Justifying and Prioritizing Roadway Lighting: A Case Study of Quebec Highways

Mahnoush Heydari

A Thesis

in

The Department

of

Building, Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Applied Science (Civil Engineering) at
Concordia University

Montreal, Quebec, Canada

August, 2016

This is to certify that the thesis prepared

By: *Mahnoush Heydari*

Entitled: **Justifying and prioritizing roadway lighting: A Case Study of Quebec Highways**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Civil Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Zhenhua Zhu ____________________ **Chair**

Dr. Ali Nazemi ____________________ **Examiner**

Dr. Ivan Contreras ____________________ **External Examiner**

Dr. Luis Amador ____________________ **Supervisor**

Approved by ____

Chair of Department or Graduate Program Director

____

Dean of Faculty

Date ____

ABSTRACT

Justifying and prioritizing roadway lighting: A Case Study of Quebec Highways

Mahnoush Heydari

Roadway lighting is an effective countermeasure capable of reducing night-time motorized collisions under the right circumstances. Its initial viability can be learnt through collision modification factors showing beneficial effects of roadway lighting on local roads. However, this requires time-series of data from several years before and after the implementation of lighting. There is a need to estimate collision modification factors of lighting from locally observed cross sectional data from as little as one year. There is also a lack of a practical method capable of replacing the complicated warrant system to support decisions of whether or not to illuminate roads. Such method should be able to identify and prioritize segments that will benefit the most from being illuminated. This research presents a method to estimate collision modification factors with as little as one year of data. In addition, this research presents a practical method that identifies and prioritizes candidate road segments for being illuminated. A case study of Quebec's highways found that lighting is an effective countermeasure and that expected benefits approximate 60% reduction in night time collisions. It was found that segment size plays an important role and that Bayesian data fusion can be used to abstract from segment size to estimate a generic collision modification factor. It was found that safety performance functions for desired land use and sites type can be used in combination with the observed number of collisions to classify those sites expected to observe benefits from being illuminated.

ACKNOWLEDGMENTS

Foremost, I would like to thank my supervisor, Professor Luis Amador, for his support, guidance, motivation, enthusiasm and inspiration. I sincerely appreciate all his contributions of time, energy and ideas to make my M.Sc. program a unique and productive experience.

I gratefully acknowledge my brother, Shahram for his helpful guidance and advice. The support and understanding of my parents and my lovely sister made this entire long journey possible, with the constant encouragement of my family is what drove me to thrive for excellence throughout these years despite difficulties of being away from home.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVIATIONS

AADT	Annual Average Daily Traffic
AASHTO	American Association of State Highway and Transportation Officials
CIE	International Commission on Illumination
CMF	Crash/Collision Modification Factor
COMT	Council of Ministers Responsible for Transportation and Highway Safety of Canada
FB	Full Bayesian
FHWA	Federal Highway Administration
GPS	Geographic Positioning System
IESNA	Illuminating Engineering Society of North America
JAGS	Just Another Gibbs Sampler
LED	Light Emitting Diode
MCMC	Markov Chain Monte Carlo
MOT	Ministry of Transportation
MTQ	Ministry of Transportation of Quebec
NCHRP	National Cooperative Highway Research Program
OpenBUGS	Open source Bayesian Using Gibbs Sampler
PDO	Property Damage Only
PFI	Potential for Improvement
SAAQ	Automobile Insurance Company of Quebec
TAC	Transportation Association of Canada
TRB	Transportation Research Board
USA	United States of America
VAR	Variance
WHO	World Health Organization
WinBUGS	Windows Bayesian using Gibbs Sampler
ZINB	Zero Inflated Negative Binomial
ZIP	Zero Inflated Poisson

CHAPTER 1: INTRODUCTION

1.1 Background

Road collisions could have negative impact on many aspects of people's lives, including infrastructure damage, labour loss, serious injuries and fatalities. Transport Canada reported 1,834 fatalities, 9,647 serious injuries, and 149,900 injuries in Canada only in 2014. During 2012; 74,574 property damage accidents; 39,105 cases of injuries and 436 fatalities reported by SAAQ on Quebec road network. Also, road fatalities were ranked ninth worldwide and one third of fatalities in Canada in 2011, which makes road injuries one of the top ten death leading causes (WHO, 2013). In developed countries with high income, the significant percentage of road deaths is associated with motorized vehicle crashes (WHO, 2013). In Quebec, 67% of the fatalities related to motorized users and 18% of the road fatalities involved non-motorized road users (*i.e.*, pedestrians and cyclists. Additionally, the number of accidents is expected to increase as more population become motorized (SAAQ, 2012). These statistics are clear evidence of the urgency to: (1) develop safety improvement programs and, (2) identify effective countermeasures capable of preventing accident occurrence and of reducing the social and economic expenses of injuries and fatalities. Reliable estimation of the level of safety (or risk) associated to each road entity (e.g., curve, ramp, segment) is crucial to guide the decision making that allocates safety improvements across any road network to obtain safer roads. In some instances, the road agency can improve the geometric design of the road or influence the operational characteristic deemed deficient, however, in others the agency is unable to, or the said improvement may not remediate the safety deficiency and hence the agency may want to follow a different route and addition a safety treatment (Hauer, 1997). The problem lies in the need to know which measure is more effective in mitigating crashes and their consequences, and to what degree.

Roadway lighting is one of such characteristics and has been thought to be a good countermeasure even in the presence of other geometric or operational deficiencies, because it increases driver visibility (CIE, 1992). The night time accident reduced 13 to 75% in presence or improvement of artificial lighting (Yanmaz-Tuzel and Ozbay, 2010). Night time accidents found to be more severe and dangerous in case of fatalities in comparison to daytime accidents because dark road could decrease a driver's visual capability and ability to manoeuvre or respond adequately to road hazardous (CIE, 1992; Rea *et al,.* 2009). However, only 35% of Quebec's roads (1208 out of 3,452 kilometres of highways) have the artificial illumination (COMT, 2013). The amount of investment for operation and maintenance of roadway lighting can be justified when local data shows a significant reduction of night-time accident after the implementation of lighting (AASHTO, 2005).

1.2 Problem Statement

Most of the previous safety studies made use of the before-after approach (Hauer, 1997) to estimate the ability of a treatment to reduce (or undesirably increase) road crashes; better known as crash modification factor (CMF) in order to count with a general justification for investments to lit highways and roads. However, this requires records of collisions for several years before and after the treatment was done and evidently the knowledge of the moment of time when the treatment was implemented. Hence the viability of utilizing the before-after approach is often unviable to most agencies trying to justify with local data the benefits of illuminating some of their roads. In most occasions, data from many roads but on the same time period can be available, this cross-sectional data provides a fixed radiography of geometric and traffic characteristics to which it is possible to add collision history as many police departments keep detailed collision records. Safety analysis for the same period, known as cross sectional studies, has been employed before, but, their results seem highly dependent on the segment size in which the available data is partitioned.

Many studies have investigated intersections avoiding the need to define a segment size. To date most researchers, tend to keep the size of road segments unaltered as received from the spatial database, which are commonly large segments defined over pavement maintenance needs. In addition, many researchers tend to believe that one should use large segment sizes to avoid the zero collisions problem. There is however an evident need to use small segment sizes when investigating the countermeasure effects of roadway lighting given that lighting is allocated in small portions of the network at strategic locations and that lighting could vary greatly in small distances.

There is a lack of a study that analyses the role of road segments in learning the crash modification ability of roadway lighting, considering the existence of uncertainty in the estimated values while studying different segment sizes. First, there is a need to justify lighting investments through collision reductions with limited local data. Second, there is a need to have a decision support method able to identify which sites will benefit the most from receiving lighting.

1.3 Objective

The main aim of this study is to **develop an approach to justify the provision of roadway lighting and prioritize road segments to receive illumination**

Specific objectives are described in more detail as follows:

1) To develop an approach to produce crash modification factors for roadway lighting from cross sectional data

2) To study the role of segment size in the estimation of CMF with cross sectional data

3) To propose an easy mechanism to identify and prioritize those road segments that will benefit the most from receiving lighting.

1.4 Scope and Limitation

Safety benefits of lighting can only be measured during nigh time. This study limits to the use of cross-sectional data and to the application of statistical methods for count data such as Negative Binomial Regression. Although, there are segments with zero collision count, the method of zero inflated Negative Binomial is not applied due to the many criticizes stated by past studies about ZINB (Lord, 2016), and experimentally insignificant gains observed during this study.

Another limitation was related to available attributes for site characteristics. For instance, the data available for median did not refer necessarily to a physical barrier but in some occasions to a painted line on the pavement, and speed data referred to posted speed. For the case of lighting and intersections binary variables were used to fix these inaccuracies and preserve the valuable information.

Also, the CMF values for lighting were different when analysing each road segment sizes separately, and CMFs from past studies. Nevertheless, to overcome these limitations, the proposed method, Bayesian data fusion, enhanced the accuracy of the CMF value for lighting by considering the information gain from specific segment-size groups and from past related studies.

1.5 Organization of the Thesis

This thesis consists of five following chapters as defined below:

Chapter 1 contains the introduction with background and problem statement, research objectives, and scope and limitations.

Chapter 2 reviews past studies to provide the necessary background knowledge of motorized accident modeling, Safety Performance Functions (SPF), road safety management and methods of measuring safety effectiveness such as crash modification factors. It also revises the applications of Bayesian data fusion on other fields. A summary of results from previous studies of roadway lighting effect is provided.

Chapter 3 introduces the methodology employed for processing and analyzing the data, the framework of the proposed method to develop SPFs, regression models and estimation of safety effectiveness of lighting among different segment sizes.

Chapter 4 describes the data used in this study to develop the proposed model to estimate crash modification factors for lighting from cross sectional data, and to prioritize road segment candidates to receive lighting. It presents the main findings of the role of segment size on CMF values.

Chapter 5 presents the conclusions of the study; also, it provides suggestions and recommendation for future research.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter reviews the main aspects of road safety and motor vehicle accident modeling process, the effect of explanatory variables on accident frequency, safety performance functions, statistical analysis based on regression models, and statistical methods applied to estimate specific countermeasure effect. Also, different approaches of estimating crash modification factor, as the cross-sectional and the before-and-after method are discussed in terms of each advantages and disadvantages. Bayesian framework and related previous studies are described.

The first section 2.2 describes roadway lighting and summarizes findings related to night time collisions. The second section 2.3 explains road safety definition and commonly applied statistical analysis for accident modeling such as Poisson, Poisson-gamma or Negative Binomial, zero-inflated Poisson and Negative Binomial models (ZIP and ZINB), which are applied to investigate the role of artificial lighting as a countermeasure for night time collisions. Also, comparison between different statistical approaches and identification of the appropriate models are discussed. The third section 2.4 provides a review of methods to estimate safety effectiveness and crash modification factor of special countermeasure. The final section 2.5 provides a brief explanation of Bayesian data fusion framework and past researches.

2.2 Roadway Lighting

The measurement of any form of light is known as Photometry. One of the main concerns of photometry is to measure the apparent brightness of a source of light to the human eye (IESNA, 2008). Photometric measurement units such as: lux, lumens and candelas, are indicators reflecting variables that determine visual responses and reaction times. According to the CIE (2007), common elements of road lighting are: Luminance and Illuminance, Luminance-based uniformities and Illuminance-based uniformity, Glare, threshold increment, and color of the light source.

Illuminance is the amount of light arriving at the surface of the pavement in units of lux. Luminance is the amount of light as perceived by the driver which also depends on the roadway surface and environmental conditions (wet, dry). Luminance is the amount of light as perceived by a driver and it is measured in candela per square meter (cd/m2). Some commercial cameras using specialized calibrated software have the ability to match the pixel grey-scale brightness to luminance targets (Jackett and Frith, 2013; Cai and Li, 2014). As mentioned by Rea *et al*. (2009) illuminance level should be the major criterion for designing and analyzing roadways that are commonly used by non-motorized users, such as pedestrians and bikers. Luminance should be considered as the lighting criterion for motorized roads with high-speed moving vehicles.

Uniformity refers to the longitudinal or transversal variation of the main lighting criterion (illuminance or luminance) over a segment of a road; there are several uniformity ratios; some divide the maximum observed value of the criterion (illuminance or luminance) over the average value; others divide over the minimum value. Neither overall (transversal) nor longitudinal uniformity has been found to be a significant predictor of accident rates reductions in previous studies (OPUS, 2012).

According to IESNA (2005), glare refers to the visual impairment or difficulty of observation of objects caused by a change in luminance between a target and its background as perceived by the human eye, there are two types: disability glare and discomfort glare. As their name indicated the discomfort glare is less aggressive than the disability glare which can potentially blind a driver for a few seconds.

In this study, luminance level is applied as lighting criterion to investigate the crash modification factor of roadway lighting. The effect of roadway lighting on night time accidents has been measured multiple times by researches. The main point of lighting is to increase visibility (Bullough *et al*., 2009). Roadway lighting as a countermeasure had shown 10 to 40% decreases in observed accident frequency and around 65% reduction in fatal crashes (TAC, 2004). In addition, Isebrands *et al*. (2010) found 37% reduction of collision counts in the presence of lighting. Older studies (IESNA, 1989) found reductions on accident frequency between 17 to 40%, for illuminated roadways compared to non-illuminated ones. Yannis *et al*. (2013) found that the existence of roadway lighting could diminish accident frequency, particularly fatalities and seriously injured accidents in urban and rural roads of Greece. Surprisingly, a cross-sectional study conducted by Box (1970) shows segments with higher lighting levels experienced more accidents suggesting that other forms of road deficiency in combination with higher speeds after improved visibility could be reason. A five years study conducted by Griffith (1994) investigated a reduction of 16% in observed accidents in the presence of continuously lighted segments on urban freeways when sections had the same traffic flow and all other factors are equal (Griffith, 1994). Table 2.1 shows the previous studies and methods applied to investigate the effect of lighting.

Table 2.1: Priors, CMF from Previous Studies

Obtained CMF	Calculated Safety Effectiveness (%)	*Method applied*	*Source*
0.70	30	Before-After	CIE, 1992
0.88	12	Cross-Sectional	Bullough *et al*., 2013
0.75	25	Cross-Sectional	Prestone *et al*., 1999
0.61	39	Cross-Sectional	Schewab *et al*., 1982
0.69	31	Before-After	Isebrand *et al*., 2004
0.62	38	Before-After	HSM, 2010
0.96	4	Before-After	Harwood *et al*., 2007
0.836	16.4	Cross-Sectional	Gross and Donnell, 2011
0.905	9.5	Cross-Sectional	Sasidharan and Donnell, 2013
0.72	28	Cross-Sectional	Donell *et al*., 2010
0.83	17	Cross-Sectional	Wanvik, 2009
0.72	28	Before-After	Elvik and Vaa, 2004

2.2.1 Roadway Accidents and Night Time Collision

Most of past studies focuses on intersections and interchanges to investigate the effect of different variables on accidents frequencies (Abdel-Aty *et al*., 2005; Santiago-Chaparro *et al*., 2010; Lord and Persuad, 2000; Lovegrove and Sayed, 2006). Also, according to the literature, 40% of all road crashes happened at intersections because of the presence of more traffic conflicting movements (Barua *et al*., 2010). However, there are few studies of road safety for road segments (Jonsson *et al*., 2007). Generally, these studies analyze roads based on lighting levels with mixed sizes of road segments. The number of crashes are mainly associated with traffic volume (Baek and Hummer, 2008; El-Basyouny and Sayed 2006; Hadayeghi *et al*., 2006) and complex geometry of roadways (El-Basyouny and Sayed, 2006), as horizontal or vertical curves (Hummer *et al.,* 2010) but none to the segment size which to date has not been studied.

From previous studies there is an agreement that the presence of roadway lighting diminishes nighttime accidents and consequently the number of fatalities and seriously injured crashes will be decreased (Yannis *et al*., 2013). Also, previous studies show a higher frequency of pedestrian collisions at sites with low levels of lighting (Zhou and Hsu, 2009). The rates of Collisions by severity for different types of highways in Montreal, Quebec, shows that probability of being involved in an accident on a rural-principal or arterial roads is two to three times more than on freeways with more severe crashes observed in principal arterial roads (Brown and Baass, 1995).

2.3 Statistical Analysis for Roadway Safety

The safety of a site is described by Hauer (1997), as the number accidents, or accident consequences, by kind and severity, expected to occur on that site during a defined time period. To investigate the impact of specific variable as a countermeasure, for road safety management purposes, two periods of time should be consider and the number of accidents compared before and after the treatment. This is known as a before-after road safety study. It must be notice that the before and after period refer to a chain of several observations and not a single one in order to capture the trend before and after the improvement. According to the literature, minimum duration for each time period is at least three years with annual observations.

The choice of appropriate regression model and reliable safety performance function (SPF) is indispensable for the analysis. To provide accurate results from number of accident as an outcome and contributing variables, it is noteworthy to choose the regression model which fits better the observed data. The prediction of an outcome of road safety analysis (crashes as integer values) could be done by different methods. The SPF associates an outcome (accidents), and the explanatory variables (traffic volume, site characteristics). The main objective of transportation engineers is to investigate the effect of different variables on accident frequencies of transportation facilities, and to develop statistical models that can, accurately, describe accident datasets (Milton *et al*, 2008). Traditionally, the regression approach for accident modeling is a Poisson distribution (Hauer, 1997). However, the Poisson method required that the mean and the variance are equal, posing an important restriction which is its main disadvantage (Anastasopoulos and Mannering, 2008; Mitra and Washington, 2006). This limitation becomes important in the presence of over-dispersion, which is a common case among count accident datasets, since the variance is commonly greater than the mean. The over-dispersion is due to the unobserved heterogeneity across sites such as intersections and road segments. To overcome this deficiency, accidents are presumed to be gamma distributed and PoissonGamma or Negative Binomial regression (NB) is commonly applied to overcome this deficiency (Poch and Mannering, 1996).

The regression models which have been applied to accident data are varies between past studies. Generally, Poisson regression, and Poisson mixtures (Poisson-Gamma and Poisson-Lognormal) have been employed for roadway safety analysis due to their ability to handle count data. As mentioned in Poisson model the mean is assumed to be equal to the variance. This assumption is not always satisfied, because most of the time, variances exceed the mean due to the presence of heterogeneity (Mitra and Washington, 2006). Therefore, the Poisson model is not applicable in case of over-dispersions, and Poisson mixtures should be applied. Poisson Gamma or Negative Binomial model is usually used in accident data analysis (Poch and Mannering, 1996; Hinde and Demetrio, 1998; Miaou and Lord, 2003; Anastasopoulos and Mannering, 2008) and the other Poisson mixture models, (i.e. Poisson-Lognormal) have been applied for accident data analysis in some researches (Lord and Miranda-Moreno, 2008; El-Basyouny and Sayed, 2010). These models account for the over dispersion problem.

2.3.1 Poisson Model

Poisson distribution assumes that the mean and the variance of occurred accidents are equal (Hauer, 1997). In Poisson model, the mean of the expected number of accidents, is only Expressed by site characteristics which is the SPF. Equations 2.1 and 2.2 show the simple equation of Poisson model:

$$K \sim \text{Poisson}\ (\theta) \qquad [2.1]$$

Where k is: the number of observed accidents during specific period of time, and

$$\theta = f\ (F_1, F_2, x; a) \qquad [2.2]$$

Where θ: is the mean value obtained from the SPF which is the function of contributing factors, F_1 and F_2 are traffic flows in both direction, and vector x represents site-specific attributes and regression parameter. The restriction of equality of the mean and variance is the main shortcoming of this model since it does not account for the heterogeneity across sites (Mitra and Washington, 2006).

2.3.2 Poisson-Gamma or Negative Binomial (NB) Model

To address over-dispersion (variance exceeds the mean) Negative Binomial model is appropriate and applicable. Accident data are Poisson distributed and unobserved accident heterogeneity across sites is assumed gamma distributed in this approach (Washington *et al*., 2003). In this case, the expected accident frequency (θ) is explained by the SPF and multiplicative random effects which differ in various sites describe the θ. Equation 2.3 and 2.7 introduce the Poisson model.

$$K \sim \text{Poisson}(\theta), \qquad \theta = \mu\, r \qquad [2.3]$$

Where k: is the observed accident frequency, and

$$\mu = f(x, a) \qquad [2.4]$$

Where μ: as a function of the contributing factors' vector x and the vector of unknown parameters.

The multiplicative random effect, r, is assumed Gamma distributed ($r \sim \text{gamma}(\varphi, \varphi)$) with the mean of 1 and the variance of $1/\varphi$. The φ known as "inverse dispersion parameter" which is assumed fixed across the sites in this case. (Anastasopoulos *et al*., 2008; Miranda-Moreno *et al*., 2007).

2.3.3 Zero-Inflated Regressions

In many cases of accident data, high number of zeros has been observed in the time period of study, this issue cannot be captured by conventional Poisson model (Miranda-Moreno, 2014). In such situation zero-inflated models such as Zero-Inflated Poisson (ZIP) and the Zero-Inflated Negative-Binomial (ZINB) are applicable. Two sources of zero counts are known for this model, one as a zero state, and the other source is a usual random process which follows the Poisson distribution (Miranda-Moreno, 2014). Following Equations shows the ZIP and ZINB model as presented by Miranda-Moreno (Lord and Miranda-Moreno, 2008):

$Y_i = 0$, with probability ε_i, [2.5]

$Y_i \mid \mu_i \sim$ Poisson (μ_i), with probability ($1-\varepsilon_i$) [2.6]

These are followed by consecutive distributions respectively:

$f(y_i \mid \mu_i, \varepsilon_i) = \varepsilon_i + (1-\varepsilon_i)$ Poisson(μi) for $yi = 0$, and [2.7]

$f(y_i \mid \mu_i, \varepsilon_i) = (1-\varepsilon_i)$ Poisson(μi) for $y_i = 1, 2, \ldots$ [2.8]

$\mu_i = f(\text{AADT}, x_i; \beta)$ [2.9]

Where, μ_i is known as a function of a site attributes vector, and ε_i is the error term which includes the zero counts that could not be processed by the conventional Poisson distribution. The error term is defined as a logistic link function as below:

$$\varepsilon_i = \frac{e^{\omega z_i}}{1+e^{\omega z_i}}; \qquad [2.10]$$

Where, ω is a vector of parameters, and z is a vector of unobserved site attributes. The vector of covariates z_i could estimate the probability of being in the zero count state. Also it might be a function of specific-site characteristics or other covariates that may be part of the vector x_i (Miranda-Moreno, 2014). In addition, the ZINB model has the following probability distribution:

$f(y_i \mid \mu_i, \varepsilon_i, \alpha) = \varepsilon_i + (1-\varepsilon_i)$ Poisson(μ_i, α) for $y_i = 0$, and [2.11]

$f(y_i \mid \mu_i, \varepsilon_i, \alpha) = (1-\varepsilon_i)$ NegBin(μ_i, α) for $y_i = 1, 2, \ldots$ [2.12]

During the past years, zero-inflated (ZI) regression models have been commonly applied by transportation safety engineers; when a preponderance of zeros exist accident data. Poisson or Negative Binomial (NB) distribution could not handle the existence of over expected number of zeroes in data, this could bias the sample mean. ZI models assume entities (e.g., intersections, segments, crosswalks, etc.) in two cases: one as a true-zero state (inherently safe) and second as a non-zero state, means zero accident counts in the observation time period. In fact, entities which experienced zero collisions during the observation period could be defined as either safe or unsafe. This assumption refers as dual state in ZI models. Recent researches argued that ZI models despite that statistically fit the data better should be avoided for vehicle accident modeling on highway entities. Although in past years, researchers have applied zero-inflated probability models, with dual-state assumption to generate the crash data (Shankar *et al*., 1997; Qin *et al*., 2004; Kumara and Chin, 2003; Lee and Mannering, 2002) Using this count models to provide the best statistical fit is no longer a challenge. The application of a dual-state model shows incapability to describe the qualitative difference between inherently safe and inherently unsafe sites (Lord *et al*., 2005). Studies shows over-fitting data could become problematic issue (Loader, 1999). Therefore, a balance must be struck between the statistical theory and capability of accident predicting models (Miaou and Lord, 2003). Most of the time rural highways include more inherently safe segments and a few dangerous ones, that makes average crash rate greater than freeway segments, or not necessary crashes follow a dual-state process, sometimes it is a single one due to the low exposure (Lord *et al*., 2005). Also, it is noteworthy to consider that the rural highway segments classified with high traffic volumes, which never found to follow a dual-state process (Hauer and Persaud, 1995; Harwood *et al*., 2000). Past studies, indicates that non-reported accidents could be the main reason for exceeds zeros in crash data (Kumara and Chin, 2003). More example of misconception between an excess zero in accident data and non-reported crashes on rural networks is exist. (Lord *et al*., 2003a & 2003b; Lord *et al*., 2004). Considering that the root of excess zeros in crash data could be caused by small time period of observation, under-or non-reporting accidents, low traffic flow or high-risk segments and omitted variables contributing the collision process. To overcome mentioned problem, NB and Poisson models with additional factor define as error term to capture unobserved heterogeneity could be applied as an appropriate model which yield similar fit as zero-inflated models (Lord *et al.*, 2005). On the other hand, the presence of zeroes in accident data might not be a proof of dual-state process. Zero-inflated models have been applied in previous studies for accident modeling in different cases: single and multi-vehicle crashes on rural two-lane roads (Miaou, 1994; Shankar *et al*., 1997; Qin *et al*., 2004), run-off-the-road crashes in rural areas (Lee and Mannering, 2002), accident occurrence at intersections (Mitra *et al*., 2002; Kumara and Chin, 2003). In these studies, except Miaou (1994), it has been assumed that crashes must follow a dual-state process (Lord *et al*., 2005). In conclusion, in this study, we applied the Negative Binomial regression model because, despite enhanced statistical fit of zero-inflated models, it is argued that the initial assumption of a dual state process basic concept of these models is not consistent with accident data.

2.3.4 Safety Performance Function

SPF is a mathematical equation which could explain the accident frequencies based on a series of site characteristics (contributing factors). The SPF demonstrates a non-linear mathematical relationship between expected number of accidents per unit of time and a vector of road attributes. Safety performance modeling is based on historical observations to calibrate a functional form which accounts for interactions between contributing factors and the safety response to local conditions in terms of accident frequency. A reliable SPF could examine the effect of different contributing factors on expected accident frequencies. The main step to develop safety performance function is the choice of an appropriate model function and regression approach, which estimates the parameters of a model. Basically, the main contributing factors are annual average daily traffic (AADT), segment length and various site characteristics which is showed as Equation 2.1:

$$\mathrm{Ln}\,(\mu) = \ln(a_0) + a_1 \ln(L) + a_2 \ln(AADT) + \sum \alpha x \qquad [2.13]$$

Where, μ = expected accident frequency;

$\ln(a_0)$ = constant;

a_1 = stochastic parameter for segment length

a_2 = stochastic parameter for traffic volume

a = vector of stochastic parameters of other site characteristics

L = segment length (km);

AADT = annual average daily traffic (vehicles per day);

x = vector of site characteristics.

To develop SPFs; first, available contributing factors (site characteristics) should be selected. Second, appropriate regression model should be applied to predict and estimate the number of accidents based on the site characteristics (independent variables).

Regression approaches, assume road accidents as random events which at first glance follow a Poisson distribution. However, Poisson model assumes the equality of the mean and the variance, this assumption is violated by many accident data because of unobserved heterogeneity across sites (Mitra and Washington, 2006). Usually, variance is greater than the mean which is referred as over dispersion. Hence, to account for this problem the

PoissonGamma (Negative Binomial) model must be used in accident data analysis (Poch and Mannering, 1996; Hinde and Demetrio, 1998; Miaou and Lord, 2003; Anastasopoulos and Mannering, 2008). Negative Binomial model accounts for over dispersion of accident data when the variance is bigger than the mean. Other Poisson mixture models which accounts for the mentioned problem like Poisson-Lognormal have been applied to model the accident data in various past studies (Lord and Miranda-Moreno, 2008; El-Basyouny and Sayed, 2010). However, Poisson-Gamma models assume over-dispersion as a fixed parameter. Moreover, some methodologies have been introduced dispersion parameter in a way that it varies across different sites as a function of some site attributes (Hauer, 2001; Miaou and Lord, 2003; Geedipally *et al*., 2009). Also, dispersion parameter could consider as a function of site characteristic on confidence intervals of SPFs evaluations; for instance, dispersion parameter as function of the minor and the major traffic volume which provided more accurate data (Geedipally and Lord, 2008).

Equation 2.2 and 2.3 shows the most commonly used Safety Performance Function (SPF) for count data (Miranda-Moreno, 2014). Accident counts are associated with causal factors$(L_i, AADT_i, x_{i1}, \cdots x_{ik})$, a linear relationship is assumed between all the independent variables and accident frequency, except traffic flow which has an exponential relationship (AASHTO, 2005). The coefficients β_n shows the correlation and the impact of each explanatory variable has on the outcome (accidents). An error term (ε) accounts for the unobserved impact in case of missing explanatory variables (Miranda-Moreno, 2014). Equations 2.14 and 2.15 could be applied to develop SPF for intersections and road segments, respectively.

$$Acc = L_i AADT_i^{\beta_1} \exp(\beta_0 + \beta_2 x_{i1} + \cdots + \beta_k x_{ik}) + \varepsilon \quad [2.14]$$

$$Acc = L_i \exp(\beta_0 + \beta_1 lnAADT_i + \beta_2 x_{i1} + \cdots + \beta_k x_{ik}) + \varepsilon \quad [2.15]$$

2.4 Measuring the Safety Effectiveness of a Treatment

In practical aspects of road safety, it is noteworthy to calculate the safety effectiveness of countermeasures. Over the past years, various statistical methods have been applied to investigate the effect of different variables on accident counts. The cross-sectional statistical model and the before-and-after method are the most common methods. In 1975, the concept of Collision Modification Factor (CMF) was introduced by Laughland *et al*., to reflect the safety gain of countermeasures which estimate by expected or observed changes in accident counts after a countermeasure was applied at specific site. Regression models have been applied by researchers to predict the changes in accident counts at specific site by different contributing variables after implementation of special treatment.

The objective of cross-sectional models is to verify changes in safety of a road among multiple sites with assume the absence of any major changes within the given year. Cross-sectional models applied in some cases of road safety to investigate the countermeasure effects (Council and Stewart, 1999; Zegeer and Council, 1995). However, in some statistical models the number of explanatory variables was not enough to describe accident counts, (Schoppert and Hoyt, 1968). In recent researches, the concepts of multi-stage cross-sectional models have been developed to do safety analysis of grade crossing with the aim to overcome the issues associated with cross-sectional accident modeling (Saccomanno and Lai 2005; Park and Saccomanno, 2005a; Park and Saccomanno, 2005b). Multi-stage models arrange the crossing data into similar category in case of physical and operational characteristic.

In Before-after models, sites with only one or more implemented treatment will be analyzed, while keeping other attributes the same. The effect of a given countermeasure is investigated by estimating the change between predicted (observed) accident counts after the treatment done, to the number of crashes when the countermeasure not exist (Hauer, 1997; Persaud, 2001). Two methods for before-after models are introduced by literature as naïve and empirical Bayesian (EB) models. Naïve method associated with the problem of regression-to-the-mean (RTM) bias, which means safety countermeasures applied to the sites with a high number of observed accidents. Consequently, due to the random inherent of crashes, although reduction in accidents is assigned to the countermeasure effect completely, it is more likely that accident decrease in previous high levels despite the introduction of countermeasures. The average accident frequency is likely to reach to the mean over the long term even if high number of accidents occurs in certain years (Council *et al*., 1980). Also, RTM known as treatment selection bias, which arise when the statistical assumption of random sampling of accident data is violated (Park and Saccomanno 2007; Pendleton, 1991). If safety effectiveness of countermeasure, estimates with before-and-after method, which the RTM bias does not take into account, the results might not be accurate. For instance, the EB before-after method is applied by Al-Masaeid (1997), Bahar *et al*., (2004), Elvik *et al*., (2001), Lyon *et al*. (2005), and Persaud *et al*. (2001) to decrease the RTM bias caused by selected treatments to provide an accurate model. The main difference between the cross-sectional and before-after study is that 'treatment' is something which changed from before period to after one in before-after study, which in cross-section study this change does not exist and units are different just in some traits of interest (Hauer, 2010). Another problem associated with before-and-after method is that it is not easy to create a truly controlled experiment completely. The main concern of both cross-sectional and before-and-after approach is the difficulty associated with the usage of these models to distinct between the effect of particular treatment on accidents and the effect of more general contributing factors (e.g. weather conditions) that could not be altered by decision makers (Park *et al*., 2005). Separation between the effect of particular countermeasures and other factors on number of accidents is one of the main concerns of predicting accident by statistical models. Therefore, the estimated coefficients might be unreliable for representing the changes in accident counts which resulted from defined countermeasures. Equation 2.16 shows the definition of CMF:

$$CMF_i = 1 - [N_{Bi} - N_{Ai} / N_{Bi}] = N_{Ai} / N_{Bi} \quad [2.16]$$

Where, N_{Bi} and N_{Ai} known as, yearly estimated or observed number of accidents at site before (or without) and after (or with) a safety countermeasure *i*, respectively (Park, 2007). The estimated value for CMF in equation 2.16 is always positive. The CMFs greater than unity, illustrates an increase in accident counts after the countermeasure implemented, and values smaller than unity reflects a reduction in number of collisions by introducing countermeasure. A set of CMFs for two-lane rural highways developed by federal highway administration (Harwood *et al*., 2000; Zegeer *et al*., 1992) and Highway Safety Manual (Hughes *et al*., 2004; Harkey *et al*., 2005) to investigate the effect of different operational system and design of highways.

2.4.1 Cross-Sectional Safety Evaluation

If the before- after data does not exist in treatment sites, to evaluate a safety effectiveness of a treatment, cross-sectional method could be alternated. In this case, accident data would compare with comparable non-treatment sites (HSM, 2010). Both treated and non-treated sites should be classified and the time periods should be defined. There is no specific sequence of calculation for presenting an equation in cross-sectional method; this model needs to be developed in single models of treated and non-treated. Contributing factors of accident counts such as traffic volume and road attributes should be analyzed separately for treated and non-treated model. There should be an indicator variable to illustrate the presence or absence of the treatment (i.e. presence or absence of standard lighting), or as a continuous variable which represents the treatment measurements (i.e. lighting level). Negative Binomial model with generalized linear model (GLM) could be applied to model the yearly crash frequencies. However, Lord and Persaud applied general estimating equations (GEE), by using software, to evaluate the treatment effectiveness and its precision (HSM, 2010). In this study, the effect of lighting on cross-sectional data estimated by using expected number of accidents among two groups of treated and non-treated segments. The first step was to define the appropriate level of lighting for treated and non-treated group separation. The same level of standard lighting defined for three different segment sizes to make the lighting data as a binary variable and separating groups with standard and non-standard lighting as treated and non-treated segments. The first method is to calculate accident modification function or θ, by fitting regression equation and SPF function to cross-section data. Calculating the ratio of expected crash counts for treated and non-treated group shows the amount of change in safety from non-standard lighting to standard level. The smaller the value of θ the more effective is the treatment (Hauer, 2009). Equations 2.17 and 2.18 illustrate the crash modification factor and crash reduction factor estimation, respectively:

$$\text{CMF}(\theta) = \frac{\text{expected accidents count (treated group)}}{\text{expected accident count (non-treated group)}} \qquad [2.17]$$

CRF= (1 - θ) (Daziano *et al* 2013) [2.18]

This method is applicable if assuming the lighting data as a continuous variable, while running different SPF for treated and non-treated groups. The other method to describe accident reduction effect of roadway lighting is to define lighting data as a binary variable (i.e. in this study 1 for segments with 1.2 cd/m^2 or more lighting level, and 0 for less than 1.2 cd/m^2 level of lighting).

2.4.2 Crash Reduction Effect

Park *et al*. estimated the effect of contributing factors by running the Negative Binomial regression and developing safety performance function. They transform SPF to additive parameters of a non-exponential form in order to reach baseline collision rate which adjusted by contributing factors that are consisted in the model such as warning devices, surface type, and train speed (Park *et al*., 2005). In this case, developing one single SPF for each segment size will show a coefficient for lighting. In this method (non-exponential form), coefficients (for special countermeasure) bigger than unity, represent increase in accident counts, while coefficient factors less than unity yield a reduction in collision counts. The estimated vector of lighting in exponential form of developed SPFs, yield the amount of change in collision counts when the road segment changes from non-illuminated to illuminated one. For instance, the conventional NB regression model for 750 meters segment size from Equation 2.19 is:

$$Log(N_{750m}) = -5.42 - 1.16\,(Lu) + 0.96\,(AADT) - 0.04\,(G) + 0.26\,(In) - 0.41\,(W) + 0.48\,(cu) + 1.86\,(Res) \quad [2.19]$$

Where:

Lu: Presence of standard lighting

AADT: Average annual daily traffic

G: Glare

In: Presence of intersection

W: Total width of both lanes

Cu: Presence of curve

Res: Residential area

This equation is expressed based on the additive parameters of a non-exponential form. If Equation 2.19 transform into a multiplicative form, number of accidents could be expressed as Equation 2.20:

Collision rate(N_{750m}) = 0.004 + (0.31^{Lu}) + (2.61^{AADT}) + (0.96^{G}) + (1.3^{In}) + (0.66^{W}) + (1.62^{Cu}) + (6.46^{Res}) [2.20]

Derived from the generic intercept term, for instance, lighting has a value of $e^{-1.16} = 0.31$.

In this case, accident rate is adjusted by factors included in the model such as luminance level, average annual daily traffic, total width of the lanes and etc. Factors greater than unity interpreted as an increase in the estimates of number of crashes, while factors less than unity decrease the rate. Based on Equation 2.21 if the luminance level of road segment change from non-standard to standard level (greater than 1.2 cd/m^2) the number of accident would experience a reduction by factor of 0.69.

$$\text{CMF}_{Lighting}: \frac{number\ of\ acidents\ with\ standard\ lighting}{\text{number of accident without standard lighting}} =$$

$$\frac{\exp(-5.42-1.16\ Lu+0.96\ AADT-0.04\ G+0.26\ In-0.41\ W+0.48\ cu+1.86\ Res)}{\exp(-5.420.96\ \text{AADT}-0.04\ \text{G}+0.26\ \text{In}-0.41\ \text{W}+0.48\ \text{cu}+1.86\ \text{Res})} = exp\ (-1.16) = 0.31$$

[2.21]

2.4.3 Potential for Improvement

Potential for improvement can be expressed as the net difference between the observed number of collisions and the expected number of collisions (Hauer, 1997). The expected number of collisions must be estimated by calibrated safety performance functions of the corresponding type of site, per instance all those road segments (not containing intersections) on residential zones not currently being illuminated. The higher the value of potential for improvement, the more desirable is to apply a countermeasure to correct the specific safety deficiency being studied. Values between zero and one refer to sites where no benefits seem possible as they are already performing better than those treated.

2.5 Bayesian Data Integration on Road Safety

Full Bayesian analysis could be another alternative approach for statistical analysis. In this type of modeling the estimation of parameters is simulated by an algorithm such as the Gibbs sampler or Markov Chain Monte Carlo simulation (Park, 2007). The core components of any Bayesian analysis is: first, a prior distribution which is based on results from past studies. However, the selection of suitable prior to apply in Bayesian Regression model has been a concern in some cases (Lunn *et al.*, 2000; Bishop, 2007). Second, a likelihood distribution from data observed, and third, a posterior distribution is perceived by an integration of priors and likelihood distribution. When sufficient data is available for the likelihood the posterior could balance the risk of error caused by biased priors or limited time data observations, in other cases non-informative priors could be used to avoid such biasing. Several software is capable of conducting Full Bayesian analysis such as R (Albert, 2007), JAGS (2014), WinBUGS (2014) and OpenBUGS (2014), (Heydari, 2013).

The effect of different countermeasures for railway grade crossing has been estimated using Bayesian data fusion by Saccomano *et al.*, 2007. Equation 2.22 and Equation 2.23 introduce the estimation of posterior probability distribution in Bayesian method:

$$P\ (\alpha|\text{Data}) \propto f\ (\text{Data}|\alpha).P\ (\alpha) \qquad [2.22]$$

$$P\ (\alpha|\text{Data}) = \frac{(\text{Data}|\alpha).P\ (\alpha)}{\int (\text{Data}|\alpha).P\ (\alpha)da} \qquad [2.23]$$

Where, P (α) represents the prior distribution of α, P (Data|α) is likelihood function of sample data given parameter α, P (α|Data) is posterior distribution of parameter a given observed data; and the denominator represents the marginal likelihood.

Two sources of priori and data likelihood input is integrated to estimate posterior distribution. The posterior as an output is represented with the mean, the variance and the probability distribution. Hence, the uncertainty associated with the estimation process of any countermeasure could be recognized. Bayesian data fusion defined as a promising tool to choose safety decisions accurately, in case of uncertainties (Sacccomanno, 2007).

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter presents the methodology employed to estimate a reliable CMF for artificial road lighting in the event of cross sectional data. In particular, this research investigates the effect of different segment sizes on the estimated CMFs and model parameters. Section (3.2) describes the overall methodological approach in order to provide the reader with a summarized overview. Section (3.3) explains the approach employed to prepare the dataset (including roadway lighting, road characteristics and nighttime accident counts) and how the road network was divided into three different segment sizes. Section (3.4) reviews the process of accident modeling and statistical analyses including Bayesian Data Fusion used to learn a unique CMF value from local data. Section (3.5), explains the approach to identify which road segments will benefit from receiving lighting.

3.2 Overview of the methodology

This research proposes a two-step method to support the decision making of lighting as a countermeasure to night time collisions: First Collision modification factors can be learn from local observations by either a Before and After approach or as suggested by this research by a cross sectional method in the event of insufficient time series to support the before-after traditional method. This first step serves to quickly justify in general that lighting is expected to improve night time road safety. In a second step this method proposes a practical way to visualize which road segments will benefit the most from receiving lighting, hence enabling transportation officials to prioritize sites by potential for improvement (Figure 3.1). In the event of cross sectional data, there is a need to partition the road segments and use Bayesian Data Fusion to estimate a collision modification factor for the road network.

A negative binomial regression in conjunction with a system of binary filters for the sites characteristics, can serve to develop safety performance functions for both illuminated (SPF1) and non-illuminated roads (SPF2). The definition of illuminated complies with standard levels recommended by IESNA (2005).

Observed night time collisions (NTC) can be compared to expected levels of collisions given by both SPF1 and SPF2. Roadway lighting will be very beneficial in those cases where observed collisions at non lit sites (NTC) are higher than expected number of collisions at non illuminated sites (SPF2) Lighting will be not beneficial at those roads segments where NTC of non-lit sites are less than expected number of night time collisions at illuminated sites (SPF1). Finally, roadway lighting could signify somewhat beneficial at sites located in between both SPF1 and SPF2 (Figure 3.1)

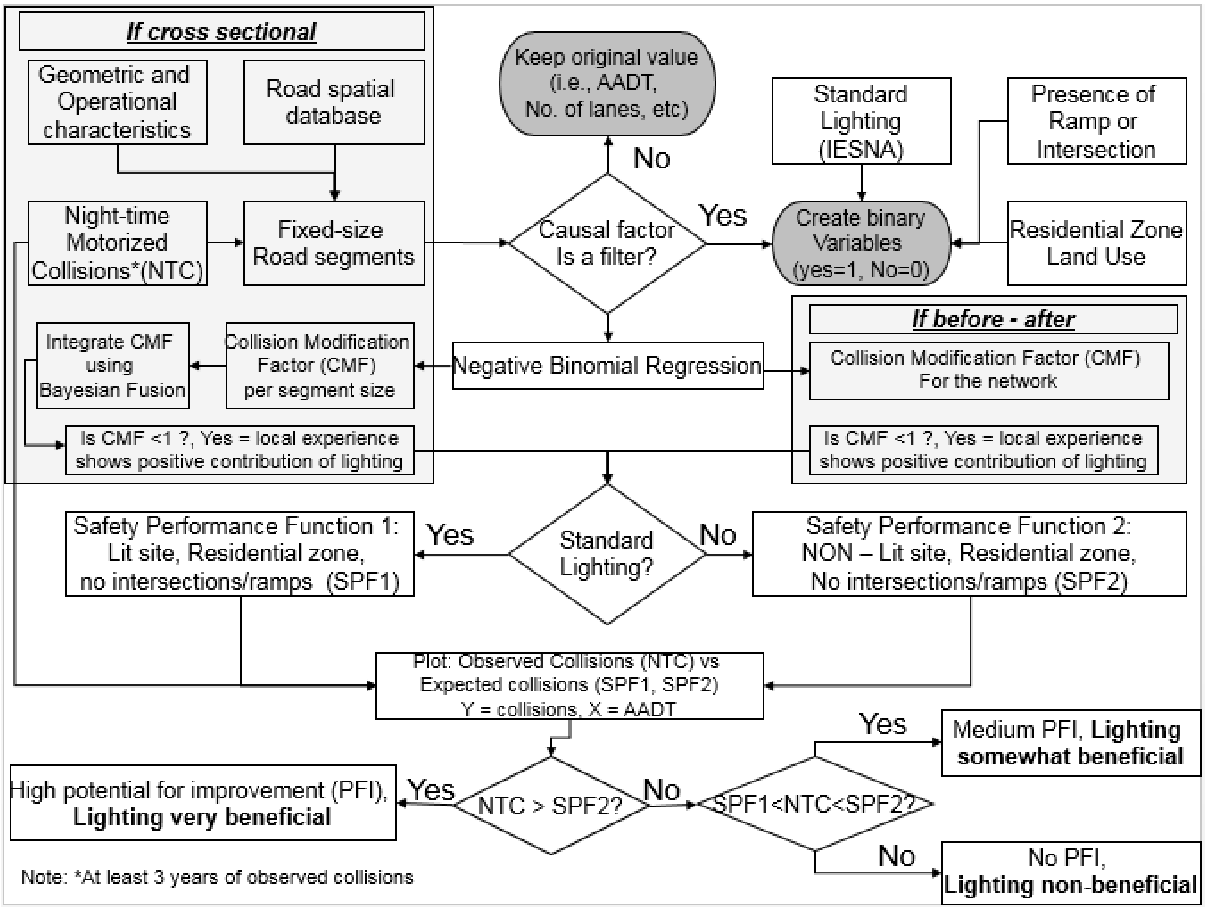

Figure 3.1: Methodology

In the event of before and after studies, there is no need to partition the segments and collision modification factors could be directly estimated from them. The rest of the method intended to identify the beneficial effects of lighting on specific segments is preserved as explained before.

The following sections provide methodological details related to the preparation of the database and the analysis conducted.

3.3 Database Preparation Procedure

The data were available in three different databases: night time collision (NTC) records, road attributes, and lighting data which was collected by other graduate students at Concordia University. The NTC was provided by the Ministry of Transportation of Quebec (MTQ) for a 5-year period, 2007 to 2011. Another dataset contained the attributes of major highways in Quebec; specifically routes 20, 40, 55, 105 and 132. The Base map of Quebec roads was provided by the Ministry of Transportation of Quebec (MTQ). Figure 3.2 shows the spatial location of major Quebec routes used in this study. Roadway lighting dataset was available in terms of luminance, illuminance and glare measurements.

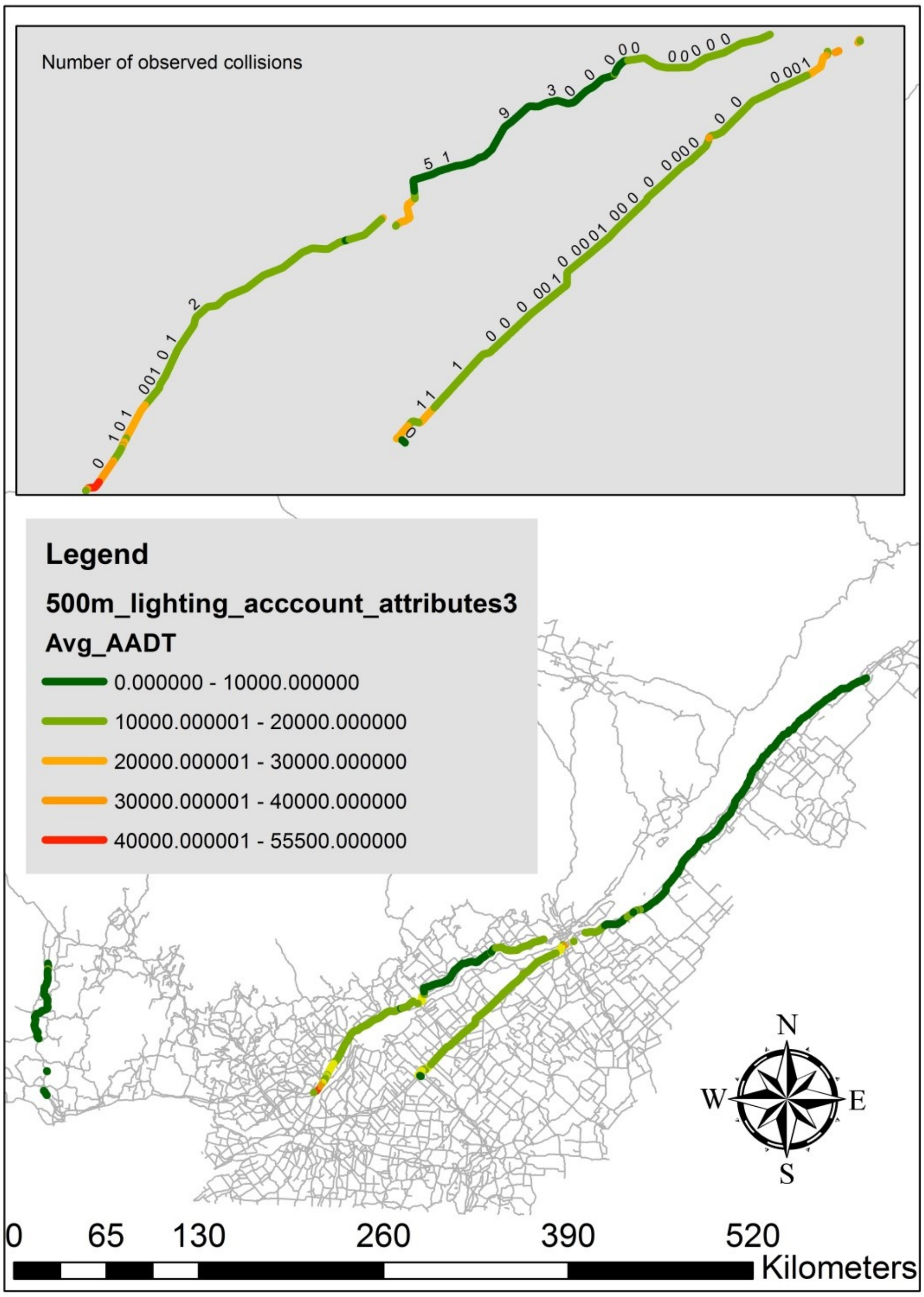

Figure 3.2: Quebec Roads on this study

The aim of the consolidated database was to combine the lighting data, accident data, and road segments data that contained geometric attributes such as total width and presence of shoulders. Lighting data was imported in *ArcMap10* using latitude and longitude coordinates from the North American Datum geographic coordinate system (NAD, 1983.prj) selected as a global coordinate system. Each displayed file was then converted into a shape file to be able to manipulate the original database when using geo-processing tools. The lighting dataset was merged with the road network by doing a spatial join and a summary of the closest lighting points was assigned to each road segment in order to reach an average of the lighting data on each route (Figure 3.3). After doing the spatial join, to incorporate nighttime crashes, the total number of nighttime accidents which is closest to each segment will be assigned to the corresponding segment size (Figure 3.4). All three datasets were merged together in ArcGIS to obtain a single database.

Only variables (i.e. columns) of interest were kept on the table of attributes on ArcGIS 10, to keep the road network dataset as clean as possible. Originally, the segments provided by the dataset are of unequal length. For the purpose of analysis and comparing, it was decided to divide the segments for each road into three different segment sizes. Routes were created for each numbered road and then split into segments of 250, 500 and 750 meters. Unfortunately, the production of routes on ARCGIS omits the original attributes allocated on each segment; hence it is needed to reintegrate those attributes once again from the original road network database. Therefore, the count of collisions is rejoined to each single segment. Also, luminance level for each segment calculated and assigned as average level of luminance. Another important variable which is considered in the analysis as a binary variable is the presence of intersection, which indicates whether the new road segment (after being split) crosses an intersection or not, because intersections had special characteristics which differ from those of mainline segments (Figure 3.5). The next step is to filter and retain the nighttime collisions to provide final database (Figure 3.6). The Data preparation procedure is illustrated in Figure 3.7.

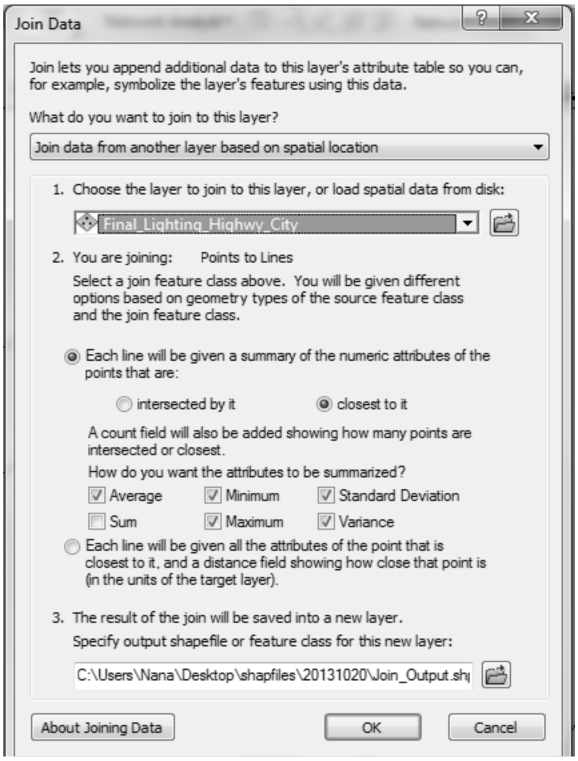

Figure 3.3: Spatial join of lighting data to roads

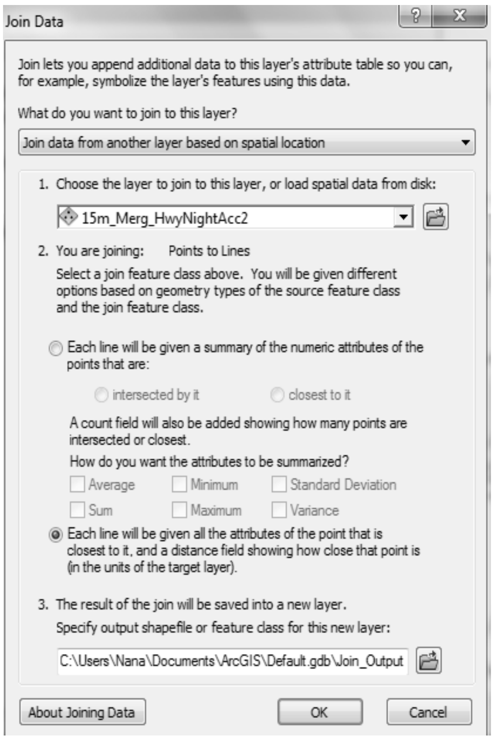

Figure 3.4: Spatial join of accidents to roads

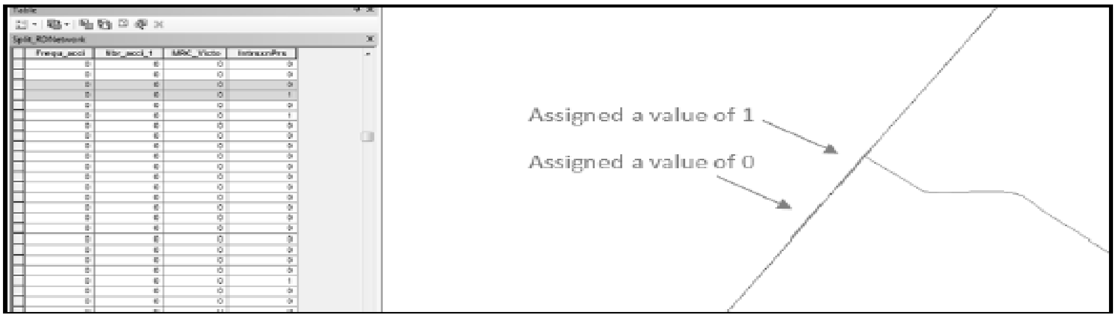

Figure 3.5: Identifying segments with crossing Intersections

Table

500m_lighting_acccount_attributes3

FID	Shape *	FID_1	RTNUMBER_2	Shape_Leng	avg_lum	Avg_glare	Sum_Acc_Nig	NBRLAN	inters	Avg_Tot_Wi	Avg_AADT	Curve
0	Polyline ZM	0	132	500.447956	0	0	0	2	0	7.185714	6900	0
1	Polyline ZM	1	20	500.606688	0	0	0	2	0	7.333333	13500	0
2	Polyline ZM	2	40	500.269723	0	0	4	2	0	7.2	11500	0
3	Polyline ZM	3	55	498.641345	0	0	2	1	0	6.133333	8296.6623	0
8	Polyline ZM	8	20	500.606688	0	0	0	2	0	7.4	13500	0
9	Polyline ZM	9	20	500.606688	0	0	0	2	0	7.266667	13500	0
10	Polyline ZM	10	20	500.606688	0	0	0	2	0	7.366667	13500	0

Figure 3.6: Sample data from ArcGIS (500 meters road segment group)

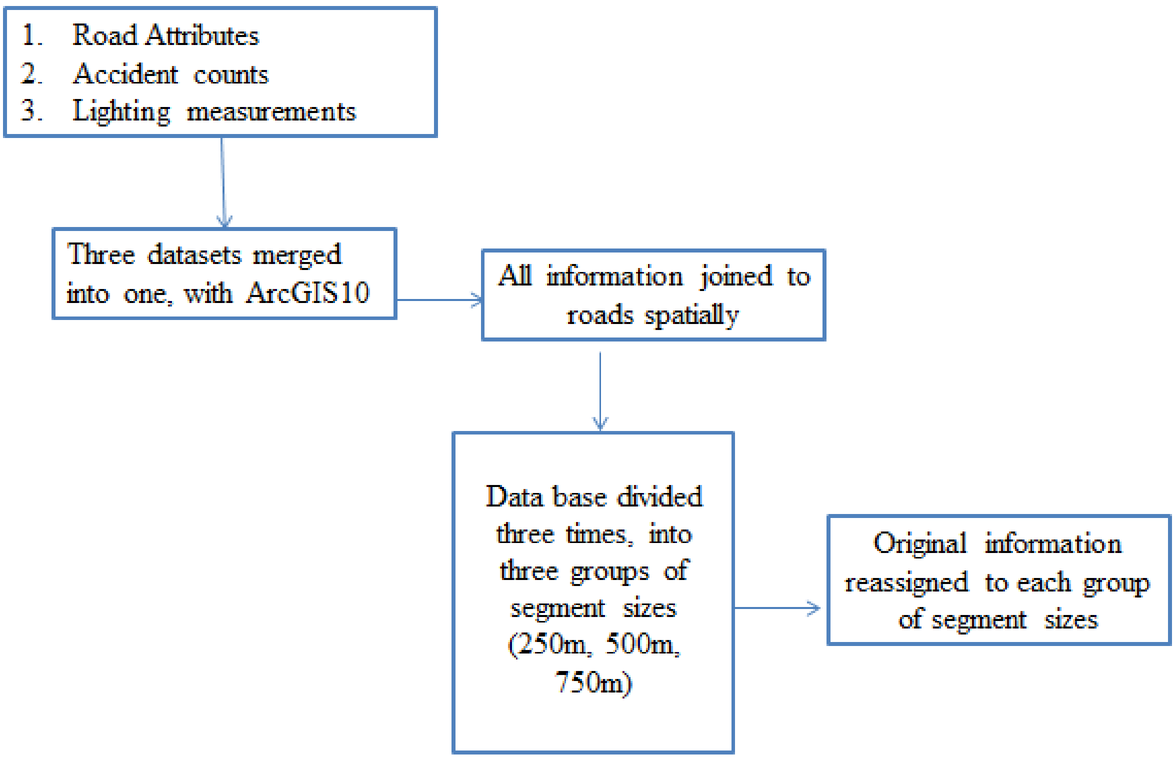

Figure 3.7: Data preparation steps

Illuminance and luminance are the two major lighting metrics used to measure the quality of the lighting on a road during nighttime (IESNA, 2005). Illuminance is the light quantity which is “the amount of light incident on the roadway surface from the roadway lighting system”. Luminance is the amount of light as perceived by the driver; it is the amount of light reflected by the pavement in the direction of the driver. It is also referred to as the “brightness” of the road (IESNA, 2005). Illuminance units are read in Lux and luminance is candela per square meter (cd/m2). As Rea *et al*. (2009) mentioned illuminance-based analysis is more appropriate to apply for roads with low levels of AADT and significant presence of pedestrians and bicyclists. In this study, luminance level is used to do the statistical analysis, because luminance level is advisable for statistical analysis of roads with higher volumes of motorized vehicles traffic.

According to the literature and analyses done in this study (Table 3.1) based on appropriate luminance level, a binary variable for luminance measurement is created.

Table 3.1: Luminance recommended level (IESNA, 2005)

Road and Pedestrian Conflict Area		Average Luminance (Lavg) (cd/m2)	Uniformity Ratio Lavg/Lmin (Maximum Allowed)	Uniformity Ratio LMax/Lmin (Maximum Allowed)	Veiling Luminance Ratio LVmax/Lavg (Maximum Allowed)
Road	Pedestrian Conflict Area				
Freeway Class A		0.6	3.5	6.0	0.3
Freeway Class B		0.4	3.5	6.0	0.3
Expressway	High	1.0	3.0	5.0	0.3
	Medium	0.8	3.0	5.0	0.3
	Low	0.6	3.5	6.0	0.3
Major	High	1.2	3.0	5.0	0.3
	Medium	0.9	3.0	5.0	0.3
	Low	0.6	3.5	6.0	0.3
Collector	High	0.8	3.0	5.0	0.4
	Medium	0.6	3.5	6.0	0.4
	Low	0.4	4.0	8.0	0.4
Local	High	0.6	6.0	10.0	0.4
	Medium	0.5	6.0	10.0	0.4
	Low	0.3	6.0	10.0	0.4

3.4 Night time collision modeling

The Negative Binomial regression is applied in this study to analyse crash frequencies while accounting for over-dispersion. Based on estimated coefficients from regression analysis safety performance functions (SPF) developed for each group of segment sizes to investigate the correlation between accident counts and independent variables (i.e. AADT, luminance level, presence of intersection and curve, land use and etc. in this study). In this study, the SPF is developed for each group of segment sizes based on Equation 3.1 presented by Miranda-Moreno, (2013).

$$Acc = L_i \ \exp(\beta_0 + \beta_1 lnAADT_i + \beta_2 x_{i1} + \cdots^{+} \beta_k x_{ik}) + \varepsilon \qquad [3.1]$$

Where:

$L_{i:}$ Segment i length

β_0 : Constant term

k : variable number (1,2,3,….)

β_k : Coefficient of explanatory variable x_k ,

Acc: Frequency or severity of night-time collisions on segments i,

$AADT_i$: Average Annual Daily Traffic of segments i,

$X_{ki:}$ Explanatory variable i

a: Coefficient of AADT at segment i

ε: Error term

The coefficient of lighting obtained from exponential form of SPFs, represents the value of lighting crash modification factor (Park *et al*., 2005).

3.4.1 The Standard Error and Confidence Intervals of CMFs

To estimate the confidence interval for safety effectiveness for a particular variable, first, standard error should be calculated. Standard error in literature is expressed as the calculated standard deviation of the difference between estimated values and values comes from sample data. The standard error of the CMF is calculated as the standard deviation of related variable coefficient which comes from the Stata table of results, multiplied by the estimated crash modification factor for that variable. In this study, based on Table 7, the standard errors are estimated from Equations 3.2, 3.3, 3.4:

250m segment size: $exp^{(-1.13)} \times 0.22 = 0.07$ [3.2]

500m segment size: $exp^{(-1.32)} \times 0.21 = 0.05$ [3.3]

750m segment size: $exp^{(-1.16)} \times 0.34 = 0.1$ [3.4]

A smaller the standard error indicates a more precise estimate. The Standard error is applied to calculate the confidence interval of crash modification factors based on Equation 3.5 and values from Table 3.2 (HSM, 2010):

$CI(X\ \%) = CMF \pm (SE \times MSE)$, [3.5]

Where CI(X %) defined as confidence interval,

CMF = crash modification factor;

SE = standard error of the CMF, and

MSE = multiple of standard error which defines from table

Table 3.2: Recommended confidence intervals and standard error. HSM (2010)

Desired Level of Confidence	Confidence Interval (probability that the true value is within the estimated intervals)	Multiple of Standard Error (MSE)
Low	65%-70%	1
Medium	90%	2
High	99.9%	3

For instance, the 90% confidence interval for crash modification factor of 250 meters segment size is calculated based on Table 3.2 and Equation 3.6:

$$CI\ (\%) = exp^{(-1.13)} \pm (0.07*2) = 0.32 \pm 0.14 \qquad [3.6]$$

3.4.2 Bayesian Data Fusion

To better capture the uncertainty related to the safety effects of countermeasures, the Bayesian data fusion technique is used to combine different sources of information. This approach is structured based on prior, likelihood and posterior. The prior distribution is the findings from past studies relating to the safety effectiveness of a variable of interest. The likelihood function is the locally observed data that could describe difference in accident occurrence at a particular location with different lighting level. The last step is to define the posterior distribution by mixing the prior and likelihood (Figure 3.8).

The countermeasure effects based on statistical analysis of accident models and observations can be combined by applying the Bayesian data fusion approach. The objective is to generate "posterior" estimates of the probability among past studies and estimated safety effectiveness of lighting for three different segment sizes and obtain a single precise value based on observed data and past studies. The posterior expression is defined as follows (Migon and Gamerman, 1999; Lee, 2004):

ρ_i (θ|x) ρ_i (θ) ρ_i (x| θ) [3.7]

Where: θ is safety effectiveness of lighting

x = Comes from accident prediction models

ρ_i (θ) = Prior probabilities of θ from past studies

ρ_i (x|θ) = Probability of observing the sample data

ρ_i (θ|x) = Posterior probability of θ give x.

In this study, different statistical models using the negative binomial regression were used for our data considering three segment sizes to calculate CMFs, and such CMF represents the likelihood distribution characterized by mean and variance which will be combine with the prior.

To reduce the computational complexity and avoid the MCMC simulations to draw the posterior, one can assume that both the prior and the data likelihood distributions are normally distributed. Suppose the prior $\theta \sim N(\mu, \tau^2)$ and the data likelihood as $[l \sim N(x, \delta^2)]$, then these distributions can be integrated analytically to estimate the posterior, which is also normally distributed with the mean of μ_0 and the variance of τ_1^2. Equations below shows how the posterior is estimated in this study (Saccomanno *et al.*, 2007):

Variance, $\tau_1^2 = (\tau^{-2} + \delta^{-2})^{-1}$ [3.8]

Mean, $\mu_0 = (\tau^{-2} \mu + \delta^{-2}{}_x) \tau_1^2$ [3.9]

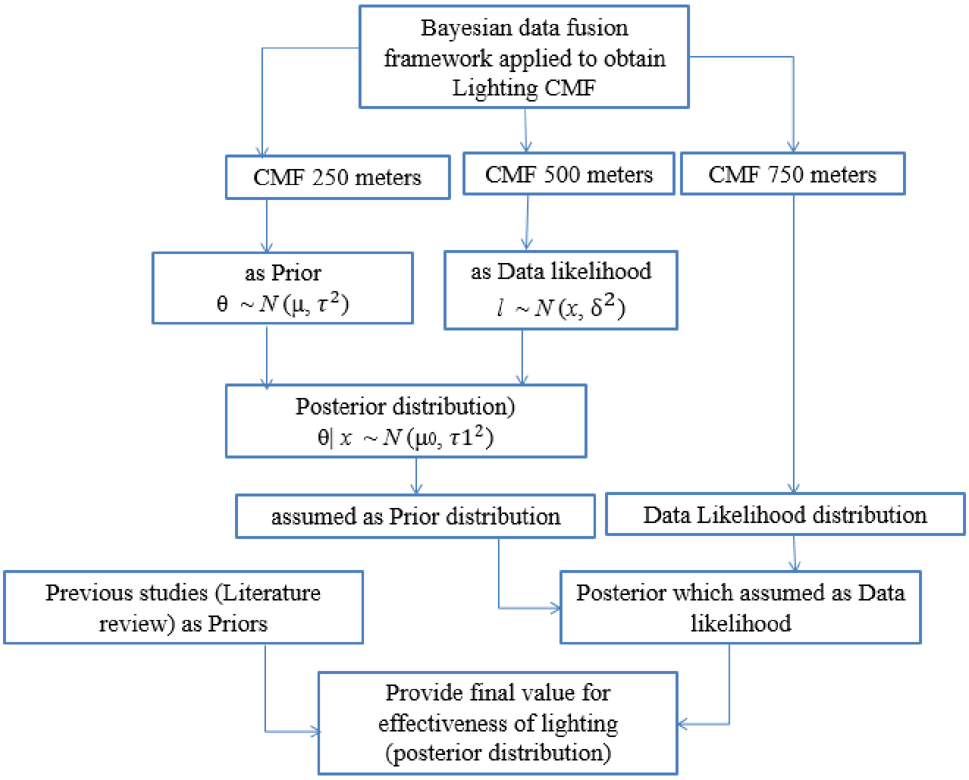

Figure 3.8: Bayesian data Fusion steps (Park, 2007)

As shown in Figure 3.8 data fusion is applied as follows: a CMF value is obtained for each segment size (250 m, 500 m, and 750 m) based on the negative binomial regression. We then combined the CMF results for the first two segment sizes: 250 m and 500 m. In this case, the CMF for the 250 m segment scenario acts as the prior and the CMF for the 500 m segment scenario acts as the likelihood. The posterior obtained from these two segment sizes was used as the prior for the 750 m segment size, being the likelihood. This way, an overall (single) CMF value was estimated. The latter estimate was obtained from the data so that it could be considered as the likelihood that was then combined with the prior obtained from previous studies. Note that different priors were considered here considering different weights since methods differ from one study to another and all studies are not the same in terms of the reliability of estimates (See Table 2.1).

According to literature the uncertainty is always associated while studying the effect of countermeasures. It is not possible to estimate a completely accurate CMF value by applying statistical accident prediction models (Button and Reilly, 2000; Leeming and Saccomanno, 1994). The variance has been used by past studies to provide a range of possible values for the CMF and the associated uncertainty. However, the Bayesian data fusion method is applied in this study to provide probability distributions to represent the CMF of road lighting while providing both the range of CMF values and the likelihood of the estimation.

As mentioned above, the CMF is assumed to follow a normal distribution when using the Bayesian data fusion technique. However, this may not be true when describing unknown CMF distributions (Park, 2007). Also, the fact that there might be an inherent uncertainty associated with estimating priors from different sources could affect the posterior distribution (Park, 2007). According to Lee (2004), based on the central limit theorem, the observations with errors (CMFs in this study) could be assumed to follow a normal distribution. The normality assumption can be useful here to avoid computationally intensive MCMC methods (Gelman *et al.*, 2004). However, this might be violated in reality; therefore, other researches assumed different distributions such as a beta distribution due to its flexibility to express the prior and posterior distribution without any assumption of symmetry (Clarke and Sarasua, 2003). We therefore verified the sensitivity of the results to the distributional assumption by assuming a beta distribution instead of a normal density. A beta distribution can be expressed as follow, where r and s are two shape parameters:

$$\Pr(X = x,r,s) = \frac{(r+s-1)!}{(r-1)!(s-1)!} X^{r-1} \ (1- X)^{r-1} \qquad [3.10]$$

Where the mean μ and the variance can be obtained from equations 3.11 and 3.12:

$$\mu = \frac{r}{r+s} \qquad [3.11]$$

$$\delta^2 = \frac{\mu\,(1-\mu)}{r+s+1} \qquad [3.12]$$

Similar to the normality assumption, using a beta density allows us to compute the posterior analytically without the need to run MCMC simulations. The parameters for the posterior (final CMF estimate) can be estimated using the following equations:

$$r_{posterior} = r_{prior} + r_{data\ likelihood} \qquad [3.13]$$

$$s_{posterior} = s_{prior} + s_{data\ likelihood} \qquad [3.14]$$

After obtaining the posterior, its mean and variance can be computed as follows:

$$r = \mu\,[\frac{\mu\,(1-\mu)}{\delta 2} - 1] \qquad [3.15]$$

$$s = [1-\mu]\,[\frac{\mu\,(1-\mu)}{\delta 2} - 1] \qquad [3.16]$$

Here, we adopted the same approach discussed for the normal assumption scenario.

3.5 When to Provide Roadway Lighting

Roadway llighting should be provided to those segments that will experience the largest amount of night time collisions reductions. A system of binary filters can be used to estimate safety performance functions specific to desired land use zoning and road sites (Figure 3.9).

A Negative Binomial regression analysis can be used to obtain coefficients to calibrate safety performance functions in which traffic volume can be used as independent variable.

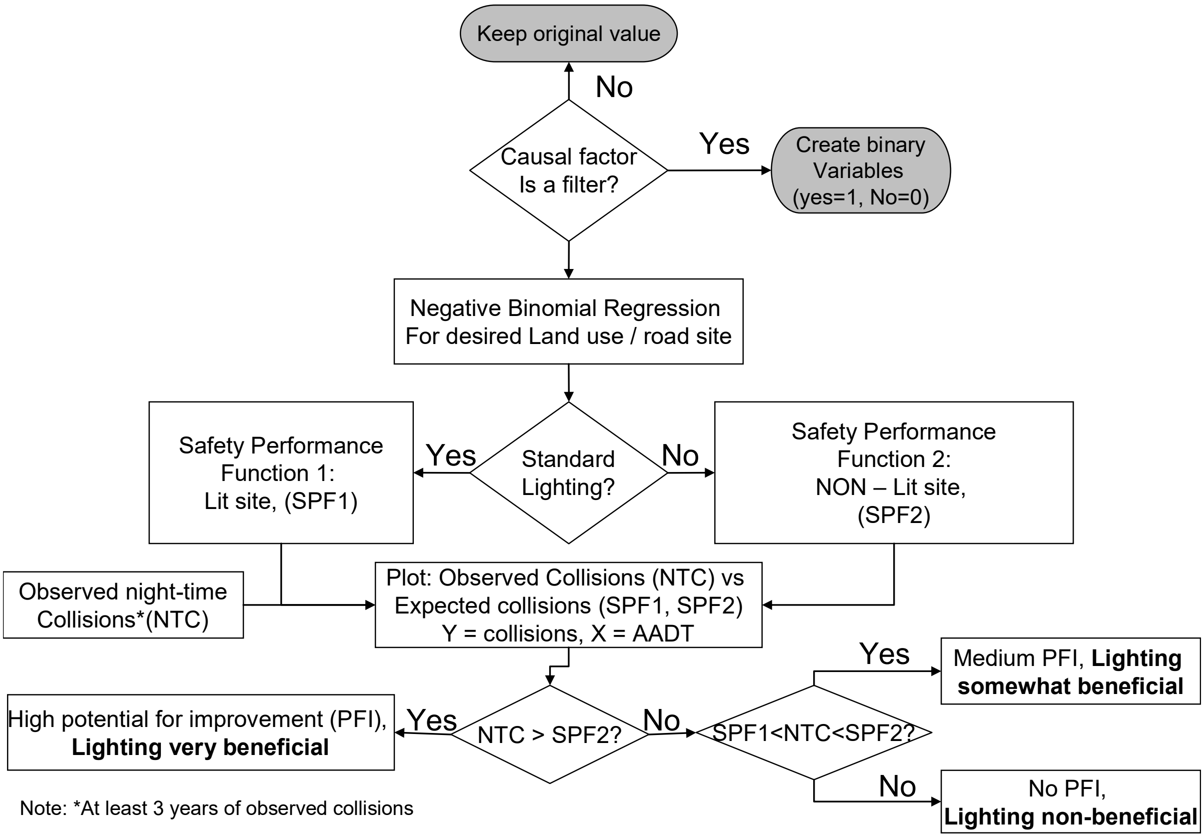

Figure 3.9: When to provide roadway lighting

Two safety performance functions can be developed: one for lit segments (SPF1) and one for non-illuminated segments (SPF2). Observed number of collisions can be plotted on the same graph. Potential for improvement can be easily learnt by comparing the position of observed sites versus expected levels for non-illuminated. Those observed sites above the non-illuminated expected trend reveal roads that will gain substantial benefits from being illuminated. Those below the SPF1 line exhibit No potential for improvement. Those segments located between SPF1 and SPF2 show little potential for improvement (Figure 3.10).

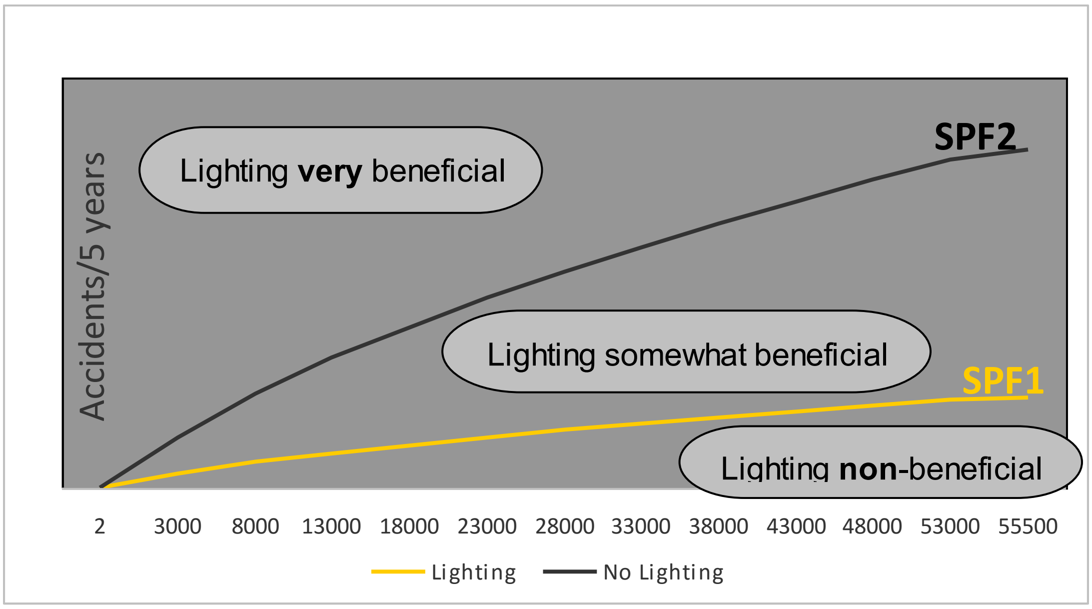

Figure 3.10: Potential for improvement and lighting beneficial effects

CHAPTER 4: RESULTS

4.1 Introduction

This chapter illustrates the results of analyses obtained from the adopted case study. First, the observed, collected and prepared cross-sectional data has been used to create different road segment samples based on different segment sizes. Second, database is used to develop SPFS and represent the road lighting CMF estimates. Finally, the analyses to obtain the CMF for roadway lighting are presented. Finally, safety performance functions are used to produce models of expected number of collisions and compared to observed night time collisions to rank road's potential for improvement.

4.2 Database Preparation

Table 4.1 represents the type of variables applied in this study to model the accident data. Some of the geometric and design characteristics of road segments, such as lighting measurements (i.e. luminance, glare) and complex geometries; for instance: the presence of a curve and an intersection is employed to develop SPFs. However, only some site characteristics were available or make sense. For example, data available for presence of median is defined based on the painted line in the middle of road lanes, but not the presence of physical barriers. Table 4.2 shows a sample summary of data observed which contains 1385, five hundred meters segments sizes.

Table 4.1: Sample row of database

Nighttime Accident	Average Luminance	Ln_AADT	Glare	Presence of intersection	Total lane width	Presence of curve	Residential zone
Discrete	Binary	Continuous	Continuous	Binary	Continuous	Binary	Binary
22	1	10.1849	3.16	1	7.462	0	1

In this study, Negative Binomial regression model is applied to group of datasets including three different segment sizes to investigate the impact of various contributing variables on the number of nighttime road collisions. The table below summarizes road characteristics and lighting measurements used as explanatory variables. Nighttime collisions counts were used as response for the regression analysis with counts varying according to each segment size. Geometry attributes included: total road width, presence of curve, segment crossing an intersection (i.e. presence of intersection), land use (i.e. residential), and average AADT. The lighting related measurements included the luminance level and glare measurement.

Table 4.2: Summary of response and explanatory factors

Variable	Mean	Std. Dev.	Min	Max
Nighttime Accident	4.0902	11.4176	0	162
Luminance	0.1241	0.2958	0	3.76
Traffic flow (AADT)	10816.22	6401.05	1375	55500
Glare	3.4394	5.2264	0	20
Intersection	0.4498	0.4914	0	1
Total lane width	7.4246	0.7446	5	12
Curve	0.1045	0.1068	0	1
Residential zone	0.0754	0.1018	0	1

A correlation matrix describing the degree of relationship between independent variables is shown in Table 4.3. If any of the variable shows the correlation of 0.7 and above, with one or more other variables, it had to be dropped from the analysis. As expected, there is a perfect correlation between luminance and illuminance levels; hence only one could be used in the analysis. Also, the correlation matrix illustrates a strong positive correlation between the number of lanes and total width (0.9), therefore the number of lanes is dropped and "total width" of the road was kept in the analysis. The presence of an intersection, traffic volume (AADT) and being at a residential zone showed values of correlation ranging from 1.1389 to 0.4183 with the response suggesting explanatory abilities. In addition, presence of an intersection and AADT were correlated as expected given that at those sites one will count more vehicles coming into the main road or highway.

Table 4.3: Correlation matrix

	Number of Nighttime Accidents	Average luminance level	Average illumination level	Glare	Number of lanes	Presence of Intersection	Total width	AADT	Speed	Presence of Curve	Residential zone
Number of nighttime accidents	1.000										
Average luminance level	-0.0021	1.000									
Average illumination level	-0.0025	1.000	1.000								
Glare	0.0822	0.0191	-0.0232	1.000							
Number of lanes	0.0555	-0.0012	-0.0014	0.0701	1.000						
Presence of Intersection	0.1389	0.0259	0.0248	0.2564	0.0322	1.000					
Total width	0.0439	0.0016	0.0014	0.0893	**0.9014**	0.0336	1.000				
AADT	0.2110	-0.0121	-0.0125	0.0712	0.1315	0.579	0.0924	1.000			
Speed	-0.0368	0.0055	0.0058	0.0709	-0.119	-0.0551	-0.064	-0.010	1.000		
Presence of Curve	0.0056	-0.0045	-0.0046	0.0165	0.1680	-0.0455	0.1553	-0.041	-0.0251	1.000	
Residential zone	0.4183	-0.0012	-0.0019	0.1634	0.1258	0.1330	0.0563	0.1971	-0.1875	-0.0140	1.000

After analyzing data, safety performance functions were developed for each group of segment sizes based on results obtained in Table 4.4, as equations below:

250 meters: *exp* $^{(-6.2-1.13*(Lu) + 0.8*(AADT) - 0.03*(g) +0.34*(in)-0.19*(w) ++2.61*(res))}$ [4.1]

500 meters: *exp* $^{(-4.89-1.32*(Lu) +0.66*(AADT) -0.03*(g) +0.39*(in)-0.13*(w) +0.33*(cu) +2.47*(res))}$ [4.2]

750 meters: *exp* $^{(-5.42-1.16*(Lu) +0.96*(AADT) -0.04*(g) +0.26*(in)-0.41*(w) +0.48*(cu) +1.86*(res))}$ [4.3]

Where: Lu = Presence of standard lighting, AADT= Average annual daily traffic, G= Glare, In= Presence of Intersection, W= Total width in both lane, Cu= Presence of curve, Res= Residential area. Table 4.4 illustrates independent variables and coefficients of road segments.

Table 4.4: Statistical analysis for explanatory variables of night time collisions

250 M Segment			**500 M Segment**		**750 M Segment**	
Variable	Coefficient	P>z	Coefficient	P>z	Coefficient	P>z
Luminance	-1.13	0	-1.32	0	-1.16	0.001
lnAADT	0.79	0	0.66	0	0.96	0
Glare	-0.03	0.002	-0.03	0.001	-0.04	0.032
Presence of Intersection	0.34	0.004	0.39	0.003	0.26	0.16
Total Width	-0.19	0.004	-0.13	0.058	-0.41	0.001
Presence of Curve			0.33	0.077	0.48	0.039
Residential area	2.61	0	2.47	0	1.86	0
Cons	-6.2	0	-4.89	0	-5.42	0.003
lnalpha	1.65		1.37		1.42	
alpha	5.21		3.96		4.15	

As seen on Table 4.4, lighting in the form of luminance is a significant ($p = 0.001$) factor capable of explaining night time road collisions; for all segment size cases the effect of more luminance is a reduction in the number of collisions with a larger value for the 500m segment. It must be noticed however that the data was not standardized and as such the explanatory power of AADT cannot be directly compared (but indirectly through natural logarithm) to the rest of the factors. For all analyses more traffic volume (lnAADT) explain more frequent night time collisions, similarly happens for glare which has a mild countermeasure effect on night time collisions as well as total width of the road. The presence of a curve only explains more frequent collisions at night for analysis with segment sizes of 500m or 750m. It is very interesting to see how being on a residential environment explain more frequent nighttime collisions but the presence of intersections only explains more frequent collisions for analyses with segment sizes of 250m and 500m with insignificant ($p=0.16$) explanatory power for larger segments given the fact that any one 750m segment with an intersection is far too long to distinguish between the actual location of the intersection and the collision.

4.3 Crash Modification Factor for Lighting

The problem with obtaining the coefficients for lighting and other explanatory variables is that they only reveal whether sites without good lighting result in more collisions than their counterparts with higher levels of luminance, but little is known, as matter of fact nothing is said in the event that a poorly lit or completely dark road segment receives roadway lighting. Given that the data available is of cross-sectional nature, safety effectiveness can be calculated with the expected number of collisions, for three different segment sizes. The collision modification factor for artificial roadway lighting can be estimated for different measured safety ratios among three segment sizes and then compared among them to gain some understanding on the role of the segment size on the value of the CMF. Then one can analytically combine three different CMFs in order to obtain a single one. Table 4.5 summarizes the obtained crash modification factor ratios which calculated from exponential form of SPFs, for the three studied segment sizes and lighting level of highways in the province of Quebec.

Table 4.5: CMF value for each group segment sizes

Segment size	250m	500m	750m
CMF	0.32	0.26	0.31
Sd	0.07	0.05	0.1
Var	0.005	0.002	0.01
Safety Effectiveness	67.7%	73.28%	0.68%

It is interesting to observe that the collision modification factor of 500m segments is lower than those of 250m and 750m (which interestingly reach almost the same value). This leads us to think that using a 500m segment is more conservative. From the CMF one can estimate the safety effectiveness of roadway lighting for both 250m and 750m segments around 67% reduction of night time collisions (Table 4.5).

Figure 4.1 shows the statistical distribution of the three values of CMF for roadway lighting. As seen those of 750 show larger variability associated to higher uncertainty, meanwhile those of 250m segments show tighter spread being surpassed by those of the 500m segment group. Hence, it appears reasonable to conclude that the CMF value for the 500 meters group is the more certain. In addition, all three distributions are placed on the positive region of collision modification factors (Figure 4.1) with the mean value of segment 5 being smaller therefore resulting in a better safety effectiveness. This finding appears to suggest that the 500 meters segment is the best performer.

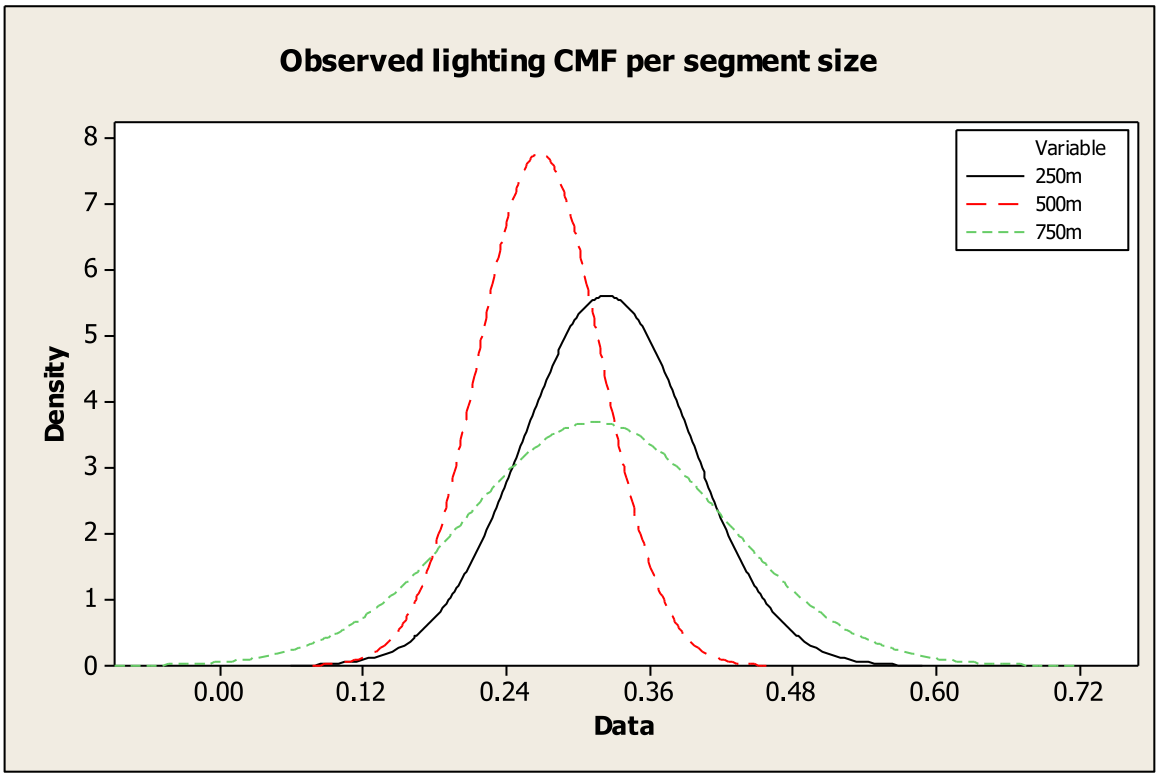

Figure 4.1: Original analysis with 100 percent of the segments

4.4 Integration of CMFs

4.4.1 Data Likelihood Distribution

At this stage of the analysis the reader may wonder which is the ideal segment size, however the response to this question depends on the accuracy of the geographical location of the roadway collisions observed. Hence, a different route can be taken; one can estimate a CMF that disregards the segment size. By integrating individual CMF as shown on figure 3.8, first the CMF obtained from studying 250 meters segment size is integrated with the one obtained from 500 meters segment size.

Second, the calculated CMF is integrated with CMF from studying 750 meters segment size. The mathematical process is showed in equations below based on Bayesian data fusion framework:

Step one: $\tau_1^2 = (0.07^{-2} + 0.05^{-2})^{-1} = 0.001$ [4.4]

$\mu_{0} = (0.07^{-2} * 0.32 + 0.05^{-2} * 0.26)\ 0.07^2 = 0.28$ [4.5]

Step two: $\tau_1^2 = (0.04^{-2} + 0.1^{-2})^{-1} = 0.001$ [4.6]

$\mu_{0} = (0.04^{-2} * 0.28 + 0.1^{-2} * 0.31)\ 0.04^2 = 0.29$ [4.7]

As a result, the estimated data likelihood distribution for lighting crash modification factor follows $N\,(0.29, 0.039^2)$. The next step is to estimate prior distribution from past studies.

4.4.2 Prior Distribution

In this research, 12 recent studies for lighting as countermeasure were combined to reach a prior distribution with defined mean and variance, (See Table 2.1). The crash modification factors estimated based on Eq. 2.18, from the lighting safety effectiveness values available for each study. Moreover, in this study, to provide accurate estimation of priors, crash modification factors from past studies combined based on the relative weight for methods applied (Saccomanno *et al.*, 2007). Equations 4.9 and 4.10 show how prior distribution is developed based on assumed weights, (see Table 4.6).

$\mu = \sum W_i\ CMF_i / \sum W_i$ [4.9]

Where, μ is weighted average effectiveness of lighting,

W_i is relative study weight for lighting in level of certainty,

CMF_i is the weighted average effectiveness for lighting from all available sources.

$$\delta = \sum W_i \delta_i / \sum Wi \qquad [4.10]$$

Where, δ is weighted standard deviation for lighting,

W_i is relative study weight for lighting in level of certainty,

δi is the weighted average standard deviation for lighting from all available sources.

The average mean and variance for prior distribution of CMFs estimated as 0.76 and 0.11, respectively as equation 4.11.

$$\text{Weighted Mean}_{\text{Prior}} = (0.72+0.96+0.62+0.69+0.7)\times 0.5 + (0.88+0.75+0.61+0.836+0.905+0.72$$

$$0.83)\times 0.33 / (5\times 0.5) + (7\times 0.33) = 0.76 \qquad [4.11]$$

Table 4.6: Relative weight based on analysis method (Saccomanno *et al*., 2007)

Level of certainty (i)	Methodology	Relative study Weight (W_i)
Medium-High	Before-After	0.5
Medium-Low	Cross-Sectional	0.33

4.4.3 Posterior Distribution

To obtain the posterior distribution, the final roadway lighting CMF from cross-sectional data is combined with past studies assuming that both distributions are normal. Since the prior distribution is estimated as $\theta \sim N\,(0.76, 0.1^2)$ and the data likelihood as $l \sim N\,(0.24, 0.039^2)$, then the posterior distribution is computed as follows, based on Equation 3.2 and Equation 3.3.

$$\tau_1^2 = (0.1^{-2} + 0.039^{-2})^{-1} = 0.037^2 \qquad [4.12]$$

$$\mu_0 = (0.1^{-2} \times 0.76 + 0.039^{-2} \times 0.29)\, 0.037^2 = 0.3 \qquad [4.13]$$

An Excel spreadsheet was used to combine prior and likelihood to produce the posterior (appendix).

Table 4.7: Final roadway lighting CMF distribution

Distribution Assumption (mean, variance)	Posterior Distribution (Final CMF)	Final Safety Effectiveness	Variance	Standard Deviation
Normal	0.34	65.32	0.001	0.038
Beta	0.33	66.04	0.001	0.040

Table 4.7 illustrates the posterior distribution of roadway lighting CMF with mean and variance which is assumed to follow once the normal distribution and secondly the beta distribution. Both results are found very similar. Based on calculated final lighting CMF, roadway lighting could reduce the number of night time collisions up to 65%, approximately. As can be seen, from Table 4.7 the final value for safety effectiveness of lighting (65%) is closer to the value obtained from analyzing the 250 m and the 750 m segment size samples.

4.5 The Role of Segment Sizes on CMF

Segment size plays an important role in statistical analysis of road safety. However, there is a lack of consensus among practitioners and researchers regarding the advisable range of values of segment sizes to partition a road network. The results of the analysis could be affected by the size chosen for the segments, and lead to obtain different results for the estimated coefficients of the SPFs, and corresponding crash modification factors and the safety effectiveness of different countermeasures.

At first one can expected that smaller segment sizes will contain more zero-accidents-counts which could cause some errors during accident modeling. Also, it becomes difficult (if not impossible) to capture geometric and operational attributes when studying small segment sizes. For example, when studying a collision reported in a short segment size, the segment length could be smaller than the actual curve's length, failing to register the changes in super-elevations experienced by the driver or any other deficient operational or geometric characteristic responsible for the collision. However, measured lighting levels benefit from smaller segment sizes as they exhibit more accurately the actual values rather than average's estimated over long segments that could experience large variations of lighting levels. The reverse statement is also true; for long or very long segment sizes; which could capture characteristics of two or more sites: per instance a curve and a straight segment will get their characteristics averaged, or a straight segment and an intersection. The average lighting level of a long segment is likely not to be accurate given the lack of relation between the specific site of the crash and the average amount of lighting estimated. Hence, it is expected that a road segment with medium size could overcome this sources of errors. The geometric and operational road characteristic could be captured clearly in medium segments sizes as they pertain to the type of site that is immediately located before the collision and so it is expected that the observe geometry and operational characteristics reveal better depiction of the deficient elements that could have led to the collision. For the case of medium segment sizes, the average lighting level would be somewhere in between those of large and short segments; that is an average value that acceptably reflects the visibility of the site.

4.6 Identifying the Ideal Segment Size

Two samples of data were withdrawn from the main data set; the first sample consisted on a random 60% sample without replacement of the observed segments containing all their corresponding characteristics. The second subset was a random sample of 40% of the observations use for validation. Statistical software state was used to create these subsets.

The previous analysis was repeated, first obtaining CMF values for each segment size as shown on Figure 4.2 for the 60% sample and on Figure 4.3 for the 40% sample. Motivated by the desire to abstract from the dependency to the segment size, all three values were combined to obtain a unique value through the Bayesian Data Fusion approach as previously done. This final result is left for the last portion of this section, first the analysis concentrates in the individual segment sizes results and the agreement to the prior expectations.

When estimating crash modification factors, one would expect that the CMF value is positive, hence, the values of the 95% of the normal curve of the CMF should lay on the positive region. In addition, it is expected that the distribution of lighting crash modification factors (CMF) are related to the segment sizes: larger segment sizes are expected not to be so accurate; meanwhile very small segments are expected to show inconclusive evidence of positive modification.

As seen on Figure 4.2 the trend observed on the previous analysis is repeated on the 60% sample results, the 500 meters segment group performs better and has less uncertainty than the 250 meters group and the 750 meters group, correspondingly. The 750 meters group not only has the poorest performance (largest CMF) but also the highest degree of uncertainty.

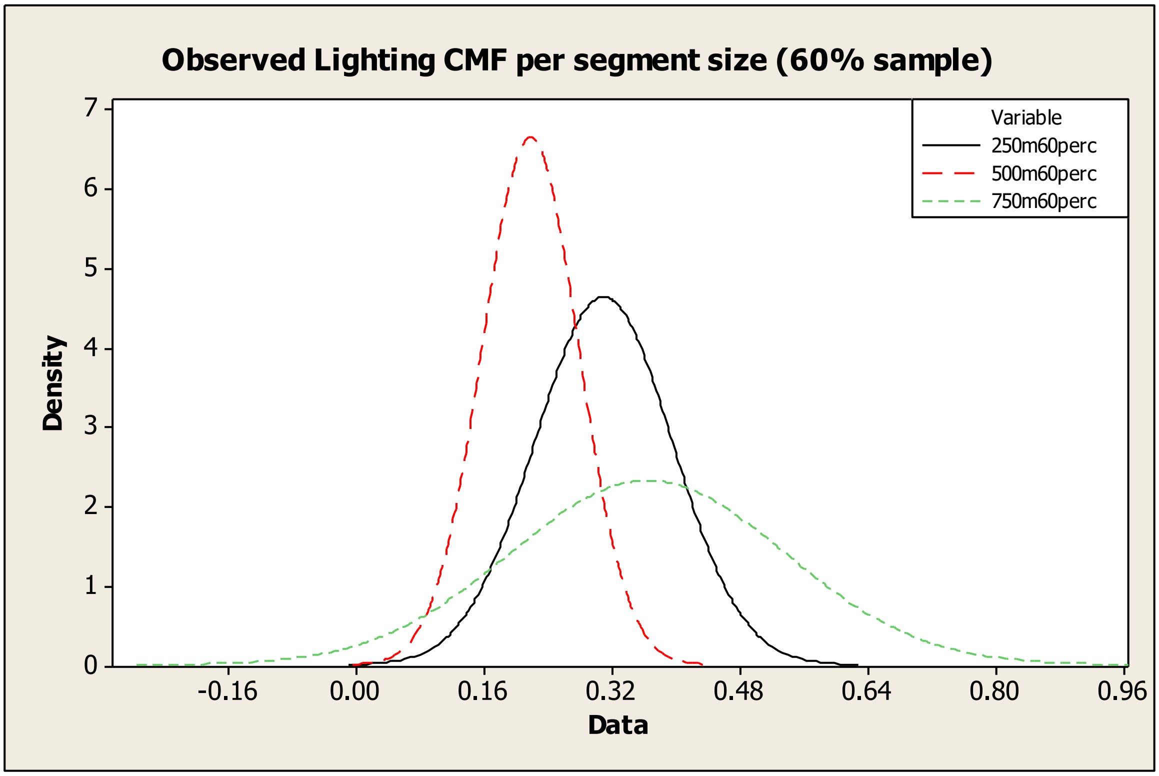

Figure 4.2: Observed distributions for lighting CMF for sample of 60% segments

For the 40% sample on Figure 4.3, one can observe different results, the 750 meters has the smallest average but the most variation as before, meanwhile the 250 meters group has slightly better certainty but the highest CMF value (worst) and the 500 meters is placed in between both of them but with the highest level of certainty.

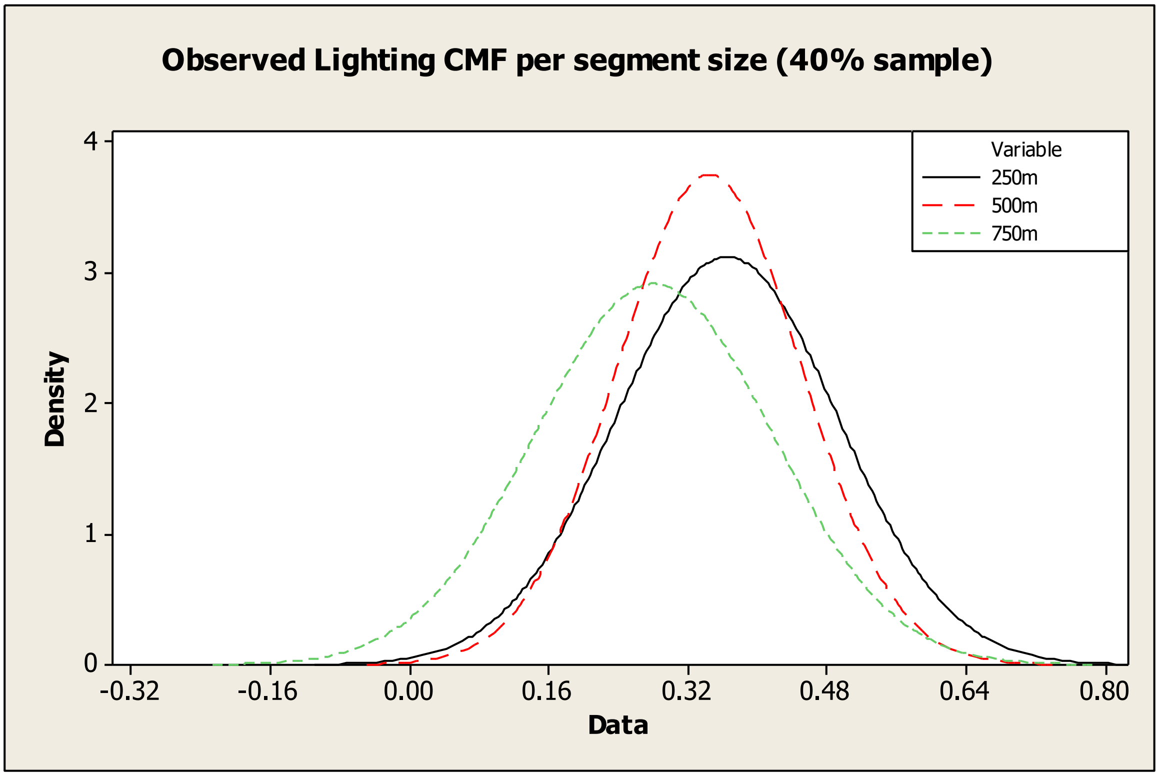

Figure 4.3: Observed distributions for lighting CMF for sample of 40% segments

Another expectation of CMF is that larger and very small segment sizes should reflect worst results in terms of uncertainty. As seen on the previous figures the CMF of smaller and larger segment sizes did fulfil such assumption. It was also expected to observe a trend of uncertainty reduction and frequency increase, as one approaches the optimal segment size. Although, medium segment size of 500 meters did showed a good performance landing on the positive values and exhibiting less variation (uncertainty), both 250 meters and 750 meters did not showed a growing or decreasing trend with their means shifting either to the left or to the right of one each other as expected, but rather resulted in the same average with similar dispersion, which leads one to question whether the CMF value of the 500 meters is correct (CMF= 0.26) or the other sizes should be considered given that their difference in length to 500 meters is not that large, possibly landing all three sizes in the same category (medium). This dilemma pushed this research to integrate all three values and consider previous studies as another source of valuable information in order to abstract from segment size. It appears that both 250 meters and 750 meters lay more in line with previous values of CMF obtained by other researchers. However, future research should attempt to establish CMF per range of segment size (1 to 5 kilometres) versus the 0 – 1 km values used herein.

The last stage of the validation was the estimation of the consolidated value for CMF for the 60% sample which reached a value of 0.41 and for the 40% sample which reached a value of 0.41.

4.7 When to provide Lighting

Decisions can be done for the need to provide roadway lighting as a countermeasure. The decisions are based on the concept of potential for improvement and can be easily visualized on Figure 3.8.

Decisions for providing lighting as a countermeasure can be studied by plotting expected number of collisions at sites with and without lighting and then comparing them with the observe number of collisions of a given site. Expected trends of night time collisions can be obtained using estimated regression coefficients and fitting regression equations to the observed cross-sectional data taking advantage of the system of filters to develop SPF for a specific type of site and road.

If the observed number of collisions is smaller than the expected number then the site is deemed as a good performer and lighting is not justifiable, otherwise the site is deemed a poor performer and lighting is expected to improve significantly the safety of it.

Figures 4.4, 4.5 and 4.6 show the expected number of collisions at road segments without intersections for residential areas under various levels of traffic volume. As it can be seen, non-lit sites with a higher observed number of night time collisions (the black trend line) are sites that will definitely benefit from being illuminated. Sites with an observed number of collisions in between the expected trends for non-lit and lit will benefit from being illuminated. Sites with an observed number of night time collisions below the trend of illuminated sites (yellow line) should not be illuminated as they are already performing better than illuminated sites and by illuminating them one could theoretically observe a detriment in the level of collisions as observed by a study of Box (1970).

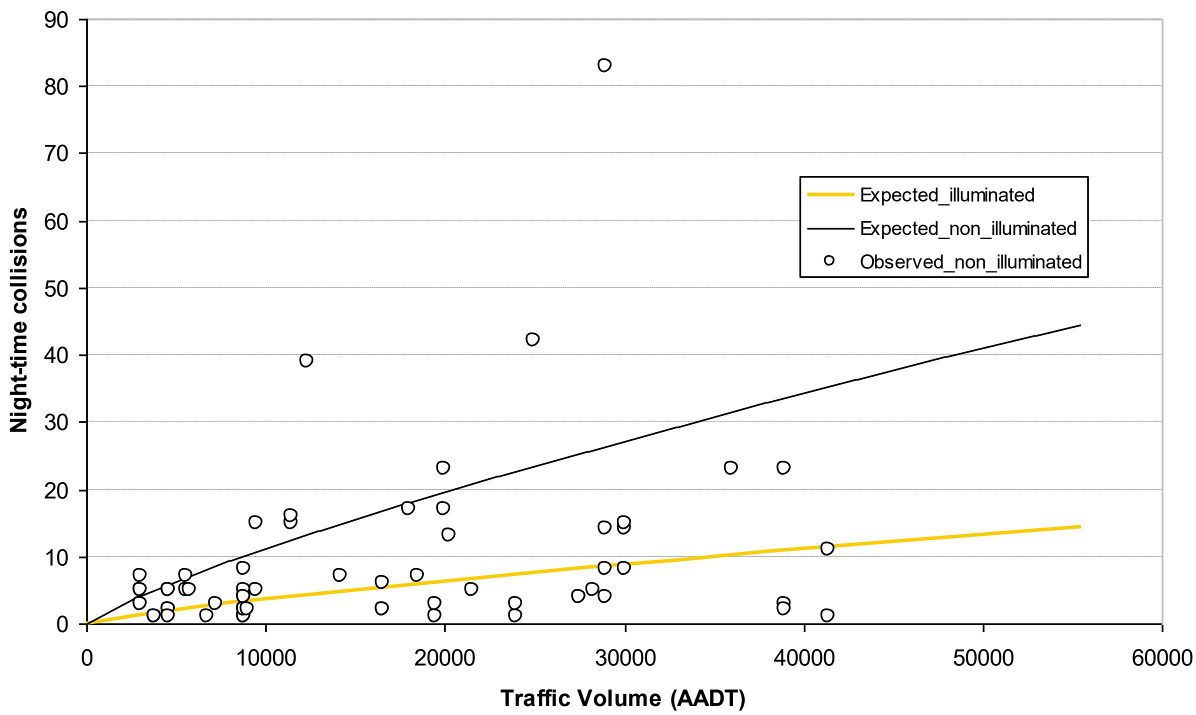

Figure 4.4: Model prediction for lighted and non-lighted road segments (250m)

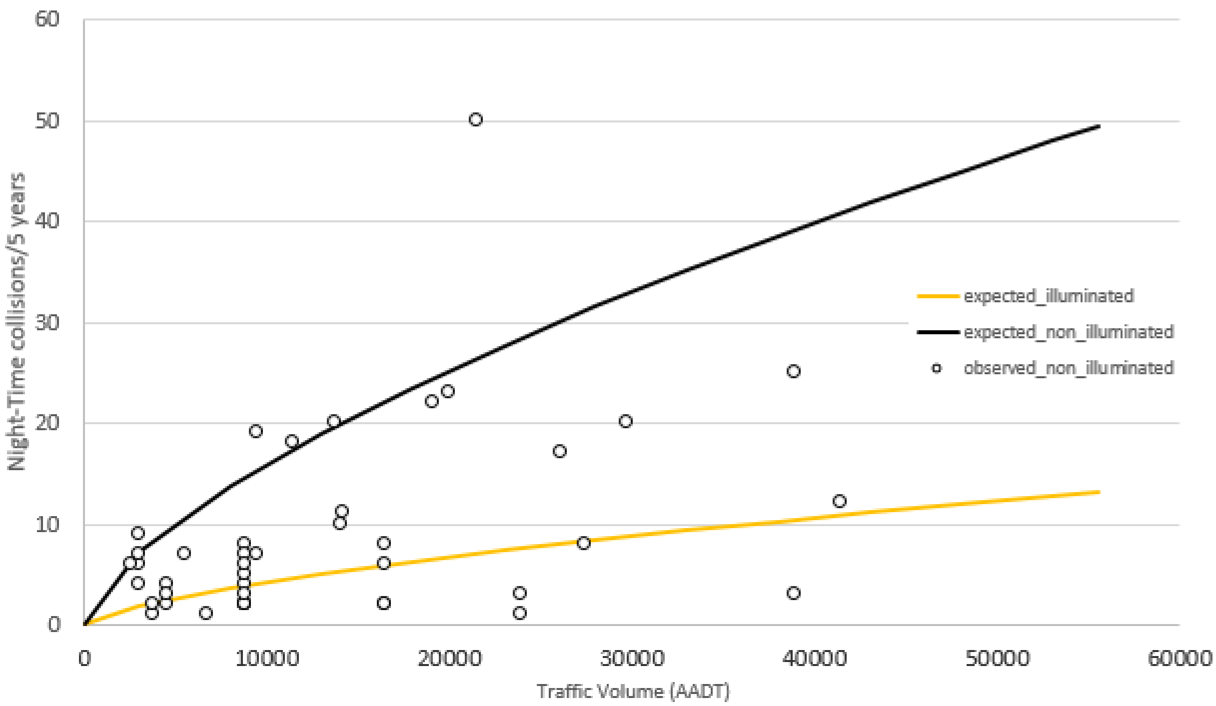

Figure 4.5: Model prediction for lighted and non-lighted road segments (500m)

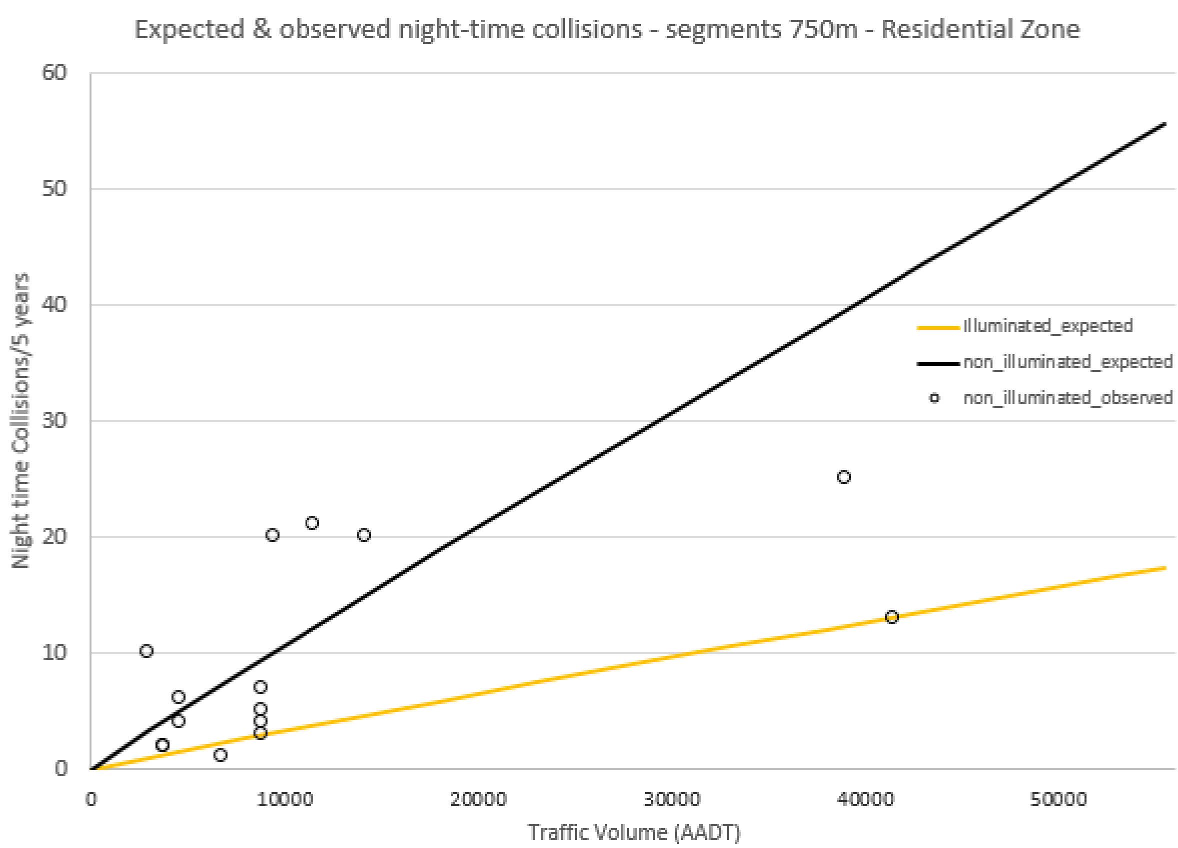

Figure 4.6: Model prediction for lighted and non-lighted road segments (750m)

CHAPTER 5: CONCLUSIONS

It is possible to estimate collision modification factors with cross sectional data; however, there is a big dependency on the segment size. Medium segment sizes around 500 meters seem more suited than other studies sizes in this research, but these results are tied to the case study data and future research is required to confirm this. Bayesian data fusion can be used to abstract from different segment size assumptions and return a sole value used to justify investing public funds in roadway lighting. However, various analyses of the data partitioned by ad-hoc segment sizes is necessary.

The case study presented found a collision modification factor of 0.34 for Quebec highways, hence justifying lighting as an effective counter measure. It was also found that the proposed model was able to produce CMF for various segment sizes. Segments of 750 meters exhibited a lot of uncertainty through their statistical distribution.

Potential for improvement supported by specific safety performance functions for desired land uses and road sites type are very useful not only to identify visually sties that will benefit from receiving artificial lighting from those that will not, but also to prioritize sites that will observe large improvement (reduction of night time collisions) versus those where lighting will have no effect and provision of it will be a waste of money.

This approach could be used in combination or as a replacement for the warrant system. Future research could explore the optimization of provision of lighting to maximize benefits to the society.

REFERENCES

AASHTO, 2005. Roadway lighting design guide. American Association of State and Highway Transportation Officials. Washington, D.C.

AASHTO, 2010. Highway Safety Manual. American Association of State and Highway Transportation Officials. Washington, D.C.

Abdel-Aty, M., Keller, J., Brady, P.A., 2005.Analysis of Types of Crashes at Signalized Intersections by Using Complete Crash Data and Tree-Based Regression. Transportation Research Record: Journal of the Transportation Research Board.1908, 37-45.

Al-Masaeid, H.R., 1997. Performance of Safety Evaluation Methods. Journal of Transportation Engineering. 123, 364-369.

Anastasopoulos, P., Mannering, F., 2008. A Note on Modeling Vehicle Accident Frequencies with Random-Parameters Count Models. Accident Analysis and Prevention. 41, 153-159.

Baek, J., Hummer, J.E., 2008. Collision Models for Multilane Highway Segments to Examine Safety of Curbs. Transportation Research Record: Journal of the Transportation Research Board. 2083, 48-56.

Bahar, G., Mollett, C., Persaud, B., Lyon, C., Smiley, A., Smahel, T., H. McGee, 2004. Safety Evaluation of Permanent Raised Pavement Markers. National Cooperative Highway Research Program, NCHRP Report 518, TRB, National Research Council, Washington, D.C.

Barua, U., Azad, A.K., Tay, R., 2010. Fatality Risk of Intersection Crashes on Rural Undivided Highways in Alberta, Canada. Transportation Research Record: Journal of the Transportation Research Board. 2148, 107-115.

Bishop, R. 2007. Camera-Controlled LED Arrays Pinpoint Danger. European Automotive Design.

Box. P.C., 1970. Relationship between Illumination and Freeway Accident. Transportation Research Record, Transportation Research Board of the National Academies. 416, 10-20.

Brown, B., Baass, K.A., 1995. Projet d'identification des Sites Dangereux sur les Routes Num´erot´ees en Mont´er´egie, vol. 1: Synth`ese. Direction de la sant´e Publique. St-Hubert, Quebec (in French).

Bullough, J., Donnell, E.T., Rea, M., 2012. To Illuminate or not to Illuminate: Roadway Lighting as it Affects Traffic Safety at Intersections. Accident Analysis and Prevention. 53, 65-77.

Bullough, J.D., Rea, M.S. and Zhou, Y., 2009. Analysis of Visual Performance Benefits from Roadway Lighting. New York.

Button, N.P., Reilly, P.M., 2000. Uncertainty in Incident Rates for Trucks Carrying Dangerous Goods. Accident Analysis and Prevention. 32, 797-804.

Cai, H. Li, L., 2014. Measuring Light and Geometry Data of Roadway Environments with a Camera. Journal of Transportation Technologies. 44-62.

CIE. 1992. Road Lighting as an Accident Countermeasure. International Commission on Illumination. Austria: CIE Publications; Research Report.

CIE. 2007. Road Transport Lighting for Developing Countries. International Commission on illumination. Austria: CIE Publications; Research Report.

Clarke, D. B., Sarasua, W.A., 2003. Fatal Crashes on Rural Secondary Highways. Columbia, South Carolina Department of Transportation, Final Report for Research Project.

Council, F., J. Stewart., 1999. Safety Effects of Conversion of Rural Two-lane to Four-lane Roadways based on Cross-Sectional Models, Journal of the Transportation Research Record, TRB. National Research Council. Washington, D.C. 1665, 35-43.

Donnell, E.T., Porter, R.J., Shankar. V.N., 2010. A framework for estimating the safety effects of roadway lighting at intersections. Safety Science. 48, 1436–1444.

Eassa, S. M., You, Q. C., 2009. Collision Prediction Models For Three-Dimensional Two-Lane Highways: Horizontal Curves. Transportation Research Record: Journal of the Transportation Research Board. 2092, 128-136.

El-Basyouny, K. Sayed, T., 2010. Full Bayes Approach to Before-and-After Safety Evaluation with Matched Comparisons: Case Study of Stop-Sign In-Fill Program. Transportation Research Record: Journal of the Transportation Research Board. 2148, 1-8.

Elvik, R., 1995. Meta-Analysis of Evaluations of Public Lighting as Accident Countermeasure. Transportation Research Record. Transportation Research Board of the National Academies, National Research Council. Washington, D.C. 1485, 112-123.

Elvik, R., Amundsen, F.H., Hofset, F., 2001. Road Safety Effects of Bypasses, Journal of the Transportation Research Record, TRB, National Research Council. Washington, D.C. 1758, 13-20.

Geedipally, S.R., Lord, D., 2008. Effects of the Varying Dispersion Parameter of Poisson-Gamma Models on the Estimation of Confidence Intervals of Crash Prediction Models. Transportation Research Record. 2061, 46-54.

Geedipally, S.R., Lord, D., Park, B.J., 2009. Analyzing Different Parameterization of the Varying Dispersion Parameter as a Function of Segment Length. Transportation Research Record. 2103, 108-118.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, 2nd Ed., CHAPMAN&HALL/CRC.

Gross, F., Donnell, E.T., 2011. Case–Control and Cross-sectional Methods for Estimating Crash Modification Factors: Comparisons from Roadway Lighting and Lane and Shoulder Width Safety Effect Studies. Journal of Safety Research. 42, 117–129.

Hadayeghi, A., Malone, B., De Gannes, R., 2006. Development of New Crash Experience Warrants for Traffic Signals in Ontario, Canada. Transportation Research Record: Journal of the Transportation Research Board. 1953, 120-127.

Harkey, D., 2005. Crash Reduction Factors for Traffic Engineering and Intelligent Transportation System Improvements: State-of-Knowledge Report, NCHRP Research Result Digest, Interim Report, NCHRP Project, National Cooperative Highway Research Program Project, Washington, D.C. 299, 17-20.

Harwood, D.W., Bauer, K.M., Richard, K.R., Gilmore, D.K., Graham, J.L., Potts, I.B., Torbic, D.J., Hauer, E., 2007. Methodology to Predict the Safety Performance of Urban and Suburban Arterials. NCHRP Web-only Document 129. TRB, Washington, DC.

Harwood, D.W., Council, F.M., Hauer, E., Hughes, W.E., A. Vogt., 2000. Prediction of the Expected Safety Performance of Rural Two-Lane Highways, Final Report, FHWA-RD-99-207. Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.

Hauer, E. 1997. Observational Before-After Studies in Safety. Washington, D.C.

Hauer, E. 2010. Cause, Effect and Regression in Road Safety: A case study. Accident Analysis and Prevention. 42, 1128–1135.

Heydari, S., Miranda-Moreno, L.F., Amador, L., 2013. Full Bayes Methods for Road Safety Studies: Does Prior Specification Matter? Journal of Transportation Research Record. 2392, 31-39.

Hinde, J., Demetrio, C.G.B., 1998. Over-dispersion: Model and Estimation. Computational Statistics and Data Analysis. 27, 151–170.

Hughes, W., Eccles, K., Harwood, D., Potts, I., Hauer, E., 2004. Development of a Highway Safety Manual, Contractor's Final Report, National Cooperative Highway Research Program Project, Washington, D.C.

Hummer J.E., Rasdor, W., Findley, D.J., 2010. Curve Collisions: Road and Collision Characteristics and Countermeasures. Journal of Transportation Safety & Security. 2, 203-220.

IESNA. 1989. Illuminating Engineering Society of North America. Value of Public Roadway lighting. Illuminating Engineering Society of North America. New York. IE SCP-31-1989.

IESNA. 2005. Illuminating Engineering Society of North America. Roadway Lighting: American Standard Practice for Roadway Lighting. New York. RP-08-00.

IESNA. 2008. lluminating Engineering Society of North America. Approved Method: Electrical and Photometric Measurements of Solid-State Lighting Products. New York. LM-79-08.

Isebrands, H.N., Hallmark, S.L., Li, W., McDonald, T., Storm, R., Preston, H., 2010. Roadway Lighting shows safety benefits at rural intersections. Journal of Transportation Engineering. 36, 949 – 955.

Jackett, M., Firth, W., 2013. Quantifying the Impact of Road Lighting on Road Safety - A New Zealand Study. International Association of Traffic and Safety Sciences. IATSS Research. 36, 139-145.

Jonsson, T., Ivan, J.N., Zhang, C., 2007. Crash Prediction Models for Intersections on Rural Multilane Highways: Differences by Collision Type. Transportation Research Record: Journal of the Transportation Research Board. 2019, 91-98.

Kumara, S.S.P., Chin, H.C., 2003. Modeling Accident Occurrence at Signalized T-Intersections with Special Emphasis on Excess Zeros. Traffic Injury Prevention. 3, 53–57.

Laughland, J., Haefner, L., Hall, J., Clough, D., 1975. NCHRP Report 162: Methods for Evaluating Highway Safety Improvements. National Cooperative Highway Research Program, Transportation Research Board. Washington, D.C.

Lee, J., Mannering, F.L., 2002. Impact of Roadside Features on the Frequency and Severity of Run off-road Accidents: an Empirical Analysis. Accident Analysis and Prevention. 34, 349–361.

Leeming, D.G., Saccomanno, F.F., 1994. Use of Quantified Risk Assessment in Evaluating the Risks of Transporting Chlorine by Road and Rail. Journal of Transportation Research Record, TRB. National Research Council, Washington, D.C. 1430, 27-35.

Loader, C., 1999. Local Regression and Likelihood. Springer-Verlag, NewYork, NY.

Lord, D., Abdou, H.A., N'Zu´e, A., Dionne, G., Laberge-Nadeau, C., 2003a. Traffic Safety Diagnostic and Application of Countermeasures for Rural Roads in Burkina Faso. Transportation Research Record. 1846, 39–43.

Lord, D., Mannering, F., 2010. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. Transportation Research Part A: Policy and Practice. 44, 291–305.

Lord, D., Miranda-Moreno, L.F., 2008. Effects of the low Sample Mean Values and the Small Sample Size on the Estimation of the Fixed Dispersion Parameter of Poisson-gamma Models: A Bayesian Perspective. Safety Science. 46, 751-770.

Lord, D., Simon P. Washington., John N. Ivan., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention. 37, 35–46.

Lord, D., Washington, S.P., Ivan, J.N., 2004. Statistical Challenges with Modeling Motor Vehicle Crashes: Understanding the Implications of Alternative Approaches. Research Report CTS-04 01, Texas Transportation Institute, College Station, Texas.

Lord, D., Persuad, B.N., 2000. Accident Prediction Models with And without Trend: Application of the Generalized Estimating Equations Procedure. Transportation Research Record: Journal of the Transportation Research Board. 1717, 102-108.

Lovegrove, G. Sayed, T., 2006. Using Macro-Level Collision Prediction Models in Road Safety Planning Applications. Transportation Research Record: Journal of the Transportation Research Board. 1950, 73–82.

Lunn. D., Thomas, A., Best, N.G., Gilks, W., Spiegelhalter, D.J., 2000. WINBUGS – A Bayesian Modeling Framework: Concepts, Structure and Extensibility. Statistics and Computing. 10, 325 – 337.

Lyon, C., Haq, A., Persaud, B.N., Kodama, S.T., 2005. Safety Performance Functions for Signalized Intersections in Large Urban Areas: Development and Application to Evaluation of Left-Turn Priority Treatment, Journal of the Transportation Research Record 1908, Washington, D.C., 165-171.

Miaou, S.P., 1994. The relationship between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions. Accident Analysis and Prevention. 26, 471–482.

Miaou, S.P., Lord, D., 2003. Modeling Traffic Crash-flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes methods. Transportation Research Record. 1840, 31–40.

Milton, J.C., Shankar, V.N., Mannering, F.L., 2008. Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. Accident Analysis and Prevention. 40, 260-266.

Miranda-Moreno, L.F., Nabavi Niaki, M., Saunier, N., Amador, L., Bruneau, J.F., 2014. Method for Road Lighting Audit and Safety Screening at Urban Intersections. Transportation Research Record: Journal of the Transportation Research Board. 2458, 27-36.

Mitra, S., Washington, S., 2006. On the Nature of Over-Dispersion in Motor Vehicle Crash Prediction Models. Accident Analysis and Prevention. 39, 459–468.

OPUS. How Does the Level of Road Lighting Affect Crashes in New Zealand - A pilot Study. Opus International Consultants Central Laboratories. New Zealand.

Park, Y.J., Saccomanno, F.F., 2005a. Collision Frequency Analysis Using Tree-Based Stratification, Journal of the Transportation Research Record TRB. National Research Council, Washington, D.C. 1908, 121-129.

Persaud, B., Lan, B., Lyon, C. Bhim, R., 2009. Comparison of Empirical Bayes and Full Bayes Approaches for Before–After Road Safety Evaluations. Accident Analysis and Prevention. 42, 38-43.

Persaud, B.N., Retting, R.A., Garder, P.E., Lord, D., 2001. Safety Effect of Roundabout Conversions in the United States: Empirical Bayes Observational Before-After Study, Journal of the Transportation Research Record, TRB. National Research Council. Washington, D.C. 1-8.

Poch, M., Mannering, F.L., 1996. Negative Binomial Analysis of Intersection Accident Frequencies. Journal of Transportation Engineering. 122, 105-113.

Preston, H., Schoenecker, T., 1999. Safety Impacts of Street Lighting at Rural Intersections. Minnesota Department of Transportation, St. Paul, MN.

Qin, X., Ivan, J.N., Ravishankar, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis and Prevetion. 36, 183–191.

R Development Core Team, 2004. R: a Llanguage and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http: //www.R-project.org.

Rea, M.S., Bullough, C.R.F., Brons, Van Derlofske, J.A., 2009. Review of the safety benefits and other Effects of Roadway Lighting.

SAAQ, 2012. *Societe de l'Assurance Automobile du Quebec*. *Bilan Routier*. C-5195.

Saccomanno, F.F., Park, P., Fu, L., 2007. Estimating Countermeasure Effects for Reducing Collisions at Highway – Railway Grade Crossings. Accident Analysis and Prevention. 39, 406–16.

Saccomanno, F.F., Lai, X., 2005. A Model for Evaluating Countermeasures at Highway-Railway Grade Crossings, Journal of the Transportation Research Record. 1918, 18-25.

Santiago-Chaparro, K.R., Qin, X., Noyce, D.A., 2010. Proposed Safety Index Based on Risk-Taking Behavior of Drivers. Transportation Research Record: Journal of the Transportation Research Board. 2147, 51-57.

Sasidharan, L., Donnell, E.T., 2013. Application of Propensity Scores and Potential Outcomes to Estimate Effectiveness of Traffic Safety Countermeasures: Exploratory Analysis using Intersection Lighting Data. Accident Analysis and Prevention. 50, 539-553.

Schoppert, D.W., Hoyt, D.W., 1968. Factors Influencing Safety at Highway-Rail Grade Crossings. National Cooperative Highway Research Program, NCHRP Report 50, Highway Research Board.

Schwab, R.N., Walton, N.E., Mounce, J.M., Rosenbaum, M.J., 1982. Synthesis of Safety Research Related to Traffic Control and Roadway Elements, Highway Lighting, Federal Highway Administration, Washington.

Shankar, V., Milton, J., Mannering, F.L., 1997. Modeling Accident Frequency as Zero-altered Probability Processes: an Empirical Inquiry. Accident Analysis and Prevention. 29, 829–837.

TAC, 2004. The Canadian Guide to in service Road Safety Reviews. Transportation Association of Canada. Ottawa.

Wanvik, P.O., 2009. Effects of Road Lighting: An Analysis Based on Dutch Accident Statistics, 1987-2006. Accident Analysis and Prevention. 41, 123-128.

WHO, 2013. World Health Organization. Global Status Report on Road Safety. Available at: http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/index.html.

Yanmaz-Tuzel, O., Ozbay, K., 2010. A Comparative Full Bayesian Before and After Analysis and Application to Urban Road Safety Countermeasures in New Jersey. Accident Analysis and Prevention. 42, 099-107.

Yannis, G. Kondyli, A. Mitzalis, N., 2013. Proceedings of the Institution of Civil Engineers, Transport, Issue TR5. 166, 271-281.

Zegeer, C.V, Stewart, J.R., Council, F.M., Reinfurt, D.W., Hamilton, E., 1995. Safety Effects of Geometric Improvements on Horizontal Curves, Journal of the Transportation Research Record, TRB, National Research Council, Washington, D.C. 1992, 11-19.

Zhou, H., Hsu, P., 2009. Effects of Roadway Lighting Level on the Pedestrian Safety. Critical Issues in Transportation Systems Planning and Management, ASCE.

APPENDICES

Negative Binomial regression for 250m segments

```
Negative binomial regression                    Number of obs   =       2494
                                                LR chi2(6)      =     408.41
Dispersion     = mean                           Prob > chi2     =     0.0000
Log likelihood = -2966.0898                     Pseudo R2       =     0.0644
```

sum_ni	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
avg_lu	-1.138903	.2217333	-5.14	0.000	-1.573493	-.7043141
avg_glare_	-.0354512	.0115897	-3.06	0.002	-.0581665	-.0127358
pre_inters	.3430718	.11982	2.86	0.004	.108229	.5779147
avg_tot_wi	-.1920988	.067526	-2.84	0.004	-.3244474	-.0597502
lnaadt	.792939	.1159532	6.84	0.000	.5656749	1.020203
zone_com_res	2.613244	.2233602	11.70	0.000	2.175466	3.051022
_cons	-6.20181	1.060229	-5.85	0.000	-8.27982	-4.1238
/lnalpha	1.651872	.0548866			1.544297	1.759448
alpha	5.216738	.2863289			4.684675	5.80923

Likelihood-ratio test of alpha=0: chibar2(01) = 6445.30 Prob>=chibar2 = 0.000

. sum sum_ni avg_lu avg_glare_ pre_inters avg_tot_wi lnaadt zone_com_res

Variable	Obs	Mean	Std. Dev.	Min	Max
sum_ni	2494	1.6251	7.39929	0	137
avg_lu	2494	.0817963	.2741091	0	1
avg_glare_	2494	-2.464531	4.869028	-29.1	0
pre_inters	2494	.2882919	.4530585	0	1
avg_tot_wi	2494	7.424835	.8365648	3.6	14.8
lnaadt	2494	9.164216	.4960441	8.006368	10.92414
zone_com_res	2494	.0585405	.2348097	0	1

Negative Binomial regression for 500m segments

```
Negative binomial regression                       Number of obs   =       1385
                                                   LR chi2(7)      =     405.08
Dispersion     = mean                              Prob > chi2     =     0.0000
Log likelihood = -2328.6137                        Pseudo R2       =     0.0800

------------------------------------------------------------------------------
      sum_ni |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     pres_lu |  -1.324006   .2120219    -6.24   0.000    -1.739561   -.9084506
   avg_glare |  -.0388782   .0121333    -3.20   0.001    -.0626589   -.0150974
 pres_inters |   .3928134   .1312821     2.99   0.003     .1355053    .6501215
  avg_tot_wi |  -.1393542   .0734302    -1.90   0.058    -.2832747    .0045663
      lnaadt |   .6624298   .1147119     5.77   0.000     .4375987     .887261
 pres_curve_ |   .3317422   .1876152     1.77   0.077    -.0359768    .6994612
 pres_zone_r |   2.470616   .2049244    12.06   0.000     2.068972    2.872261
       _cons |  -4.899084   1.090086    -4.49   0.000    -7.035614   -2.762554
-------------+----------------------------------------------------------------
    /lnalpha |   1.377777   .0616339                      1.256977    1.498578
-------------+----------------------------------------------------------------
       alpha |   3.966077   .2444447                      3.514781    4.475319
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) = 7046.62 Prob>=chibar2 = 0.000

. sum sum_ni pres_lu avg_glare pres_inters avg_tot_wi lnaadt pres_curve_ pres_zone_r

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      sum_ni |      1385    4.090253    15.50708          0        162
     pres_lu |      1385    .1241877    .3299147          0          1
   avg_glare |      1385    -3.43945    5.456997      -26.8          0
 pres_inters |      1385    .4498195    .4976552          0          1
  avg_tot_wi |      1385    7.424644    .8382055   4.233333   13.48333
-------------+--------------------------------------------------------
      lnaadt |      1385    9.147205    .5253949   7.226209   10.92414
 pres_curve_ |      1385    .1068592    .3090458          0          1
 pres_zone_r |      1385    .1018051     .302501          0          1
```

Negative Binomial regression for 750m segments

```
Negative binomial regression                      Number of obs   =        728
                                                  LR chi2(7)      =     150.80
Dispersion     = mean                             Prob > chi2     =     0.0000
Log likelihood = -1259.081                        Pseudo R2       =     0.0565

------------------------------------------------------------------------------
      sum_ni |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avg_lumi |   -1.16113   .3395963    -3.42   0.001    -1.826726   -.4955334
   avg_glare |  -.0400022    .018607    -2.15   0.032    -.0764712   -.0035333
 pres_inters |   .2630578   .1877822     1.40   0.161    -.1049886    .6311041
  avg_tot_wi |  -.4136653   .1255187    -3.30   0.001    -.6596774   -.1676532
      lnaadt |   .9619114   .2046623     4.70   0.000     .5607806    1.363042
  pres_curve |   .4859147   .2355391     2.06   0.039     .0242664    .9475629
     zone_rc |   1.865843   .3115042     5.99   0.000     1.255306     2.47638
       _cons |   -5.42382   1.810985    -2.99   0.003    -8.973285   -1.874355
-------------+----------------------------------------------------------------
    /lnalpha |    1.42535   .0816401                      1.265339    1.585362
-------------+----------------------------------------------------------------
       alpha |   4.159314   .3395668                      3.544292    4.881057
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) = 3282.76 Prob>=chibar2 = 0.000

. sum sum_ni avg_lumi avg_glare pres_inters avg_tot_wi lnaadt pres_curve zone_rc

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      sum_ni |       728    3.465659    13.45608          0        267
    avg_lumi |       728    .0961538    .2950048          0          1
   avg_glare |       728   -3.198684    5.502219      -30.7          0
 pres_inters |       728    .5082418    .5002758          0          1
  avg_tot_wi |       728    7.418917    .7468957      4.075      12.95
-------------+--------------------------------------------------------
      lnaadt |       728     9.15519     .480338   8.006368   10.63151
  pres_curve |       728    .1414835    .3487592          0          1
     zone_rc |       728    .0769231    .2666526          0          1
```

Bayesian Data Fusion Calculator

Priors

Safety Effectiveness	CMF
30	0.7
12	0.88
25	0.75
39	0.61
31	0.69
38	0.62
4	0.96
16.4	0.836
9.5	0.905
28	0.72
17	0.83
28	0.72
Avarege	0.768417
Sd	0.112473
Var	0.01265
weighted prior	
0.76304158	

Likelihood

250 m		500 m		750m	
exp(Coeff Lighting)=CMF	Sd Table	exp(Coeff Lighting)=CMF	Sd Table	exp(Coeff Lighting)=CMF	Sd Table
0.323033256	0.22	0.267135302	0.21	0.313486181	0.34
Sd of CMF	Safty Effectiveness	Sd of CMF	Safty Effectiveness	Sd of CMF	Safty Effectiveness
0.071067316	67.69667436	0.056098413	73.2864698	0.106585302	68.65138191
Variance	Confidence Interval (95%)	Variance	Confidence Interval(95%)	Variance	Confidence Interval(95%)
0.005050563	0.819101376	0.003147032	0.845061525	0.011360426	0.899684422
	0.534832111		0.620667871		0.473343216

 # 1 (250m,500m)

197.9977101	317.759719	
		CMF
515.7574291		0.288594357
VAR	0.001938896	
Sd	0.044032896	

#2

515.7574291	88.02486424
603.7822933	
VAR	0.001656226
Sd	0.040696758
CMF	
0.292223313	

Posterior

#3

79.05086264	603.7822933
682.8331559	
VAR	0.001464487
Sd	0.038268611
CMF	
0.346729439	

Safety Effectiveness
65.32705609

Priors

Safety Effectiveness	CMF
30	0.7
12	0.88
25	0.75
39	0.61
31	0.69
38	0.62
4	0.96
16.4	0.836
9.5	0.905
28	0.72
17	0.83
28	0.72
Avarege	0.768416667
Sd	0.112472589
Var	0.012650083
α	β
10.04112964	3.026168448
weighted priors	0.76304158

Likelihood

250 m		500 m		750m	
exp(Coeff Lighting)=CMF	Sd Table Stata	exp(Coeff Lighting)=CMF	Sd Table stata	exp(Coeff Lighting)=CMF	Sd Table stata
0.323033256	0.22	0.267135302	0.21	0.313486181	0.34
Sd of CMF	Safty Effectiveness	Sd of CMF	Safty Effectiveness	Sd of CMF	Safty Effectiveness
0.071067316	67.69667436	0.056098413	73.2864698	0.106585302	68.65138191
Variance	Confidence Interval (95%)	Variance	Confidence Interval(95%)	Variance	Confidence Interval(95%)
0.005050563	0.819101376	0.003147032	0.845061525	0.011360426	0.899684422
	0.534832111		0.620667871		0.473343216
α	β	α	β	α	β
13.66388293	28.63480509	16.35111182	44.85798971	5.625214677	12.31884481

	# 1 (250m,500m)	
α	30.01499475	
β	73.49279479	CMF
		0.289978125
VAR	0.0019701	
Sd	0.04438581	

#2	
α	β
35.64020943	12.31884481
VAR	0.003898845
Sd	0.062440733
CMF	
0.743138287	

Posterior

#3	
α	β
45.68133907	88.83780805
VAR	0.001654885
Sd	0.040680276
CMF	
0.339589865	

Safety Effectiveness
66.04101346

76

76