_{11} * SHDW + \beta_{12} * LU + \beta_{13} * Lnn}$$, with

$E(\mu_i)$ = mean number of motorcycle crashes on intersection type *i*

β_j = parameters to be estimated and indicate the effect of explanatory variable *j* on the mean number of motorcycle crashes

The remaining abbreviations in the model are defined in Table 4 above.

4.2.3 Results

In this section, only results from the negative binomial model are presented. The scaled deviance was used to test the fit of the model. Its value (4.52 on 10 degrees of freedom) indicated a good fit.

All the variables in the traffic flow and traffic control categories increased the number of motorcycle crashes. They included non-motorcycle flow on the major road, non-motorcycle flow on the minor road, motorcycle flow on the major road, motorcycle flow on the minor road and approach speed on the major and minor roads.

All intersection geometry characteristics reduced the number of motorcycle crashes. They included average lane width on major road, average lane width on minor road, number of lanes on major road, number of lanes on minor road and average shoulder width on major and minor roads.

4.2.4 Conclusion and recommendation

The model developed in this study can be used to determine appropriate intervention measures for intersections with respect to motorcycle crashes. Using this model, suitable design measures of unsignalized intersections can be specified. The treatment could be the provision of non-exclusive motorcycle lane facilities at intersections. However, this model might only be valid in developing countries like Malaysia where the proportion of motorcycles using unsignalized intersections constitutes 50% of all vehicles.

4.3 Study 3

Bauer and Harwood (2002) developed a log linear regression model for three-arm/STOP-controlled urban at-grade intersections in California. The accident data used were collected from collision types from 1990 to 1992. Their aim was to study the relationships between the number of accidents and, geometric characteristics, traffic control and traffic volume.

4.3.1 Explanatory variables

The explanatory variables used were collected on traffic flow, traffic control, geometric design and the environment of an intersection. Engineering judgement was used to decide on which explanatory variables to feed in the model. The variables in each category and the direction of their effects on the mean number of accidents are given in Table 5 below.

Table 5: Explanatory variables in each category

Category of explanatory variables	Variable levels	Effect	Causes more accidents
1. Total number of accidents			
Traffic flow			
AADT on major & minor roads	continuous	+	Increased AADT
Geometric features			
Average lane width on major road	continuous	-	Narrow width
Presence of median on major road	Yes = 1, no = 0	-	Absence of median
Traffic control			
Major road left-turn prohibition	Left turn prohibited vs. Left turns permitted	-	Left turns permitted
Minor road right-turn channelization	No free right turns vs. free right turns	-	Free right turns
Design speed of major roads	continuous	+	High speed
2. Fatal and injury accidents			
Traffic flow			
AADT on major & minor roads	continuous	+	Increased AADT
Traffic control			
Design speed of major roads	continuous	+	High speed
Major road left-turn prohibition	Left turn prohibited vs. Left turn permitted	-	Left turn permitted
Access control on the major road	None vs. Partial	-	Partial access control on the major road
Geometric design features			
Outside shoulder width on major roads		/	
Average lane width on major road	continuous	-	Narrow width
Functional class of the major road	Major collector vs. Principal arterial	/	
	Minor arterial vs. Principal arterial	/	
Lighting	yes =0, no = 1	-	Presence of lights
Major road left-turn channelization	Discrete	/	
Major road right-turn channelization	No free right turns vs. free right turns	-	Free right turns
Minor road left-turn channelization	Discrete	/	
Minor road right-turn channelization	No free right turns vs. free right turns	-	Free right turns
Presence of median on major road	yes =1, no = 0	-	Absence of median
Number of lanes on minor road	≤ 3 vs. ≥ 4	-	≥ 4
Number of lanes on major road	≤ 3 vs. ≥ 6	-	≥ 6
	4 or 5 vs. ≥ 6	-	≥ 6
Environmental factors			
Terrain		/	

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

4.3.2 Model structure

A log linear regression model which assumes the number of accidents to follow a Poisson or a negative binomial distribution was fitted and the coefficients were estimated using maximum likelihood method. For this study, the accidents were assumed to follow a negative binomial distribution. Separate models were fitted for the total number of accidents (fatal, injury and non-injury) and, fatal and injury accidents. Both models were of the form below:

$E(\mu_i) = e^{\beta_0 + \beta_1 x_{i1} + ... + \beta_j x_{ij}}$, with

$E(\mu_i)$ = mean number of accidents at intersection type *i*

x_{ij} = vector of explanatory variables from the j^th characteristic at intersection type *i* other than traffic flow

β_j = parameters describing the effect of explanatory variables, j, on the mean number of accidents at intersection *i*

Generally, x_{i1} and x_{i2} represent the log of the traffic flow in AADT on minor and major roads at intersection *i* respectively. Therefore, the model can be written as:

$$EA(\mu_i) = e^{\beta_0 + (AADT_{minor})\beta_1 + (AADT_{major})\beta_2 + \beta_3 x_{i3} + ... + \beta_j x_{ij}}$$

4.3.3 Results

In this section, the obtained results from the fitted models are described. First, we present the analysis of the total number of accidents and in the next section results from the analysis of fatal and injury accidents are explained. The scaled deviance and Pearson Chi-square statistics were used to test the fit of the models. The scaled deviance and Pearson Chi-square statistics of the model for the total number of accidents were 1 and 1.08 respectively while those of the model for fatal and injury accidents were both 1. These statistics indicate a good fit.

4.3.3.1 Analysis of total number of accidents

The results from the model fitted on the total number of accidents are in Table 5 under the subheading of total number of accidents**.** Models for all accidents (total) revealed positive relationships between the total number of accidents and AADT on the major and minor roads and, design speed of the major road. In contrast, wider average lane width on major road, major road left-turn prohibition, none free minor road right-turn channelization and presence of median on major road reduced the total number of accidents.

4.3.3.2 Analysis of fatal and injury accidents

The results from the model fitted on the fatal and injury accidents are under the subheading of fatal and injury accidents in Table 5. The fatal and injury accident model showed that increased AADT on major and minor road and the design speed of major roads increased the number of accidents. However, wider average lane width on major road, lighting, non-free major road right-turn channelization, non-free minor road right-turn channelization, the presence of median on major road, having three or fewer than three lanes compared to four or more than four lanes on minor roads, having three or fewer than three lanes compared to six or more than six lanes on the major roads and having four or five lanes on the major roads compared to six or more than six lanes reduced the number of accidents.

4.3.4 Conclusion and recommendation

The results from this study indicate that increased AADT on major and minor roads and higher design speeds of major roads increased the number of accidents. Thus, reducing design speeds of major roads and controlling AADT on major and minor roads may result in a reduced number of accidents.

4.4 Study 4

Bauer and Harwood (2000) used a negative binomial model to investigate the relationship between traffic accidents and geometric design, traffic control and traffic volume variables for at-grade intersections on rural single carriageway two-lane roads. Data were collected from 1,434 four-leg intersections and 2,692 three-leg intersections. They studied 5,631 accidents on four-leg intersections of which 2,759 were non-fatal injury accidents and the remaining fatal. A total of 6,399 accidents were considered on three-leg intersections including 2,905 injury accidents. Accident and traffic data used in this study for analysis were collected for the period 1990 to 1992. The response variable was total number of accidents.

4.4.1 Explanatory variables

The categories of explanatory variables considered included traffic flow, traffic control, geometric design and environmental factors. The variables in each category and the direction of their effects on the number of accidents are presented in Table 6 below.

Table 6: Explanatory variables in each category

Category of explanatory variables	Effect
Traffic flow	
AADT on major and minor roads	*
Traffic control	
Design speed of major road	*
Major road left-turn prohibition	*
Type of access control on major road	*
Geometric features	
Number of lanes on major roads	*
Width of major road outside shoulder	*
Presence of major road and crossroad right-turn channelization	*
Presence of a median	*
Presence of a major road left-turn channelization	*
Average lane width on major road	*
Functional class of major road	*
Presence of night lighting at intersection	*
Environmental factors	
Type of terrain	*

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

4.4.2 Model structure

Models were fitted using a Poisson and negative binomial regression. However only the results from the negative binomial model were presented. The general model form is stated by the following equation:

$$E(\mu_i) = e^{\beta_o + AADT_{\min or}^{\beta_1} + AADT_{major}^{\beta_2} + \sum \beta_j x_{ij}}$$ with

$E(\mu_i)$ = expected number of accidents on intersection type *i*

AADT $_{major}$ = incoming annual average daily traffic on the major road

AADT $_{\min or}$ = incoming annual average daily traffic on the minor road

x_{ij} = vector of explanatory variables, *j*, on intersection i other than AADT

β_j = vector of coefficients to be estimated and indicate the effect of explanatory variable, *j*, on the number of accidents

4.4.3 Results

The model explained between 34 – 39% of the observed variability in the number of accidents. Statistically significant explanatory variables (the direction of the effects not given) from this model included: number of lanes on major roads, presence of major road left-turn prohibition, type of access control on major road, width of major road outside shoulder, presence of major road and crossroad right-turn channelization, design speed of major road, presence of a median, presence of a major road left-turn channelization, average lane width on major road, type of terrain, functional class of major road and presence of night lighting at intersection.

4.4.4 Conclusion

From this study, it appears that AADT in the major and minor road accounted for most of the variability in the expected number of accidents. Geometric design variables accounted for a small proportion of the observed variability.

It was noted that in this study, the directions of the effects of explanatory variables were not indicated in this study. This limits the potential for direct use of these results for researchers interested in the direction of the effect. Nevertheless, for the objective of this review these results provide information on which variables have a significant effect on the number of accidents.

4.5 Study 5

Vogt (1999) describes the development of a negative binomial regression model for three types of intersections on rural roads in California and Michigan. Three types of intersections were studied and they were: three-leg intersections between four-lane major roads and two-lane stop controlled minor roads; four-leg intersections between four-lane major and two-lane stop controlled minor road; and signalized intersections between two-lane roads. Traffic and accident data were collected for the period 1993 to 1995.

The study involved 84 three-leg intersections, 72 four-leg intersections and 49 signalised intersections. Data collected on accidents included the total number of accidents, the number of accidents by severity class (fatal, injury and property damage only) and by accident type (head-on, sideswipe, rear end, overturned, pedestrian, hit object and other accidents).

Two criteria were used to define an accident as an intersection accident:

1. Any accident occurring at the intersection or within 76m of the intersection centre along the major road or within 30.5m (76m in California) of the intersection centre along the minor road.
2. Any accident that occurs within 76m of the intersection centre and that meets one of the following additional control conditions:
   a. Vehicle-pedestrian accident,
   b. An accident in which one of the vehicles involved in a crash is making a left turn, a right turn, or a U-turn.
   c. A multiple-vehicle accident, in which the accident type is either sideswipe, rear-end, broadside or angle.

Four response variables were considered: the total number of intersection accidents (both criteria) and the total number of intersection injury accidents (both criteria).

4.5.1 Explanatory variables

Data were collected on more than 40 explanatory variables covering important issues such as, traffic volume, roadside characteristics, channelization, intersection geometry, characteristics of major and minor road alignments and other characteristics. Examples are posted speed limits, type of signal, number of signal phases, road lighting, type of terrain and state. Table 7 below provides all variables and the direction of their effects on the total number of accidents.

Table 7: Explanatory variables in each category

Category of explanatory variables	Effect	Causes more accidents
1. Three leg intersections		
Traffic flow		
AADT on major and minor roads	+	Increased volumes
Percentage of traffic turning left on the major road	*	
Geometric design features		
Median width on major road	-	Narrow medians on major road
Number of driveways in major road	+	Many driveways on major road
Angle between road & Vertical alignments	*	
2. Four-leg intersections		
Traffic flow		
AADT on major and minor roads	+	Increased volumes
Percentage of incoming major traffic turning left during peak hours	+	Increased incoming major traffic turning left during peak hours
Geometric design features		
Presence of left-turn lanes on major road	-	Absence of left-turn lanes on major road
Grade and sight distance	*	
Traffic control		
Posted speed limit on the minor road	*	
3. Signalized intersections		
Traffic flow		
AADT on major and minor roads	+	Increased volumes
Average absolute percent grade change per 100m along the major and minor road approaches, within 244m of distance from the centre of the intersection	+	Increased % grade change per 100m along major & minor road approaches
Percentage of all incoming truck traffic in peak hours	+	More incoming truck traffic in peak hours
Percentage of incoming minor traffic turning left during peak hours	-	Lower percentage of incoming minor traffic turning left during peak hours
Geometric features		
Vertical curves on major and minor road	*	
Number of driveways on the major road	*	
Traffic control		
Protected left turn lanes on major road	*	

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

4.5.2 Model development

Negative binomial models were developed for each intersection type and the general model form is given by the following equation:

$$E(\mu_i) = e^{\beta_o + \sum \beta_j x_j}$$, with

$E(\mu_i)$ = expected number of accidents on intersection *i*

x_j = vector of explanatory variables, *j*

β_j = vector of coefficients to be estimated and indicate the effect of the explanatory variable, *j*, on the number of accidents

4.5.3 *Results*

In the following paragraphs, a brief explanation of the results found by applying the model developed in section 4.5.2 is given. We shall first present the results from three-leg intersections analysis, followed by four leg intersections and finally findings from signalized intersections.

4.5.3.1 Three-leg intersections analysis

Results from three-leg intersection analysis indicated that AADT on major and minor roads and the number of driveways on the major road increased the total number of accidents. On the other hand, the median width of the major road had a negative effect on the total number of accidents. Also the percentage of traffic turning left on the major road had a significant effect on the total number of accidents. Models for the number of injury accidents revealed that the angle between road alignments and vertical alignments had a significant effect on the number of accidents.

4.5.3.2 Four-leg intersection analysis

From four-leg intersection analysis, AADT on each road and the percentage of incoming major traffic turning left during peak hours had a positive effect while the presence of left-turn lanes on major road had a negative influence on the total number of accidents. Grade and limited sight distance had an effect on the total number of accidents. Models for the number of injury accidents showed that posted speed limit on the minor road was a significant variable.

4.5.3.3 Signalized four-leg intersections

It was observed from the model for total accidents at signalized four-leg intersections that AADT on major and minor road, the percentage of all incoming truck traffic in peak hours and average absolute percent grade change per 100m along the major and minor road approaches (within 244m of distance from the centre of an intersection) increased the total number of accidents while the presence of major road left turn lane and the percentage of incoming minor traffic turning left during peak hours reduced the total number of accidents. Further, protected left turn lanes on the major road, vertical curves on major and minor road and the number of driveways on the major road were significant.

Injury accident models indicated that the percentage of truck traffic and the percentage of turning movements were significantly associated with the number of injury accidents.

4.5.4 *Conclusion and recommendation*

Increased AADT on major and minor roads, many driveways on the major road, increased incoming major traffic turning left during peak hours, increased % grade change per 100m along major and minor road approaches and total incoming truck traffic in peak hours influenced accident occurrence. The reduction of AADT on major and minor roads, for instance by providing public means such as buses which carry many passengers, may result in a reduced number of vehicles on roads.

Although the above models were used to draw inferences on the number of accidents, the information on how good they are was not provided. Thus, the users of these results need to be cautious with their interpretation. As a suggestion, the deviance and Pearson Chi-square statistics are examples of the tests we can use to assess the fit of these models.

4.6 Study 6

Sayed and Rodriguez (1999) developed Poisson and negative binomial regression models for urban unsignalized but stop-sign controlled intersections. The study was done in

Greater Vancouver Regional District, Vancouver Island and British Columbia. Data were available for the years 1993 to 1995 for each intersection. There were 186 three-arm and 233 four-arm intersections. The number of accidents was recorded for each intersection. An accident was considered to be an intersection accident if it occurred within 30m of the intersection.

4.6.1 Explanatory variables

Data were collected on traffic volumes only. They included annual average daily traffic on major and minor roads.

4.6.2 Model structure and results

Separate models for each intersection type were fitted using multiplicative models in which the number of accidents was assumed to vary around the expected value according to the negative binomial distribution. The model relates the number of accidents to the product of traffic flows entering the intersection raised to a certain power and takes the form below.

$E(\mu_i) = \beta_o * AADT_{minor}^{\beta_1} * AADT_{major}^{\beta_2}$, with

$E(\mu_i)$ = expected accident frequency at intersection type *i*

$AADT_{major}$ = major road traffic volume

$AADT_{minor}$ = minor road traffic volume

β_j = parameters to be estimated and represent the effect of traffic volume*, j,* on the expected accident frequency at intersection type *i*

The Pearson Chi-square was used to test the fit of the models. For both types of intersections, the Pearson Chi-square statistics (205 and 214 for three-arm intersections) and (246 and 265 for four-arm intersections) were smaller than the tabulated ones. This implied that the null hypotheses stating that the models fit the data could not be rejected. Both major and minor road traffic volume increased the expected accident frequency at three-arm and four-arm intersections.

4.7 Study 7

Kulmala (1995) studied the safety of rural intersections on Finnish main roads using Poisson and negative binomial regression models. There were 915 three-arm and 847 four-arm intersections on single carriageway rural roads. A total of 1749 accidents in the three-arm intersections, of which 566 were injury accidents (912 victims) and 2,325 accidents in the four-arm intersections, of which 826 were injury accidents (1,325 victims) were studied. Traffic and police-reported accidents were collected from 1983 to 1987. During this period, the operational and geometric characteristics of the selected intersections were not subject to change or modification. The intersection area was defined as the 200m length of major road adjacent to the centre of the intersecting alignments and the 100m length of the minor road.

Response variables included the number of all injury accidents, single accidents, crossing accidents, turning accidents, rear-end accidents, injury accidents involving motor vehicles only, injury accidents involving unprotected road users and accident victims.

4.7.1 Explanatory variables

Explanatory variables were collected on geometric characteristics, traffic control and traffic volumes. Table 8 below presents the explanatory variables in each category and their effects on the frequency of accidents. The author did not distinguish between which variables were considered for the three-arm or four-arm intersections. However, the significant variables for each type of intersection were mentioned.

Table 8: Explanatory variables in each category

Category of explanatory variables	**Variable levels**	**Effect**	**Causes more accidents**
Three-arm intersections			
Traffic flow			
Total number of vehicles	continuous	+	Number of increased vehicles
Proportion of traffic from minor road		+	Increased volumes
Traffic control			
Speed limits on major and minor roads	Continuous (km/hr)	+	High speed >70km/hr
Geometric design features			
Minor road descends to intersection	Yes = 1, no = 0	+	Presence of minor road descends to intersection
Minor approach width	Continuous (m)	+	Wider approach width ≥15m
Intersection on hilltop along major road	Yes = 1, no = 0	+	Presence of intersection on hilltop along major road
Existence of right-turn lane on major road	Yes = 1, no = 0	-	Absence of right-turn lane on major road
Sight distance from the minor road	Continuous (m)	+	Short sight distance <100m
Existence of pedestrian facility on minor road	Yes=1, no = 0	+	Presence of a pedestrian facility on minor road
Four-arm intersections			
Four-arm intersections			
Traffic flow			
Total number of vehicles	continuous	+	Increased vehicles
Proportion of traffic from minor road		+	Increased volumes
Traffic control			
No variable was significant			
Geometric design features			
Presence of traffic islands	Yes = 1, no = 0	-	Absence of traffic islands
Minor approach width	Continuous (m)	+	Wider approach width ≥15m
Intersection on hilltop along major road	Yes = 1, no = 0	+	Presence of intersection on hilltop along major road
Existence of right-turn lane on major road	Yes = 1, no = 0	-	Absence of right-turn lane on major road
Sight distance from the minor road	Continuous (m)	+	Short sight distance <100m
Existence of painted island on the major road	Yes = 1, no = 0	-	Absence of painted island on the major road
Curve on minor road before intersection	Yes=1, no = 0	+	Presence of curve on minor road before junction

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

4.7.2 Model structure

The safety of rural intersections on Finnish main roads was studied using Poisson and negative binomial regression models but only results from the negative binomial models were presented. Multiplicative models were fitted to the data and were of the form below:

$E(\mu_i) = \beta_o * AADT_{\min or}^{\beta_1} * TT_i^{\beta_2} * e^{\beta_1 x_{i1} + \ldots + \beta_j x_{ij}}$ with

$E(\mu_i)$ = expected accident frequency on intersection *i*, for a given period

TT_i = total number of vehicles (millions) entering an intersection, i, in a given period

$AADT_{\min or}$ = traffic on the minor road, expressed as a percentage of total intersection flow

x_{ij} = vector of explanatory variables, j, describing the geometry, environment and operation of intersection *i*

β_j = parameters to be estimated and represent the effect of the geometric, environmental and operational characteristics, j, at intersection i on the expected accident frequency

4.7.3 Results

In this section we present the results from three-arm and four-arm intersections. First, results from three-arm intersections will be presented and results from four-arm intersections will follow.

4.7.3.1 Three-arm intersection analysis

Results from three-arm intersections showed that the variables total number of vehicles, proportion of traffic from minor road, intersection on hilltop along major road, minor approach width ≥ 15m, speed limit on minor road > 70km/h, speed limit on major road > 70km/h, minor road descends to intersection, sight distance from minor road at < 100m and existence of pedestrian facility on minor road increased the number of injury accidents. On the contrary, the existence of a right-turn lane on the major road had a negative effect on the number of injury accidents at three-arm intersections.

4.7.3.2 Four-arm intersection analysis

Findings from four-arm intersections indicated that the total number of vehicles, proportion of traffic from minor road, intersection on hilltop along major road, minor approach width ≥ 15m, sight distance from minor road at < 100m and curve on the minor road before intersection (straight, bend) increased the number of injury accidents. In contrast, the existence of painted island on the major road and existence of a right-turn lane on the major road reduced the number of injury accidents.

4.7.4 Conclusion and recommendation

Results from this study showed the importance of minor road traffic operational characteristics on intersection safety. Two conditions closely related to high accident rates at intersections were revealed: too high approaching speeds and very short sight distances before the intersection on the minor road.

The study lacked data on turning movements and pedestrians. This resulted in insufficient disaggregation of exposure data. It is recommended to collect data on turning movements and pedestrians for future research.

Secondly, it is recommended to examine hundreds of road elements, intersections in this case, in order to detect systematic patterns in the road safety picture. Further more if the amount of data is not large enough, the standard errors of the parameter estimates will be too large for variables to be statistically significant.

While the results from this study give an insight into which variables have an effect on the expected accident frequency, it was not mentioned how good the fitted models were. Hence, the quality of the results should not be overestimated. The deviance and Pearson Chi-square tests could be utilized as measures for goodness of fit.

4.8 Study 8

Kumara and Chin (2003) used the zero-inflated negative binomial (ZINB) model to identify variables affecting accident occurrence at signalized T-intersections in Singapore while taking into account the presence of excess zeros. A total of 104 three-legged signalized tee intersections over a period of 9 years (1992 - 2000) were used to develop the model.

4.8.1 Explanatory variables

Data were collected on traffic volume, traffic control measures and geometric design elements. Table 9 below contains explanatory variables in each category and their effect on the occurrence of accidents.

Table 9: Explanatory variables in each category

Category of explanatory variables	**Variable levels**	**Effect**	**Causes more accidents**
Traffic flow			
Total approach volume in AADT	continuous	+	Increased volumes
Total left-turn volume in AADT	continuous	+	Increased volumes
Right-turn volume in AADT	continuous	/	
Traffic control			
Surveillance camera	Yes = 1, no = 0	/	
Median railings	Yes = 1, no = 0	-	Absence of median railings
Number of signal phases per cycle	continuous	+	Larger number of signal phases per cycle
Permissive right turn	Yes = 1, no = 0	+	Presence of permissive right turn
Geometric design features			
Sight distance (m)	>100m = 1, else = 0	-	< 100m
Existence of 5% gradient	Yes = 1, no = 0	+	5% gradient
Number of approach lanes	continuous	/	
Median width	continuous(m)	/	
Horizontal curve	Yes = 1, no = 0	+	Horizontal curve
Right-turn channelization	Yes = 1, no = 0	-	Absence of right-turn channelization
Uncontrolled left-turn channelization	Yes = 1, no = 0	/	
Uncontrolled left-turn slip road	Yes = 1, no = 0	+	Uncontrolled left-turn slip road
Length of left-turn slip road	continuous (m)	/	
Acceleration section on left-turn lane	Yes = 1, no = 0	-	Absence of acceleration section on left-turn lane
More than 5% approach gradient	Yes = 1, no = 0	-	Grades ≥ 8%

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

4.8.2 Model development

Zero inflated models are based on the assumption that there are two states for the accident generating process: a normal state corresponding to the usual assumption of a constant expected number of accidents per unit of time and a safe state in which accidents will not occur (Reurings et al., 2005). The resulting probability distribution for the number of accidents will be a distribution of the Poisson family (the negative binomial distribution) and a distribution containing zero outcomes only (no accidents recorded). Then, the resulting probability distribution will contain an excessive number of zeros compared to the standard Poisson or negative binomial distribution. This motivates the use of two distributions: a standard count distribution and a distribution for the zero case. In the case of accident studies, this dual state system can be expressed as a probability function by assuming q_i as the probability that intersection approach i will exist in the zero accident state and $(1-q_i)$ as the probability that a zero accident observation actually follows a suitable count distribution such as Poisson or negative binomial. Therefore,

$$P(n_i) = \begin{cases} q_i + (1-q_i)R_i(0) \ldots\ldots & n_i = 0 \\ (1-q_i)R_i(n_i) \ldots\ldots & n_i = 1,2,\ldots \end{cases}$$

With

n_i = the annual number of recorded accidents on intersection approach *i*

R_i = the probability mass function of occurrence of n_i accidents corresponding to the standard distribution to be modified by the zero state

$q_i = \dfrac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}$ with

β = vector of regression coefficients

x_i = vector of explanatory variables describing the characteristics of the approach *i*

In the zero inflated negative binomial (ZINB) models, R_i corresponds to the negative binomial distribution with mean λ_i, that is,

$R_i(n_i) = (\Gamma(\theta+n_i)/[n_i!\Gamma(\theta)])^* u_r^{\theta}\ [1-u_i]^{n_i}$ with

$\theta = \dfrac{1}{\alpha}$, $u_i = \dfrac{\theta}{[\theta+\lambda_i]}$ and α the overdispersion parameter

Zero-inflated models can be formulated in one of two ways either assuming a single parameter vector for q_i and λ_i or separate parameter vectors. For this study, only the former approach is presented (see Lambert, 1992, for more extensive literature).

From $p(n_i)$ the model parameters β, α and τ are estimated. The probability density function for the observed random variable n_i is $p(n_i) = (1-q_i)R_i(n_i) + z_i q_i$ with $z_i = 1$ when n_i is observed to be zero and $z_i = 0$ for all other values of n_i. The indicator variable z_i eases the maximization of the loglikelihood function given by $L(\beta,\alpha,\tau) = \sum_{i=1}^{N} \log p(n_i)$. Since the negative binomial distribution is specified, the maximum likelihood method is used for parameter estimation. The overdispersion parameter α was found to be statistically significant, which seems to imply at first glance that the negative binomial distribution is appropriate. However, when tested against the alternative ZINB, (τ), by the Vuong statistic , it could be concluded that the ZINB model was a better representative model than the negative binomial model in detecting excess zeros after controlling for overdispersion (Kumara and Chin, 2003).

The Vuong statistic distinguishes between the overdispersion of the negative binomial model and the force of the splitting mechanism in the ZINB part of the model (Vuong, 1989). The Vuong statistic is distributed as a standard normal distribution, hence, its value is compared to the critical value for standard normal distribution. Therefore it is taken that V > 1.96 distinctly favors the zero-inflated count models while $V \leq -1.96$ distinctly favors the parent negative binomial model otherwise the test is indecisive. The selection of the relevant model is therefore established by evaluating the Vuong statistic, V and the *t*- statistic for the negative binomial overdispersion parameter α as shown in Table 10 (Shankar et al., 1997). The complementary condition of selecting the zero-inflated Poisson or ZINB model is that the shape parameter τ must be statistically significant.

Table 10: Decision rule for model selection using the Vuong statistic and negative binomial overdispersion parameter criteria

	***t* Statistic of α**	
Vuong statistic	**<2**	**>2**
V < -1.96	Poisson	NB
V > 1.96	ZIP models	ZINB models

NB = negative binomial, ZINB = zeroinflated negative binomial, ZIP = zero-inflated Poisson

The primary advantage is that Vuong's statistic makes use of information about the entire distribution, not just the zero outcomes. However, this statistic makes probabilistic statements about two models. It tests the null hypothesis that the two models are as close to the actual model against the alternative that one model is closer. It cannot make any decision whether the closer model is the true model. Other methods have been used for testing standard Poisson and negative binomial regression models against zero inflated alternatives. For example, Ridout et al. (2001) applied the score and log-likelihood ratio tests.

4.8.3 Results

Results indicated that increased total approach volume, increased left-turn volume, existence of a horizontal curve, permissive right-turn phase, short sight distances, large number of signal phases and uncontrolled left-turn slip road increase accident occurrence. On the other hand however, right-turn channelization, acceleration section on the left-turn lane, median railings and more than 5% approach gradient reduce the accident occurrence.

Increased volumes imply greater interaction between vehicles and perhaps more conflicts. Furthermore, as volumes increase, there are fewer gaps in traffic for right-turning as well as left-turning drivers. This results in increased accidents due to greater exposure. The existence of a horizontal curve increased accident occurrence. Another study by Poch and Mannering (1996) also indicated that the horizontal curve on an intersection approach would increase all types of accidents. The permissive right-turn phase also increased accident occurrence. Permissive right-turn phasing makes right-turning vehicles proceed with through vehicles when there is a gap. This increases the likelihood of collisions resulting from the right-turning vehicles failing to give way to oncoming vehicles. Shorter sight distance (shorter than 100 m) increased accident frequency. Short sight distances reduce the drivers' abilities to judge the traffic conditions at the intersection. In his study about safety at rural three- and four-arm junctions, Kulmala (1995) also concluded that the short sight distances increase accident occurrence. The results indicated that accident occurrence increases with more phases per cycle. In general the number of phases is higher for busy intersections with more conflicting demands on the intersection (Poch & Mannering, 1996). The result is not surprising since most accidents occur during the phase-change period.

The right-turn channelization, acceleration section on the left-turn lane, median railings and more than 5% approach gradient were found to reduce accident occurrence.

The provision of right-turn slip road may reduce accident occurrence and increase the zero accident state. This exclusive lane lowers speed among approaching vehicles; hence, a possible reduction in conflicts. In their study of intersection accident frequencies, Poch and Mannering (1996) concluded that, without an exclusive lane, right-turning vehicles that are required to slow in a lane shared with speed-maintaining vehicles may cause a speed differential that tends to cause rear-end accidents. The provision of an acceleration section in the left-turn lane reduced accident occurrence as it enables drivers to merge more easily and reduce accidents. There was evidence that the presence of a median railing reduced accident occurrence. Median railings are installed to prevent pedestrians from crossing the road where there are no designated pedestrian crossings. Although grades with 8% and above are known to be more hazardous (Thagesen, 1996), grades from 5% to 8% may be safer because of reduced speed. In addition, there are improvement factors such as signs and markings at these gradients. All approach gradients at the intersections considered in the current study were less than 8%.

The reduction in accident occurrence over time is attributed to engineering advances in vehicles, roads, lighting systems, more advanced driver training and vehicle checking systems that have been introduced in Singapore over the years.

4.8.4 Conclusion and recommendation

The purpose of this study was to identify the factors affecting road accident occurrence at signalized tee intersection approaches in Singapore, taking into account cases of records of no accidents. The results indicated the significance of several highway geometric characteristics, traffic control measures and traffic characteristics and also the effects of excess zeros on accident modeling. It was found that the presence of right-turn channelization, an acceleration section on the left-turning lane, median railing and existence of a more than 5% gradient reduce accident occurrence while increased total and left-turn volumes, an uncontrolled left-turn slip road, signal phases per cycle, existence of a horizontal curve and a permissive right-turn phase would increase accident occurrence.

From this study, the provision of right-turn slip road at intersections is recommended since right-turn slip roads reduce accident occurrence and increase the zero accident. Also, it was observed that the uncontrolled left-turn lane, allows left-turning vehicles to merge into the cross traffic stream which may also increase the likelihood of sideswipe and head-to-side accidents since drivers fail to yield to oncoming traffic. However, if an acceleration section is provided in the left-turn lane, drivers may be able to merge more easily and reduce accidents. Hence, acceleration sections are recommended in the left-turn lanes at intersections. Further, median railings were found to reduce accident occurrence. Therefore, the installation of median railings is recommended as one of the most important accident control devices at intersections.

4.9 Summary of Poisson and negative binomial models

The Poisson regression model is used when discrete response variables have counts as possible outcomes, for instance, the number of deaths on motorways in Flanders in a particular year. There is no fixed upper limit for the outcome. Since the outcome must be a nonnegative integer, its distribution should cover a nonnegative range and the simplest of such a distribution is a Poisson distribution. The Poisson distribution is used for counts of events that occur randomly over time or space. A key feature of the Poisson distribution is that its variance equals its mean, which implies that sample counts vary more when their mean is higher. In a standard Poisson model, the variance to mean ratio is one. However, in practice, it is a common phenomenon for this ratio to exceed one. In the given example of death, a standard Poisson model assumes that each person has the same probability of dying in a motorway accident in Flanders in a particular year. In reality, these probabilities vary due to factors such as amount of time spent driving, whether a person wears a seatbelt, etc. Such variation causes death counts to display more variation than predicted by the Poisson model. This phenomenon is called overdispersion and has implications on the results. If the variance-mean ratio exceeds one, the analysis which assumes a standard Poisson model underestimates standard errors and thus wrongly inflates the test statistics and the level of significance. However, in the absence of overdispersion, a Poisson model is a credible choice for count data. We can measure overdispersion by the scaled deviance and scaled Pearson Chi-square statistics. When the values of these statistics are much larger than one, the data are said to exhibit overdispersion. In cases of overdispersed data, the negative binomial is a distribution for count data that permits the variance to exceed the mean (Agresti, 2002). That way, the mean is permitted to depend on explanatory variables.

There are intersections at which zero accidents are recorded on a number of occasions. When there is a zero accident record over a period of time, it may indicate either that the intersection is nearly safe, or that the zero record is a chance occurrence or accidents are not reported. Since the standard Poison and negative binomial models do not help to identify accident contributory factors in this case, it becomes necessary to model the two states. Moreover if the two states are modelled as a single state, the estimated models may be biased as there may be an overrepresentation of zero accidents. Hence the presence of excess zeros in the accident count data may be mistakenly regarded as the presence of overdispersion in the data set, which arises because of an incorrectly specified model. To handle count data with excess zeros, the zero-inflated negative binomial model was employed (Kumara and Chin, 2003).

The results from all models indicated that atleast one variable in each category of traffic flow, traffic control, geometric, environmental and land use characteristics has a significant effect on accident occurrence. Thus, one can conclude that all these categories should be given attention in accident risk analysis at intersections. Although eight studies have been used to illustrate these techniques, the Poisson and negative binomial regression models are still applied in other studies, for example Wong et al. (2007) and Wang and Abdel-Aty (2006).

Despite the desirable properties of negative binomial regression models, they presuppose that the accident counts at any intersection are independent. However, as the data contain location-specific effects and are likely to be serially correlated, it is suggested to consider techniques which adjust for variations in accident counts due to locations. Such techniques are random effects models (Kim et al., 2007) presented in the following chapter.

5. RANDOM EFFECTS MODELS

Models presented in the preceding sections (multiple logistic regression, multiple linear regression, negative binomial and Poisson regression) assume independent residuals across the number of accidents. These models are to some extent problematic to estimate when the data structure is characterized by correlated responses within clusters (intersections). The correlation within clusters violates the assumption of residual independence made by earlier statistical methods. Due to serial correlation in the accident data, non-hierarchical models seem to be inappropriate since accident data variables are likely to have location specific effects. Further, if significant correlation within clusters is not modelled, the consequence is attenuation of effects (parameter estimates tend toward zero), biased parameter estimates, under-estimated standard errors and incorrect statistical inferences. To overcome these problems, a more suitable alternative is random effects models which account for correlation within clusters by introducing random effects in the population based models (Kim et al., 2007). As a result, we describe random effects models in this section.

In this chapter, three studies in which the random effects models have been applied are discussed. In the first study, we study the associations between different types of crashes and, crash-level characteristics and intersection-level characteristics. In the second study, the relationships between accident occurrence and, traffic volume, traffic control and geometric characteristics of an intersection will be investigated. The third study describes a hierarchical binomial logistic model to identify the significant factors affecting the severity level of driver injury and vehicle damage in traffic crashes at urban signalized intersections in Singapore.

5.1 Study 1

Kim et al. (2007) fitted a random effects binomial logistic regression model (random intercept model) on intersection accident data in the state of Georgia to identify variables that affect the probability that certain types of crashes will occur by exploiting the hierarchical structure of intersection crashes. There were 548 motor vehicle crashes collected from 91 two-lane rural intersections from 38 counties in the state of Georgia for 2 years (1996–1997). Crashes were coded as intersection related when they occurred within 76m (250 ft) of an intersection. Crash prediction models were estimated for angle, head-on, rear-end and sideswipe (both same direction and opposite direction) crashes. The crashes represent the lower level of the hierarchy, while the intersection at which the crash occurred represents the higher-level of hierarchy or cluster. All dependent variables were dummies and are presented in Table 11 below.

Table 11: Dependent variables

Dependent variable	Levels of dependent variables
Angle	1 if angle crash, 0 otherwise
Head on	1 if head-on crash, 0 otherwise
Rear-end	1 if rear-end crash, 0 otherwise
Side same direction	1 if sideswipe crash (same direction), 0 otherwise
Side opposite direction	1 if sideswipe(opposite direction) crash, 0 otherwise

5.1.1 Explanatory variables

Data on crash-level and intersection-level characteristics were collected. The crash-level characteristics included a clear weather indicator, a surface condition indicator, a daylight indicator, a curve indicator and a grade indicator. The intersection-level characteristics comprised of a shoulder indicator, a signal indicator, a driveway indicator and an intersection angle indicator. The variables collected in each category and the directions of their effects are presented in Table 12 below.

Table 12: Explanatory variables in each category

Category of explanatory variables	Variable levels	Effect	Causes more accidents
Angle crash model			
Clear weather indicator	1 if crash occurred during clear weather condition, 0 otherwise	+	Clear weather
Surface condition indicator	1 if crash occurred on a wet road-surface, 0 otherwise	-	Not wet surface condition
Daylight indicator	1 if crash occurred during daylight, 0 otherwise	+	Daylight
Curve indicator	1 if crash occurred on a horizontal curve, 0 otherwise	+	Horizontal curve
Signal indicator	1 if signalized intersection, 0 unsignalized intersection	-	Unsignalized intersection
Rear-end crash model			
Daylight indicator	1 if crash occurred during daylight, 0 otherwise	+	Daylight
Signal indicator	1 if signalized intersection, 0 unsignalized intersection	+	Signalized intersection
Sideswipe same direction model			
Curve indicator	1 if crash occurred on a horizontal curve, 0 otherwise	-	Non horizontal curve
Shoulder indicator	1 if shoulder exists on either major or minor roads, 0 otherwise	-	Absence of shoulder exists on either major or minor roads
Intersection angle indicator	1 if the degree of intersection angle is 90, 0 otherwise	+	Intersection angle = 90
Sideswipe opposite direction model			
Clear weather indicator	1 if crash occurred during clear weather condition, 0 otherwise	-	Not clear weather
Surface condition indicator	1 if crash occurred on a wet road-surface, 0 otherwise	-	Not wet surface condition
Daylight indicator	1 if crash occurred during daylight, 0 otherwise	-	Night time
Curve indicator	1 if crash occurred on a horizontal curve, 0 otherwise	-	Non horizontal curve
Grade indicator	1 if crash occurred on a vertical curve, 0 otherwise	-	Non vertical curve

Table 13 below provides summary statistics of the intersection crash data. A total of 548 crashes represent the lower level of hierarchy: 274 (50%) occurred at unsignalized intersections and 274 (50%) occurred at signalized intersections while 91 rural intersections represent the higher level of hierarchy: 64 unsignalized intersections and 27 signalized intersections of two-lane roads.

Table 13: Summary statistics of the intersection crash data

		Signalized intersections		Unsignalized intersections	
Variable	**Total**	**Number**	**%**	**Number**	**%**
Number of crashes	548	274	50.0	274	50.0
Number of intersections	91	27	32.3	64	67.7
Dependent variables					
Angle	239	107	19.5	132	24.1
Headon	16	8	1.5	8	1.5
Rearend	144	104	19.0	40	7.3
Sidesame	33	19	3.5	14	2.5
Sideopposite	20	8	1.4	12	2.2
Crash-level characteristics					
Clear	333	177	32.3	156	28.5
Surface	116	49	8.9	67	12.2
Daylight	424	229	41.8	195	35.6
Curve	504	263	48.0	241	44.0
Grade	208	77	14.1	131	23.9
Intersection-level characteristics					
Shoulder	67	22	24.2	45	49.4
Signal	27	27	29.7	0	0.0
Driveway	26	14	15.4	12	13.2
Intersection angle	12	10	11.0	2	2.2

5.1.2 Model structure

Five separate random intercept binomial logistic models were estimated: one for each crash type. The models fitted to the data are well described in Guo and Zhao (2000) and are of the form below:

$$\ln\left(\pi_{ij} / (1-\pi_{ij})\right) = \beta_{00} + \sum_{q=1}^{Q} \beta_{0q} w_{qj} + \sum_{p=1}^{P} \beta_{p0} x_{pij} + u_{0j}$$, with

π_{ij} = probability that a type of crash *i* will occur at intersection *j*

β_{00} = intercept

w_{qi} = vector of intersection-level characteristic *q* at intersection *j*

x_{pij} = vector of crash-level characteristic *p* at crash level *i* at intersection *j*

β_{0q} = regression coefficients associated with intersection-level characteristics

β_{p0} = regression coefficients associated with crash-level characteristics

u_{0j} = random effect due to intersection level *j*

$u_{0j} \sim N\left(0, \sigma_u^2\right)$

5.1.3 Results

The findings obtained are presented in this order: first, results from head-on crash model, followed by results from angle crash model, rear-end crash model, sideswipe same direction model and lastly the sideswipe opposite direction model.

5.1.3.1 Head-on crash model

In the head-on crash model no statistically significant variables were discovered. This is attributed to the small number of head-on crashes in the population of crashes (16 head-on crashes). This was too small to yield statistically meaningful results.

5.1.3.2 Angle crash model

Results from the angle crash model revealed that angle crashes were more likely to occur during clear weather conditions relative to other weather conditions such as rainy, snowy days, more likely to occur during the daytime compared to the night-time and were more likely to occur on horizontal curves compared to straight sections. More angle crashes during daylight might be due to higher traffic volumes especially during peak hours compared to other times of the day. Angle crashes were less likely on wet road surface and at intersections with signals. The random intercept was significant, implying that there was evidence of variation in angle crashes between intersections.

The effects of clear weather conditions and daytime are likely to be capturing exposure since greater numbers of vehicles pass through the intersections during daytime and clear weather. The horizontal curve effect is due to the restricted sight distance. Angle crashes in this study represent vehicles turning in the intersection (before or after the intersection into driveways) or vehicles entering near the intersection (with horizontal curves) from driveways and thus restricted sight distance. On wet road surfaces, drivers tend to reduce speeds and drive carefully which reduces crash occurrence.

5.1.3.3 Rear-end crash model

Rear-end crashes were more likely during daytime and at intersections with signals. The significant random intercept indicated that variation in rear-end crashes exists between intersections.

The probability of rear-end crashes is significantly greater at signalized intersections compared to unsignalized intersections. This finding is consistent with the results of a previous study (Greibe, 2003). Wang et al. (2003) suggested that this might be due to the combination of the leading vehicle’s unexpected deceleration by a signal and the ineffective response of the following vehicle’s driver to this deceleration. Another explanation would be sight restrictions caused by large leading trucks, buses, etc. which result in rear-end collisions from a sudden stopping of lead vehicles.

5.1.3.4 Sideswipe same direction model

The sideswipe same direction model revealed that crashes occurring on horizontal curves (near an intersection) and intersections with shoulders were less likely to result in sideswipe same direction crashes. The likely reason is that same direction sideswipe crashes are associated with lane changing, which may arise from avoidance of conflicts in a busy intersection such as sudden right- and left-turning movements by nearby vehicle drivers. Further, intersections with shoulders provide more room for vehicle collision avoidance. In contrast, same direction sideswipe crashes were more likely to occur at right-angled intersections (i.e. the angle of intersection is 90°) than at skewed intersections. This is probably due to higher volume intersections having less probability of being skewed and thus greater exposure. The random intercept was significant, indicating variation in sideswipe same direction crashes between intersections. The remaining variables were not significant.

5.1.3.5 Sideswipe opposite direction model

The sideswipe opposite direction model showed that all crash level characteristics were statistically significant variables and were less likely to result in sideswipe opposite direction crashes. On the other hand, none of the intersection-level characteristics was significant. The random intercept was significant, indicating variation in sideswipe opposite direction crashes between intersections.

All of the crash-level characteristics revealed negative relationships sideswipe with opposite direction crashes. Sideswipe opposite direction crashes were less likely to occur during daytime and clear weather conditions. It is assumed that drivers can determine potential conflicts during the daytime more easily, that is, an opposing vehicle’s trajectory is more readily determined and can be constantly tracked. At night-time, in contrast, vehicle headlights prohibit the constant tracking of opposing vehicle trajectories and so conflicts are reasonably more likely (compared to daylight conditions and compared to same-direction sideswipe crashes). In addition, crashes occurring on wet roads, horizontal curves and vertical curves were less likely to be involved in sideswipe opposite direction crashes. This might be due to the fact that drivers travelling on wet roads, horizontal curves, and vertical curves are forced to decelerate their speeds by posted speed limit signs or for safety.

5.1.4 Conclusion and recommendation

The findings of this study suggest that all types of crashes at rural intersections in Georgia were indeed hierarchical in structure. This was observed from the significant random intercepts. However, it is not mentioned how well the models fitted the data. This means that users of the findings from this study need to be cautious with the interpretations since the quality of these models is not known. However, with respect to the objective of this review, the results throw light on which variables have a significant effect on the number of accidents while taking location specific effects in account. To have an idea of the goodness of fit, we suggest comparing the plots of the observed and predicted number of accidents (Reurings et al., 2005).

Angle crashes were more likely to occur during clear weather conditions, daytime and on horizontal curves compared. Angle crashes were less likely on wet road surface and at intersections with signals. Rear-end crashes were more likely during daytime and at intersections with signals. The horizontal curves and intersections with shoulders were less likely to result in sideswipe same direction crashes and more likely to occur at right-angled intersections than at skewed intersections. Sideswipe opposite direction crashes were less likely to occur during daytime, clear weather conditions, on wet roads, on horizontal curves and on vertical curves.

There is need to introduce the hierarchical structure of the data in future investigations since random intercepts suggest significant variations in the probability of specific types of crashes occurring at intersections. Also, it is recommended to compare effects for crash level and intersection level characteristics across crash types rather than assessing the effects independently since exposure is constant across crash types.

In addition to variables employed in this study, it is believed that particular types of crash outcome probabilities may also be associated with personal characteristics (e.g. driver attentiveness, reaction times, vision, aggressiveness, etc.) and vehicle characteristics (braking characteristics, mass, steering characteristics, condition of tires, etc.). Thus, including these variables into the models may improve the accuracy of the prediction models (Kim et al., 2007).

5.2 Study 2

Chin and Quddus (2003) applied a random effects negative binomial regression model to investigate the relationship between accident occurrences and, geometric, traffic and control characteristics of signalized intersections in the Southwestern part of Singapore. Accident data from 1992 to 1999 were used in the analysis. They studied a total of 52 four-legged intersections and 3000 accidents in which 90 were fatal and 150 accidents with serious injuries. The response was the total number of annual accidents on each road.

5.2.1 Explanatory variables

Data were collected with respect to traffic volumes, geometric elements and regulatory control measures. The variables considered in each category and their effects are provided in Table 14 below.

Table 14: Explanatory variables in each category

Category of explanatory variables	Variable levels	Effect	Causes more accidents
Traffic flow			
Total approach volume from loop detectors	Thousands of vehicles (1.416-53.84)	+	Increased total approach volume from loop detectors volume
Right-turn volume from loop detectors	Thousands of vehicles (0.5-28.78)	+	Increased right-turn volume from loop detectors
Cycle time	Seconds	*	
Red-time in pedestrian crossing	Seconds	*	
Number of phases per cycle	continuous (2-5)	+	Large number
Traffic control			
Approach speed limit	continuous (km/hr)	*	
Number of bus stops at approach road	continuous (0-4)	+	More stops
Signal control type	Adaptive=1, pre-timed=0	-	Pre-timed signal control
Existence of surveillance camera	yes=1, no=0	+	Existence of surveillance camera
Geometric design features			
Curvature on the approach road	Yes=1, otherwise=0	*	
Approach median width greater than 2m	Yes=1, otherwise=0	+	Wider approach median width >2m
Acceleration section on left turn lane	Yes=1, otherwise=0	-	Absence of acceleration section on left turn lane
Existence of median	yes=1, otherwise=0	*	
Pedestrian refuges	yes=1, otherwise=0	*	
Exclusive right-turn lane	yes=1, otherwise=0	*	
Uncontrolled left-turn lane	yes=1, otherwise=0	+	Uncontrolled left-turn lane
Number of bus bays	continuous (0-4)	-	Fewer bus bays
Approach road width	continuous (7.2-36m)	/	
Distance of pedestrian crossing from the intersection	continuous (64.75-500m)	/	
Distance of downstream bus stop from intersection	continuous (45.73-495.5m)	/	
Distance of upstream bus stop From intersection	continuous (5.36-400m)	/	
Diagonal distance on left-turn slip road	continuous (0-52.5m)	/	
Intersection sight distance	continuous (65-400)m	+	

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

5.2.2 Model structure

A random effects negative binomial model was adopted by introducing a random location-specific effects term into the relationship between the expected number of accidents (μ_{it}) and explanatory variables, X_{it} of an intersection *i* in a given time period, t i.e. $\mu_{it} = \mu_{it}\delta_i$ where δ_i is a random location-specific effects term. To ensure a positive value, the term, μ_{it} can be rewritten as $\mu_{it} = \mu_{it}\delta_i = \exp(\beta X_{it} + u_i)$ where, β is the coefficient vector to be estimated and u_i the random effects across locations.

In this random effects model, the random effect is added to the negative binomial model by assuming that the overdispersion parameter is randomly distributed across groups. This formulation is better able to account for the unobserved heterogeneity across locations and time.

To avoid multicollinearity between highly correlated variables, only the most significant variable was taken from among the highly correlated variables in the model. Multicollinearity biases the standard error of the coefficients and as a result wrong signs or unlikely magnitudes in the coefficients are obtained. For example, the presence of curvature and the intersection sight distance were found to be correlated; the former was dropped as it was less significant than the latter. The random effects negative binomial regression of the total annual accident frequency on the intersection geometric, traffic and regulatory control characteristics was estimated by the maximum likelihood algorithm. All explanatory variables in the data set were put in the model and the *t*-statistic was used to test the hypotheses that $\beta = 0$ for each variable. The insignificant ones were dropped one by one. In addition, odds ratios were calculated to facilitate the interpretation of the variables.

5.2.3 Results

The log-likelihood ratio index and log-likelihood ratio were used to evaluate the model. The log-likelihood ratio index was used to test if the model had a good fit. This index compares the log-likelihoods of the fitted model against that of the zero model. The log-likelihood ratio was used to measure how much the explanatory variables in the model explained the variation in the response variable (Chin and Quddus, 2003). The log-likelihood ratio index was found to be 0.318. This indicates that the model fits well. The value of the log-likelihood ratio was 0.65, implying that the variables in the model explained about 65% of the variations in the number of accidents at the intersections.

Findings indicated that the variables total approach volume in thousand, right-turn volume in thousand, uncontrolled left-turn lane, intersection sight distance (m), median width greater than 2m, number of bus stops, number of phases per cycle and existence of surveillance camera had a positive impact while acceleration section on left-turn lane, number of bus bays and signal control type had a negative impact on the total annual number of accidents on each road. Also, engineering judgments confirmed the validity and practicality of these results.

As total approach volume increases, there are fewer available gaps for the right-turn opposing traffic as well as left-turn merging traffic. As a result of fewer turning opportunities, drivers may be more willing to take risks when making the turn. In the same way, higher right-turn volumes result in more conflicts between right-turn and straight-through vehicles. Several studies, e.g. Kulmala (1995), Abdel-Aty and Radwan (2000), Poch and Mannering (1996), have also indicated that traffic volume increases accident occurrence.

The uncontrolled lane allows left-turn vehicles to merge into the cross traffic stream. This increases the likelihood of accidents, possibly of the sideswipe and head-to-side types. One reason is that the uncontrolled lane gives greater opportunities for drivers to merge into the cross-stream traffic when the latter has right-of-way.

Intersection sight distance, which ranges from 65 to 400m in the sample, is associated with higher accidents. This may seem surprising at first since higher risks are expected with restrictive sight distances, as suggested in studies, such as Poch and Mannering (1996). However, such cases may be confined to very short sight distances, possibly below the observed range in the sample. For the range covered, increases in the sight distances may allow drivers to have greater freedom of manoeuvre and may increase their vehicle speeds thus resulting in possibly greater accident frequencies and severity risks. Kulmala (1995) also found similar results in his study on four-legged intersections.

Approach median widths greater than 2m were associated with high accidents. This is because wider approach median widths allow greater degrees of freedom for right-turning vehicles. Near the stop line, wider median widths may also create more conflicts as the number of conflict points is higher and movements of through vehicles are less channelized.

The presence of bus stops near the intersection was found to increase the total accident frequency at the intersection. The presence of bus stops decreases the width of the approach road, resulting in higher conflicts because standing buses become obstacles to the moving traffic. On the other hand, there is a decrease in accident occurrence when bays are provided at the bus stops. This is because bus bays separate the stationary buses from the other moving vehicles and exposure to accidents is reduced or eliminated.

The number of phases per cycle had the biggest positive effect among the traffic control factors that affect the safety of intersections. Having a higher number of phases per cycle increases the number of accidents. This is not surprising since most accidents occur during the phase change periods. In a similar study, Poch and Mannering (1996) found that eight-phase signal controls increase rear-end and approach turn accidents.

The presence of a surveillance camera along the approach was found to associate with an increase in the total accident frequency on the approach. It may be reasoned that the positive correlation may be because these intersections already have high accident rates even before the cameras were installed. Further, the surveillance camera may increase head-to-rear accidents on the approach due to greater traffic instability. Thus, an overall higher accident rate may have resulted from the presence of the camera.

When an intersection is under signal control, the number of long gaps in the opposing traffic may be reduced and in turn this forces the traffic to follow a more regular discharge pattern with well-defined right-of-way. Consequently, conflicts between traffic streams may be reduced and these are generally limited to periods during phase changes. This may contribute to reducing cross-traffic accidents and appears to support the findings of Poch and Mannering (1996), who found that intersections with signal control are safer than those without signal control.

5.2.4 Conclusion and recommendation

The log-likelihood ratio index used in this study has the undesirable characteristic that for the same data set, it will increase whenever new variables are added to the model. This index is intended to identify a point where adding more variables to the model is not worthwhile. As a solution, the number of variables in the model could be considered through the degrees of freedom to give the adjusted log-likelihood ratio index. This index seeks to find subsets of variables for which it is at maximum or so close to the maximum that adding more variables is not worthwhile (Kutner et al., 2005).

It was found that increased total approach volumes, increased right-turn volumes, the presence of uncontrolled left-turn lane, median widths above 2m, the presence of bus stops, intersection sight distance together with the presence of a surveillance camera and the number of phases per cycle are associated with a higher total accident occurrence. On the other hand, the presence of an acceleration section and the provision of bus bays as well as the use of adaptive signal control reduce accident occurrence.

Longer sections on the left-turn lane for acceleration prior to the merge should be provided. This is because the acceleration section reduces the speed differential between the left-turn merging and cross traffic. This allows vehicles to take up better positions prior to the merge which leads to improvements in safety. Narrower approach median widths (≤ 2m) are recommended. This will reduce the degree of freedom for right-turning vehicles and reduce conflicts near the stop line, thus, a reduction in accident occurrence. Bus bays should be provided at the bus stops since bus bays separate the stationary buses from the other moving vehicles and exposure to accidents is reduced or eliminated. Signal controls should be provided at intersections since they allow the traffic to follow a more regular discharge pattern with well-defined right-of-way. This reduces conflicts between traffic streams and cross-traffic accidents.

5.3 Study 3

Huang et al. (2007) developed a hierarchical binomial logistic model to identify the significant explanatory variables affecting the severity level of driver injury and vehicle damage in traffic crashes at urban signalized intersections. They used Bayesian inference to estimate the effects of the variables. Although these authors claim to have used the Bayesian methods, their application is not clearly explained in their study. They just used the Bayesian confidence intervals to test the significance of fixed effects' parameters. The significance of the fixed effects could also be obtained without applying the Bayesian methods. We feel that a small part of the Bayesian methods were applied and we propose that this study should not be considered to a detailed extent as a Bayesian application.

The study was carried in Singapore, a heavily urbanized island country with an area of about 700 km^2 and 3235 km of roads (in 2005). Crash data used in the analysis were collected from 2003 to 2005. A total of 19,832 crashes were reported in this period. Of these, 4,095 cases occurred at signalized intersections and were used in the analysis. A total of 7,840 driver–vehicle units were involved.

Two categorical severity indicators were of interest and they included driver injury severity and vehicle damage severity. Driver injury severity was coded as: (a) fatal or serious injury, DI(A), (b) slight or no injury, DI(B); and vehicle damage severity was coded as: (a) extensive damage, VD(A), (b) slight or no damage, VD(B). To yield a net effect estimate of each potential factor on individual severity, a binary dependent variable was defined by combining the two severity indicators: (a) DI(A) and VD(A) denoted as IS(A), representing high individual severity (b) otherwise is low individual severity denoted as IS(B). A summary of severity statistics is given for three years in Table 15.

Table 15: Summary of crash severity at signalized intersection by years

Year	DI(A)	DI(B)	% of DI(A)	VD(A)	VD(B)	% of VD(A)	IS(A)	IS(B)	% of IS(A)
2003	39	2622	1.49	491	2170	22.63	508	2153	23.59
2004	37	2885	1.28	398	2524	15.77	412	2510	16.41
2005	36	2221	1.62	173	2084	8.30	192	2065	9.30
Total	112	7728	1.45	1062	6778	15.67	1112	6728	16.53

DI(A): driver with fatal/serious injury; VD(A): vehicle with extensive damage; IS(A): DI(A) and VD(A); DI(B): driver with slight or no injury; VD(B): vehicle with slight or no injury; IS(B): otherwise.

5.3.1 Explanatory variables

Data were collected on geometric features, traffic conditions, driver and vehicle characteristics. A total of 25 variables were coded for each intersection crash. To avoid multi-collinearity as well as wrong signs or implausible magnitudes in the estimated coefficients, a correlation matrix for all variables which were hypothesized to relate to the severity levels was checked. For the highly correlated variables, only the most significant variable was retained in the analysis; for example, weather condition was excluded because of its high correlation with road surface. Finally, a total of 10 explanatory variables in the crash-level were used, i.e. day of week, time of day, intersection type, nature of lane, road surface, street lighting, road speed limit, vehicle movement, presence of red light camera and pedestrian involved.

It should be noted that the correlation coefficients used to check for multicollinearity in this study were misused as these coefficients are used when pairs of explanatory variables are continuous. For discrete variables as is the case here, other tests such as the Fisher's exact and Chi-Square tests should have been used. Further, the ordinality of the variables could have been taken into account, as done by the linear trend test (Agresti, 2002).

To explore how driver–vehicle characteristics affected the severity levels, vehicle type, driver age, driver gender, involvement of offending party and passenger involved were selected. Unfortunately, several vehicle safety features such as airbags and anti-lock brakes are not included in the crash dataset. Although those variables may be important to affect the individual severity, they are not so useful in Singapore since most vehicles were less than 6 years old and are hence equipped with the latest protective features in modern cars. Moreover, the strict compulsory annual inspection on all vehicles to ensure they are road worthy means that these features are in serviceable conditions.

The definitions of the selected variables are presented in Table 16. All these variables were split as groups of dummy variables based on the engineering experiences or existing findings in previous studies. For example, vehicle type was categorized as three groups of two-wheel vehicle, light vehicle and heavy vehicle, since the vehicle weight had been identified relevant to injury severity (Evans and Frick, 1994). Time of the day was classified as: day time; if crash occurred at 10 a.m.–5 p.m., Night time; if crash occurred at 8 p.m.–7 a.m. and Peak time; if crash occurred at 7 a.m.–10 a.m. or 5 p.m.–8 p.m.

Table 16: Explanatory variables and their summary statistics

Variable	Variable levels	Effect	Reduces chances of drivers' severity
Day of week	If crash at weekend = 1, otherwise = 0	/	
Time of day	Night time vs. Day time	+	Night time
	Peak time vs. Day time	-	Day time
Intersection type	X vs. Other types	/	
	T/Y vs. Other types	+	T/Y intersections
Nature of lane	Single lane vs. Centre lane	/	
	Left-most lane vs. Centre lane	/	
	Right-most lane vs. Centre lane	+	Right-most lane
Road surface	If road surface is dry = 0, otherwise = 1	/	
Street lighting	If street lighting is fine = 0, otherwise = 1	-	Bad condition of street lighting
Presence of red light camera	If a red light camera is present = 1, otherwise = 0	+	Red light camera at risky intersections
Pedestrian involved	If passengers involved = 1, otherwise = 0	-	Passengers not involved
Vehicle type	Two-wheel vehicle vs. light vehicle	+	Two-wheel vehicle
	Heavy vehicle vs. light vehicle	-	Light vehicle
Driver age	≤25 vs. 26–45	+	≤25
	46–65 vs. 26–45	/	
	>65 vs. 26–45	+	>65
Involvement of offending party	If driver is likely at fault = 1, otherwise = 0	+	Drivers at fault

+ = Positive effect, - = Negative effect, * = Significant (direction of effect unknown), / = Not significant

5.3.2 Model structure

In this section, the model structure and inference are described. A description of the hierarchical binomial logistic model is given first, this is followed by the Bayesian inference and an assessment of random effects using intra-class correlation coefficient is finally discussed.

5.3.2.1 Hierarchical binomial logistic model

Crash was considered as cluster (level 2) and there were a number of sub-clusters (level 1) per cluster, i.e. driver–vehicle units involved in a crash. For the development of the hierarchical binomial logistic model, a similar methodology was used as in section 5.1 study 1 (Kim et al., 2007). The only difference is that in the present study crashes are at the higher-level of hierarchy (clusters) and driver–vehicle units are at the lower-level of hierarchy (sub-clusters).

5.3.2.2 Bayesian inference

In Bayesian models, given model assumptions and parameters, the likelihood of the observed data are used to modify the prior beliefs of the unknowns, resulting in the updated knowledge summarized in posterior densities. Hence, the distinctions between fixed and random effects disappear since all effects are now considered random. In case prior information for the model unknowns is not known, uninformative priors were assumed for all regression coefficients. These priors were assumed to follow normal distributions, (0, 1000), and the variance of the normally distributed random effects was assumed to follow an inverse Gamma distribution (0.001, 0.001). The model was computed via the Gibbs sampler, a Markov chain Monte Carlo (MCMC) technique (Gilks et al., 1995) using Win BUGS software (Spiegelhalter et al., 2003a). The 95% Bayesian confidence interval (95% BCI) was used to examine the significance of explanatory variables, which provides probability interpretations with normality assumption on unknowns and confidence interval estimations (Gelman et al., 2003). The coefficient estimations were considered significant if the 95% BCIs of their odds ratios did not include "1". In addition, engineering judgment confirmed the validity and practicality of the sign of each variable.

5.3.2.3 Assessment of random effects using intra-class correlation coefficient

An intra-class correlation coefficient ρ (ICC) examines the proportion of specific crash-level variance (level 2) in overall residual variance (Jones and Jorgensen, 2003; Kim et al., 2007). Since the logistic distribution for the individual-level (level 1) residual implies a variance of $\pi^2/3 = 3.29$, it is inferred that for a two-level logistic random intercept model with an intercept variance of $\tau^2{}_{20}$, the ICC for between-crash residual is given by the expression below.

$$\rho = \frac{\tau_{20}^2}{\tau_{20}^2 + \pi^2/3}$$

The ICC is an indicator of the magnitude of the within-crash correlation. A value of ρ close to zero means that there is a very small variation between the different crashes, indicating that an ordinary binomial model may be adequate for the data. On the other hand, a relative large value of ρ implies a favour for hierarchical model.

In this study, the magnitude of the between-crash variance, π_{20}^2 was 1.34. Hence, the ICC was obtained as 0.3. This means that 30% of the unexplained variations in individual severity resulted from between-crash variance, which suggests the use of a hierarchical model structure. If an ordinary binomial model was used, the results would be biased.

5.3.3 Results

The results obtained from the technique applied in section 5.3.2 are explained in this section. The results indicated nine significant variables and they included time of the day, intersection type, nature of lane, street lighting, presence of red light camera, pedestrian involved, vehicle type, driver age and involvement of offending party.

The time of crash occurrence was categorized into three periods, i.e., day time (10 a.m.–5 p.m.), night time (8 p.m.–7 a.m.) and peak time (7 a.m.–10 a.m. or 5 p.m.–8 p.m.). Compared with crash occurrences during day time, crashes which occur at night time had a higher chance of high severity. This finding is consistent with Simoncic (2001) who found that crashes at night were more serious than those during daytime. This may be expected since speeding and alcohol use resulting in higher crash severity are more likely at night. Also, the high probability of high severity in night time is consistent with previous studies for severities of motorcycle crashes (Quddus et al., 2002) and single vehicle crashes (Rifaat and Chin, 2005) in Singapore. Furthermore, chances of high severity during peak time were found to be lower compared to crashes during day time. It can be reasoned that due to the higher traffic volume, the vehicle speeds during peak time are considerably reduced compared to off-peak time, hence resulting in lower crash severity.

It was found that crashes occurring at T/Y type intersections increase the chances of high severity in contrast to other type of intersections. Vehicles merging into the major road from the minor road at T/Y type of intersections have a higher probability to collide with going-through vehicles on the major road, which may result in serious crashes. In addition, a shorter sight distance, commonly associated with T/Y type intersections, may also cause more severe crashes.

Another significant geometric factor was the nature of lane, where the right-most lane (left driving) increased the chance of severe crashes compared to the central lane. This result is consistent with the Khorashadi et al. (2005) who found that for right driving, if the location of collision is on the left lane, the likelihood of injury severity increased. The higher severity risk may be caused by higher speed on right-most lane than on other lanes.

Also, bad street lighting condition increased the chances of severe crash. This result was expected because drivers may have more reaction time and better perception ability on crash risk in good street lighting environments. This finding implies that improving the street lighting can substantially improve the safety condition at intersections.

The presence of red light camera was associated with higher severity level. This may seem surprising compared to findings in many studies in which the red light camera was found to reduce crash frequencies, as well as relieving the crash severity e.g. Huang et al., (2006). Although red light camera itself may not increase the risk of severe crashes, the result from the current study is associated with high risk intersections. In particular, intersections with red light camera may have already been placed in sites with more severe crashes since traffic authorities always install cameras at hazardous intersections. Moreover, this result supports the findings by Chin and Quddus (2003), where the presence of a surveillance camera was found to be associated with an increase in the total crash frequency at intersections.

The involvement of pedestrians reduced the chances of drivers' severity. This is reasonable since pedestrians, are easier to injure than drivers in the collisions. It is also supported by Chang and Wang (2006), who found that pedestrians were more likely to have higher risks of being injured than other types of vehicle drivers in traffic crash.

Vehicle type was classified as two-wheel vehicle, light vehicle and heavy vehicle. It was found that two-wheel vehicles increased while heavy vehicles reduced individual severity compared to light vehicles. The severity risk in two-wheel vehicle (e.g. motorcycles) is expected as two-wheel riders do not have the facility of safety protections that are available in light vehicle (e.g. cars), such as seatbelt, airbag etc. Again the two-wheel riders may be thrown off from the vehicle at the time of collision while in the case of car crashes this may rarely happen. This result agrees with that of Kocklelman and Kweon (2002) who found that riding a motorcycle is causing more severe injury than driving a car.

Driver age was found to be significant. Both the young group (≤ 25 years) and the aged group (> 65 years) increased individual severity. It is likely because young drivers drive more recklessly (Rifaat and Chin, 2005; Kocklelman and Kweon, 2002) while aged drivers have relatively weak risk detecting and reacting abilities. Another reason for young drivers to be involved in severe crashes may be that they represent a large proportion of riders of two-wheel vehicles, which have been proven to be associated with a higher risk of being involved in more severe crashes (Rifaat and Chin, 2005; Quddus et al., 2002). Furthermore, as indicated by Rifaat and Chin (2005), decrease of visual power, deterioration of muscle strength and reaction time may be responsible for the aged drivers to be involved in severe crashes. Yan et al. (2005) also found that drivers under 26 years of age had a higher risk of accident involvement because they are willing to take on risks.

Lastly, the involvement of an offending party significantly affected crash severity. The at-fault driver–vehicle unit had a higher chance of individual severity than the not-at-fault party.

5.3.4 Conclusions and recommendation

In this study, no statistical test on the goodness of fit of the fitted model was provided. However, Huang et al. (2007) suggest that the adequacy of the model can be checked by assessing the underlying assumptions of the hierarchical binomial logistic model. The model is considered fitting if the probability plots do not deviate strongly from normality, if there is a constant variance, if the residuals average to zero or close to zero and if the independence assumption is satisfied. Further, we also suggest the comparison of the observed and predicted accident counts (Reurings et al., 2005).

The estimation of random effects using intra-class correlation showed that 30% of the unexplained variation in severity level resulted from between-crash variance. Therefore, it is useful to account for the severity correlation of driver–vehicle units involved in the same multi-vehicle crashes.

The study identified nine significant variables using 95% Bayesian confidence interval. Among these were time of the day, intersection type, nature of lane, street lighting, presence of red light camera and pedestrian involved.

In particular, it was found that crashes occurring in peak time, in good street-lighting condition and in the case of pedestrians involved were associated with lower severity while those occurring in night time, at T/Y type intersections, on right-most lane and in the presence of red light cameras had higher chances of being severe. Vehicle type, driver age and involvement of offending party were also found to affect severities of driver injury and vehicle damage significantly. Specifically, results indicated that heavy vehicles have a better resistance on serious injury or extensive damage, while two-wheel vehicles, young or aged drivers, with the involvement of an offending party have a higher risk of being severe.

The higher chance of the at-fault driver–vehicle unit to be involved in individual severity provides evidence for educating drivers to keep away from risk-taking manoeuvres. Also, the lower severity of crashes occurring in good street-lighting condition implies that improving the street lighting can substantially improve the safety condition at intersections.

5.4 Summary of random effects models

Random effects models assume that intersection accidents' data are hierarchical in nature, with accident-level and intersection-level hierarchies. Hierarchical structure is basically a statistical description of a data structure that is characterized by correlated responses within hierarchical clusters. Random effects models are justified by the presence of correlation within clusters; otherwise population average based modeling methods are appropriate (traditional logistic regression, negative binomial, Poisson, linear regression, etc.). The hierarchy in these intersection accident data are proposed as follows. The accidents themselves represent the lowest level of the hierarchy, while the intersection at which the accidents occurred represents the higher-level hierarchy or cluster. It is reasonable to claim that correlation exists among accidents occurring at the same intersection, since these accidents may share unobserved or unrecorded characteristics of the intersection. These unobserved factors might include poor pavement condition, low pavement friction, poor reflectivity of road signs or lane striping, excessive distractions at the site, nearby drinking establishments, heavy animal populations, etc. Because of such factors, accident frequencies and types observed at a particular location may be correlated. The correlation within clusters (higher-level units) violates the assumption of residual independence assumed in population average based methods. If significant correlation within clusters is left unchecked, that is, without considering hierarchy, the consequence is biased parameter estimates and biased standard errors (Kim et al., 2007). This leads to misleading conclusions on accident occurrence. To overcome these problems, random effects models are used to capture all unobserved heterogeneity.

From the previous studies, it seems that traffic flow, traffic control, geometric, driver characteristics, vehicle types and environmental characteristics contribute to the risk of accidents at intersections. Thus, it is advantageous that future research of risk analysis at road intersections considers these characteristics.

6. CART

The CART technique is used to group accidents according to the intersection and accident types. This technique involves splitting the data into branches on a tree diagram based on given data. This chapter briefly describes the classification and regression tree technique (CART). One study has been described to illustrate this technique.

6.1 Study 1

Lau et al. (1989) developed a three-level accident prediction model based on a grouping and classifying technique called Classification and Regression Trees (CART) for urban intersections. This methodology included a three-level prediction procedure with a “tree” structure for easy interpretation and applications. It also involved an adjustment procedure to adjust for various reporting levels of property damage only accidents in different police authorities. Table 17 below presents an overview of the application of APMs developed in this study.

Table 17: An overview of the application of accident prediction models

	Models		
	Injury	PDO	Fatal
Adjust under-reporting	N.A	YES	N.A
Level 1 Traffic intensity	Linear relation by least square		N.A
Level 2 Control, Design & Environmental factors	Tree by CART		
Level 3 Individual accident history	Linear combination history + group		

N.A = Not applicable, PDO = Property damage only

Level 1: Generation of the base model
At this level of analysis, only traffic intensity expressed in millions of vehicles entering an intersection from all approaches is required for the injury and PDO accident models. A constant is used for fatal accidents at this level of analysis. The following equation was derived to estimate the forecasted number of injury accidents per year (FIACCYR) at an intersection with the number of vehicles entering an intersection from all approaches per year (MVYR).

FIACCYR = 0.61856 + 0.16911*MVYR.
Results from this equation only provide a crude estimate of the safety of intersections.

Level 2: Grouping intersections by CART
For this analysis, detailed data such as design, control and environmental features of the intersection is required. Estimates from this analysis are refinements of estimates from Level 1 using this detailed data.

The CART program is used to analyse the residuals of the base model of injury and PDO accidents and to group intersections with similar accident patterns. This grouping of intersections is referred to as Level 2 prediction.

Level 3: Adjustment by Accident History
In addition to information from Levels 1 and 2, the individual accident history of an intersection is also required for an analysis done at this level.

Results from Level 3 represent future safety estimates of existing intersections. These estimates are based on the concept of a linear combination of the accident history of an intersection and the accident history of a group of intersections. The following equation is used to predict the level of safety of an individual intersection:

$Z = a * E(m) + (1-\alpha) * x$ with

$$\alpha = \left[\frac{(1+Var(m)}{E(m)} \right]^{-1}$$

$E(m)$ = the expected number of accident
x = the accident count

In summary, the CART technique is used to cluster the number of accidents based on a particular criterion. For example, the technique can be used to obtain the number of accidents for each injury severity, the number of accidents for each intersection type and etc.

7. DISCUSSION AND CONCLUSION

The aim of this report was to review APMs used in literature to identify which variables have a significant effect on accident occurrence so that we can have a starting point for further research. Having identified the significant variables, our next step is the development of an appropriate accident prediction model for road intersections in Flanders. Several techniques have been reviewed including multiple logistic regression, multiple linear regression, Poisson and negative binomial regression models, random effects models and the CART technique. In this chapter, these are discussed in order to derive the most appropriate technique and significant explanatory variables in assessing the safety of road intersections.

The multiple logistic regression technique described in chapter two is used to analyze only accident binary outcomes. However, there are many studies in which accident outcomes are continuous. In such cases, multiple linear regression analysis which describes relationships between continuous outcomes and explanatory variables are more credible. Although multiple linear regression models are used widely in traffic accident studies, they have limitations to describe adequately the random, discrete and nonnegative accident events. These include the presence of undesirable statistical properties, such as the possibility of negative accident counts, and the lack of distributional properties, such as the condition of normally distributed accident occurrence (Chin and Quddus, 2003). For such reasons, there is need to employ techniques which can sufficiently describe discrete and nonnegative accident events. Such techniques include Poisson regression and negative binomial regression.

Since accident occurrences are unavoidably discrete and more likely random events, the Poisson regression models appear to be more suitable than the multiple linear regression models. However, these models have potential problems; one constraint is that the mean must be equal to the variance. If this assumption is not valid, that is, the accident data are significantly overdispersed (the variance is much greater than the mean), the standard errors usually estimated by the maximum likelihood method will be biased and the test statistics derived from the model will be incorrect. This results in incorrect estimation of the likelihood of accident occurrence (Chin and Quddus, 2003).

To solve the problem of overdispersion, the negative binomial distribution has been employed instead of the Poisson. To establish the negative binomial regression model, an overdispersion parameter is introduced into the relationship between the mean and the variance. By relaxing the condition of mean equal to variance, the negative binomial regression models have more desirable properties than the Poisson to describe the relationship between accident occurrence and road characteristics (Chin and Quddus, 2003). However, if there is no overdispersion, the Poisson regression models are also suitable.

There are intersections at which zero accidents are recorded on a number of occasions. When there is a zero accident record over a period of time, it may indicate either that the intersection is nearly safe, or that the zero record is a chance occurrence or accidents are not reported. Since the standard Poison and negative binomial models do not help to identify accident contributory factors in this case, it becomes necessary to model the two states. Moreover if the two states are modelled as a single state, the estimated models may be biased as there may be an overrepresentation of zero accidents. Hence the presence of excess zeros in the accident count data may be mistakenly regarded as the presence of overdispersion in the data set, which arises because of an incorrectly specified model. To handle count data with excess zeros, the zero-inflated negative binomial or Poisson models are employed (Kumara and Chin, 2003). This helps to distinguish safe intersections with little probability of accident occurrence from those with zero accident record due to chance or when accidents are not reported.

Despite the desirable properties of negative binomial regression models, they presuppose that the accident counts at any intersection are independent. Negative binomial models ignore the correlation and treat within-intersection accidents the same as between-intersection accidents, thereby producing biased results. However, as the data contain location-specific effects and are likely to be serially correlated, it is suggested to consider techniques which adjust for variations in accident counts due to locations. Such techniques are random effects models (Kim et al., 2007). In addition, multiple logistic regression, multiple linear regression, negative binomial and Poisson regression models assume independent residuals across units. These models are to some extent problematic to estimate when the data structure is characterized by correlated responses within clusters (intersections). The correlation within clusters violates the assumption of residual independence made by earlier statistical methods. Due to serial correlation in the accident data, non-hierarchical models seem to be inappropriate since accident data variables are likely to have location-specific effects. Further, if significant correlation within clusters is not modelled, the consequence is attenuation of effects (parameter estimates tending toward zero), biased parameter estimates, under estimated standard errors and incorrect statistical inferences. To overcome these problems, a more suitable alternative are random effects models which account for correlation within clusters (Kim et al., 2007). As a result, random effects models are a more plausible choice if data are serially correlated.

Further, random effects models assume that intersection accidents' data are hierarchical in nature, with accident-level and intersection-level hierarchies. Hierarchical structure is basically a statistical description of a data structure that is characterized by correlated responses within hierarchical clusters. The hierarchy in these intersection accident data is proposed as follows. The accidents themselves represent the lowest level of the hierarchy, while the intersection at which the accidents occurred represents the higher-level hierarchy or cluster. It is reasonable to claim that correlation exists among accidents occurring at the same intersection, since these accidents may share unobserved or unrecorded characteristics of the intersection. These unobserved factors might include poor pavement condition, low pavement friction, poor reflectivity of road signs or lane striping, excessive distractions at the site, nearby drinking establishments, heavy animal populations, etc (Chin and Quddus, 2003). Because of such factors, accident frequencies and types observed at a particular location may be correlated. To take into account this correlation, random effects models are used to capture all unobserved heterogeneity (Chin and Quddus, 2003).

The techniques discussed so far are not used to classify accidents according to a particular criterion. This necessitates the use of the CART technique. This is used to classify the number of accidents in different groups based on a given rule, for example, according to intersection type.

Different APMs for different intersection types and accident types have been developed by several authors. Generally, it is recommended that fitting separate models for different intersection types and accident types gives a better fit and description of the data than one model for all intersection types. Provided data on intersection types and accident types are available, it is recommended to fit disaggregated models rather than aggregated models (Reurings et al., 2005; Turner and Nicholson, 1998). For example Vogt (1999) fitted separate models for three-leg intersections and four-leg intersections (chapter 4, section 4.5).

Several authors considered different types of intersections on rural and urban roads. Among the considered intersections on urban roads were four-arm stop controlled, three-arm stop controlled, four-arm signalized intersections, four-arm unsignalized, three-arm unsignalized, four-arm unsignalized but stop controlled, three-arm unsignalized but stop controlled intersections and roundabouts. For rural roads, four-leg, three-leg, signalized, two-lane intersections and roundabouts were studied.

Although similar techniques were applied on rural and urban road intersections, a different model structure was used. Nevertheless, the majority of the models discussed on rural and urban road intersections were of the form given below and based on this report, we would prefer a model for intersections to be of this form.

$\mu_i = \beta_o * Q_{MI}^{\beta_1} * Q_{MA}^{\beta_2} * e^{\sum \beta_i x_i}$ with

μ_i = expected number of accidents at intersection type *i*

Q_{MA} = number of vehicles entering an intersection from the major road

Q_{MI} = number of vehicles entering an intersection from the minor road

x_i = vector of values of risk factors, i, other than number of vehicles

β_o = intercept

β_1, β_2 = effect of traffic volume on the expected number of accidents and is modelled as elasticity

β_i = parameters to be estimated and represent the effect of risk factors, *i*, on the expected number of accidents other than traffic volume

The elasticity shows the percentage change in the expected number of accidents associated with a 1% change in traffic volume. The effects of risk factors that influence the probability of accidents given exposure are modelled as an exponential function, that is as e (the base of natural logarithms) raised to the sum of the product of coefficients, β_i, and values of the explanatory variables, x_i, denoting risk factors. The choice of an exponential form is logical in the view of the characteristics of the Poisson distribution since accident counts are positive and rare events at intersections (Reurings et al., 2005). The preferred form of the model is population average based but it can be extended to a hierarchical form if random effects are believed to provide valuable information. This can be done by incorporating a random intercept in the model or a random slope if the study is longitudinal.

Moreover, the choice of the model depends on the nature of the response and the objective of the research. If interest is in making inference on the entire population, population average based models (chapters 2, 3 and 4) are suitable. In contrast, researchers interested in location specific inference would opt for random effects models (chapter 5). Researchers who wish to group accidents in different types the CART technique is a plausible choice.

Different methods were used to assess the goodness of fit of the reviewed APMs models. They include the deviance statistic, the Pearson Chi-square statistic, the Freeman-Tukey index, the adjusted R-square, log-likelihood ratio index and log-likelihood ratio. The deviance and the Pearson Chi-square statistics compare the fitted model versus a saturated model. The degree of freedom for these statistics is the difference between the parameters in the saturated model and the fitted model. The fitted model fits the data if these statistics are greater than the values from the Chi-square tables (Agresti, 2002). The log-likelihood ratio index was used to test if the model had a good fit. This index compares the log-likelihoods of the fitted model against that of the zero model. The log-likelihood ratio was used to measure how much the explanatory variables in the model explained the variation in the response variable (Chin and Quddus, 2003). The Freeman-Tukey index and the adjusted R-square are used to describe how much of the variation in the response is explained by the explanatory variables in the model. It is hard to compare the goodness of fit of the discussed models because different measures are used and the fitted models had different objectives. Further, the tests for some models were not mentioned. However, all models for which goodness of fit tests were mentioned fitted well. We would like to comment that tests for goodness of fit are vital and should be done and reported for all models since they provide information on how good the fitted model is and the quality of the results.

There are several possibilities for explanatory variables. The variables annual average daily traffic (AADT) on minor roads, AADT on major roads, total vehicle counts and pedestrians crossing all arms (AADT), lighting and signal timing were used in almost all models and were statistically significant. Therefore, it is desirable that APMs for intersections include these variables. Generally traffic flow, traffic control, geometric characteristics, driver characteristics, land use and, vehicle types and features were statistically significant.

The selection of explanatory variables appears to depend on data availability. A first step is to include variables that have been found in previous studies to exert a major influence on the number of accidents, are not very highly correlated with other explanatory variables included and can be measured in a valid and reliable way (Reurings et al., 2005).

It appears that in majority of the reviewed studies, the explanatory variables were included in the models without a variable selection procedure. This implies that the selection of the variables is done on a subjective basis which might lead to biased results. Thus, for future research, we would like to recommend the use of variable selection procedures to minimize such bias and misleading results. A number of variable selection procedures are used including adjusted R-square, Mallow's Cp, the prediction sum of squares (PRESS), etc. (Kutner et al., 2005), the likelihood ratio test, the Wald test, etc. (Agresti, 2005) and many other tests can be used depending on the kind of model being fitted and its objective.

The list below provides all explanatory variables used in one of the discussed models. The number in brackets indicates the number of studies in which the variable was used and was statistically significant. A list of explanatory variables for urban intersections is presented first in section 7.1 after which that of rural intersections is given in section 7.2.

7.1 Significant explanatory variables for urban intersections

The variables have been grouped into traffic flow, traffic control, geometric characteristics, driver characteristics, vehicle type/features, environmental factors and land use.

7.1.1 Traffic flow

AADT on minor road (5)
AADT on major road (5)
Total vehicle counts and pedestrians crossing all arms (AADT) (4)
AADT on intersecting roads per lane (2)
Total left-turn volume in AADT (1)
Right-turn volume from loop detectors (1)
Turning vehicles (2)
Pedestrian-bicycle interaction (1)
Pedestrian involved in accident (1)

7.1.2 Traffic control

Major road left-turn lane road (controlled) (1)
Major road left-turn lane road (uncontrolled) (1)
Major road left-turn prohibition (2)

Permissive right turn (1)
Access control on major road (2)
Design speed of major road (1)
Spot speeds of vehicles approaching the intersection along the major road (1)
Approach speed on minor road (1)
Approach speed limit at intersections (2)
Signal timing (4)
Signal phasing (3)
Camera installed (1)
Signal control type (1)
Signal indicator (2)
Number of bus stops at approach road (1)
Traffic control of level 2 (yield on minor road)
Traffic control of level 3 (No control on minor road)

7.1.3 Geometric characteristics

Average lane width on major road (3)
Average lane width on minor road (1)
Average width of the minor road (1)
Total number of lanes on all intersecting roads (2)
Number of lanes on major road (3)
Number of lanes on minor road (2)
Presence of median on major road (3)
Average width of median on major road (1)
Approach median width >2m (1)
Major road right-turn channelization (2)
Uncontrolled left-turn road slip (1)
Minor road right-turn channelization (2)
Right-turn channelization (1)
Controlled/exclusive right-turn lane (1)
Lighting (4)
Outside shoulder width of major road (1)
Outside shoulder width of minor road (1)
Pedestrian facility/ refuges (1)
Sight distance from stop-line (3)
Horizontal curve (2)
5% gradient (1)
More than 5% approach gradient (1)
Acceleration section on left-turn lane (2)
Number of bus bays (1)
Divided/undivided highway (1)
Intersection type (1)
Nature of the lane (1) i.e. single lane, left-most lane, right-most lane, centre lane.

7.1.4 Driver characteristics

Alcohol/drug use (1)
Age (1)
Residence of driver (1)
Gender (1)

7.1.5 Vehicle type/features

Passenger car (1)
Van (1)
Light trucks (2)
Heavy vehicles (2)
Two-wheel vehicle (1)

7.1.6 Environmental factors

Accident time (2)
Wet road surface condition (1)
Urban/rural roads (1)

7.1.7 Land use

Land use category (1)

7.2 Significant explanatory variables for rural intersections

Similar categorization as in section 7.1 has been done for explanatory variables of rural intersections, that is, traffic flow, traffic control, geometric characteristics, driver characteristics, vehicle type/features and environmental factors.

7.2.1 Traffic flow

AADT on minor road (3)
AADT on major road (2)
Total vehicle counts and pedestrians crossing all arms (AADT) (1)
Percentage of traffic turning left on the major road (1)
Percentage of incoming minor traffic turning left during peak hours (1)
Percentage of all incoming traffic during peak hours (1)
Average absolute percent grade change per 100m along the major and minor road approaches, within 244m of distance from the centre of the intersection (1)

7.2.2 Traffic control

Speed limit on major road (2)
Speed limit on minor road (2)
Design speed of major road (1)
Type of access control on major road (1)
Major road left-turn prohibition (1)

7.2.3 Geometric characteristics

Average lane width on major road (1)
Skew angle (1)
Intersection angle = 90^o (1)
Number of lanes on all intersecting roads (2)
Number of lanes on major road (2)
Presence of right-turn lane on major road (1)
Presence of left-turn lanes on major road (2)
Sight distance (3)
Existence of a horizontal curve (2)
Functional class of major road (1)
Outside shoulder width on major roads (2)
Outside shoulder width on minor roads (1)
Lighting (1)
Intersection on hilltop along major road (1)
Minor approach width (1)
Minor road descends to intersection (1)
Vertical curve on major and minor roads (3)
Existence of pedestrian facility on minor road (1)
Median width of major road (1)
Major road right-turn channelization (1)
Crossroad right-turn channelization (1)
Presence of major road left-turn channelization (1)

Presence of traffic islands (1)
Existence of painted islands on major road (1)

7.2.4 Driver characteristics

Alcohol/drug use (1)
Age (1)
Driver residence (1)
Gender (1)

7.2.5 Vehicle type/features

Passenger car (1)
Van (1)
Light trucks (1)
Heavy vehicles (1)

7.2.6 Environmental factors

Type of terrain (1)
Clear weather (1)
Daylight (1)
Wet surface condition (1)
Divided/undivided highways (1)
Accident time (1)
Urban/rural roads (1)

Despite the significance of some measures in the reviewed studies, their effectiveness is not consistent: it might increase, decrease or have no impact at all on the number of accidents. Below, we explain the likely cause of this inconsistency in some of the results across the studies. Nevertheless, the genuine reasons of these results are not well-known.

AADT
Increased volumes imply greater interaction between vehicles and perhaps more conflicts. Furthermore, as volumes increase, there are fewer gaps in traffic for right-turning as well as left-turning drivers. This results in increased accidents due to greater exposure. On the other hand, putting more vehicles on roads which are already characterised by many vehicles results in limited space to drive. The drivers tend to reduce speed and this results in fewer chances to collide and thus fewer accidents. This might explain why AADT increases accidents in some studies while in others it decreases them. Nonetheless, increased AADT on roads has been revealed to increase accidents in majority of the reviewed studies.

Presence of surveillance and light cameras
The camera may be installed at a location where the probability of an accident to happen is small. In such cases, the effectiveness of a camera may not be realized. On the other hand, a camera can be placed at a location characterised by a high probability of accident occurrence which might result in increased accidents. All these scenarios lead to different directions of this variable in terms of accidents. As stated before in the text, it is important that the locations are chosen at random without considering the accident history as this might bias the results and lead to inconsistencies.

Number of lanes
A possible explanation of the inconsistency is that more traffic lanes lead to higher speeds and that changing lanes represents a new hazard. Increased speed can occur particularly where the capacity of the road was previously small, but which then becomes adequate when the number of traffic lanes is increased. On roads with intersections, the crossing becomes wider and more complicated when the number of traffic lanes

increases and this results in more accidents at intersections connecting roads with more lanes than those connecting roads with fewer lanes. In most of the studies reviewed in this report, a higher number of lanes on roads joined at an intersection resulted in more accidents.

Curves on intersections
Horizontal curves cause sight obstruction on the inside of the curves such as buildings and a wall which limits stop sight distance and increases the chances of accidents. On the contrary, drivers travelling on horizontal curves may experience some uncertainty due to a limited sight and reduce speed (Kim et al., 2007). In that case, the occurrence of accidents is reduced.

Lighting (daylight and lighting of roads at night)
A possible reason for the conflicting results of lighting is that during the daylight, there is increased exposure due to more traffic. This increases accident occurrence. Further, road lighting can increase the number of accidents where collisions with lampposts are involved. In addition, some drivers tend to drive fast and aggressive since they can see the road clearly, which might increase accident occurrence. On the contrary, seeing the road clearly can result in reduced number of accidents as drivers can clearly see the obstacles and take appropriate action, hence, the inconsistency in the effect of lighting. Daylight and lighting of roads at night increased the number of accidents in almost all the studies reviewed in this report.

Presence of median
The existence of a median on some roads, reduces vision, hinders overtaking and can, if constructed using concrete barriers increase exposure resulting in more accidents. In urban areas, medians prevent turning manoeuvres at intersections. Also, crossing traffic can be reduced in urban areas through the construction of medians. This lowers the chances of accidents and as a result the effect of the presence of medians on the number of accidents becomes inconsistent. In this report, the presence of medians reduced the number of accidents in most of the studies.

8. 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**, pp. 447-456.

Abdel-Aty, M., and Keller, J., (2005). Exploring the overall and specific crash severity levels at signalized intersections. *Accident Analysis & Prevention*. **37**, pp. 417-425.

Agresti, A., (2007) *An Introduction to Categorical Data Analysis,* Wiley Series in Probability and Statistics. 2nd ed. New York, John Wiley & Sons.

Agresti, A., (2002) *Categorical Data Analysis,* Wiley Series in Probability and Statistics. 2nd ed. New York, John Wiley & Sons.

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

Andersen, J. A., McLellan, B. A., Pagliarello, G., and Nelson, W. R., (1990). The relative influence of alcohol and seat belt usage on severity of injury from motor vehicle crashes. *The Journal of Trauma*. **30,** pp. 415-417.

Arndt, O.K, and Troutbeck, R.J., (1998). Relationship between roundabout geometry and accident rates. *International Symposium on Highway Geometric Design Practices, Boston Massachusetts*. **28**(1), pp. 1-16.

Bauer, K.M., and Harwood, D.W., (2000). Statistical models of at-grade intersections accidents-addendum. Publication FHWA-RD-99-094, FHWA, Washington D.C.

Bedard, M., Guyatt, G. H., Stones, M. J., and Hirdes, J. P., (2002). The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. *Accident Analysis & Prevention*. **34**, pp. 717-727.

Chang, L.Y., and Wang, H.W., (2006). Analysis of traffic injury severity: an application of non-parametric classification tree techniques. *Accident Analysis & Prevention*. **38** (5), pp. 1019-1027.

Chin H.C., and Quddus, M.A., (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis & Prevention*. **35**, pp. 253-259.

Eby, D.W., and Molnar, L.J., (1998). Matching safety strategies to youth characteristics: a literature review of cognitive development. Technical Report. US Department of Transportation, Washington, DC.

Evans, L., (1993). Mass ratio and relative driver fatality risk in two-vehicle crashes. *Accident Analysis and Prevention.* **25**, pp. 213-224.

Evans, L., and Frick, M., (1994). Car mass and fatality risk: has the relationship changed? *Am. J. Public Health*. **84** (1), pp. 33-36.

Gelman, A., Carlin, J.B., and Stern, H.S., (2003). Bayesian Data Analysis, 2nd ed. Chapman & Hall, New York.

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J., (1995). Markov Chain Monte Carlo methods in Practice. Chapman & Hall, New York.

Greibe, P., (2003). Accident prediction models for urban roads. *Accident Analysis & Prevention*. **35**, pp. 273-285

Guo, G., and Zhao, H., (2000). Multilevel modeling for binary data. *Annu. Rev. Sociol*. **26**, pp. 441-462.

Harnen, S., Radin, U. R. S., Wong, S. V., and Hashim, W. W. I., (2003). Motorcycle crash prediction model for non-signalized intersections. *In: IATSS Research.* **27**(2), pp. 58-65.

Harwood, D.W., Council, F.M., Hauer, E., Hughes, W.E., and Vogt A., (2000). Prediction of the Expected Safety Performance of Rural Two-Lane Highways. FHWA-RD-99-207.

Huang, H., Chin, H.C., and Md. Mazharul, H., (2007). Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis. *Accident Analysis & Prevention*. **40**(1), pp. 45-54.

Huang, H.L., Chin, H.C., and Heng, H.H., (2006). Effect of red light camera on accident risk at intersections. In: TRB 2006 Annual Meeting CD-ROM, Transportation Research Board, National Research Council, Washington, DC.

Wang, X., Abdel-Aty, M., Nevarez, A., and Santos, J.B., (2007). Investigating of safety influence area for four-legged signalized intersections: Nationwide survey and empirical inquiry. TRB 2008 Annual Meeting CD-ROM, Transportation Research Board.

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

Kaiser, S. H., (1994). Urban intersections that work for pedestrians: A new definition for level of service. Presented at the 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1994.

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

Kim, D., Lee, Y., Washington, S., and Choi, K., (2007). Modeling crash outcome probabilities at rural intersections. *Application of hierarchical binomial logistic models*. **39**, pp. 125-134.

Kocklelman, K.M., and Kweon, Y.J., (2002). Driver injury severity: an application of ordered probit models. *Accident Analysis & Prevention*. **34** (3), pp. 313–321.

Kulmala, R., (1995). Safety at rural three and four-arm junctions; Development and application of accident prediction models. Technical Research Centre of Finland, VTT 233, Espo.

Kumara, S. S. P., and Chin, H. C., (2003). Modeling Accident Occurrence at Signalized Tee Intersections with Special Emphasis on Excess Zeros. *Traffic Injury Prevention*. **4**(1), pp. 53-57.

Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W., fifth edition (2005). Applied Linear Statistical Models.

Lambert, D., (1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics.* **34**, pp. 1–14.

Levine, E., Bedard, M., Molloy, D. W., and Basilevsky, A., (1999). Determinants of driver fatality risk in front impact fixed object collisions. *Mature Medicine Canada*. **2**, pp. 239-242.

Lund, A. K., & Ferguson, S. A., (1995). Driver fatalities in 1985–93 cars with air bags. *The Journal of Trauma*. **38**, pp. 469-475.

O'Neill, B., and Lund, A. K., (1993). The effectiveness of airbags in preventing driver fatalities in the United States. Proceedings of a Conference on Air Bags and Belts: Evaluation and Implications. Montreal: Health and Welfare Canada.

Poch, M., Mannering, F., (1996). Negative Binomial Analysis of Intersection Accident Frequencies. *J. Transportation Eng.* **122**, pp. 105-113.

Quddus, M.A., Noland, R.B., and Chin, H.C., (2002). An analysis of motorcycle injury and vehicle damage severity using ordered probit models. *Journal of. Safety Research*. **33**(4), pp. 445-462.

Reurings, M., Janssen, T., Eenink, R., (SWOV), Elvik, R., (TOI), Cardoso, J., (LNEC), and Christian, S., (KfV), (2005). Accident Prediction Models and Road safety Impact Assessment; a state-of art.

Rifaat, S.M., and Chin, H.C., (2005). Analysis of severity of single-vehicle crashes in Singapore. In: TRB 2005 Annual Meeting CD-ROM, Transportation Research Board, National Research Council, Washington, DC.

Ridout, M., Hinde, J., and Demetrico, C.G.B., (2001). A Score Test for Testing a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alternatives. *Biometrics*. **57**, pp. 219-223.

Salifu, M., (2004). Accident prediction models for unsignalized urban junctions in Ghana. *In: IATSS Research*. **28**(1)**,** pp.68-81.

Sayed, T., and Rodriguez, F., (1999). Accident prediction models for urban unsignalized intersections in British Columbia. *In: Transportation Research Record*. **1665**, pp. 93-99.

Shankar, V., Milton, J., and Mannering, F., (1997). Modeling Accident Frequencies as Zero-Altered Probability Process: An Empirical Inquiry. *Accident Analysis & Prevention.* **29**, pp. 829-837.

Simoncic, M., (2001). Road fatalities in Slovenia involving a pedestrian, cyclist or motorcyclist and a car. *Accident Analysis & Prevention*. **33**(2), pp. 147–156.

Spiegelhalter, D.J., Thomas, A., Best, N.G., and Lunn, D., (2003a). WinBUGS version 1.4.1 User Manual. MRC Biostatistics Unit, Cambridge, UK.

Stamatiadis, N., and Deacon, J.A., (1995). Trends in highway safety: effects of an aging population on accident propensity. *Accident Analysis & Prevention*. **27**, pp. 443-459.

Stamatiadis, N., and Deacon, J.A., (1997). Quasi-induced exposure: methodology and insight. *Accident Analysis & Prevention*. **29**, pp. 37–52.

Stimpson, W. A., Zador, P. L., and Tarnoff, P. J., (1980). The influence of the time duration of yellow traffic signals. *ITE Journal*. **50**, pp. 22-29.

Storie, V., (1977). Male and female car drivers: drivers observed in accidents, Report no. 761. Transport and Road Research Laboratory, Crowthorne, UK.

Thagesen, B., (1996). Highway and Traffic Engineering in Developing Countries, 1st ed., E & FN Spon, London.

Transportation Research Board. Highway Capacity Manual 2000. National Research Council, Washington, D.C., 2000.

Turner, S., and Nicholson, A., (1998). Intersection accident estimation: The role of intersection location and non-collision flows. *Accident Analysis & Prevention*. **30**, (4), pp. 505-517.

Ulleberg, P., (2001). Personality subtypes of young drivers. Relationship to risk-taking preferences, accident involvement and response to a traffic safety campaign. Transport Res. Part F: *Traffic Psychol. Behav*. **4**, pp. 279-297.

Vogt, A., (1999). Crash models for rural intersections: four-lane by two-lane stop-controlled and two-lane signalized. Report FHWA-99-128, FHWA, Washington D.c.

Vuong, Q., (1989). Likelihood Ratio Tests for Model Selection and Non-nested Hypothesis. *Econometrica*. **57**, pp. 307-334.

Waller, P. F., Blow, F. C., Maio, R. F., Singer, K., Hill, E. M., and Schaefer, N., (1997). Crash characteristics and injuries of victims impaired by alcohol versus illicit drugs. *Accident Analysis & Prevention*. **29**, pp. 817-827.

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

Wang, Y., Idea, H., and Mannering, F., (2003). Estimating rear-end accident probabilities at signalized intersections: occurrence-mechanism approach. *J. Transport. Eng*. **129**(4), pp. 377-384.

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

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**, pp. 983-995.

Yates, D.S., David, S., Moore, D.S., and Starnes, (2008). *The Practice of Statistics, 3rd Ed.* Freeman. ISBN 978-0-7167-7309-2.

Zador, P. L., and Ciccone, M. A., (1993). Automobile driver fatalities in frontal impacts: Air bags compared with manual belts. *American Journal of Public Health*. **83**, pp. 661-666.