ESTIMATING PEDESTRIAN VOLUMES AND CRASHES AT URBAN SIGNALIZED INTERSECTIONS

JASON F. KENNEDY

Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Civil Engineering

Dr. Hesham Rakha, Chair
Dr. Kathleen Hancock
Dr. Patches Johnson Inge
Dr. Shinya Kikuchi
Dr. Pamela Murray-Tuite

December 10, 2008
Falls Church, VA

Keywords: Pedestrian crashes, Crash modeling, Exposure to risk, Signalized intersections, Crash risk estimation

ESTIMATING PEDESTRIAN CRASHES AT URBAN SIGNALIZED INTERSECTIONS

JASON F. KENNEDY

ABSTRACT

Crash prediction models are used to estimate the number of crashes using a set of explanatory variables. The highway safety community has used modeling techniques to predict vehicle-to-vehicle crashes for decades. Specifically, generalized linear models (GLMs) are commonly used because they can model non-linear count data such as motor vehicle crashes. Regression models such as the Poisson, Zero-inflated Poisson (ZIP), and the Negative Binomial are commonly used to model crashes. Until recently very little research has been conducted on crash prediction modeling for pedestrian-motor vehicle crashes. This thesis considers several candidate crash prediction models using a variety of explanatory variables and regression functions. The goal of this thesis is to develop a pedestrian crash prediction model to contribute to the field of pedestrian safety prediction research. Additionally, the thesis contributes to the work done by the Federal Highway Administration to estimate pedestrian exposure in urban areas. The results of the crash prediction analyses indicate the pedestrian-vehicle crash model is similar to models from previous work. An analysis of two pedestrian volume estimation methods indicates that using a scaling technique will produce volume estimates highly correlated to observed volumes. The ratio of crash and exposure estimates gives a crash rate estimation that is useful for traffic engineers and transportation policy makers to evaluate pedestrian safety at signalized intersections in an urban environment.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Hesham Rakha, for his guidance and support in my research. His encouragement and pursuit of academic excellence has helped push me to succeed while at Virginia Tech. Additionally I would like to thank Dr. Patches Johnson Inge, Dr. Pamela Murray-Tuite, Dr. Shinya Kikuchi, and Dr. Kathleen Hancock, my other committee members, for constructive comments and advice through the entire research process. I also would like to thank Science Applications International Corporation (SAIC) for their financial and professional support of my graduate studies at Virginia Tech.

DEDICATIONS

Most importantly I would like to dedicate this thesis to my best friend and wife, Beth. Without her love and support I would not have been able to complete this long, but rewarding journey at Virginia Tech. Additionally this thesis is dedicated to my two children, Calvin and Casey. The joy they bring to my life inspired me when I was away from them during my studies and I hope to be able to return the favor ten-fold. I would also like to dedicate this thesis to my parents, Forrest and Kathleen Kennedy. They always have supported me in my academic pursuits and taught me the importance of having a strong work ethic. Without their love and encouragement, both in the past and present, I could not have achieved this goal. I also dedicate this thesis to my wife's parents, Mike and Simone Onder. Their love and support during my studies have helped make this possible. Finally, I would like to dedicate this thesis to all the people in my life who have always believed in me. Your unwavering belief in me has given me the ability and confidence to go forward each day with purpose.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

1.1 PROBLEM OVERVIEW

Pedestrian exposure research has been conducted for decades, however there isn't a consensus on how to ultimately define the metric and the best methods to implement the metric. The Federal Highway Administration (FHWA) and the National Highway Traffic Safety Administration (NHTSA) have both indicated that establishing an acceptable pedestrian exposure measure is a high priority so that crash rates can be objectively compared (Molino et al, 2008).

The highway safety community has used modeling techniques to predict vehicle-to-vehicle crashes for decades. Specifically, generalized linear models (GLMs) are commonly used because they can model non-linear count data such as motor vehicle crashes. Until recently relatively little research has been conducted on crash prediction modeling for pedestrian-motor vehicle crashes. The National Cooperative Highway Research Program (NCHRP) had initiated an effort to develop a section in the Highway Safety Manual (HSM) to address pedestrian safety prediction models however this effort is not completely developed and there is still a need for additional research (Harwood et al, 2008).

The thesis presents two techniques to estimate pedestrian volumes at signalized intersections and other locales in an urban environment. The research presented in the thesis investigates the pedestrian exposure (volumes) estimation techniques specifically at signalized intersection by comparing estimated pedestrian volumes with observed pedestrian volumes. The research presented in this thesis also investigates the development of pedestrian crash prediction models using generalized linear models (GLMs) in an urban environment as a contribution to the relatively-nascent field of pedestrian safety prediction. A comparison between two sets of predicted pedestrian volumes in the final pedestrian crash prediction model is also made.

1.2 RESEARCH OBJECTIVES

There are two major objectives to this thesis. The first objective is to estimate pedestrian volumes at urban signalized intersections using two different techniques and limited empirical data. The second objective is to determine if a pedestrian crash prediction model can be developed from exposure data at urban signalized intersections using road width, and to evaluate the final model by comparing against statistical tests and results from previous research.

1.3 THESIS CONTRIBUTION

Although research has been done with crash prediction models at signalized intersections most of the work has focused on vehicle-to-vehicle crashes. Limited information exists on pedestrian crash prediction models at signalized intersections. The first goal of this thesis is to develop pedestrian exposure (volume) estimation methodologies that could be universally used and accepted. The second goal is to investigate and present a pedestrian crash prediction model that will add to the growing field of pedestrian safety prediction research.

1.4 THESIS ORGANIZATION

The thesis is organized into five chapters. Chapter 1 presents the problem overview and research objectives of the this thesis. Chapter 2 provides a review of research related to pedestrian exposure, vehicle exposure, and pedestrian crash prediction models. Chapter 3 presents the two pedestrian exposure estimation techniques developed. Chapter 4 describes the data collection and analysis methodologies used in the creation of the pedestrian crash prediction model and the results of the crash model development process. Chapter 4 also compares estimated pedestrian volumes that were created with the two methodologies described in Chapter 3. Chapter 5 presents the study conclusions and recommendations for further research. The SAS® statistical package will be used to develop the crash prediction models analyzed in this thesis. A complete SAS output for the final pedestrian crash prediction model will presented in the attached Appendix.

CHAPTER TWO: LITERATURE REVIEW

2.1 INTRODUCTION

This section presents some background information on some of the available literature related to pedestrian crash statistics, crash prediction models, and various types of exposure measures for both vehicles and pedestrians. The first section deals with pedestrian crash statistics on a national and local level. The second section focuses on previously developed crash prediction models. The third section deals with the FHWA's official measure of vehicle exposure as well as some of the various alternative measures of vehicle exposure. Finally, the fourth section deals with a sample of the leading measures of pedestrian exposure.

2.2 PEDESTRIAN CRASH STATISTICS

Pedestrian crash data are available from both a national and local level. The National Highway Traffic Safety Administration (NHTSA) requires States to report all motor vehicle fatalities, including those involving pedestrians, so that they can be included in the Fatality Accident Reporting System (FARS). Each fatal crash has more than 100 elements related to the crash, vehicle, and people involved in each crash event (NHTSA 2008). For all crashes involving pedestrians (including non-fatal), the General Estimating System (GES) is the most thorough database in the United States.

Locally, pedestrian crash data are available from the District of Columbia's Department of Transportation (DDOT). DDOT collects pedestrian-related crash data from police reports made by the Metropolitan Police Department (MPD). Types of information available in the DDOT crash database include crash location, date, possible cause, type of vehicles involved, number of pedestrians involved, presence and type of traffic control device, and several other variables.

2.2.1 National Statistics

It is estimated that only about 10% of all intersections in the United States are signalized (FHWA 2007). However, in 2006 over 46% of all intersection and intersection-related pedestrian fatalities occurred at signalized intersections (NHTSA 2006a). This means that of the estimated 3 million intersections in the United States, only 300,000 signalized intersections accounted for almost half of all intersection pedestrian fatalities. Additionally, the 518 pedestrian fatalities at signalized intersections accounted for over 10% of all pedestrian fatalities in 2006 (NHTSA 2006a).

2.2.2 Washington DC Statistics

Between 2000 and 2005, over 2,600 pedestrian-motor vehicle crashes occurred in Washington, DC (DC GIS Office 2007). 77 pedestrians involved in those crashes were fatally injured. In 2006, 15 pedestrians were fatally injured in a motor-vehicle crash. Of the 15 pedestrian fatalities, 5 were at or near a signalized intersection. Over 62% (5 of 8) of intersection or intersection-related pedestrian fatalities occurred at a signalized intersection, of which there are approximately 1,600 signalized intersections in the

District of Columbia (NHTSA 2006b). In addition, there were approximately 2,600 pedestrian-motor vehicle crashes in the Washington DC between 2000 and 2005. Many of these crashes occurred at or near an intersection, and many of those were signalized. It is not unreasonable to expect a higher proportion of signalized intersection or intersection-related fatalities in an urban environment they tend to have a higher proportion of signalized intersections.

2.3 CRASH PREDICTION MODELING

Crash data are very important for conducting safety analyses because they show what has already happened. Trends and patterns can be analyzed to improve potentially unsafe locations. However historical crash data can be used to predict future crashes when used in crash prediction models. Crash prediction models are important tools because they can be applied to, once validated, similar locations that have little or no crash data. This section of the literature review presents a selection of some of the current models that are popular with researchers and practitioners.

2.3.1 Poisson Model

Previous crash predictive modeling work used multiple linear regression techniques based on assumptions that crash data were normally distributed and were homeoscedastic. However researchers discovered that crash data were not typically linear and did not usually follow normal distributions (Maher, 1996).

Miaou et al, (1992, 1993) found that the Poisson regression model was more effective at predicting truck crashes when compared to standard linear regression models. Dionne et al. (1993) and Jovanis and Chang (1986) used a Poisson model to relate exposure variables to crashes.

Calieno et al, (2006) compared the Poisson, Negative Binomial, and Negative Multinomial distributions in developing crash models for multilane highways in Italy. The researchers found that the Poisson was inappropriate to model crashes due to the over-dispersion of the data. Maycock and Hall (1984) showed that a negative binomial structure was superior to a Poisson model due to the "excess zeros" that can characterize crash data.

The Poisson distribution has been used for modeling count data however it is limited in its ability to predict outcomes based on over-dispersed data. While the research in this thesis will investigate the Poisson distribution, it is anticipated that the final model will not use this distribution function.

2.3.2 Negative Binomial (Poisson-gamma) Model

Lord (2006) investigated a model that utilized the negative binomial (or Poisson-gamma) distribution. Crash data tends to be characterized by having low sample mean values and small sample sizes. The objectives of this study were: 1) determine if the "low mean problem" (LMP) affects the dispersion parameter of the model and 2) determine the effects of an unreliable dispersion parameter on highway safety analysis. Several Poisson-gamma distributions were created using a variety of values for the sample size, dispersion parameter, and the mean. Three estimators commonly used by highway safety modelers were studied. Lord (2006) concluded that crash data characterized by low sample mean and small sample size affects the estimation of the dispersion parameter. Additionally, the three estimators, (method of moments, weighted regression, and maximum likelihood method) were more affected in the extreme conditions. As sample size decreases the mis-estimation of the dispersion parameter increases quite significantly. The study also found that when the dispersion parameter is mis-estimated, highway safety analyses might be flawed if erroneous modeling outputs are used.

2.3.3 Zero-Inflated Poisson (ZIP) Model

The zero-inflated Poisson model (ZIP) assumes data come from two distinct sources or distributions (Lord, 2005). Traffic data, specifically crash data, commonly has more zeros than would be expected in a normal Poisson distribution. The ZIP model has been used to model crash data because crashes are fairly rare events and often times there are high numbers of locations or samples that have zero (0) crashes. The model generates data using two processes. The first process generates only zero counts and the second process generates counts from a Poisson model.

Shankar et al. (2002) developed crash prediction models for pedestrians and motorized traffic. Two models, the negative binomial and the ZIP, were considered. Issues such as the presence of excess zeros and unobserved heterogeneity in pedestrian crash distributions were discussed. Pedestrian crashes from Washington State between 1991 and 1994 were sampled for consistently available data and included in the models. The study compared the application of both the negative binomial and the ZIP models to the empirical pedestrian crash data and found that the ZIP model is the most suitable for analyzing the pedestrian crash contexts.

2.3.4 Linear Regression Model

Rakha et al. (2008) demonstrated that a least square Linear Regression Model (LRM) could be successfully validated to predict crashes on access roads by spacing and AADT. This study also looked at two General Linear Models (GLM), the Poisson and negative binomial models, to compare against the results of the LRM. Using data from the Virginia Department of Transportation (VDOT), Rakha et al. (2008) found that the LRM approach was superior to both the Poisson and negative binomial models because it accounted for the high number of access road sections that had zero crashes.

2.4 Vehicle Exposure

The Federal Highway Administration (FHWA) is responsible for estimating the annual vehicle-miles traveled (VMT) in the United States. This estimate is the result of multiplying the vehicular volume and the centerline miles for all public roads in the U.S. While the monitoring of vehicular traffic is important for distributing Federal-aid funding, it is also important to gauge the performance of the Nation’s highways. Vehicle crash and fatality rates and vehicle delay are performance measures the FHWA can estimate using the annual VMT.

Currently there is no analogous estimate for pedestrian-miles traveled (PMT) in the United States. Several research studies have been conducted over the years using different exposure methodologies to address this problem. However to date, no single exposure method has been universally accepted by transportation professionals and until one is, it is difficult to quantitatively assess and analyze the crash exposure rate of pedestrians in the United States. Crash data alone does not give the entire picture for pedestrian safety and therefore it is important to have accurate pedestrian exposure data.

Collecting vehicular counts has become automated with traffic counters and tubes. It is not as easy to collect pedestrian volumes. Pedestrian counts, for the most part, still require humans to manually count pedestrians in the field or conduct post-hoc video analysis. Both of these data collection methods are time consuming and expensive. Due to the more complex nature of estimating pedestrian volumes and exposure rates the research presented in this thesis focuses on urban signalized intersections. The feasibility of such an estimate should be tested at urban signalized intersections first. Signalized intersections tend to have higher volumes of both vehicles and pedestrians compared to non-signalized intersections. Additionally local transportation agencies maintain better asset management records for signalized intersections and therefore the estimation technique may be easier to test and improve when tested within a known population.

2.4.1 Exposure Method: FHWA's Vehicle-Miles Traveled

Vehicular travel (exposure) is officially reported in the United States by the Federal Highway Administration (FHWA) (FHWA 2006). FHWA estimates the vehicle-miles traveled, also known as VMT, in the United States each year on all public roads. FHWA's Office of Highway Policy Information oversees the estimate of the VMT through the Highway Performance Monitoring System (FHWA 2005). State Departments of Transportation are required to report the vehicle-miles traveled in their jurisdiction to FHWA so that it can be included in the Highway Performance Monitoring System (HPMS) (FHWA 2005).

FHWA's vehicle-miles traveled is calculated by taking the product of the annual average daily traffic (AADT) and the centerline length of the section associated with the AADT (FHWA 2005). The centerline length of all public roads is reported to FHWA by Departments of Transportation annually. The requirements for the reporting of AADT data vary by functional system. FHWA requires States to report AADT for each section of Interstate, National Highway System (NHS), and other principal arterial. However for lower classes of functional system, such as minor arterials, rural major collector, and urban collector systems, travel is estimated from roadway section samples through the HPMS. The estimates are calculated from the samples using the centerline length for each section sampled, the AADT, and sample expansion factors from the HPMS. Table 2-1 summarizes the HPMS data by functional system.

HPMS Data	Rural Functional Systems					
	Interstate	Other Principal Arterials	Minor Arterial	Major Collector	Minor Collector	Local
Interstate Lane Miles Interstate VMT	Universe Universe					
Non-Interstate PAS Lane Miles Non-Interstate PAS VMT		Universe Universe				
FA Highway Lane Miles 1/ FA Highway VMT 1/	Universe Universe	Universe Universe	Universe Sample 2/	Universe Sample 2/		
NHS Lane Miles	Universe	Universe	Universe	Universe	Universe	Universe
Miles Lane Miles VMT	Universe Universe Universe	Universe Universe Universe	Universe Universe Sample 2/	Universe Universe Sample 2/	Universe Universe 3/ Summary 4/	Universe Universe 3/ Summary 4/
Total Public Road Miles	Certified Mileage ---					

HPMS Data	Urban Functional Systems					
	Interstate	Other Freeways & Expressways	Other Principal Arterial	Minor Arterial	Collector	Local
Interstate Lane Miles Interstate VMT	Universe Universe					
Non-Interstate PAS Lane Miles Non-Interstate PAS VMT		Universe Universe	Universe Universe			
FA Highway Lane Miles 1/ FA Highway VMT 1/	Universe Universe	Universe Universe	Universe Universe	Universe Sample 2/	Universe Sample 2/	
NHS Lane Miles	Universe	Universe	Universe	Universe	Universe	Universe
Miles Lane Miles VMT	Universe Universe Universe	Universe Universe Universe	Universe Universe Universe	Universe Universe Sample 2/	Universe Universe Sample 2/	Universe Universe 3/ Summary 4/
Total Public Road Miles	Certified Mileage ---					

1/ Universe data are used to estimate lane-miles & VMT for the few miles of NHS that are on the minor collector & local functional systems.
2/ Expanded sample data are used.
3/ Universe miles times 2 (lanes) are used. States are not required to report number of through lanes on these systems, except for any NHS sections.
4/ Summary data are used. States are not required to report section level AADT on these systems, except for any NHS sections.

Definitions:
Universe: Data reported for all roadway links in the system.
Sample: Data reported for a randomly selected sample of roadway links in the system.
Summary: Data reported in aggregated form by functional system.
PAS: Principal arterial system made up of interstate, other freeways & expressways, and other principal arterial systems.
VMT: Vehicle miles of travel.
FA: Federal-aid.
NHS: National highway system.

Figure 2-1. Sources of Selected HPMS Data by Functional System (FHWA 2005)

2.4.2 Exposure Method: Alternative Measures

Although the FHWA's VMT is the standard exposure measure used in the United States, there are several other methods that have been used by researchers and practitioners to measure exposure to motor vehicle crashes. Table 2-1 below provides a selection of alternative vehicle exposure measures commonly used (Carroll, 1973).

Table 2-1. Selection of Vehicle Exposure Measures (Carroll, 1973)

Category of Exposure Measure	Explanation and Types of Metric
Distance	Average miles driven, per vehicle, per day Total distance of a vehicle in a year
Population	Number of registered vehicles Number of licensed drivers Fuel consumption, in gallons Total population or sections of the population (gender, age, race, ethnicity) Person-trips
Vehicle Trips	Average number of vehicle trips made by members of a population per day, week, or year Proportion of driving trips taken for a particular purpose
Time Traveling	Average time driven, per person, per day or year Total time traveled by a driver or passenger (vehicle hours, passenger hours)

Using the National Household Travel Survey (NHTS) researchers investigated the motor vehicle crash injury rates by mode of travel (Beck, 2007). Beck et al. analyzed travel exposure data, in person-trips, in addition to data from the NHTSA's FARS and GES databases to calculate exposure-based fatal and nonfatal traffic injury rates in the U.S. The results of the study found that the overall fatal injury rate was 10.4 per 100 million person-trips and the nonfatal injury rate was 754.6 per 100 million person-trips. As one of the first studies to quantify these two types of rates for all modes of travel, Beck et al. found that motorcyclist, pedestrians, and bicyclists had an increased injury risk. Additionally, this study showed that the elderly, males, and adolescents have a higher risk of traffic injury.

Keall and Frith (1999) examined three main measures of exposure to risk and discussed the ways these measures could be used to improve highway safety. The first estimate examined was traffic flow on road sections. They used New Zealand's National Traffic Database, which is a database of all road sections and their respective traffic counts, to assess the social cost of crashes by road type. According to the results of the analysis, travel on country roads (compared to town and city roads) is about 50% more risky in terms of total societal cost. The social costs, which are a weighted sum of crashes, (measured as cents per kilometer) include all costs that a community may face such as property damage, pain, and suffering (Keall and Frith 1999).

The second exposure risk measure examined was the travel of people. Roadside alcohol and household travel surveys were two methods used. The roadside surveys are conducted by police to measure alcohol impaired driving. The results, when compared to vehicle flow and time of night, seem to indicate that higher impaired drivers (over 120mg / 100ml) tend to drive on lower volume roads after midnight. Keall and Frith hypothesize that the more impaired drivers may avoid higher volume roads to avoid being stopped by police. Another tool used to examine travel of people was the New Zealand Travel Survey. This survey measured people's travel behavior through household-based interviews. From this information, personal exposure to risk in terms of distance traveled or hours traveled can be estimated. Since these surveys are taken periodically changes in risks for different gender and age groups can be observed.

The third estimate described by Keall and Frith (1999) is based on the travel of vehicles. Odometer reading data from New Zealand's National Motor Vehicle Registration System can be compared between inspections and vehicle-based distances can be estimated.

2.5 PEDESTRIAN EXPOSURE

Pedestrian crashes, vehicular crashes, and the number of vehicles traveling on public roads are all well documented in the United States. Vehicular exposure has been well documented and is officially estimated each year by the FHWA. Pedestrian and vehicular crash data are collected by local, State, and the Federal governments. However what is not well documented is the exposure of pedestrians to motor vehicle crashes or the development of pedestrian crash prediction models. There is no national database or data clearinghouse that has an accurate estimate of the exposure to crashes for pedestrians in the United States and very little research, as compared to vehicle-vehicle prediction models, has been done to date.

Pedestrian exposure can be measured in several ways. Although researchers and practitioners have tested several methods to measure the crash risk pedestrians expose themselves to, nothing to date has been officially adopted by FHWA or any other federal agency with national jurisdiction.

In general, pedestrians choose the location where they will cross the road. Urban road networks with high volumes, or at least constant volume of motor vehicles, however will typically force pedestrians to cross at intersections, especially those that are signalized.

Table 2-2. Leading Pedestrian Exposure Measures

Category of Exposure Measure	Explanation and Types of Metric
Hazardous Pedestrian Behavior	Pedestrian Volume times Vehicular Volume Pedestrian Volume times Intersections Crossed
Distance	Average miles walked, per person, per day Total aggregate distance of pedestrian travel across an intersection
Population	Number of residents of a given area or in a demographic group
Pedestrian Trips	Average number of walking trips made by members of a population per day, week, or year Proportion of walking trips taken for a particular purpose Number of pedestrians observed in a given area over a fixed interval
Time Traveling	Average time walked, per person, per day or year Total aggregate time traveled by all pedestrians or total time traveled by individual pedestrian

Pedestrian volumes vary even at signalized intersections, depending on several key variables such as time-of-day, day-of-week, weather, land-use characteristics and employment (Green-Roesel 2007). Some existing pedestrian exposure methods utilize volume data along with some other variable. Table 2-2 shows a sample of some of the more common pedestrian exposure measures developed to-date (Center for Applied Research 2003).

Many exposure methods exist but none are as analogous to the FHWA's vehicular exposure measure (VMT) as is the proposed pedestrian exposure measure of pedestrian-miles traveled (PMT). A portion of the research for this thesis investigated this proposed measure. The objective of that research was to consider the feasibility of the methodology and the ability to estimate pedestrian exposure to crash risk for an entire large city in the United States (Molino et al, 2008).

2.6 Summary of Literature Review

The literature review presented in this section provided some basic background information regarding pedestrian crash statistics, crash prediction models, and vehicle and pedestrian exposure measures.

Pedestrian fatality data is available on a nationwide level through the NHTSA's FARS database. Information on the location of fatal crashes, such as signalized intersections, can be searched through FARS to identify the specific crashes related to this research. Locally, DDOT collects and monitors both fatal and nonfatal pedestrian crashes in the city. Pedestrian crash data are available in a GIS layers by location.

Crash prediction models examined in this literature review are examples that have been used recently to assess highway safety. Depending on the model, some can be used at specific locations, such as signalized intersections, and can be used for pedestrians-vehicle or vehicle-vehicle crashes.

A review of several prevailing exposure measures was presented. Measures for both vehicles and pedestrians were examined, including the standard used by the FHWA for vehicle exposure. The literature review reveals a need to develop pedestrian exposure and crash prediction models.

Chapter Three: Pedestrian and Bicyclist Exposure to Risk: A Methodology for Estimation in an Urban Environment

John A. Molino
Jason F. Kennedy
Patches Johnson Inge
Pascal A. Beuse
Science Applications International Corporation
Turner-Fairbank Highway Research Center
6300 Georgetown Pike, F-215
McLean, VA 22101
Phone (202) 493-3381; Fax (202) 493-3390
John.A.Molino@saic.com
Jason.F.Kennedy@saic.com
Patches.L.Johnson@saic.com
Pascal.A.Beuse@saic.com

Amanda K. Emo
Ann Do
Federal Highway Administration
Turner Fairbank Highway Research Center
6300 Georgetown Pike, T-210
McLean, VA 22101
Phone (202) 493-3395; Fax (202) 493-3374
Amanda.Emo@fhwa.dot.gov
Ann.Do@fhwa.dot.gov

November 2008

Total Number of words: approx. 7,550
Number of tables: 3
Number of figures: 1
Number of words without title page, tables and figures: approx. 6,930
Corresponding author: John A. Molino

3.1 ABSTRACT

There is currently no commonly accepted or adopted measure of pedestrian and bicycle exposure to risk. Consequently, a large portion of the field of pedestrian and bicycle safety is lacking an adequate means to evaluate the effectiveness of its efforts. The present paper presents a proposed metric for measuring pedestrian and bicycle exposure to risk: hundred million pedestrian or bicycle miles traveled on facility shared with motor vehicles. A method for implementing the proposed exposure metric is described for 8 shared facility types characteristic of the urban environment of Washington, DC. These facilities include three types of intersections, mid-block road segments, driveways, alleys, parking lots, parking garages, school areas and areas with playing/dashing/working in the roadway. The methodology is then used to calculate the annual pedestrian and bicycle exposure for the city for the calendar year 2007. The results of these calculations revealed 0.82 hundred million pedestrian miles traveled for pedestrian exposure and 0.37 hundred million bicyclist miles traveled for bicyclist exposure. In this way both the feasibility and scalability of the proposed metric were successfully demonstrated for a relatively large urban environment. Thus the proposed metric has the potential to eliminate one of the major obstacles in the pedestrian and bicycle safety field, the lack of adequate exposure data. While further refinement and validation are still needed, the proposed metric provides a possible initial foundation to develop a national unit of risk exposure for pedestrians and bicyclists.

3.2 INTRODUCTION

In order to know whether any safety countermeasure is effective in reducing pedestrian and bicycle crashes, it is essential to have information not only on the number of pedestrian and bicycle crashes, fatalities and injuries; but also on the relative exposure of the pedestrians and bicyclists at risk. In roadway safety, risk is generally defined as the ratio of crashes to exposure. Unfortunately there is no commonly accepted and adopted measure of pedestrian and bicycle exposure (*1*). Consequently, a large portion of the field of pedestrian and bicycle safety is lacking an adequate means to evaluate the effectiveness of its efforts.

3.2.1 Background

Pedestrian fatalities resulting from traffic crashes in the United States have dropped over the past decade from 5,449 in 1996, to 4,784 in 2006 (*2*). This decrease in fatalities could be a result of various factors but, without knowing the exposure of pedestrians and bicyclists, it is difficult to know what is driving this trend. That is, the reduction in fatalities could be due to improved safety countermeasures, but it could also be the result of fewer people walking and bicycling.

In the case of motor vehicle safety, the Federal Highway Administration (FHWA) and National Highway Traffic Safety Administration (NHTSA) are the primary agencies in the United States responsible for determining crash risk. Crash data (numerator) are available through NHTSA's Fatality Analysis Reporting System (FARS) (*3*) and General Estimates System (GES) (*4*), and exposure data (denominator), in terms of hundred million vehicle miles traveled (VMT), are available through FHWA's Highway Performance Monitoring System (HPMS)(*5*).

Currently pedestrians and bicyclists are not accounted for in the denominator of this ratio. However, to assess pedestrian and bicycle crash rates, it is essential to determine their exposure. In 2000, NHTSA and FHWA conducted a series of Pedestrian and Bicycle Strategic Planning Workshops. Out of a total of 57 pedestrian and 57 bicycle research needs, the lack of adequate pedestrian and bicyclist exposure data ranked in the top category among the four highest priority research needs for both pedestrian and bicyclist research (*6*).

A variety of pedestrian/bicycle exposure measures have been suggested and tried in the past, but these measures are mutually inconsistent, and none of them has achieved widespread acceptance. A partial list of some of these earlier metrics includes:

- Population (*7*)
- Number of roads crossed (*8*)
- Time spent walking (*7*)
- Pedestrian volume times vehicular volume (*9*)
- Pedestrian volume crossing the street (*10*)
- Bicycle trips (*11*)

- Bicycle miles traveled (*12)*.

The research described herein is directed at establishing a metric of exposure for use in determining the effectiveness of pedestrian/bicycle safety programs at all levels. In this context, the present paper has two major goals: 1) to describe a methodology for realizing the new metric in practice, and 2) to employ this methodology to calculate the annual pedestrian and bicycle exposure for a relatively large urban environment in terms of the proposed metric.

3.2.2 Proposed Exposure Metric

The proposed pedestrian and bicycle exposure metric has much in common with its motor vehicle crash rate analog. This measure is hundred million pedestrian or bicycle miles traveled on a facility shared with motor vehicles. In its more precise form, this measure may be defined as hundred million pedestrian or bicycle miles traveled on a shared facility, including parking lots, driveways, alleys, and parking garages, as well as other facilities where pedestrians and bicyclists are allowed to share the same space with motor vehicles. The proposed exposure metric is closest to the pedestrian volume crossing the street measure that has previously been explored (*10*), since both concentrate on pedestrians walking only on the roadway, and not on sidewalks, trails and other sheltered places. In addition to including a bicycle component, the proposed metric adds the concept of distance traveled to obtain a more accurate estimate of individual exposure to a potentially hazardous environment.

While the amount of pedestrian and bicyclist travel, regardless of location, is important for understanding the level of outdoor activity or mobility of a population, the present research was focused on the amount of walking/bicycling engaged in while at risk of being involved in a motor vehicle crash. The pedestrian crash rate is dependent on having a quantity of exposure in the denominator that corresponds with the numerator. It would not be appropriate to include pedestrian and bicyclist travel on non-shared facilities, because there is a negligible risk of being involved in a motor vehicle crash for that type of travel. Any walking or bicycling engaged in while on a non-shared facility would ultimately be accounted for by means of before and after crash and exposure studies, if a trail, sidewalk, or other sheltered facility were being installed.

FHWA successfully tested this proposed metric for methodological feasibility in the fall of 2006 in Washington, DC. This early testing indicated the measure was a viable contender for a pedestrian and bicycle exposure metric. However, these tests only employed one measurement site for each of seven unique types of pedestrian and bicycle facilities. The present project measures multiple sites of each kind of facility, and combines the data from all sites into an overall estimate of pedestrian and bicycle exposure to the risk of a motor vehicle crash for the city of Washington, DC, for the calendar year 2007. This estimate of annual exposure for a moderately large American city is regarded as the first step in demonstrating the scalability of the proposed metric for consideration at a national level.

3.2.3 Research Approach

Measurements of pedestrian and bicycle volumes and travel distances were made in Washington, DC, during the summer and early fall of 2007, generally at a peak time of the year for tourist visits (*13*). The data collection procedure was completely passive using personnel who observed pedestrian and bicyclist movements while standing on the sidewalk or while sitting in a parked vehicle. These observers counted the number of pedestrians and bicyclists who traveled in the street, or other motor-vehicle shared facility, during 15-minute intervals. There was no interference with the flow of pedestrian, bicycle, or motorized traffic, and no personal contact or interviews were conducted with pedestrians, bicyclists, or drivers.

The observers also estimated the length of crosswalks, roadways, driveways and parking lots, in most cases by means of known lane widths, car lengths and other indirect means. Sometimes more precise measurements were made with tape measures, distance wheels, or remote distance measuring equipment as a validation check for lane width estimates. For safety reasons, the observers always worked in pairs, and no direct observations were made between 10 PM and 6 AM. All nighttime measurements were made using the District of Columbia Department of Transportation's (DDOT) traffic cameras accessed via the Internet (*14*).

3.2.4 Sampling Procedure

Altogether, 122 locations were sampled throughout Washington, DC, resulting in a total of 364 unique continuous 15-minute counts. Each of the 122 locations, with the exception of three, was observed at between one and four different times, usually during the same day. Three locations were selected for observation between 18 and 38 times over several days of the week to investigate temporal variation. The three locations selected for temporal variation were signalized intersections representing two different Land Use areas (residential and commercial) and three different Wards. Observations of pedestrian/bicycle volumes and distances were sampled for the following 8 types of facilities:

1. Signalized intersections (39 locations)
2. Stop-controlled (all-way) intersections (27 locations)
3. Partially stop-controlled intersections (18 locations)
4. Mid-block locations with no crosswalks (10 locations)
5. Blocks with a large number of driveways and/or alleys (8 locations)
6. Parking lots and parking garages (10 locations)
7. Locations with playing, darting, dashing, auto repair, etc. in the roadway (8 locations)
8. School crossing areas, sampled when school was in session (2 locations).

A stratified random sampling procedure was devised so that an adequate spatial and temporal distribution of measurements was obtained. The 6 sampling variables used in the procedure are listed below, along with the number of categories per variable, and the range and mean for the number of observation locations per category:

1. Hour of Day, 24 categories: Range = 1-27 locations, Mean = 12.33
2. Time Period, 7 categories: Range = 13-20 locations, Mean = 16.72
3. Day of Week, 7 categories: Range = 9-25 locations, Mean = 17
4. Land Use Type, 7 categories (Low Density Residential; Medium Density Residential; High Density Residential; Public/Open Space; Federal/Local/Mixed Use/Public-Institutional; Industrial; Commercial): Range = 1-44 locations, Mean = 17.43
5. Political District (DC Ward), 8 categories: Range = 13-19 locations, Mean = 15.25
6. Zoning Type, 7 categories: Range = 1-44 locations, Mean = 17.43.

As can be seen in the above list of sampling variables, Time Period, Day of Week, and Political District (apportioned for roughly equal population) all had relatively uniform sampling distributions. Hour of Day had substantially fewer locations sampled at night, since there are fewer pedestrians present at night, and the security of the observers was an issue. Land Use Type and Zoning Type also had substantially fewer locations in Industrial and Manufacturing Areas, since Washington, DC, is primarily not an industrial city.

3.3 METHOD

3.3.1 Materials and Equipment

Observers traveled to designated measurement locations with the exception of the counts taken via traffic camera on the Internet. Each observer used a clipboard to hold a site-specific data collection form for each 15-minute count. Three mechanical counters were attached to the top of the clipboard. At most locations these were employed to count: #1 pedestrians in the crosswalk (one foot in the crosswalk for at least half of the crossing distance), #2 pedestrians not in the crosswalk (jaywalkers) and #3 bicycles traversing the data collection zone. An average diagonal crossing distance was applied to certain jaywalking counts. The observers also used a stopwatch to record time.

Several data sources were required to estimate the total population of each type of facility in the city. DDOT provided a list of the locations of all signalized intersections in the city (*15*). Satellite imaging software was used to measure and confirm the width of roads, and to provide estimated counts of all stop-controlled intersections, partially-controlled intersections, parking lots, and driveways in the city. Geographic Information System (GIS) software was used to estimate the total number of alleys, and the Yellow Pages telephone book was used to estimate the total number of parking garages. A standard statistical software package was used to perform general linear statistical modeling on the signalized intersection data.

3.3.2 Data Collection Procedures

Observers recorded the address, date, shift time, day of the week, weather, and observers' names on a form, and drew a sketch of the data collection location. Pedestrian and bicycle volumes were recorded on the mechanical counters. Crossing distances for all roadways, intersection legs, driveways, etc. were also recorded. Facility-specific data collection procedures were as follows:

3.3.2.1 Signalized and Stop-Controlled (All-Way) Intersections

One observer stood at one corner of the intersection, and the second observer stood at the diagonally opposite corner. Both observers faced the center of the intersection and were responsible for both legs of the intersection (road and crosswalk) to their immediate left, creating two separate zones split diagonally down the middle of the intersection. The range of observation extended to 50 ft (15.3 m) beyond the intersection box for each leg.

3.3.2.2 Partially Stop-Controlled Intersections

Data collection procedures for partially stop-controlled intersections (one-way and two-way stops) were similar, with one exception. Observers noted which roads were controlled by stop signs and which roads were uncontrolled. Additionally, observers noted if there were differences in road size and vehicular use for the intersecting roads. Observers classified the roads as primary, secondary, or equal based on judgments of size and vehicular use.

3.3.2.3 Mid-Block Locations with No Crosswalk

At an approximate mid-block location, one observer stood on one side of the road and the other observer stood on the opposite side. Both observers faced the center of the road and covered the area from their immediate left to the nearest intersection. Pedestrians and bicyclists entering the roadway from the mid point location were counted by the observer whose side they entered, regardless of where they exited the roadway. The observers counted all pedestrians crossing directly across the road at various angles, and all bicyclists riding in the road in the data collection zone. Appropriate average diagonal crossing distance estimates were applied to the pedestrian crossing counts.

3.3.2.4 Blocks with Driveways/Alleys

The observer locations and coverage areas were similar to the Mid-Block locations described above. Driveways widths were measured. Alternatively, a representative sample was taken, if there were numerous driveways at a given location, and the average width was recorded. The observers counted all pedestrians crossing the driveway(s) along the road or along the sidewalk, all pedestrians crossing the road in the assigned zone, and all bicyclists riding in the road or on the sidewalk in the assigned zone.

3.3.2.5 Parking Lots / Parking Garages

In parking lots and parking garages, one observer could only monitor 2-3 driving lanes at one time, thus the parking lots were split into two quadrants if there were between 3-6 driving lanes. When there were more than 6 driving lanes, the parking lot was split into 4 quadrants. The quadrant measurements were split into three zones. The average walking distance for each zone was assigned to each pedestrian traversing that zone. Parking lots and parking garages were generally the largest facility type observed, and presented the most measurement challenges. Therefore, bicycles were not counted at this facility type.

3.3.2.6 Playing / Working In Roadway

The observer locations and coverage areas were similar to the Driveways/Alleys locations described above. As pedestrians/bicyclists entered a shared facility (street, driveway, alley, or other shared facility) their time of entry was recorded. As the pedestrians/bicyclists completed their activity in the facility, the observers recorded the distance traveled, the number of pedestrians/bicyclists in the group, the type of activity, and their exit time. At the end of the 15-minute data collection period, for each activity recorded, the observers calculated the total time by subtracting the entry times from the exit times.

3.3.2.7 School Crossing Areas

Data were collected at one elementary school and one high school during the standard school day between the hours of 8AM and 3PM. In this case the observers were concerned only with the roads adjacent to the block on which the school was situated. The observers measured or estimated the width and length of all roads adjacent to the schools, as well as of any school entrance/exit driveways. The mechanical counters were used in a manner appropriate for each particular intersection in the vicinity of the school.

3.3.3 Data Analysis Procedures

Data from the 15-minute counts were multiplied to estimate hourly counts. Data were multiplied by two if there were two 15-minute counts and by 4 if there was one 15-minute count. When empirical data were not available, they were estimated using an expansion technique based on the temporal distribution of the entire data set for all locations observed (*16*). In this process the data set was first collapsed across all measurement locations and facility types to develop hourly adjustment factors. In the case of pedestrians, these hourly adjustment factors are shown in Figure 3-1, both for the present Washington, DC, data and for data derived from several American cities by Zegeer et al. (*16*). The 21st hour of the DC adjustment factor curve was estimated due to a lack of data. The 20th and 22nd hours were averaged, and the result was used to represent an average hourly total for the 21st hour of the day.

A similar set of hourly adjustment factors was computed for the DC bicycle counts, only these data were more variable. The Washington, DC, adjustment factors were then employed to estimate data for the missing hours at each location. All of the hourly data were subsequently summed over the 24-hour period to obtain an estimate of daily pedestrian and bicycle volume at each location. Walking/riding distances were obtained from the empirical data collected at the location. The average walking or riding distance was multiplied by each estimated hourly count, and then summed over 24 hours to obtain a daily distance estimate at each location. However, for bicycle exposure, the bicycle volumes for intersections and mid-block locations were multiplied by an average city block length of 500 feet (153 m), since bicyclists primarily ride the entire length of the block.

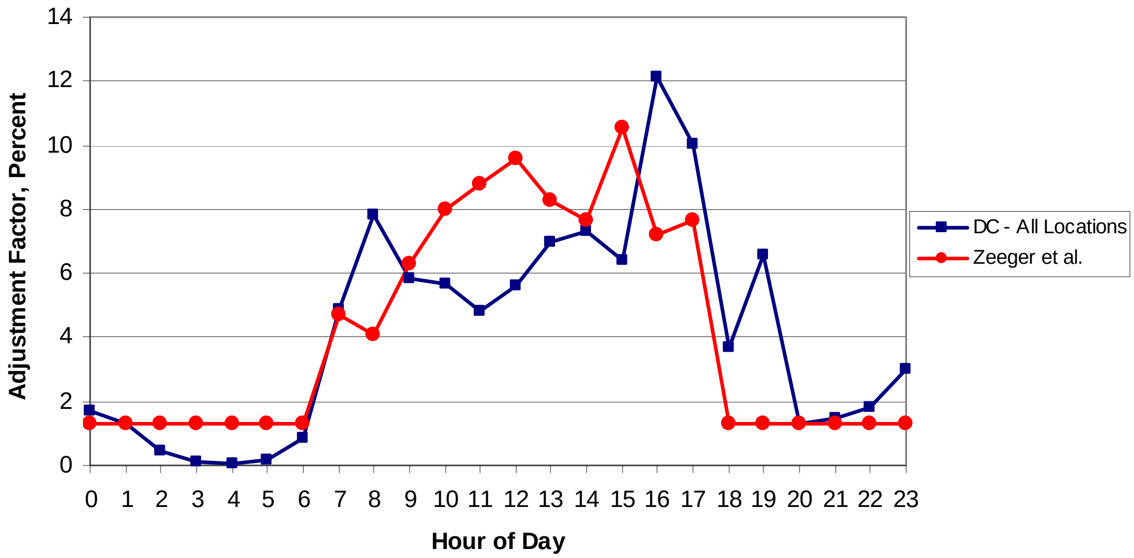

Figure 3-1 Adjustment Factor (Percent) as a Function of Hour of Day.

Within a given facility type, for example signalized intersections, the daily volume and distance estimates for each location were converted to logarithms and averaged over the number of facilities in the sample, in this case 39 locations. A logarithmic transform was applied to account for the skewed nature of the underlying volume and distance measurement distributions, approaching more closely a Poisson than a Gaussian form. The resulting geometric mean daily volume and distance calculations were taken as the best parameter estimates to characterize the daily activity at a typical signalized intersection in the city. To obtain the daily volume and distance for all signalized intersections, the geometric mean daily volume and distance estimates for a typical signalized intersection were multiplied by the total number of signalized intersections in the city. A similar process was used for the other types of facilities, except methods varied for determining the total population of that facility type. The annual volume and distance totals from all eight facility types were summed to obtain the estimated pedestrian and bicycle exposure for the city of Washington, DC, for the calendar year 2007. Facility-specific data analysis procedures were as follows:

3.3.3.1 Signalized Intersections

Data were collected at 39 signalized intersections. According to the DDOT, as of February 2008, there were 1,581 signalized intersections in Washington, DC (*15*).

3.3.3.2 Stop-Controlled (All-Way) Intersections

Data were collected at 27 all-way stop-controlled intersections. Satellite images were employed to determine the number of stop-controlled intersections. First, satellite images were used to obtain the total number of intersections of all types in Washington, DC. An estimate of the number of partially-controlled intersections was made (see section below). The number of signalized (1,581) and partially-controlled intersections (926) were subtracted from the total number of intersections, and the balance represented the number of stop-controlled intersections (3,654).

3.3.3.3 Partially Stop-Controlled Intersections

Data were collected at 18 partially stop-controlled intersections. Satellite images were used to determine the number partially stop-controlled intersections that exist in Washington, DC. Intersections with one road with no stop bars and one road with stop bars were counted as partially stop-controlled intersections. By this method the total number of partially-controlled intersections was determined to be approximately 926.

3.3.3.4 Mid-Block Locations with No Crosswalk

Data were collected at 10 sites at mid-block locations with no marked crosswalks. Mid-block locations were segments of road uninterrupted by an intersection. The number of mid-block locations in the city was estimated using satellite images. It was assumed that each intersection had four legs (road sections), with the exception of the intersections on the border of the city. The ratio of unique road sections to intersections in a standard grid system is 1.5-to-1. The total number of intersections in the city was estimated using satellite images, and multiplied by 1.5 to obtain the total number of mid-block with no crosswalk locations (9,242).

3.3.3.5 Blocks with Driveways/Alleys

Data were collected at 8 sites that had at least one driveway or alley that intersected the sidewalk and the road. Satellite images were used to estimate the total number of driveways (936) and alleys (960) in the city.

3.3.3.6 Parking Lots / Parking Garages

Data were collected at 8 sites that consisted of specialized parking facilities (lots and garages) where pedestrians and vehicles share the environment. Many parking facilities are closed during the late night and early morning hours. Consequently, for this facility type, the calculation of daily estimates was restricted to the hours between 6AM and 9PM, for a total of 15 hours. Satellite images and the Yellow Pages telephone book were employed to obtain population estimates of the total number of parking lots (904) and parking garages (242), respectively.

3.3.3.7 Playing / Working In Roadway

Data were collected at 8 sites in typically residential areas. There are on average 13 daylight hours in any given day of the year in Washington, DC (*17*), including approximately 30 minutes prior to sunrise and 30 minutes after sunset. Thus data were collected between the hours of 6AM and 7PM, when pedestrians and bicyclists might be found playing, working, riding or spending extended periods of time in a shared environment with motor vehicles. A 13-hour adjustment factor was developed and applied to the data from each location. A DC Government Land Use Map was used to estimate the percentage of residential land use for each Ward. Each Ward's percentage of residential land use was multiplied by the number of mid-block sections (as previously described) to obtain the number of potential play / work in roadway locations. Ward subtotals were summed to obtain the total number of potential locations (6,464).

3.3.3.8 School Crossing Areas

An inventory of all K-12 schools in the city was obtained from the DC Public School System and from observing satellite images. Each school was classified as an elementary, middle, or high school. A sample of schools was viewed via satellite images to determine their block structure, and categorized either as being in a half-block or whole-block environment. Based on the sample percentages, approximately 243 schools occupy a half block, and 162 schools occupy a whole block. Pedestrian and bicycle volumes and distances from the two schools were multiplied by the number of schools in each respective category (half block or whole block) to obtain the daily estimate for all schools in the city. These daily estimates were multiplied by 180 days to account for the number of schools days in a calendar year. No peak estimates were made since the typical school day was the same regardless of time of year.

3.3.4 Linear Modeling

The above data analysis procedures were based upon aggregating and summing empirical data on pedestrian and bicycle exposure to generalize from isolated samples of a few locations at a few times during a certain day to an estimate of the exposure for all locations for a calendar year. This empirical summation technique employed only one of the 6 sampling variables (Hour of Day) to refine the estimating procedure. This technique also employed one extrapolation variable, the seasonal correction factor, which was not a part of the empirical data set collected by the observers, because the correction factor was estimated from the visitor's bureau statistics.

To investigate the possible effect of some of the other sampling variables on pedestrian and bicycle volumes and distances, general linear statistical modeling methods were employed on a subset of the intersection data. The modeling followed a technique similar to the one used by Qin and Ivan (*10*). All 6 of the sampling variables were considered, along with one additional variable, Week Category (two categories: weekday, weekend). Models were run for pedestrian data only, and separate models were run for each of the three different types of intersections.

3.4 RESULTS

3.4.1 Pedestrian Exposure

For the city of Washington, DC, for the calendar year 2007, Table 3-1 shows the derivation of the annual pedestrian volumes and travel distances estimated for each type of facility shared with motor vehicles. Column 1 gives the name of the facility type, with the sample size in parentheses. Column 2 gives the approximate population of that facility type in the city. The estimated mean daily pedestrian volumes and distances for a typical facility of the given type are shown in Columns 3 through 6. The arithmetic means are given in bold numerals in Columns 3 and 4, with the upper and lower bounds for one standard error (alpha = 0.05) shown above and below the mean. Since the distribution of pedestrian volumes and distances was not Gaussian, but closer to Poisson, a logarithmic transform was applied to the underlying daily mean values. This transformation made the distribution of daily means much closer to normal. Thus the geometric means are given in bold numerals in Columns 5 and 6, with the upper and lower bounds equivalent to one standard error shown above and below the mean, as an indication of variability.

As expected, the upper and lower variability bounds for the geometric mean were asymmetric. Likewise, the geometric mean was lower than the arithmetic mean, since the arithmetic mean tended to overestimate the central tendency of the data due to the positive skew of the underlying measurement distribution. The geometric mean was also generally much closer to the median than the arithmetic mean. Thus the geometric mean was used to characterize the daily pedestrian volumes and distances for each facility type. These geometric means (Columns 5 and 6) were aggregated and summed across locations, days and seasons as described in the Method Section to obtain the Annual Volume and Distance estimates shown in Columns 7 and 8. The upper and lower variability estimates in Columns 7 and 8 were aggregated and summed in the same way as their corresponding means. The relatively high variability observed in the data was the result of the small sample sizes employed in the present study, as is readily apparent in a comparison of the *n* in Column 1 with the corresponding population in Column 2. As can be seen at the bottom of Column 8, the final result of the present research effort was an estimated annual mean pedestrian exposure of 0.82 hundred million pedestrian miles traveled (on a facility shared with motor vehicles) for the entire city of Washington, DC, for the calendar year 2007.

Table 3-1 Estimated Pedestrian Exposure for Washington, DC, 2007.

Facility Type (*n*)	Number of Facilities	Arithmetic Mean		Geometric Mean		Annual Pedestrian Volume (millions)	Annual Pedestrian Distance (millions of miles)
		Volume	**Distance (feet)**	**Volume**	**Distance (feet)**		
Signalized Intersection (39)	**1,581**	6,799	371,522	3,005	160,726	1,717	17.4
		5,603	**304,512**	**2,403**	**127,552**	**1,373**	**13.8**
		4,407	237,502	1,921	101,226	1,098	10.96
Stop-Controlled (All-Way) Intersection (27)	**3,654**	844	33,313	537	20,695	709.6	5.2
		722	**28,128**	**431**	**16,627**	**569.1**	**4.2**
		600	22,943	346	13,358	456.4	3.34
Partially-Controlled Intersection (18)	**926**	1,158	47,649	634	25,017	212.1	1.6
		916	**37,102**	**471**	**18,518**	**157.6**	**1.2**
		673	26,555	350	13,707	117.1	0.87
Mid-block Location with No Crosswalk (10)	**9,242**	1,935	68,184	916	41,153	3,058	26.0
		1,274	**49,762**	**641**	**28,410**	**2,142**	**18.0**
		612	31,340	449	19,613	1,500	12.41
Driveway/Alley (8)	**1,896**	213	2,129	52	516	35.4	0.0670
		126	**1,257**	**26**	**257**	**17.6**	**0.0334**
		38	384	13	128	8.79	0.0167
Parking Lot / Parking Garage (10)	**1,146**	9,194	499,077	8,388	430,019	3,475	33.7
		7,960	**419,373**	**6,836**	**340,176**	**2,832**	**26.7**
		6,725	339,669	5,572	269,104	2,308	21.11
Play/Work in Roadway (8)	**6,464**	661	99,060	501	57,912	1,170	25.6
		517	**68,747**	**337**	**38,720**	**787**	**17.1**
		373	38,435	227	25,889	530	11.46
School Crossing Area (2)	**405**	4,192	176,248	4,192	176,248	305.6	2.43
		3,038	**118,228**	**2,810**	**103,012**	**204.9**	**1.42**
		1,884	60,208	1,884	60,208	137.3	0.83
					Total	**8,084**	**82.4**

Initially, this estimate of annual pedestrian exposure may seem somewhat higher than expected. For example, the 2006 estimate for annual motor vehicle traffic exposure in Washington, DC, was only 36.2 hundred million vehicle miles traveled (VMT) for all roadway functional classes, a factor of only about 44 times greater. However, in a large city, although pedestrians do not travel as far as motor vehicles, in certain areas of the city, the density of pedestrians far outweighs the density of motor vehicles. Examination of Table 1 also reveals that four of the facility types contributed most to the overall pedestrian exposure (between 14 and 27 million miles traveled each). These were Parking Lots and Parking Garages, Mid-Block Locations, Play/Work in Roadway, and Signalized Intersections. The Driveway/Alley facility type contributed the least to overall pedestrian exposure in the city.

Qin and Ivan (*10*) collected daily pedestrian volumes at selected sites in rural Connecticut using a similar procedure, which also counted only pedestrian road crossings. Although their data were from rural regions, 10 out of the 32 sites sampled were classified as Downtown Areas, and 18 out of the 32 sites represented signalized intersections. Their weekday daily pedestrian volumes for a single site ranged from a minimum of 19 to a maximum of 2,788. In the present effort, the geometric mean daily pedestrian volumes for a typical site of a particular facility type ranged from 26 to 2,403. This correspondence may be regarded as a partial validation of the proposed methodology.

3.4.2 Bicycle Exposure

For the city of Washington, DC, for the calendar year 2007, Table 3-2 shows the derivation of the annual bicycle volumes and travel distances estimated for each type of facility shared with motor vehicles. Table 2 was constructed in the same manner as Table 1, with one exception. In the aggregation and summation process, no seasonal correction factor was used in the bicycle exposure estimates, because seasonal pattern data for bicycles in Washington, DC, were not presently available to justify and develop such a correction. It is likely that fewer bicycle trips are made during days with cold or inclement weather, and therefore the bicycle exposure estimates presented are probably somewhat overestimated.

As can be seen at the bottom of Column 8, in the case of bicycles, the final result was an estimated annual mean bicycle exposure of 0.37 hundred million bicycle miles traveled (on a facility shared with motor vehicles) for the entire city of Washington, DC, for the calendar year 2007. This is a little less than half of the annual pedestrian exposure. The grand total does not include the totals for Mid-Block Locations, because they were accounted for in the three types of intersection locations. However, the Mid-Block totals are given in the Table as an example of empirically-based estimates for that type of facility. As was the case for pedestrian exposure, the relatively high variability observed in the data was the result of the small sample sizes employed in the present study. The biggest contributor to annual bicycle exposure was the category of Signalized Intersection (17.5 million miles traveled). In certain ways, this major contribution is an artifact of the estimation procedure, which concentrated on bicycle counts from intersections, since more sites of this type were sampled for observations. In fact, most of the actual exposure occurred at Mid-Block Locations (37.3 million miles traveled), but this estimate was excluded from the total, as explained above. If Mid-block Locations had served as the basis for the calculations, instead of intersections, the estimated annual bicycle exposure would have been 0.40 hundred million miles traveled, less than a 10 percent discrepancy.

Table 3-2 Estimated Bicycle Exposure for Washington, DC, 2007.

Facility Type (*n*)	Number of Facilities	Arithmetic Mean		Geometric Mean		Annual Bicycle Volume (millions)	Annual Bicycle Distance (millions of miles)
		Volume	Distance (feet)	Volume	Distance (feet)		
Signalized Intersection (39)		478	238,863	363	181,280	209	19.8
	1,581	**424**	**211,771**	**319**	**159,694**	**184**	**17.5**
		369	184,680	281	140,678	162	15.38
Stop-Controlled (All-Way) Intersection (27)		253	126,528	116	58,168	155	14.7
	3,654	**182**	**90,976**	**99**	**49,337**	**132**	**12.5**
		111	55,425	84	41,847	112	10.57
Partially-Controlled Intersection (18)		273	136,362	186	93,082	62.9	6.0
	926	**222**	**111,220**	**151**	**75,679**	**51.2**	**4.8**
		172	86,079	123	61,530	41.6	3.94
Mid-block Location with No Crosswalk (10)		279	99,869	210	76,734	708	49.0
	9,242	**220**	**79,752**	**161**	**58,453**	**544**	**37.3**
		161	59,634	124	44,527	417	28.45
Driveway/Alley (8)		73	727	66	664	45.97	0.0871
	1,896	**62**	**616**	**53**	**528**	**36.6**	**0.0692**
		51	506	42	420	29.1	0.0550
Play/Work in Roadway (8)		145	7,242	109	5,431	256	2.4
	6,464	**110**	**5,501**	**85**	**4,231**	**200**	**1.9**
		75	3,760	66	3,296	156	1.47
School Crossing Area (2)		748	123,028	748	123,028	54.5	1.70
	405	**392**	**64,764**	**164**	**28,279**	**11.96**	**0.39**
		36	6,500	36	6,500	2.6	0.09
					Total	**615**	**37.2**

3.4.3 Linear Modeling

In order to investigate the possible effect of some of the sampling variables, general linear statistical modeling was employed on the pedestrian data for signalized intersections. The modeling followed a technique similar to the one used by Qin and Ivan (*10*). The 15-minute pedestrian counts (mean = 95.63, median = 16.00) served as the dependent variable. Independent variables were: (1) Land Use Group (LUG: commercial, residential, park open space [POS], federal, etc.); (2) Hour of Day (HOUR) or Time Period (PER (seven time-of-day categories: morning, mid-morning, noon, afternoon, early evening, late evening and overnight)), and (3) Day of the Week (DOW) or "week category" (WKCAT: weekday and weekend). The collected data contained 112 complete observations for all variables.

Forward selection and backward elimination techniques, using all independent variables and all possible two-way interaction terms, were used to determine the final model. The pedestrian counts were not normally distributed, but instead followed a Poisson distribution. Thus, linear modeling techniques requiring normality assumptions could not be used. The use of generalized linear models to analyze non-linear count data has been well-documented and noted in other similar pedestrian studies (*10*). In the present study a Poisson regression was used with a log linear model. Significant terms (p<0.0001) in the model were: PER, LUG, DOW and the two-way interaction of PER with LUG. Table 3-3 shows the main effect parameter estimates for 15-minute pedestrian counts at signalized intersections.

Table 3-3 Main Effect Parameter Estimates for 15-Minute Pedestrian Counts

Variable	Category	Estimate	Standard Error
Intercept	n/a	2.883	0.178
PER	Morning	1.905	0.196
	Mid-Morning	0.892	0.239
	Noon	0.819	0.19
	Afternoon	3.046	0.262
	Early Evening	1.085	0.218
	Late Evening	3.678	0.234
	Overnight	0	0
LUG	Residential	0	0
	Commercial	2.506	0.181
	Federal	-0.391	0.391
	POS	1.387	0.308
DOW	Monday	-3.789	0.097
	Tuesday	-1.193	0.096
	Wednesday	0	0
	Thursday	-1.16	0.043
	Friday	-1.497	0.043
	Saturday	-3.102	0.194
	Sunday	-2.507	0.067

To arrive at yearly counts, the above parameter estimates were substituted into the appropriate model equation (*10*). For a given PER, the 15-minute predicted count was multiplied by the number of 15-minute intervals in that PER. Next, this estimate was multiplied by the number of facility type examples in a given LUG. Similarly, predictions for the remaining PER in a given DOW were computed. This method was repeated for all other DOW and LUG categories. Those computations were then summed to attain an estimated pedestrian count for one week at all signalized intersections. From this weekly estimate, the yearly estimate was calculated by applying the seasonal correction factor adjusted for weekly periods. The total number of miles traveled was estimated by multiplying the total number of pedestrians by the mean width of all the sampled signalized intersections (mean = 51.3 ft, sd = 11.5 ft, median = 50 ft; 15.6, 3.51 and 15.3 m, respectively). This distance was then converted to miles. The result was an estimated pedestrian exposure for signalized intersections of 0.355 hundred million miles (hmm) of roadway traveled (sd = 0.0055 hmm). This mean was about 2.6 times greater than the estimate derived from the scaling method. Such a discrepancy is not completely unexpected, given the small samples sizes and early stages of model development represented in the present study.

3.5 DISCUSSION

The present paper presents a new metric for measuring pedestrian and bicycle exposure to risk. The proposed metric is the distance traveled by pedestrians or bicyclists on a shared facility, in hundred million miles walked or biked. The paper had two major goals: 1) to describe the methodology for realizing this new metric in practice, and 2) to employ this methodology to calculate the annual pedestrian and bicycle exposure for a relatively large urban environment in terms of the proposed metric. Both of these goals were accomplished. First, the method for implementing the proposed exposure metric was described in considerable detail for the 8 pedestrian/vehicle and bicycle/vehicle shared facility types characteristic of the urban environment of Washington, DC. Other cities may have different characteristic shared facility types which have not yet been explored. Second, the methodology was used to calculate the annual pedestrian and bicycle exposure for Washington, DC, for the calendar year 2007. The result was 0.82 hundred million miles traveled for pedestrian exposure, and 0.37 hundred million miles traveled for bicyclist exposure. To achieve this result a scaling technique was employed to generalize from mean daily pedestrian and bicycle volumes at a single example of a given facility to annual exposures for an entire city. This feature allows the methodology to be used for daily and even hourly (as well as monthly and seasonal) exposure calculations for the purpose of effectively comparing before and after crash data in the evaluation of specific pedestrian and bicycle safety countermeasures. This feature also allows the methodology to be used for comparisons across different locations in a given city or area, to track changes in the spatial patterns of pedestrian and bicycle activity. In addition, a linear regression model was tried as a possible approach to enhance the efficiency of estimating exposure.

By using the estimated pedestrian and bicycle exposure as denominators, and pedestrian and bicycle crashes as numerators, the respective crash risks can be calculated. The estimated crash risk for pedestrians in Washington, DC, for 2007 was 617 (crashes) / 0.82 (hundred million miles traveled), or 752 crashes per hundred million miles traveled (*13*). The estimated crash risk for bicyclists in Washington, DC, for 2007 was 289 (crashes) / 0.37 (hundred million miles traveled), or 781 crashes per hundred million miles traveled (*13*). When first comparing the crashes by themselves, it would appear that pedestrians are at a higher risk of being in a motor-vehicle related crash compared to bicyclists. However, when exposure is taken into consideration, the crash risks are actually similar.

Thus the proposed metric has the potential to eliminate one of the major obstacles in the pedestrian and bicycle safety field, the lack of adequate exposure data (*1*). In its general form the proposed metric can handle both pedestrians/bicyclists crossing the roadway as well as traveling on and along the roadway. The metric captures the concept of sharing the roadway (or other facility) with motorized traffic. The distance component makes it sensitive to the amount of individual exposure to a potentially hazardous environment on a single crossing or travel segment. The metric can also be easily converted to hundred million skater miles or scooter (or other mode) miles traveled (on a shared facility) for special applications. In its most general form, the proposed metric becomes hundred million non-motor vehicle miles of shared transportation facility traveled. Thus parking lots, parking garages, driveways and alleys can also be accommodated and aggregated, along with intersections and mid-block locations. Special variations of the methodology have been elaborated to handle the additional contribution of school areas and playing, working, darting, and dashing behaviors in the roadway. Thus the proposed metric shows promise of being able to serve as the basis for a universal measure of pedestrian and bicycle exposure.

These promising features of the proposed exposure metric do not imply that no more work need be done. As mentioned before, the small temporal and spatial sample sizes employed in the present study resulted in relatively large variability. More research is needed to estimate and test appropriate sampling variables and sample sizes. However, it is envisioned that yearly counts would not be required at all locations. Similar to the counting procedure used to estimate VMT, the pedestrian and bicyclist counts would take place on a rotating basis so that many locations would be covered, but over a 3-year or 5-year cycle. The length of this cycle would depend upon the rapidity of anticipated changes in exposure estimates over time. Improved sampling procedures need to be developed to enhance usefulness to both the empirical aggregation approach and to the linear regression modeling approach. Moreover, for this metric to be effective on a national scale, tests need to be conducted in different cities of differing sizes and geographical locations, as well as in small towns and rural areas. New variations of the measurement methodology may need to be developed to accommodate these different environments. In all of these efforts, scalability needs to be maintained so that the metric can serve the practicing engineer determining the effectiveness of a given single pedestrian/bicycle safety countermeasure at a given location in a given city or town, as well as planners and policy makers concerned with city-wide, regional or national trends. In addition, increased involvement from the local area and community needs to be enlisted to evaluate the potential for propagating the metric on a self-sustaining basis. At the same time, the technical details of creating and fielding a national database for pedestrian and bicycle exposure need to be elaborated, and funding and maintenance issues need to be explored. If these steps are taken, the potential return on investment could be significant. The widespread implementation of such a metric for pedestrian and bicycle exposure could have far-reaching implications for reducing pedestrian and bicycle fatalities and injuries, since it could form the basis for selecting optimal engineering safety countermeasures on a broad scale, and allocating limited fiscal resources in a cost-effective manner.

3.6 ACKNOWLEDGEMENTS

The FHWA sponsored this work. The primary contractor was Science Applications International Corporation (SAIC). This research could not have been completed without the assistance and advice of numerous people, including contractors and government employees: Mary Anne Bertola, Elliott Tibor, Sheryl Miller, Gabriel Rousseau, Dana Duke, Vaughan Inman, Vernese Edghill, Wanda Parham, Nicole Fowler, Tiana Petit, Megan Pavlock, and Erin Dagnall. George Branyan and Bill McGuirk, from the DC Department of Transportation, were valuable resources during the planning and data collection phases of the study. Stephen Fleger (SAIC) and Thomas Granda (FHWA) managed the overall Support Services Contract under which the work was performed.

3.7 REFERENCES

1. Raford, N., and D. Ragland. Space Syntax: Innovative Pedestrian Volume Modeling Tool for Pedestrian Safety. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1878*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 66-74.
2. National Center for Statistics and Analysis. *Traffic Safety Facts: Pedestrians.* Publication DOT HS 810 810. NHTSA, U.S. Department of Transportation, 2006.
3. Fatality Analysis Reporting System. NHTSA, U.S. Department of Transportation.
4. General Estimates System. NHTSA, U.S. Department of Transportation.
5. Federal Highway Administration. *Highway Statistics*. Publication FHWA-PL-08-018. U.S. Department of Transportation, 2006.
6. NHTSA/FHWA Pedestrian and Bicycle Strategic Planning Research Workshops. Final Report. NHTSA/FHWA, U.S. Department of Transportation, April 2000.
7. Chu, X. The Fatality Risk of Walking in America: A Time-Based Comparative Approach. Presented at 4th International Conference on Walking in the 21st Century, Portland, OR, 2003.
8. Keall, M. D. Pedestrian Exposure to Risk of Road Accident in New Zealand. *Accident Analysis and Prevention*, Vol. 27, No. 27, 1995, pp. 729-740.
9. Cameron, M. A Method of Measuring Exposure to Pedestrian Accident Risk. *Accident Analysis and Prevention*, Vol. 14, No. 5, 1982, pp. 397-405.
10. Qin, X., and J. Ivan. Estimating Pedestrian Exposure Prediction Model in Rural Areas. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1773,* Transportation Research Board of the National Academies, Washington, D.C., 2001, pp. 89-96.
11. FHWA. *National Household Survey.* U.S. Department of Transportation, 2001.
12. Kaplan, J. Characteristics of the Regular Adult Bicycle User. Master's Thesis. University of Maryland, 1975. Accessed www.bicyclinglife.com/Library on November 6th, 2008.
13. *Washington, DC's 2006 Visitor Statistics.* Destination DC. http://www.washington.org/images/marketing/2006_visitorstats.pdf. Accessed July 15, 2008.
14. DDOT Traffic Cameras Location Map. http://app.ddot.dc.gov/services_dsf/traffic_cameras/index.asp. Accessed July 15, 2008.
15. DDOT Office of Traffic Services Administration. Personal communication. November 2008.
16. Zegeer, C., R. Stewart, H. Huang, P. Lagerwey, J. Feganes, and B. Campbell. *Safety Effects of Marked versus Unmarked Crosswalks at Uncontrolled Locations.* Publication FHWA-HRT-04-100. FHWA, U.S. Department of Transportation, 2005.
17. *The Old Farmer's Almanac.* http://www.almanac.com/astronomy/. Accessed July 15, 2008.

CHAPTER FOUR: PEDESTRIAN CRASH PREDICTION MODELS

Jason F. Kennedy, Hesham A. Rakha, and Patches Johnson Inge

ABSTRACT

Crash prediction models have been used by the highway safety community to predict vehicle-to-vehicle crashes over the past several decades using Generalized Linear Models (GLMs) because they can model non-linear discrete data such as motor vehicle crashes. Poisson, Zero-inflated Poisson (ZIP), and the Negative Binomial models are commonly used to model crashes. However, until recently, very little research has been conducted on crash prediction modeling for pedestrian-motor vehicle crashes. This paper considers several candidate crash prediction models using thirteen explanatory variables and three regression functions. Data from Washington, DC are used in the development of a pedestrian crash prediction model for signalized intersections. The results of the crash prediction analyses indicate the pedestrian-vehicle crash model is similar to models developed in earlier research efforts. Furthermore, two pedestrian volume estimation methods were compared demonstrating that a scaling technique produces volume estimates highly correlated with observed pedestrian volumes. The ratio of crashes and exposure estimates gives a crash rate estimation that is useful for traffic engineers and transportation policy makers to evaluate pedestrian safety at signalized intersections in an urban environment.

4.1 INTRODUCTION

In order to evaluate the safety impacts associated with various traffic operational projects it is essential that a normalized crash rate be computed. The computation of a crash rate requires estimating some normalized measure of pedestrian-vehicle exposure. It is common to use pedestrian-miles traveled as a measure of exposure, however this assumes that the crashes are linearly proportional to the pedestrian volume, which in most cases is not true.

The research presented in this paper develops a pedestrian-vehicle crash prediction model that accounts for the pedestrian-vehicle exposure. The database used to develop the model is initially described in terms of the data elements and how the data were collected and obtained. Subsequently, the crash prediction model development is discussed followed by a description of the crash prediction results. Additionally, the observed pedestrian volumes are compared to the two pedestrian volume-estimating method estimates described in Molino et al (2008). The results from this task (predicted pedestrian volumes) are used in the crash prediction model to obtain predicted crashes and then compared to the predicted crashes obtained with the observed pedestrian volumes. Finally, the study conclusions are presented.

4.2 SIGNALIZED INTERSECTION DATABASE

Before the analysis could begin, the data had to be collected, reduced, and organized. A database of 200 signalized intersections from Washington, DC was developed. The methodology of how and where the data were collected and obtained is described below.

4.2.1 Exposure Data

The exposure data used in the model development were pedestrian and vehicle volumes and crosswalk length (pedestrian travel distance) at each intersection. Pedestrian and vehicle volumes were obtained from the District of Columbia's Department of Transportation (DDOT). Crosswalk length (feet) for each leg was estimated using Google Earth™ satellite imagery.

DDOT conducts pedestrian counts and vehicle-turning movement counts at approximately 100 intersections each year (Schneider et al, 2005). DDOT has over 29 years of counts for signalized and stop-controlled intersections throughout the city of Washington, DC. The 200 signalized intersections that were used in the research were chosen from between 2003 and 2006, to coincide with the pedestrian crash data (described in a subsequent section) that was provided by DDOT. A larger database would have been preferred to further reduce variability however, 200 locations were deemed acceptable given the limited availability of location-specific pedestrian crash data. While previous research has used larger datasets the availability of pedestrian crash data limited the temporal range that could be considered and thus the database size.

The two requirements each location had to meet for consideration was that it was a signalized intersection and the traffic and pedestrian volume data collection took place between 2003 and 2006. The candidate intersections were then chosen based on the available data provided by DDOT. Each year's DDOT's intersection count inventory contained varying numbers of signalized intersections. While an attempt was made to include an equal number of signalized intersections in the database per year (50) it was not possible. All of the signalized intersection counts were included from years that had less than 50 available candidates (years 2004 – 2006). For 2003, the intersections that were included in the database were chosen at random because there were more than 50 possible candidates. In the event any intersection was included twice in the original selection process, the count from the later date was retained and the count with the earlier date was replaced with a new, unique intersection so as to have 200 unique intersections represented in the database. Table 4-1 shows the number of intersections by year.

Table 4-1. Number of signalized intersections by year

Year	# of Intersections
2003	72
2004	37
2005	44
2006	47
Total	200

Pedestrian and vehicle counts were made by DDOT data collectors in teams of two or three, depending on the volume intensity of the intersection being observed. The counts were taken in 15-minute intervals continuously over the course of a 10-hour day with an hour break for lunch. Pedestrian counts were made for each crosswalk at every leg of the intersection regardless of direction of travel. Vehicle counts were made for each leg of the intersection by turning movement. Counts were taken with mechanical clicker boards. Following each 15-minute data collection interval, the pedestrian and vehicle counts were recorded on a paper data collection form (see Figure 4-1).

10/2/2003 **16TH STREET, NW & FULLER STREET**

VEHICLE AND PEDESTRIAN COUNT
DISTRICT OF COLUMBIA
DEPARTMENT OF PUBLIC WORKS
BUREAU OF TRAFFIC SERVICES

		Time	South Side Motor Vehicles Left	South Side Thru	South Side Right	South Side TOTAL	South Side COUNTER-FLOW	Peds A	North Side Motor Vehicles Left	North Side Thru	North Side Right	North Side TOTAL	North Side COUNTER-FLOW	Peds B	West Side MV COUNTER-FLOW	Peds C	East Side Motor Vehicles Left	East Side Thru	East Side Right	East Side TOTAL	East Side COUNTER-FLOW	Peds D	TOTALS MV	TOTALS Peds
			W	N	E		S	11	E	S	W		N	11	W		S	W	N		E			
DAY -	THURS	7:00		82		82	521	1	5	519	11	535	83	15	11	4	2		1	3	5	17	617	37
WEATHER -	CLEAR	7:15		96		96	499		6	496	14	516	98	27	15	7	3	1	2	6	6	26	612	60
POSTED BY -	JONES	7:30		75		75	544	4	5	542	9	556	81	30	9	4	2		6	8	5	32	631	70
INTERSECTION ID	10160280	7:45		117	1	118	569		8	569	11	588	120	31	11	3			3	3	9	29	706	63
CONTROLLED BY		Hr Tot		370	1	371	2133	5	24	2126	45	2195	382	103	46	18	7	1	12	20	25	104	2566	230
STOP SIGN ON FULLER ST		8:00		126		126	576		7	576	7	590	130	19	7	6			4	4	7	41	716	66
		8:15	2	97		99	550	5	4	549	6	559	102	29	10	2	1	2	5	8	4	27	658	63
		8:30	3	133	2	138	563	3	6	562	6	574	140	25	12	3	1	3	7	11	8	30	712	61
		8:45	3	93	2	98	563	8	4	559	11	574	98	15	18	9	4	4	5	13	6	9	672	41
		Hr Tot	8	449	4	461	2252	16	21	2246	30	2297	470	88	47	20	6	9	21	36	25	107	2758	231
		9:00	2	99	7	108	544	8	13	543	9	565	100	7	13	10	1	2	1	4	20	15	673	40
		9:15	6	105	1	112	412	4	3	411	12	426	109	5	20	12	1	2	4	7	4	9	538	30
		9:30		79		79	380	7	13	380	6	399	86	9	6	9			7	7	13	13	478	38
		9:45		73		73	308	8	2	308	10	320	82	3	10	7			9	9	2	8	393	26
		Hr Tot	8	356	8	372	1644	27	31	1642	37	1710	377	24	49	38	2	4	21	27	39	45	2082	134
		10:00	1	82		83	294	6	9	294	5	308	87	8	6	9			5	5	9	10	391	33
		10:15		112		112	270	11	11	270	9	290	116	7	9	6			4	4	11	5	402	29
		10:30	1	120		121	221	10	5	221	6	232	124	9	7	8			4	4	5	8	353	35
		10:45		134	1	135	199	13	2	199	3	204	137	7	3				3	3	3	8	339	28
		Hr Tot	2	448	1	451	984	40	27	984	23	1034	464	31	25	23			16	16	28	31	1485	125
		11:00	5	115	1	121	196	8	4	194	3	201	117	7	10	3	2	2	2	6	5	6	322	24
		11:15	1	124		125	177	12	6	177	7	190	124	5	10	1		2		2	6	9	315	27
		11:30	2	112		114	241	8	5	241	7	253	113	7	9	3			1	1	5	3	367	21
		11:45	2	115		117	227	9	1	227		228	115	2	3			1		1	1	12	345	23
		Hr Tot	10	466	1	477	841	37	16	839	17	872	469	21	32	7	2	5	3	10	17	30	1349	95
		12:00	4	110		114	221	4	3	219	2	224	110	6	6	2	2			2	3	6	338	18
		12:15	3	122		125	212	5	1	209	1	211	124	4	4	3	3		2	5	1	15	336	27
		12:30	2	124		126	183	3	2	183	3	188	124	11	5	4					2	13	314	31
		12:45	1	114	2	117	229	9	3	227	7	237	115	3	8	2	2		1	3	5	6	354	20
		Hr Tot	10	470	2	482	845	21	9	838	13	860	473	24	23	11	7		3	10	11	40	1342	96
PEDESTRIANS (10hr)	1326	2:00	1	158	1	160	185	4	7	184	3	194	160	3	4		1		2	3	8	14	354	21
10- HR FACTORS	(default=.6)	2:15	1	149	3	153	165	4	5	162	5	172	150	2	6	3	3		1	4	8	7	325	16
X =	0.60	2:30	1	208	4	213	211	9	4	210	4	218	209	11	5		1		1	2	8	8	431	28
Y =	0.60	2:45	3	180		183	198		7	195	4	206	183	6	7	2	3		3	6	7	8	389	16
Z =	0.60	Hr Tot	6	695	8	709	759	17	23	751	16	790	702	22	22	5	8		7	15	31	37	1499	81
W=	0.60	3:00	1	192	2	195	217	4	6	216	3	225	194	9	4	2	1		2	3	8	6	420	21
PEAK HOURS SUMMARY:		3:15	1	222	3	226	224	5	3	220	5	228	223	8	6	1	4		1	5	6	5	454	19
AM PkHr. (starting)	7:45 AM	3:30	2	210	3	215	236	9	4	235	4	243	212	9	6	3	1		2	3	7	8	458	29
AM PkHr Vol =	2818	3:45		250	3	253	241	10	5	241	3	249	253	7	3				3	3	8	7	502	24
AM PkHr Factor =	.97	Hr Tot	4	874	11	889	918	28	18	912	15	945	882	33	19	6	6		8	14	29	26	1834	93
AM LOS =		4:00	4	223	1	228	252	5	6	252	8	266	226	5	12	4			3	3	7	15	494	29
AM V/C =		4:15	2	199	2	203	251	10	6	251	1	258	203	9	3	5			4	4	8	13	461	37
AM PkHr %OF24 =	9.0%	4:30		216	1	217	249	6	5	249	2	256	221	4	2	4			5	5	6	10	473	24
PM PkHr (starting)	3:45 PM	4:45		210		210	242	14	3	240	3	246	223	6	3	5	2		13	15	3	13	456	38
PM PkHr Vol =	1945	Hr Tot	6	848	4	858	994	35	20	992	14	1026	873	24	20	18	2		25	27	24	51	1884	128
PM PkHr Factor =	.96	5:00		221		221	248	8	4	247	4	255	229	7	4	10	1		8	9	4	15	476	40
PM LOS =		5:15		217		217	230	4	5	230	5	240	219	9	5	6			2	2	5	14	457	33
PM V/C =		5:30		211	1	212	237		6	237	4	247	215	6	4	5			4	4	7	12	459	23
PM PkHr %OF 24 =	6.2%	5:45		179	3	182	229		6	229	5	240	182	4	5	4			3	3	9	9	422	17
		Hr Tot		828	4	832	944	12	21	943	18	982	845	26	18	25	1		17	18	25	50	1814	113
10 HOUR TOTAL			54	5804	44	5902	12314	238	210	12273	228	12711	5937	396	301	171	41	19	133	193	254	521	18613	1326
A M PEAK HOUR			5	473	3	481	2258	8	25	2256	30	2311	492	104	40	14	2	5	19	26	28	127	2792	253
P. M PEAK HOUR			6	888	7	901	993	31	22	993	14	1029	903	25	20	13			15	15	29	45	1930	114
24 Hr (estimate)			90	9673	73	9837	20523	397	350	20455	380	21185	9895	660	502	285	68	32	222	322	423	868	31022	2210

Figure 4-1. DDOT intersection count data collection form.

DDOT did not collect crosswalk length data. The crosswalk lengths were estimated using the Google Earth™ ruler tool. Distances were measured in the middle of the crosswalk from curb to curb in feet. Figure 4-2 shows a typical crosswalk length measurement using Google Earth™. This figure is unaltered and used for illustrative purposes.

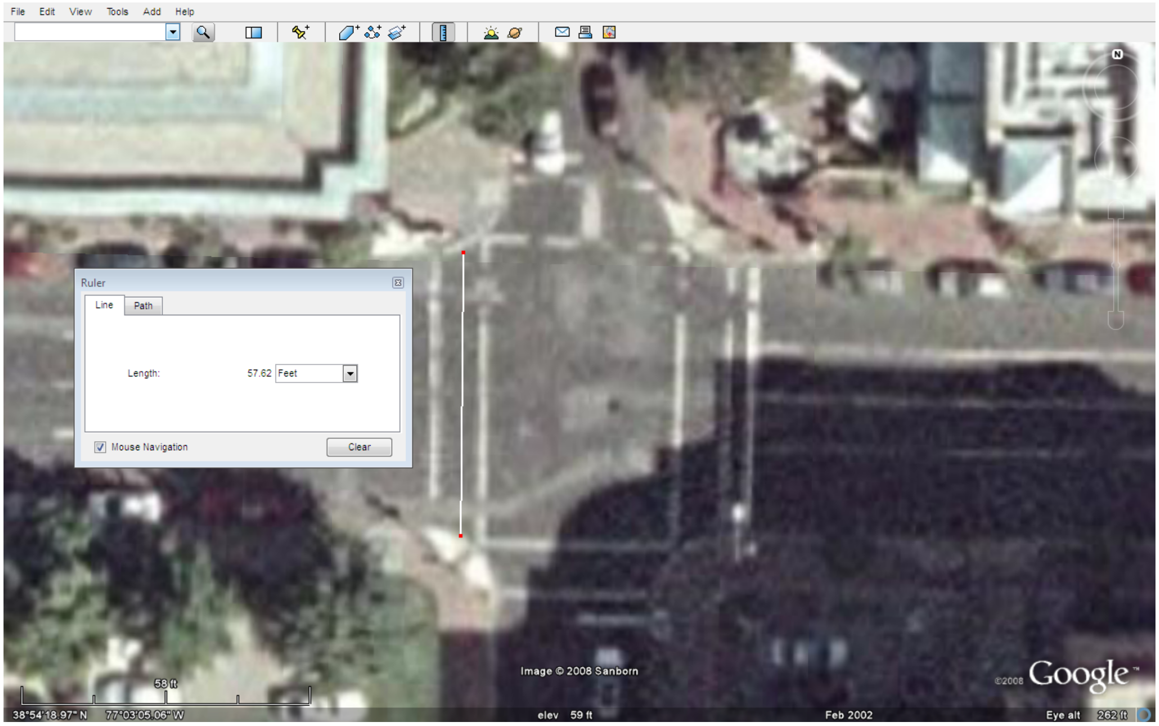

Figure 4-2. Google Earth ruler tool used to estimate intersection crosswalk lengths. (This figure is unaltered and for illustrative purposes only.)

The pedestrian and vehicle counts and crosswalk lengths for each leg were entered into an Excel® spreadsheet for all 200 signalized intersections included in the crash prediction model database.

4.2.2 Crash Data

The Metropolitan Police Department (MPD) is the lead agency responsible for investigating and documenting motor vehicle crashes, including those involving pedestrians, in the city of Washington, DC. The MPD typically documents motor vehicle crashes using an Accident Report (PD-10). DDOT receives these PD-10s and maintains the information as part of their highway and traffic safety statistics databases.

The crash data used in the analysis were obtained from a pedestrian-motor vehicle crash database created and maintained by DDOT. Figure 4-3 shows the frequency distribution of the crash data used in the analysis.

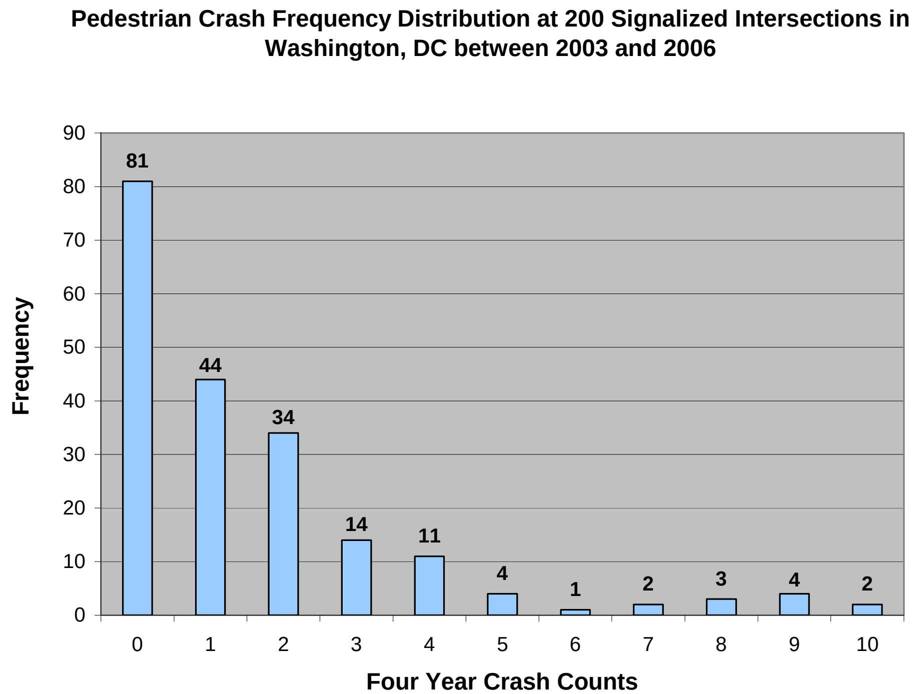

Figure 4-3. Frequency distribution of pedestrian crashes at signalized intersections.

The total number of pedestrian-motor vehicle crashes between 2003 and 2006 (4-year counts), for each of the 200 signalized intersections were used as the response or dependent variable. Table 4-2 gives the descriptive summary statistics for the pedestrian crash data. The frequency distribution (Figure 4-3) and descriptive statistics (Table 4-2) provide strong evidence that the crash data do not follow a normal (Gaussian) distribution.

Table 4-2. Descriptive statistics of signalized intersection pedestrian crash data

Descriptive Statistics	
Mean	1.59
Standard Error	0.15
Median	1.00
Mode	0
Standard Deviation	2.16
Sample Variance	4.66
Kurtosis	4.29
Skewness	2.04
Range	10
Minimum	0
Maximum	10
Sum	318
Count	200
Confidence Level (95.0%)	0.30

4.2.3 Intersection Characteristics Data

Supplemental data were collected on the characteristics of the 200 signalized intersections. The supplemental data included intersection geometry, vehicular traffic directional flow (one-way or two-way), and number of legs. All three of these data types were obtained through the use of Google Earth™ satellite imagery and maps and then categorized into two levels. Table 4-3 shows the data type, level, and number of intersections per level.

Table 4-3. Number of intersections by data type and level

Data Type	Level	# of Intersections
Geometry	90 degree	108
	Not 90 degree	92
Vehicle Traffic Flow	Two-Way (all legs)	111
	One-Way (at least one leg)	89
Legs	3 or 4 legs	177
	5 or 6 legs	23

4.3 Model Development

This section describes the initial model that was used as the basis of the analysis as well as the other models investigated during the analysis. The models investigated included various combinations of explanatory variables and regression functions to produce the best possible fit to the data.

4.3.1 Initial (Hypothesized) Model

Since traffic crashes are compared to some type of exposure measure to establish crash rates, the following equation was used to derive the initial model (hypothesis) used in the analyses (Equation 1). The hypothesized explanatory (exposure) variables are listed in the denominator.

$$CR = \frac{C_p}{V_p^{\,a} V_v^{\,b} L_t^{\,c}} \tag{1}$$

Where:

CR = Crash Rate
C_p = Pedestrian Crashes
V_p = Pedestrian Volume
V_v = Vehicle Volume
L_t = Total distance of intersection crosswalks for all legs
a,b,c = exposure coefficients

The SAS® 9.2 statistical package was used to conduct the regression analysis. Generalized Linear Models (GLMs) were developed using the GENMOD program within SAS®. Equation 1 was reformulated into a linear form to be suitable for the statistical programming syntax. Equation 2 is the initial model used in SAS®.

$$\ln C \;=\; a \ln V_p \;+\; b \ln V_v \;+\; c \ln L_t \;+\; \ln CR \tag{2}$$

The full model including other parameters can be cast as Equation 3.

$$C = V_p^{\,a} V_v^{\,b} L_t^{\,c} e^{\sum_i B_i x_i} + E \tag{3}$$

Where B_i are coefficients of the various explanatory coefficients x_i and E is the error term. It should be noted the $V_p^a\, V_v^b\, L_t^c$ accounts for the exposure while the $B_i x_i$ are the various factors accounted for in the model.

4.3.2 Regression Models

This section provides a brief review of the mathematical formulas that are behind the regression models explored during the model development.

4.3.2.1 Poisson Model

The Poisson regression model, as the literature has noted, possesses statistical properties desirable for predicting traffic crashes. Crash data are count data which, by nature, are non-negative, discrete, and have a variance that increases with the mean of the distribution (SAS, 2008). The Poisson model uses a Poisson distribution function, which is defined in Equation 4 as:

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \tag{4}$$

where the parameter λ¸ is the mean value of the random variable Y, E(Y) = λ. This random variable takes on values as integers from zero to infinity. Larger values of the mean parameter will produce greater random variable values. The Poisson distribution assumes that the mean equals the variance.

4.3.2.2 Zero-Inflated Poisson (ZIP) Model

The zero-inflated Poisson regression was developed to model count data with excess zeros. The zero-inflated Poisson (ZIP) regression mixes two statistical processes by using a Logit model to determine if a count is from an "always-zero group" or a "not-always-zero group" and a Poisson model to model the outcomes in the "not-always zero group" (SAS, 2008). The probability density function of the ZIP model is shown in Equation 5.

$$P(N_i) = \begin{cases} P_i + (1 - P_i)\mathrm{e}^{-\mu_i} & (N_i = 0) \\ (1 - P_i)\dfrac{\mathrm{e}^{-\mu_i}\mu_i^{N_i}}{N_i!} & (\text{otherwise}) \end{cases} \tag{5}$$

P_i is used to represent the additional probability of intersection *i* to have no crashes while 1 - P_i is the probability that intersection follows the Poisson distribution (Qin et al, 2004). The probability that intersection *i* will have no crashes is represented by $e^{-\mu i}$.

4.3.2.3 Negative Binomial Model

Data modeled with a Poisson distribution can sometimes be over dispersed (variance exceeds the mean). When data are over dispersed, a Negative Binomial (NB) model may be more appropriate. The Negative Binomial probability function, as shown in Equation 6 takes the form:

$$P(Y = y) = \frac{\Gamma(y + 1/k)}{\Gamma(y + 1)\Gamma(1/k)} \frac{(k\mu)^y}{(1 + k\mu)^{y+1/k}} \quad \text{for y = 0, 1, 2,…} \tag{6}$$

where: μ = expected number of crashes
k = dispersion parameter
y = actual or observed number of crashes at an intersection during a period of time

As k approaches zero, the negative binomial regression yields the Poisson regression.

4.3.3 Explanatory Variables

Prior to developing the crash models, the variables were analyzed for possible collinearity. The results indicate only a small handful of possible combinations (in bold) of two variables that produced correlation coefficients above .50 at a significance level of .05. Table 4-4 shows the correlation matrix. The explanatory variables that were used to

Table 4-4. Correlation matrix of raw explanatory variables

		V_p	V_v	L_t	Tf	G	Lg	R_v	Rr_v	L_v	Rl_v	Min_v	Maj_v	MiMaRv
V_p	**Correlation Coef.**	1.0000												
	p-value	---												
V_v	**Correlation Coef.**	0.1736	1.0000											
	p-value	0.0140	---											
L_t	**Correlation Coef.**	0.2052	**0.5612**	1.0000										
	p-value	0.0036	**<.0001**	---										
Tf	**Correlation Coef.**	0.0011	0.1833	0.1959	1.0000									
	p-value	0.9874	0.0094	0.0054	---									
G	**Correlation Coef.**	0.0009	0.1226	0.3658	0.1401	1.0000								
	p-value	0.9896	0.0837	<.0001	0.0479	---								
Lg	**Correlation Coef.**	0.0308	0.0710	0.2362	0.2449	0.2962	1.0000							
	p-value	0.6649	0.3176	0.0008	0.0005	<.0001	---							
R_v	**Correlation Coef.**	0.1737	0.4928	0.3966	0.1100	0.1787	0.2255	1.0000						
	p-value	0.0139	<.0001	<.0001	0.1209	0.0113	0.0013	---						
Rr_v	**Correlation Coef.**	0.0825	0.0716	0.0680	0.0026	0.2331	0.2590	**0.7039**	1.0000					
	p-value	0.2456	0.3137	0.3390	0.9712	0.0009	0.0002	**<.0001**	---					
L_v	**Correlation Coef.**	0.0893	0.4765	0.3956	0.1839	0.1608	0.1682	**0.6558**	0.3173	1.0000				
	p-value	0.2086	<.0001	<.0001	0.0091	0.0230	0.0173	**<.0001**	<.0001	---				
Rl_v	**Correlation Coef.**	0.0609	0.0048	0.1620	0.1778	0.1969	0.1550	0.3963	0.4781	**0.7550**	1.0000			
	p-value	0.3916	0.9458	0.0219	0.0118	0.0052	0.0284	<.0001	<.0001	**<.0001**	---			
Min_v	**Correlation Coef.**	0.1171	**0.5890**	0.3377	0.1221	0.1051	0.1183	0.3221	0.0086	0.2596	0.0129	1.0000		
	p-value	0.0986	**<.0001**	<.0001	0.0851	0.1387	0.0954	<.0001	0.9033	0.0002	0.8562	---		
Maj_v	**Correlation Coef.**	0.0770	**0.9306**	0.4225	0.1660	0.1040	0.0248	0.3326	0.1988	0.3670	0.1043	0.3738	1.0000	
	p-value	0.2783	**<.0001**	<.0001	0.0188	0.1427	0.7271	<.0001	0.0048	<.0001	0.1415	<.0001	---	
MiMaRv	**Correlation Coef.**	0.0814	0.0678	0.0248	0.0761	0.1756	0.1228	0.0561	0.0868	0.0151	0.0500	**0.6311**	0.2971	1.0000
	p-value	0.2519	0.3400	0.7275	0.2843	0.0129	0.0832	0.4300	0.2216	0.8321	0.4824	**<.0001**	<.0001	---

predict the dependent variable (four-year pedestrian crashes) in the models are described below in Table 4-5. The table describes the variables, the reasoning why they were chosen, and the model hypothesis. The explanatory variables related to pedestrian volumes and vehicular volumes are totals for the four-year period between 2003 and 2006. The totals were derived by multiplying the 200 individual daily pedestrian and vehicular volumes by 365 to get annual totals and then by 4 to get four-year totals. It was assumed that the individual counts were "average" daily pedestrian and vehicular volume therefore no adjustments (i.e. seasonal variation) were made to the estimate. Tables 4-6 and 4-7 provide some basic descriptive statistics for each of the explanatory variables initially considered. The statistics in Table 4-6 and Table 4-7 are based off of the raw data and natural logarithm of the raw data, respectively.

Table 4-5. Description of explanatory variables

Variable	Description	Reason for Consideration	Model Hypothesis
lnV_p	Natural logarithm of the total volume of pedestrians crossing the road at each intersection for 4 years.	Pedestrian volumes are a primary exposure measure when considering crashes with vehicles. (Qin and Ivan, 2001; Lyon and Persaud, 2002)	An increase in volume will result in some increase in crashes.
lnV_v	Natural logarithm of the total volume of vehicles entering an intersection for 4 years	Vehicular volumes are a primary exposure measure when considering crashes with pedestrians. (Qin et al, 2004; Zegeer et al, 2005)	An increase in volume will result in some increase in crashes.
lnL_t	Natural logarithm of the total length of all the crosswalks at the intersection.	Distance exposed is typically used in vehicle exposure measures therefore pedestrian distance is a reasonable analogy. (Molino et al, 2008; Harwood et al, 2008)	Larger distances will result in some increase in crashes.
Tf	The vehicular traffic flow on the intersection approaches. This is a categorical variable that has two levels; 1.All 2-way traffic, 2. Not all 2-way	The vehicular traffic flow may have an influence on the pedestrian crashes and the volumes. Pedestrians may or may not prefer to cross at intersections with one-way traffic when compared to two-way traffic. (Lea and Associates, 1978)	One-way traffic flow will have fewer crashes than with two-way traffic.
G	Geometric shape of the intersection. This is a categorical variable that has two levels; 1. 90-degree 2. Not 90-degree	The line of sight for both pedestrians and vehicles approaching an intersection could be affected by the intersection geometry, and have an impact on exposure and crashes. (Carter et al, 2006)	Non-90 degree intersections will have a higher number of crashes.
Lg	Number of intersection legs. This is a categorical variable that has two levels; 1. 3 or 4 legs, 2. 5+ legs	The number of legs in which vehicles and pedestrians can interact may have an impact on crashes. (Harwood et al, 2008)	Intersections with more legs will have a higher number of crashes.
lnR_v	Natural logarithm of the total volume of right turning vehicles at the intersection for 4 years.	Pedestrians who cross on green are potentially at risk from right-turning vehicles. The total number of right-turning vehicles or the proportion of all vehicles that are right-turning were considered. (Leden, 2002)	As the number of right turning vehicles increases, crashes will increase.
$lnRr_v$	Natural logarithm of the percentage of the total vehicle volume turning right at the intersection.		As the percentage of right turning vehicles increases, crashes will increase.
lnL_v	Natural logarithm of the total volume of all left turning vehicles at the intersection for 4 years.	Pedestrians who cross on green are potentially at risk from left-turning vehicles. Vehicles turning left during an unprotected move may not be aware of pedestrians crossing because they are watching on-coming vehicle traffic. The total number of left-turning vehicles or the proportion of all vehicles that are left-turning were considered. (Lord, 1996; Lyon and Persaud, 2002)	As the number of left turning vehicles increases, crashes will increase.
$lnRl_v$	Natural logarithm of the percentage of the total vehicle volume turning left at the intersection.		As the percentage of left turning vehicles increases, crashes will increase.
$lnMin_v$	Natural logarithm of total vehicle volume on minor road entering the intersection for 4 years.	Intersections with roads that have substantially different vehicle volumes may have different crash patterns than those that are homogenous. The minor and major road vehicle volumes are considered as well as the ratio of minor to major volumes. (Harwood et al, 2008)	As volume on the minor road increases, crashes will increase.
$lnMaj_v$	Natural logarithm of total vehicle volume on major road entering the intersection for 4 years.		As volume on the major road increases, crashes will increase.
lnMiMaRv	Ratio of minor road vehicle volume to major road vehicle volume entering the intersection.		As the ratio increases, crashes will increase.

Table 4-6. Descriptive statistics of explanatory variables using the raw data

Variable	Mean	Standard Error	Median	Standard Deviation	Sample Variance	Kurtosis	Skewness	Minimum	Maximum	Confidence Level (95.0%)
V_p	2960070	434297.52	1176030	6141894	3.772E+13	39.07	5.56	16060	56665520	981512
V_v	40569115	1578900.16	38381940	22329020	4.986E+14	2.80	1.29	7859180	136076380	3568314
L_t	211.230	4.14	202	58.509	3.423E+03	-0.50	0.33	78	370	9.350
Tf	1.445	0.04	1	0.498	2.482E-01	-1.97	0.22	1	2	0.080
G	1.460	0.04	1	0.500	2.496E-01	-1.99	0.16	1	2	0.080
Lg	1.115	0.02	1	0.320	1.023E-01	3.95	2.43	1	2	0.051
R_v	4274413	304464.03	2823640	4305772	1.854E+13	9.74	2.58	52560	28488980	688089
Rr_v	0.109	0.01	0	0.090	8.183E-03	4.03	1.85	0	1	0.014
L_v	3671564	280313.51	2534560	3964232	1.572E+13	10.65	2.79	77380	27287400	633509
Rl_v	0.091	0.01	0	0.074	5.456E-03	3.69	1.78	0	0	0.012
Min_v	5706169	357561.03	4428180	5056677	2.557E+13	4.37	1.76	33580	30502320	808088
Maj_v	17799305	751449.44	16492160	10627100	1.129E+14	3.57	1.40	3120020	71532700	1698276
$MiMaR_v$	0.368	0.02	0	0.267	7.136E-02	-0.47	0.65	0	1	0.043

Table 4-7. Descriptive statistics of explanatory variables using the natural logarithm of the raw data

Variable	Mean	Standard Error	Median	Standard Deviation	Sample Variance	Kurtosis	Skewness	Minimum	Maximum	Confidence Level (95.0%)
$\ln V_p$	13.89235	0.1045893	14	1.479115	2.1877824	0.27575	-0.23348	10	18	0.2363717
$\ln V_v$	17.36864	0.0400213	17	0.565987	0.320341	-0.3293	-0.29239	16	19	0.0904481
$\ln L_t$	5.313451	0.020233	5	0.286138	0.081875	-0.2292	-0.30357	4	6	0.0457266
$\ln R_v$	14.8302	0.0709382	15	1.003218	1.006447	0.66831	-0.48697	11	17	0.1603204
$\ln Rr_v$	-2.53845	0.0612179	-3	0.865752	0.7495261	2.69274	-0.80551	-7	-1	0.1383524
$\ln L_v$	14.62743	0.0756196	15	1.069423	1.1436648	0.36587	-0.50008	11	17	0.1709003
$\ln Rl_v$	-2.74122	0.0665194	-3	0.940727	0.8849673	2.81273	-1.18318	-6	-1	0.1503339
$\ln Min_v$	15.07165	0.0853419	15	1.206916	1.4566472	2.4912	-1.35952	10	17	0.1928726
$\ln Maj_v$	16.51742	0.0437703	17	0.619005	0.3831674	-0.343	-0.31292	15	18	0.0989208
$\ln MiMaR_v$	-1.44577	0.0863958	-1	1.221821	1.492846	3.34989	-1.68064	-7	0	0.1952545

4.3.4 Candidate Models

Several models were tested using combinations of 13 variables and three regression model types. A backward stepwise selection process was used to refine the models. All of the variables of interest were included initially in the model. Variables that were not significant at the .05 level were then removed one at a time. Two final models were obtained for each regression model type. The final candidate models are presented in Table 4-8.

Table 4-8. Results of candidate models

Model Number	Regression Model Type		Regression coefficient						Predictor of Excess Zeros	Over dispersion parameter K (standard error)	R^2_{LR}	R^2_K (based on over dispersion)	Scaled Deviance / DF	Scaled Pearson X^2 / DF
			Intercept	a	b	c	d	e						
1	Poisson	Variable	----	$\ln V_p$	$\ln V_v$	Tf	Lg	*$\ln L_t$*	N/A	Scaled at 1	.18	N/A	1.84	1.96
		Coefficient	-17.4	.34	.93	.43	-.42	*-.55*						
		Standard Error		.04	.15	.13	.19	*.29*						
		p-value	< .0001	< .0001	< .0001	.0008	.03	*.06*						
2	Poisson	Variable	----	$\ln V_p$	$\ln V_v$	Tf			N/A	Scaled at 1	.17	N/A	1.85	2.03
		Coefficient	-17.8	.30	.79	.30								
		Standard Error	2.04	.04	.12	.12								
		p-value	< .0001	< .0001	< .0001	.009								
3	ZIP	Variable	----	$\ln V_p$	$\ln Maj_v$	Tf	Lg		$\ln Maj_v$	Scaled at 1	.13	N/A	N/A	1.37
		Coefficient	-11.7	.34	.47	.44	-.50		-.93					
		Standard Error	2.24	.04	.13	.14	.19		.37					
		p-value	< .0001	< .0001	.0002	.001	.01		.03					
4	ZIP	Variable	----	$\ln V_p$	Tf				$\ln R_v$	Scaled at 1	.10	N/A	N/A	1.32
		Coefficient	-3.86	.31	.36				-.61					
		Standard Error	.66	.04	.13				.19					
		p-value	< .0001	< .0001	.004				< .0001					
5	Negative Binomial	Variable	----	$\ln V_p$	$\ln L_v$	Tf			N/A	.72 (.16)	.07	.39	1.04	.96
		Coefficient	-8.84	.36	.27	.39								
		Standard Error	1.58	.06	.09	.18								
		p-value	< .0001	< .0001	.004	.03								
6	Negative Binomial	Variable	----	$\ln V_p$	$\ln V_v$				N/A	.59 (.14)	.09	.50	1.04	1.11
		Coefficient	-20.01	.33	.90									
		Standard Error	3.00	.06	.17									
		p-value	< .0001	< .0001	< .0001									

Note: Variables in italics were not significant at the .05-level.
Note: $R^2_{LR} = 1 - \text{Log Likelihood}_{FullModel} / \text{Log Likelihood}_{NullModel}$; $R^2_K = 1 - K / K_{max}$
Note: 1-standard error in parentheses

4.3.6 Final Pedestrian Crash Prediction Model

Table 4-9 shows the final pedestrian-vehicle crash prediction model at signalized intersections in Washington, DC. Pedestrian volume and vehicle volume were the only two independent variables that made the final model. Both variables were statistically significant at the 0.05 confidence level. This model was chosen based on an evaluation of the coefficients, goodness-of-fit measures, and how these compared to previous research. As indicated by the positive coefficients, the model indicates that pedestrian crashes increase with increasing pedestrian volume and total vehicle volume. One standard error values are provided in parentheses. Four goodness-of-fit measures are presented for the negative binomial model results. Since there is no universal definition of R-squared in nonlinear models, having more than one measure of a pseudo R-square allows for flexibility for evaluating the results (Cameron et al, 2007).

Table 4-9. Final pedestrian crash prediction model

Number of Sites	Regression coefficient			Over dispersion parameter			Scaled Deviance / DF	Scaled Pearson X^2 / DF
	Intercept a	$\ln V_p$ b	$\ln V_v$ c	K	R^2_{LR}	R^2_K		
200	-20.00 (3.00)	.33 (.06)	0.90 (.17)	0.59 (.14)	0.09	0.50	1.04	1.11

Note: 1-standard error in parentheses

The goodness-of-fit measures, R^2_{LR} and R^2_K had values of .09 and .50, respectively. The formulas for R^2_{LR} and R^2_K are indicated in Equations 6 and 7, respectively (Harwood, Zegeer, Lyon, et al, 2008; Vogt and Bared, 1998). These values indicate that the model explains more variance in the pedestrian-vehicle crash frequency than an intercept-only model. Additionally, the scaled deviance and scaled Pearson chi-square goodness-of-fit measures were 1.04 and 1.11, respectively. Values near 1 indicate the model is a good fit to the data.

$$R^2_{LR} = 1 - \text{Log Likelihood}_{FullModel} / \text{Log Likelihood}_{NullModel} \quad (6)$$

$$R^2_K = 1 - K / K_{MAX} \quad (7)$$

where: K = over dispersion parameter estimated in the negative binomial model
K_{MAX} = over dispersion parameter estimated in the negative binomial model with only a constant term and an over dispersion parameter

The hypothesized model included a distance variable, L_t for the crosswalk length. This variable was not statistically significant at the 0.05 confidence level in any of the variations of the models. The reason for this lack of significance could be due to the fact that vehicle volumes and roadway width tend to be correlated. A wider road is more likely to have higher vehicular volumes. The correlation coefficient (r) between vehicle volume and crosswalk length (~ roadway width) was 0.56 (p<.0001). While not a substantially high linear relationship, it is moderately high and does provide some explanation for the results. Vehicle volume, which did make it into the final model, could be accounting for crosswalk width indirectly. It is also completely feasible that distance (length of crosswalk) is not a significant factor in pedestrian-motor vehicle crashes. Duration in the road (time-based exposure) may be more significant, however that was not explored in this paper.

The model coefficients demonstrate that the average crash rate (exponent of the intercept) is approximately 2.06 crashes/10^9 pedestrian-vehicles. The coefficients also demonstrate that vehicle crashes are more sensitive to increases in the vehicle volume in comparison to pedestrian volume (0.90 vs. 0.33).

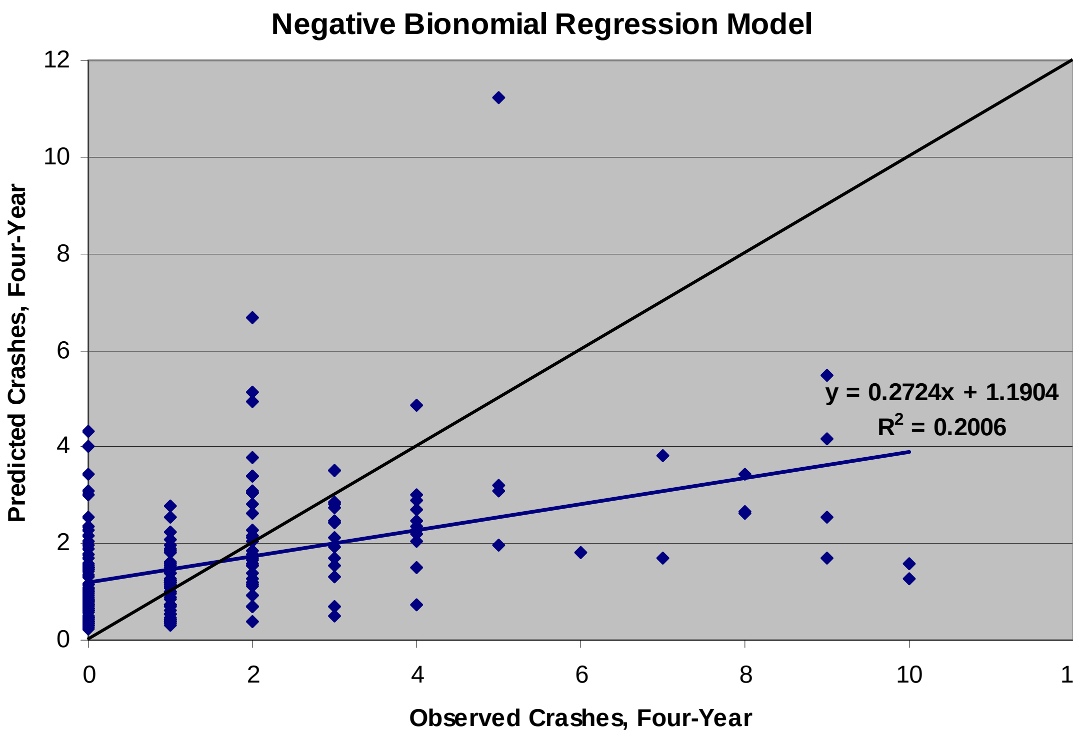

Figure 4-4. Comparison of observed and predicted crashes using final crash prediction model.

Figure 4-4 shows a comparison of the observed crashes and the predicted crashes from the final pedestrian crash prediction model in Table 4-9. The pattern of the graphed data highlights the fact that the observed data are discrete values and the predicted data are continues values. The trendline and associated linear equation reveals a reasonable level of correlation (Pearson correlation coefficient of .20). The 45-degree line shows how the relationship of the data compares against a perfect linear relationship. The level of correlation is comparable to previous research (Rakha et al, 2008) and is acceptable for the purposes of this paper.

4.4 Comparison of Pedestrian Volumes

Daily pedestrian volumes for each of the signalized intersections were estimated using two exposure techniques employed by Molino et al (2008). The first technique utilized a scaling function that is based on an hourly distribution for a 24-hour day. The technique was first developed by Zegeer et al (2005) and slightly modified by Molino et al (2008). Zegeer et al (2005) used a single 1-hour count (100%) at each location to estimate daily pedestrian volumes where as Molino et al used two 15-minute counts from consecutive hours. Additionally, Zegeer et al (2005) used 12-hour counts at a few locations to derive the distribution, whereas Molino et al (2008) collapsed all locations to create a generic 24-hour distribution that was assumed for an entire city regardless of the specific location. The hourly adjustment factors developed by Molino et al (2008) are shown in Table 4-10. The estimation of pedestrian exposure was for the entire city and not specific locations and one main goal of that research was to produce safety measures using limited empirical data.

Each location had 100% counts from 7AM to 6PM (except for 1PM – 2PM), however only two (2) 15-minute counts in consecutive hours were needed for this particular estimation technique. The two consecutive hours were chosen for each of the 200 intersections so that all hours were accounted for as equally as possible. The 15-minute counts were multiplied by four (4) to estimate hourly counts and then those two hourly estimates were each divided by their respective hourly adjustment factor. These two numbers could each be considered potential daily total estimates however depending on which two hours were assigned for that intersection, the counts could vary significantly due to the adjustment factor.

Therefore to address the possibility of drastically different results (depending on which adjustment factor was used) the two estimated totals were averaged. This averaged number was then used as the basis for the daily estimate by multiplying each of the 24 hourly adjustment factors by this estimated daily total. The 24 hourly estimates were then summed to compute a daily pedestrian volume at each intersection. Since the intersection count data were taken at various times of the year no seasonal adjustment was made to any of the daily totals.

Table 4-10. Adjustment factors by time of day (Molino et al, 2008)

Hour of Day	Adjustment %
0:00	1.7
1:00	1.3
2:00	0.5
3:00	0.1
4:00	0.1
5:00	0.2
6:00	0.9
7:00	5.0
8:00	7.9
9:00	5.9
10:00	5.7
11:00	4.9
12:00	5.7
13:00	7.1
14:00	7.4
15:00	6.5
16:00	12.3
17:00	10.1
18:00	3.7
19:00	6.7
20:00	1.3
21:00	0.1
22:00	1.8
23:00	3.0

The second pedestrian volume estimating technique utilized general linear statistical modeling (Molino et al, 2008). The independent variables were: (1) Land Use Group (LUG: commercial, residential, park open space (POS), federal, etc.); (2) Hour of Day (HOUR) or Time Period (PER (seven time-of-day categories: morning, mid-morning, noon, afternoon, early evening, late evening and overnight)), and (3) Day of the Week (DOW) or "week category" (WKCAT: weekday and weekend). Molino et al (2008) used a Poisson regression with a log linear model. Significant terms (p<0.0001) in the model were: PER, LUG, DOW and the two-way interaction of PER with LUG. The main effect parameter estimates for 15-minute pedestrian counts at signalized intersections are listed in Table 4-11 (Molino et al, 2008).

Table 4-11. Main effect parameter estimates for 15-minute pedestrian counts (Molino et al, 2008)

Variable	Category	Estimate	Standard Error
Intercept	n/a	2.883	0.178
Time Period (PER)	Morning	1.905	0.196
	Mid-Morning	0.892	0.239
	Noon	0.819	0.19
	Afternoon	3.046	0.262
	Early Evening	1.085	0.218
	Late Evening	3.678	0.234
	Overnight	0	0
Land Use Group (LUG)	Residential	0	0
	Commercial	2.506	0.181
	Federal	-0.391	0.391
	POS	1.387	0.308
Day of Week (DOW)	Monday	-3.789	0.097
	Tuesday	-1.193	0.096
	Wednesday	0	0
	Thursday	-1.16	0.043
	Friday	-1.497	0.043
	Saturday	-3.102	0.194
	Sunday	-2.507	0.067

The hourly counts for each of the intersections were grouped into time categories as described in Molino et al, 2008. Land use and day of the week categories were assigned to each location as appropriate. The pedestrian volumes estimates for time category by land use type and day of week were derived from the coefficients (parameter estimates). The time category estimates were summed to compute the daily estimate for each day of the week and land use combination (7 x 4 = 28). The 28 daily estimates were assigned to the 200 signalized intersections with the same land use and day of the week category. The predicted volumes were compared to the observed volumes at the 200 signalized intersections in the database developed for this paper. Figure 4-5 shows the correlation between the observed and predicted pedestrian volumes using the generalized linear model method. It should be noted that the predicted data appear to be plotted in categories, or bins. This is a result of the finite number of outputs (predicted volumes) that can occur with the categorical variables used in the GLM modeling procedure. Figure 4-6 shows the correlation between the observed and predicted pedestrian volumes using the scaling technique.

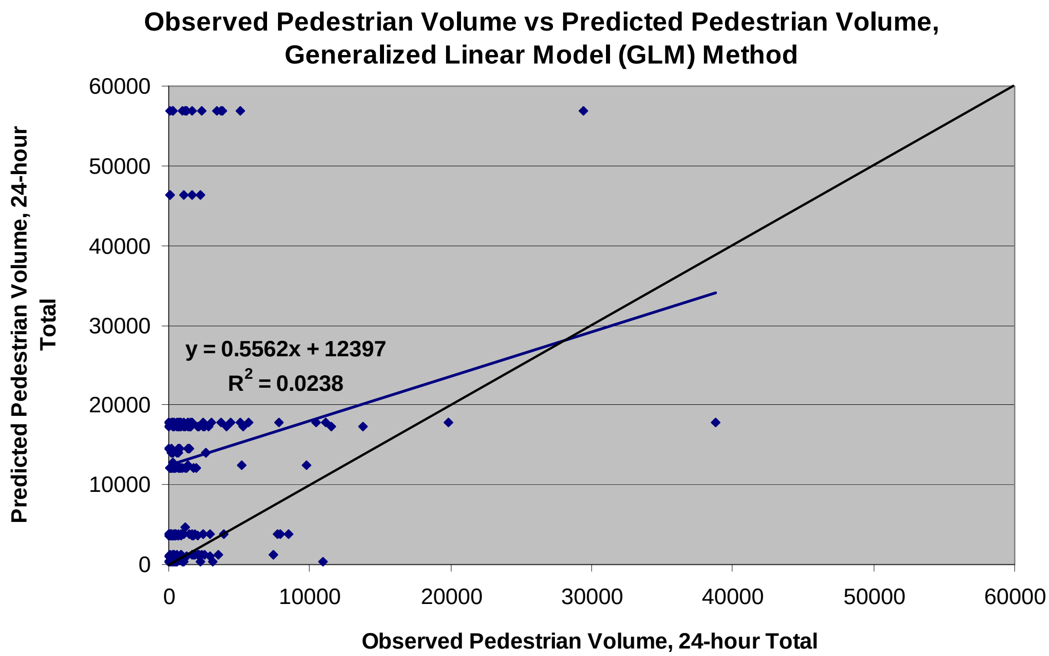

Figure 4-5. Comparison of observed pedestrian volume versus predicted for 200 signalized intersections using the GLM method

A trendline was created and a linear equation with an R-square value was used to assess the strength of the relationship between the two variables. The R^2 value of 0.0238 indicates there is practically no relationship between the predicted and observed pedestrian volumes.

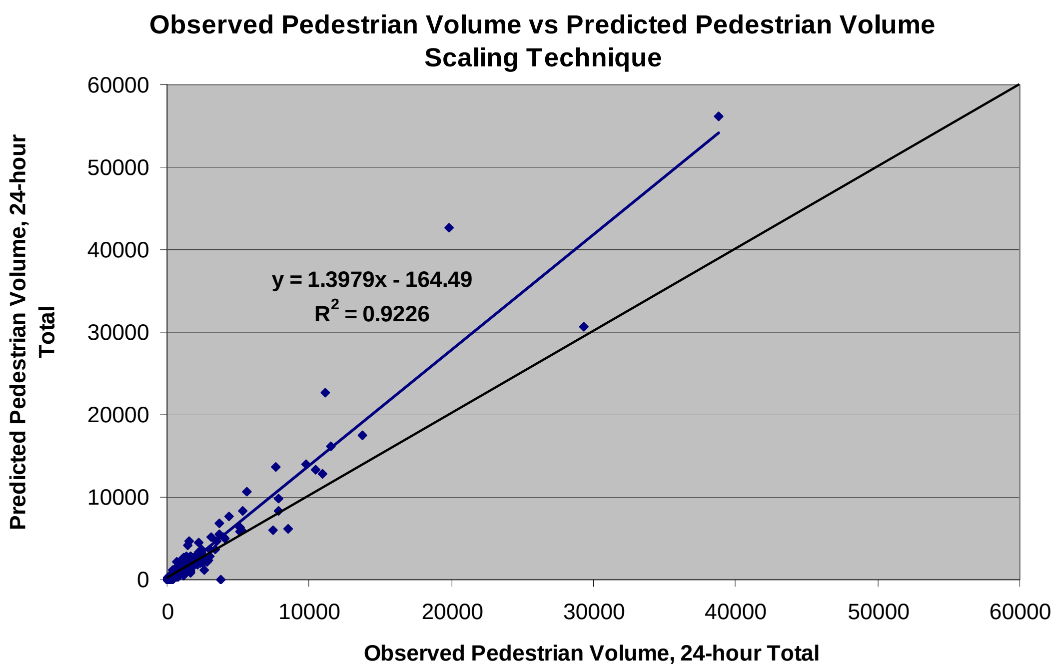

Figure 4-6. Comparison of observed pedestrian volume versus predicted for 200 signalized intersections using the scaling technique

A trendline was created and a linear equation with an R-square value was used to assess the strength of the relationship between the two variables. The R^2 value of 0.9226 indicates that approximately 92% of the variance in the response variable (predicted pedestrian volumes) is accounted for by the explanatory variable (observed pedestrian volume). This is evidence that a very strong, positive relationship exists between the predicted and observed pedestrian volumes when the scaling technique was used.

4.5 Comparison of Predicted Crashes

The final crash prediction model was run twice using the two sets of predicted pedestrian volumes instead of the observed pedestrian volumes. Figure 4-7 compares the observed crashes and the predicted crashes using the predicted pedestrian volumes obtained with the GLM estimating method. The trendline and associated linear equation indicate the correlation of the observed and predicted crashes. The correlation coefficient (R^2 =.1096) indicates that there is a weaker relationship between the observed and predicted crashes when the observed pedestrian volumes were used in the crash prediction model.

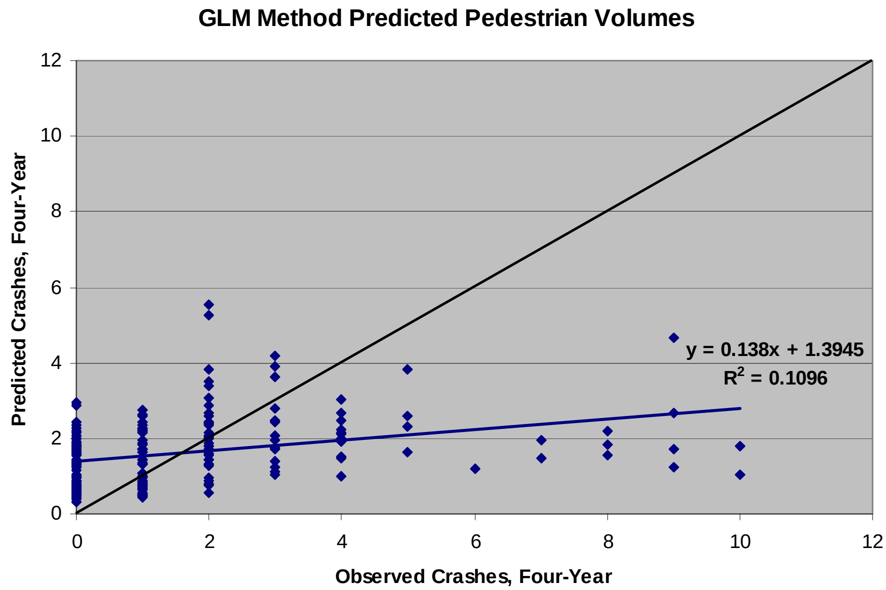

Figure 4-7. Crash comparison using predicted pedestrian volumes by the GLM method

Figure 4-8 compares the observed crashes and the predicted crashes using the predicted pedestrian volumes obtained with the scaling estimating method. The trendline and associated linear equation indicate the correlation of the observed and predicted crashes. The correlation coefficient (R^2 =.187) indicates that there is a stronger relationship between the observed and predicted crashes when the pedestrian volumes obtained with the scaling method were used in the crash prediction model.

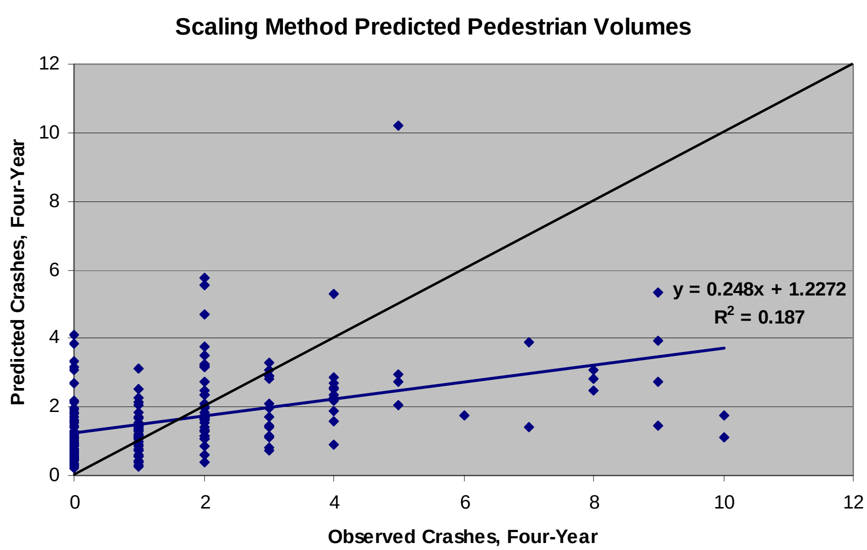

Figure 4-8. Crash comparison using predicted pedestrian volumes by the aggregate method

4.7 Discussion of Findings and Conclusions

The results of the crash prediction analyses indicate that the pedestrian-vehicle crash prediction model is similar in structure and performance to models developed in earlier research efforts (Lyon and Persaud, 2002; Zegeer, et al., 2005, Harwood, et al, 2008). Pedestrian volume and vehicular volume, which have been previously identified as strong predictor variables, were highly significant in the model developed. One specific difference that should be noted is that all of the variables in the final model presented in this paper were significant at the 0.05 confidence level. The final models from previous work by Lyon and Persaud (2002) included variables that met less strict confidence level of 0.10. The models from Zegeer et al (2005) included variables that had significance levels that ranged between < 0.0001 and 0.984. Another difference that should be noted is that the sample size for the dataset used in this paper was smaller than the ones used in Lyon and Persaud (2002) and Zegeer et al (2005). The database of signalized intersections described in this paper had 200 locations, whereas the two studies cited had over 350 and 2,000 sites, respectively. Additionally, the Lyon and Persaud (2002) differentiated between 3-leg and 4-leg intersections. This paper included intersections with 3 to 6 legs and used that characteristic as a possible explanatory variable. The results obtained in this paper are as strong, if not stronger, than those cited studies with much larger sample sizes.

The results from the pedestrian volume estimation analysis show that when using the data used in the crash prediction model, scaling technique provided stronger results when compared to the GLM method. The scaling technique produced predicted pedestrian volumes that were highly correlated to the observed pedestrian volumes. The GLM method appeared to have very little correlation. As noted in the paper by Molino et al (2008) the linear model developed to predict pedestrian volumes was based off of a fairly small sample size, which would explain the weak relationship of predicted to observed volumes. The scaling technique seems to have done a good job in predicting pedestrian volumes even with a small sample size (2 counts per intersection). This was the first successful attempt to validate the methodologies used in Molino et al (2008). Additionally, when the predicted pedestrian volumes obtained from the scaling technique were used in conjunction with the final pedestrian crash prediction model, the results were stronger than when the predicted volumes were obtained from the GLM method.

The pedestrian volume estimates from Molino et al (2008) are only one half of the total exposure denominator used in the model developed in this paper, however they fill a void that was desperately needed for crash model development. The pedestrian crash model developed in this paper requires accurate exposure data for both vehicles and pedestrians. While vehicle volume data are fairly easy to obtain, extensive pedestrian volume data are not. The technique originally proposed by Zegeer et al (2005), modified and improved in Molino et al (2008), and validated in this paper offers traffic engineers a fairly reliable way to estimate pedestrian volumes at signalized intersections with relatively minimal empirical data. Had DDOT only collected data for two hours at each of 200 locations (equal distribution of hour of day represented) the results would have similar to the 100% observed. The cost savings for data collection would have been substantial without any loss in model accuracy.

By having a reliable pedestrian exposure measure to feed into the crash prediction model, traffic engineers can utilize the model to assess the safety of signalized intersections throughout an urban environment with more confidence. Although the exposure coefficients developed from the research described in this paper indicate that vehicle volumes are more influential in predicting pedestrian crashes the pedestrian volume variable was still highly significant. The pedestrian exposure coefficient seems to indicate that since there is not a linear relationship (i.e. = 1, or near 1) the focus for safety may need to lie on the vehicles side. If it is indeed true that vehicular volumes have such a stronger influence in pedestrian crash prediction then it may be useful to focus more on safety countermeasures related to the vehicle than the pedestrian. Safety countermeasures can be evaluated more accurately with the addition of pedestrian volumes to the exposure denominator because there is now a known coefficient.

The pedestrian crash prediction model using the scaling technique for pedestrian volume estimation offered the best results of all the combinations considered in this research. While more additional validation and calibration needs to be investigated, the initial analyses and results described seem promising for advancing the prediction and evaluation of pedestrian exposure to crash risk.

4.8 Recommendations for Further Research

Although there were 200 signalized intersections in the database used to develop the pedestrian crash prediction models, only 4 years worth of crash and exposure (pedestrian and vehicular volumes) data were available. Crash and exposure data over a longer time period of time would provide a more robust database to develop better models. It is also recommended that the database include more signalized intersections because previous research used larger data sets (Lyon and Persaud, 2002; Zegeer et al, 2005).

The research described in this paper focused on signalized intersections. It is recommended that data from non-signalized (stop-controlled, uncontrolled) intersections be used to test the breadth of the predictive capabilities of the signalized intersection pedestrian crash prediction model. Additionally it is recommended that the model be calibrated and validated against a completely different set of signalized intersections from Washington, DC as well as different urban areas to test the predictive capabilities even further.

Additional work needs to be conducted to refine the pedestrian volume estimation techniques described in Molino et al (2008) so that the exposure estimates can be more reliable when compared to crash estimates. Additionally, a sensitivity analysis should be conducted to determine the minimum number of signalized intersections that empirical data are needed to develop a reliable hourly distribution of adjustment factors. The GLM method still could also be a useful tool once it becomes more robust and is refined to the point it can predict more accurately than the more simplified scaling technique.

CHAPTER FIVE: SUMMARY OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the results, presents conclusions, and makes recommendations for future research.

5.1 SUMMARY OF FINDINGS AND CONCLUSIONS

This thesis first described a data collection and pedestrian volume estimation methodology for urban signalized intersections. Estimates for one calendar year for an entire city were made using two techniques and were compared. The estimates of 1,373,000,000 pedestrians (obtained with the scaling technique) and 3,569,800,000 pedestrians (obtained with the general linear modeling technique) were not compared for accuracy since actual counts were unavailable. Although aggregated results were not compared, single site estimates were compared to previous research. Previous research indicated similar results for daily counts at a single site and therefore this could provide partial validation of the estimation methodologies.

Pedestrian-motor vehicle crash prediction models were developed using a variety of explanatory variables and regression models. The results of the crash prediction analyses indicate that the final pedestrian-vehicle crash prediction model chosen is similar in structure and performance to models developed in earlier research efforts (Lyon and Persaud, 2002; Zegeer, et al., 2005, Harwood, et al, 2008). Pedestrian volume and vehicular volume, which have been previously identified as strong predictor variables, were highly significant in the model developed. All of the variables in the final model presented in this thesis were significant at the 0.05 confidence level. The final models from previous work by Lyon and Persaud (2002) included variables that met less strict confidence level of 0.10. The models from Zegeer et al (2005) included variables that ranged between < 0.0001 and 0.984. Additionally, the sample size for the dataset used in this thesis was smaller than the ones used in Lyon and Persaud (2002) and Zegeer et al (2005). The database of signalized intersections described in this thesis had 200 locations, whereas the two studies cited had over 350 and 2,000 sites, respectively. Additionally, the Lyon and Persaud (2002) differentiated between 3-leg and 4-leg intersections. This thesis included intersections with 3 to 6 legs and used this characteristic as a possible explanatory variable. The results obtained in this thesis are as strong, if not stronger, than those cited studies with much larger sample sizes.

Following the crash modeling analyses, an analysis was done to assess the accuracy of the pedestrian volume estimates using the two techniques described in Chapter 3. The analyses showed that when using the data used in the crash prediction model, the scaling technique provided stronger results when compared to the GLM method. The scaling technique produced predicted pedestrian volumes that were highly correlated to the observed pedestrian volumes. The GLM method appeared to have very little correlation. As noted in Chapter 3 the linear model developed to predict pedestrian volumes was based off of a fairly small sample size, which would explain the weak relationship of predicted to observed volumes. The scaling technique seems to have done a good job in predicting pedestrian volumes even with a small sample size (2 counts per intersection). This was the first successful attempt to validate the results from the pedestrian volume estimation methodologies Additionally, when the predicted pedestrian volumes obtained from the scaling technique were used in conjunction with the final pedestrian crash prediction model, the results were stronger than when the predicted volumes were obtained from the GLM method.

Pedestrian volumes are only one half of the total exposure denominator used in the model developed in this thesis; however they fill a void that was desperately needed for crash model development. The pedestrian crash model developed in this thesis requires accurate exposure data for both vehicles and pedestrians. While vehicle volume data are fairly easy to obtain, extensive pedestrian volume data are not. The technique originally proposed by Zegeer et al (2005), modified and improved in Chapter 3, and validated in this thesis offers traffic engineers a fairly reliable way to estimate pedestrian volumes at signalized intersections with relatively minimal empirical data. Had DDOT only collected data for two hours at each of 200 locations (equal distribution of hour of day represented) the results would have similar to the 100% observed. The cost savings for data collection would have been substantial without any loss in model accuracy.

By having a reliable pedestrian exposure measure to feed into the crash prediction model, traffic engineers can utilize the model to assess the safety of signalized intersections throughout an urban environment with more confidence. Although the exposure coefficients developed from the research described in this thesis indicate that vehicle volumes are more influential in predicting pedestrian crashes the pedestrian volume variable was still highly significant. The pedestrian exposure coefficient seems to indicate that since there is not a linear relationship (i.e. = 1, or near 1) the focus for safety may need to lie on the vehicles side. If it is indeed true that vehicular volumes have such a stronger influence in pedestrian crash prediction then it may be useful to focus more on safety countermeasures related to the vehicle than the pedestrian. Safety countermeasures can be evaluated more accurately with the addition of pedestrian volumes to the exposure denominator because there is now a known coefficient.

The pedestrian crash prediction model using the scaling technique for pedestrian volume estimation offered the best results of all the combinations considered in this research. While more additional validation and calibration needs to be investigated, the initial analyses and results described seem promising for advancing the prediction and evaluation of pedestrian exposure to risk.

5.2 Recommendations for Future Research

Additional work needs to be conducted to refine the pedestrian volume estimation techniques described in Chapter 3 so that the exposure estimates can be more reliable when compared to crash estimates. A sensitivity analysis should be conducted to determine the minimum number of signalized intersections that empirical data are needed to develop a reliable hourly distribution of adjustment factors. More data are needed to develop the GLM technique to bring it closer to the scaling techniques predictive capabilities.

Crash and exposure data over a longer time period of time would provide a more robust database to develop better models. Although there were 200 signalized intersections in the database used to develop the pedestrian crash prediction models, only 4 years worth of crash and exposure (pedestrian and vehicular volumes) data were available. It is also recommended that the database include more signalized intersections because previous research used larger data sets (Lyon and Persaud, 2002; Zegeer et al, 2005).

The pedestrian volume estimation techniques were used at other types of intersections (Molino et al, 2008). However the crash prediction modeling research described in this thesis focused on signalized intersections. It is recommended that data from non-signalized (stop-controlled, uncontrolled) intersections be used to test the breadth of the predictive capabilities of the signalized intersection pedestrian crash prediction model. Additionally it is recommended that the model be calibrated and validated against a completely different set of signalized intersections from Washington, DC as well as different urban areas to test the predictive capabilities even further.

References

Beck, L., A. Dellinger, M. O'Neil. (2007). "Motor Vehicle Crash Injury Rates by Mode of Travel, United States: Using Exposure-Based Methods to Quantify Differences." American Journal of Epidemiology, Vol. 166, No. 2., pp. 212-218.

Caliendo, C., Guida, M., Parisi, A. (2007). "A Crash-Prediction Model for Multilane Roads." Accident Analysis and Prevention. Vol. 39, No. 4. pp. 657-670.

Cameron, A. and Trivedi, P. (1998). Regression Analysis of Count Data, Cambridge University Press, Cambridge, MA.

Carroll, P. (1973). Symposium on Driving Exposure, University of Michigan.

Carter, D., Hunter, W., Zegeer, C., Stewart, J.R. (2006) Pedestrian and Bicyclist Intersection Safety Indices. Final Report. FHWA-HRT-06-130. USDOT/FHWA. Washington, DC.

Center for Applied Research, Inc. (2003). Identification of Alternative Methodologies for Collecting Pedestrian Exposure and Crash Data. Contract DTNH22-99-D-15099, Task Order 4. DRAFT.

Davis, S., H. Robertson, L. King. (1989). "Pedestrian / Vehicle Conflicts: An Accident Prediction Model." Transportation Research Record. Washington, DC. No. 1210, pp 1 – 11.

Dionne, G., Desjardins, D., Laberge-Nadeau, C., Maag, U. (1993). Medical conditions, risk exposure and truck drivers' accidents: an analysis with count data regression models. Presented at the 37th Annual Meeting of the Association for the Advancement of Automotive Medicine, San Antonio, Texas.

Federal Highway Administration (FHWA). (2005). Highway Performance Monitoring System Field Manual. U.S. Department of Transportation.

Federal Highway Administration (FHWA). (2005). Highway Statistics. U.S. Department of Transportation.

Federal Highway Administration (FHWA). (2007) Safety Compass, Vol. 1, Issue 2.

Green-Roesel, M. Diogenes, and D. Ragland. (2007). "Estimating Pedestrian Accident Exposure: Protocol Report." UC Berkeley Traffic Safety Center, Berkeley, CA.

Harwood, D., C. Zegeer, C. Lyon, et al. (2008). NCHRP 129: Pedestrian Safety Prediction Methodology. Transportation Research Board, Washington, D.C.

Jovanis, P. P.; Chang, H. L. (1986). “Modelling the relationship of accidents to miles travelled.” Transportation Research Record. Washington, DC. No.1068 pp. 42-51.

Keall, M., W. Frith. (1999). “Measures of Exposure to Risk of Road Crashes in New Zealand,” IPENZ Transactions. Vol. 26, No 1, pp 7-12.

Keall, M. (1995). “Pedestrian Exposure to Risk of Road Accident in New Zealand,” Accident Analysis and Prevention. Vol 27, No. 5, pp 729-740.

Lea, N.D. and Associates. (1978). *A Study of Measures to Improve Pedestrian and Bicycle Safety,* Report No. TP 1004, Canadian Department of Transport.

Leden, L. (2002). “Pedestrian risk decreases with pedestrian flow. A case study based on data from signalized intersections in Hamilton, Ontario.” Accident Analysis and Prevention. Vol. 34, pp. 457-464.

Lord, D. (2006). “Modeling Motor Vehicle Crashing Using Poisson-Gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter,” Accident Analysis and Prevention, Vol. 38, No. 4, pp. 751-766.

Lord, D. (1996). “Analysis of Pedestrian Conflicts with Left-Turning Traffic.” Transportation Research Record. Washington DC. No. 1538. pp 61-67.

Lord, D., Washington, S., Ivan, J. (2005). Poisson, Poisson-gamma, and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. Accident Analysis and Prevention, Vol. 37, pps. 35-46.

Lyon, C., B. Persaud. (2002). “Pedestrian Collision Prediction Models for Urban Intersections.” Transportation Research Record. No. 1818. pp 102-107.

Maher, M., Summersgill, I. (1996). “A Comprehensive Methodology for the Fitting of Predictive Accident Models.” Accident Analysis and Prevention, Vol 28, No 3, pp 281-296.

Maycock, G., Hall, R. D. (1984). Accidents at 4-arm Roundabouts. Laboratory Report LR1120. Crowthorne, Berks, U.K.: Transport Research Laboratory.

Miaou, S-P., Hu, P. S., Wright, T., Rathi, A. K., Davis, S. (1992). “Relationships between truck accidents and highway geometric design: a Poisson regression approach.” Transportation Research Record No. 1376 pp. 10-18.

Miaou, S-P, Lum, H. (1993). “Modeling vehicle accidents and high-way geometric design relationships.” Accidident Analysis and Prevention. pp.689-709.

Molino, J., Kennedy, J., Johnson, P., Beuse, P., Emo, A., Do, A. (2008). Pedestrian and Bicyclist Exposure to Risk: A Methodology for Estimation in an Urban Environment. 88th Annual Meeting of the Transportation Research Board CD-ROM.

National Highway Traffic Safety Administration (NHTSA) (2006a). Fatal Accident Reporting System (FARS). U.S. Department of Transportation.

National Highway Traffic Safety Administration (NHTSA) (2006b). *Traffic Safety Facts*. U.S. Department of Transportation.

Qin, X., and J. Ivan. (2001). Estimating Pedestrian Exposure Prediction Model in Rural Areas. Transportation Research Record: Journal of the Transportation Research Board, No. 1773, Transportation Research Board of the National Academies, Washington, D.C., pp. 89-96.

Qin, X., J. Ivan, N. Ravishanker. (2004). "Selecting Exposure Measures in Crash rate Prediction for Two-Lane Highway Segments." Accident Analysis & Prevention. No. 36. pp. 183-191.

Rakha, H., M. Arafeh, A. Abdel-salam, A. Flintsch. (2008). "Linear Regression Crash Prediction Models: Issues and Proposed Solutions," Transportation Research Board 87th Annual Meeting. CD-ROM. Transportation Research Board, Washington, D.C.

SAS Institute Inc. (2008). *SAS/STAT® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

Schneider, R., R. Patton, J. Toole and C. Raborn. (2005). Pedestrian and Bicycle Data Collection in United States Communities: Quantifying Use, Surveying Users, and Documenting Facility Extent. FHWA, Office of Natural and Human Environment, U.S. Department of Transportation, Washington, D.C.

Shankar, V., Ulfarsson, G., Pendyala, R., Nebergall, M. (2003). "Modeling Crashes Involving Pedestrians and Motorized Traffic." Safety Science. Vol. 41, pp 627-640.

Vogt, A., Bared, J. (1998). Accident Models for Two Lane Rural Roads: Segments and Intersections. FHWA-RD-98-133. Washington, DC.

Zegeer, C., R. Stewart, H. Huang, P. Lagerwey, J. Feganes, and B. Campbell. (2005). *Safety Effects of Marked versus Unmarked Crosswalks at Uncontrolled Locations.* FHWA-HRT-04-100. FHWA, U.S. Department of Transportation.

World Wide Web:

District of Columbia Geographic Information Systems, Accessed January 2008. http://dcatlas.dcgis.dc.gov/catalog.

National Highway Traffic Safety Administration (NHTSA), Accessed February 2008. http://www.nhtsa.dot.gov

APPENDIX: FINAL PEDESTRIAN CRASH PREDICTION MODEL OUTPUT

Generalized Linear Regression Model Analysis

Source: SAS® Software

```
                                The SAS System                                 539
                                                  17:21 Sunday, November 2, 2008

                            The CONTENTS Procedure

Data Set Name        THESIS.SIGNALIZED                  Observations          200
Member Type          DATA                               Variables             21
Engine               V9                                 Indexes               0
Created              Sunday, November 02,               Observation Length    168
                     2008 05:23:04 PM
Last Modified        Sunday, November 02,               Deleted Observations  0
                     2008 05:23:04 PM
Protection                                              Compressed            NO
Data Set Type                                           Sorted                NO
Label
Data Representation  WINDOWS_32
Encoding             wlatin1  Western (Windows)


                        Engine/Host Dependent Information

Data Set Page Size          16384
Number of Data Set Pages    3
First Data Page             1
Max Obs per Page            97
Obs in First Data Page      77
Number of Data Set Repairs  0
Filename                    E:\School\THESIS\Working
                            Copy\Signalized_All\Analysis\signalized.sas7bdat
Release Created             9.0201M0
Host Created                W32_VSHOME


                  Alphabetic List of Variables and Attributes

             #    Variable            Type    Len    Label

             2    Crashes             Num       8    Crashes
             3    Crashes4            Num       8    Crashes4
            20    Pred4YearPed        Num       8    Pred4YearPed
            21    Pred4YearPed2       Num       8    Pred4YearPed2
            19    Tot4Min_MajRatio    Num       8    Tot4Min_MajRatio
            12    Tot4YearLeftVeh     Num       8    Tot4YearLeftVeh
            18    Tot4YearMaj         Num       8    Tot4YearMaj
            17    Tot4YearMin         Num       8    Tot4YearMin
            16    Tot4YearPed         Num       8    Tot4YearPed
            13    Tot4YearRightVeh    Num       8    Tot4YearRightVeh
            11    Tot4YearVeh         Num       8    Tot4YearVeh
             8    geo_binary          Num       8    geo_binary
             4    geo_code            Num       8    geo_code
             5    legs                Num       8    legs
             9    legs_binary         Num       8    legs_binary
             1    loc_id              Num       8    loc_id
            14    percVehLeft         Num       8    percVehLeft
            15    percVehRight        Num       8    percVehRight
             7    totrdwidth          Num       8    totrdwidth
            10    traf_binary         Num       8    traf_binary
```

```
           6    traf_code           Num       8    traf_code

                                The SAS System                              541
                                                 17:21 Sunday, November 2, 2008

                            The CONTENTS Procedure

Data Set Name        WORK.LOGS                          Observations          200
Member Type          DATA                               Variables             33
Engine               V9                                 Indexes               0
Created              Sunday, November 02,               Observation Length    264
                     2008 06:25:38 PM
Last Modified        Sunday, November 02,               Deleted Observations  0
                     2008 06:25:38 PM
Protection                                              Compressed            NO
Data Set Type                                           Sorted                NO
Label
Data Representation  WINDOWS_32
Encoding             wlatin1  Western (Windows)


                      Engine/Host Dependent Information

Data Set Page Size          16384
Number of Data Set Pages    4
First Data Page             1
Max Obs per Page            61
Obs in First Data Page      44
Number of Data Set Repairs  0
Filename                    C:\Users\Jason\AppData\Local\Temp\SAS
                            Temporary Files\_TD1868\logs.sas7bdat
Release Created             9.0201M0
Host Created                W32_VSHOME


                  Alphabetic List of Variables and Attributes

          #    Variable              Type    Len    Label

          2    Crashes               Num       8    Crashes
          3    Crashes4              Num       8    Crashes4
         20    Pred4YearPed          Num       8    Pred4YearPed
         21    Pred4YearPed2         Num       8    Pred4YearPed2
         19    Tot4Min_MajRatio      Num       8    Tot4Min_MajRatio
         12    Tot4YearLeftVeh       Num       8    Tot4YearLeftVeh
         18    Tot4YearMaj           Num       8    Tot4YearMaj
         17    Tot4YearMin           Num       8    Tot4YearMin
         16    Tot4YearPed           Num       8    Tot4YearPed
         13    Tot4YearRightVeh      Num       8    Tot4YearRightVeh
         11    Tot4YearVeh           Num       8    Tot4YearVeh
          8    geo_binary            Num       8    geo_binary
          4    geo_code              Num       8    geo_code
          5    legs                  Num       8    legs
          9    legs_binary           Num       8    legs_binary
         30    lnPred4YearPed        Num       8
         33    lnPred4YearPed2       Num       8
         29    lnTot4Min_MajRatio    Num       8
         32    lnTot4YearMaj         Num       8
         31    lnTot4YearMin         Num       8
         27    lnpercVehLeft         Num       8
         28    lnpercVehRight        Num       8
         26    lnrdwid               Num       8
         22    lntot4yearleftveh     Num       8
         23    lntot4yearped         Num       8
         24    lntot4yearrightveh    Num       8
         25    lntot4yearveh         Num       8
          1    loc_id                Num       8    loc_id
         14    percVehLeft           Num       8    percVehLeft
         15    percVehRight          Num       8    percVehRight
          7    totrdwidth            Num       8    totrdwidth
         10    traf_binary           Num       8    traf_binary
```

```
       6    traf_code             Num       8    traf_code

                          The SAS System                              543
                                           17:21 Sunday, November 2, 2008

                       The GENMOD Procedure

                        Model Information

     Data Set                      WORK.LOGS
     Distribution          Negative Binomial
     Link Function                       Log
     Dependent Variable             Crashes4    Crashes4


                 Number of Observations Read         202
                 Number of Observations Used         200
                 Missing Values                        2


                      Class Level Information

                   Class           Levels    Values

                   geo_binary           2    1 2
                   legs_binary          2    1 2
                   traf_binary          2    1 2
                   traf_code            4    1 2 3 4


                 Criteria For Assessing Goodness Of Fit

       Criterion                    DF           Value        Value/DF

       Deviance                    197        205.1526          1.0414
       Scaled Deviance             197        205.1526          1.0414
       Pearson Chi-Square          197        218.2180          1.1077
       Scaled Pearson X2           197        218.2180          1.1077
       Log Likelihood                         -74.7598
       Full Log Likelihood                   -314.3795
       AIC (smaller is better)                636.7590
       AICC (smaller is better)               636.9641
       BIC (smaller is better)                649.9523

                          The SAS System                              544
                                           17:21 Sunday, November 2, 2008

                       The GENMOD Procedure

   Algorithm converged.


             Analysis Of Maximum Likelihood Parameter Estimates

                                  Standard   Wald 95% Confidence        Wald
   Parameter       DF   Estimate     Error         Limits           Chi-Square

   Intercept        1   -20.0056    2.9932   -25.8722   -14.1389        44.67
   lntot4yearped    1     0.3274    0.0605     0.2089     0.4459        29.33
   lntot4yearveh    1     0.9025    0.1667     0.5757     1.2292        29.30
   Dispersion       1     0.5895    0.1403     0.3145     0.8645

                       Analysis Of Maximum
                       Likelihood Parameter
                            Estimates

                    Parameter       Pr > ChiSq

                    Intercept           <.0001
                    lntot4yearped       <.0001
                    lntot4yearveh       <.0001
```

Dispersion

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
lntot4yearped	1	29.13	<.0001
lntot4yearveh	1	29.68	<.0001