ABSTRACT

DEVELOPMENT OF A MULTIVARIATE LOGISTIC MODEL TO PREDICT BICYCLE ROUTE SAFETY IN URBAN AREAS

by
Cheryl Allen-Munley

In response to the renewed appreciation of the benefits of bicycling to the environment and public health, public officials across the nation are working to establish new bicycle routes. During the past two decades, a number of methods have been endorsed for the selection of "suitable" bicycle routes. These methods are limited in that they do not explicitly address bicycle safety nor do they reflect urban conditions.

The purpose of this research is to develop an objective bicycle route safety rating model based on injury severity. The model development was conducted using a logistic transformation of Jersey City's bicycle crash data for the period 1997-2000. The resulting model meets a 90% confidence level by using various operational and physical factors (traffic volume, lane width, population density, highway classification, the presence of vertical grades, one-way streets and truck routes) to predict the severity of an injury that would result from a crash that occurred at a specific location. The rating of the bicycle route's safety is defined as the expected value of the predicted injury severity. This rating is founded on the premise that safe routes produce less severe accidents than unsafe routes.

The contribution of this research goes beyond the model's predictive capacity in comparing the safety of alternative routes. The model provides planners with an understanding, derived from objective data, of the factors that add to the route's safety, the factors that reduce safety and the factors that are irrelevant. The model often confirms widely held beliefs as evidenced by the finding that highways with steep grades, truck routes and

poor pavement quality create an unfavorable environment for bicyclists. Conversely, the model has found that increased volume and reduced lane width, at least in urban areas, actually reduce the likelihood of severe injury. Planners are encouraged to follow the lead of experienced bicyclists in choosing routes that travel through the urban centers as opposed to diverting bicyclists to circuitous routes on wide, low volume roads at the periphery of cities.

DEVELOPMENT OF A MULTIVARIATE LOGISTIC MODEL TO PREDICT BICYCLE ROUTE SAFETY IN URBAN AREAS

by
Cheryl Allen-Munley

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Transportation

Interdisciplinary Program in Transportation

May 2003

APPROVAL PAGE

DEVELOPMENT OF A MULTIVARIATE LOGISTIC MODEL TO PREDICT BICYCLE ROUTE SAFETY IN URBAN AREAS

Cheryl Allen-Munley

Dr. Janice R Daniel, Dissertation Advisor — Date
Assistant Professor of Civil and Environmental Engineering, NJIT

Dr. Rongfang (Rachel) Liu, Committee Member — Date
Assistant Professor of Civil and Environmental Engineering, NJIT

Dr. Athanassios Bladikas, Committee Member — Date
Associate Professor, Industrial and Manufacturing Engineering, NJIT

Dr. Kyriacos C. Mouskos, Committee Member — Date
Research Professor, CCNY-CUNY

Gerald J. Nissen, PE, PP, Committee Member — Date
Jersey City Municipal Engineer

BIOGRAPHICAL SKETCH

Author: Cheryl Allen-Munley

Degree: Doctor of Philosophy

Date: May 2003

Undergraduate and Graduate Education

- Doctor of Philosophy in Transportation,
  New Jersey Institute of Technology, Newark, NJ, 2003

- Master of Business Administration,
  Columbia University School of Business, New York, NY, 1983

- Bachelor of Science in Civil Engineering,
  Massachusetts Institute of Technology, Cambridge, MA, 1978

Major: Transportation

Presentation and Awards:

Allen-Munley, Cheryl
"Using the North Jersey Regional Transportation Model for Transportation Decision Making, Transaction, Atlantic City, NJ, 1997.

Allen-Munley, Cheryl
The National Council on Transportation and Industrial Productivity, Student of the Year Award, Transportation Research Board, Washington, DC, 2003.

Professional Licenses and Committees

- Professional Engineer, New Jersey, 1987, Pennsylvania, 2002.

- Professional Planner, New Jersey, 1994.

- Member, Transportation Research Board Subcommittee on Bicycle Facility Design

To my family,
I have dedicated my life to deserving their high regard.

If you wish to know whether society is stagnant, learning scholastic, religion a dead formality, you may learn something by going into universities and libraries; something also by the work that is doing on cathedrals and churches, or in them; but not quite as much by looking at the roads. For if there is any motion in society, the Road, which is the symbol of motion, will indicate the fact. When there is activity or enlargement, or a liberalizing spirit of any kind, then there is intercourse and travel, and these require roads. So if new ideas are abroad and new hopes rising, then you will see it by the roads that are building. Nothing makes an inroad without making a road. All creative action, whether in government, industry, thought or religion, creates roads.

Horace Bushnell

ACKNOWLEDGMENT

Words cannot express my gratitude to Dr. Janice Daniel, who served as my dissertation advisor, a treasured friend and a role model. Her insight, her intelligence and her ability to never lose sight of our objective has carried me through to this happy end. In addition to my appreciation for his participation on my committee, I want to thank Kyriacos Mouskos for his early encouragement and outstanding instruction. Special thanks are given to Dr. Rachel Liu, Dr. Athanassios Bladikas and Gerald Nissen. After working together for over a decade, I sincerely doubt I will ever have a finer co-worker and mentor than Gerald Nissen.

I would like to thank Dr. Sunil Dhar, NJIT Assistant Professor of Mathematics for his assistance in logistic modeling theory, Guilin Li for his help with SAS software and Keir Opie for his help with TRANSCAD. Thanks to Arun Raj for help collecting and compiling data.

Without the accident data there would be no model. A thousand thanks to Lieutenant Theodore Wilczwski, Orlando Cortez, Director of the Jersey City Police records room and Joao D'Souza, Director of Jersey City Traffic Engineering Division for being so generous with their time and resources.

Thank you to Dr. Ronald Kane, Dean of Graduate Studies Program. This would not have been possible without your continued support and the NJIT Presidential Fellowship.

In closing, I reserve the highest praise for my husband, Dean Munley. He may not have been the fastest typist nor the most cheerful mapper. Still, when the going got rough, he was pulling beside me at every turn in the road. As he shared my burdens, I share the honor of this degree with him.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

CHAPTER 1

INTRODUCTION

Public policy encourages increasing bicycle and pedestrian modes of travel. In contrast to motor vehicles, bicycles do not have adverse impacts on air quality and road congestion. Their ability to maneuver in small places allows bicyclists to avoid the delays traffic jams impose on other motorists. In recognition of these benefits, the National Bicycling and Walking Study (1999), sponsored by the U.S. Department of Transportation (USDOT), designed an action plan to: 1. Double the number of trips made by walking and bicycling; and 2. Reduce the number of pedestrians and bicyclists killed or injured by 10%.

While pedestrian and bicycle injury and fatality rates continue to fall each year, the absolute number of pedestrian and bicycle accidents are rising with the increase in total trips. In 2001, bicyclists suffered 690 fatalities and 51,000 injuries resulting from traffic crashes. Bicycle fatalities represented 2% of all traffic fatalities, which is a large percentage when considering that bicyclists only represent 0.7% of all trips. The fact that bicyclists and pedestrians as a group are over represented in traffic fatalities is of great concern. In 2000, bicyclists and pedestrians represented only 0.9% of all trips, but suffered 2.0% of all traffic fatalities. If the goal of increased non-motorized travel is to be met, safe facilities must be provided to the walking and bicycling public.

The Institute of Transportation Engineering (ITE) Pedestrian and Bicycle Task Force (2000) identified "Retrofitting Facilities" as a priority issue. Achieving this goal will require care in the selection of routes for sidewalks and bike lanes. Poor choices can increase exposure to accidents causing increased fatalities and severe injuries. A credible model for use in providing a safety index for bicycle routes would offer state and local officials with the means to assess the safety of these routes. This rating could be used to compare the relative safety of alternative bicycle routes, as well as, flag areas which must be improved to make the route suitable for use by bicyclists. The goal of this research effort is to develop such a model.

In the past decade, a wealth of research has been conducted on bicycle safety. Researchers have classified accidents according to events preceding the crash (Hunter, 1990). Risk factors of various bicycling facilities such as wide curb lane (Hunter, 1990), sidewalks (Aultman-Hall, 1998), urban main roads (Sharples, 1999), intersections (Wachtel, 1994), shared-use facilities (Wachtel, 1997) and highway shoulders (Khan, 1995) have been investigated. Adult bicyclists have been surveyed to understand their characteristics and to determine their preferences (Aultman-Hall, 1998; Antonakos, 1994; Moritz, 1996). Clarke (2000) studied bicycle friendly factors from key areas around the country. The injuries resulting from bicycle crashes were studied Rodgers (1995). To the benefit of the bicycling commuter, this vast body of knowledge must be distilled into a practical tool that can be used by local officials for the planning and design of bicycle facilities.

Many attempts have been made to develop models to rate bicycle routes (Landis, 1994; Epperson, 1994, 1997; Sorton, 1994 and Harkey, 1998). These models typically offer indices that rate roadways based on a group of factors that affect their "suitability" for use by bicyclists, but do no directly address safety.

For commuter bicycle routes, factors that do not pertain to safety should be considered of limited importance. The location of rest facilities and bicycle repair shops may be of interest to recreational cyclists, but the commuter bicyclist is known to select his route to minimize trip time (Aultman-Hall, 1997). Safety is the only valid reason for diverting the commuting cyclist from his preferred minimum path.

One shortcoming of existing bicycle route selection models is that previous bicycle route selection models limited their application to highway sections, basing their assessments on factors that have limited relevance in urban settings. Yet, the majority of bicycle/motor vehicle crashes occur in urban areas. Cross and Fisher (1977) in their landmark study on bicycle/motor vehicle crashes found that while 86% of the road system is in rural areas, rural areas are the location of only 11% of bicycle/motor vehicle collisions. The bicycle/motor vehicle collision rate per road mile is 42 times higher in urban than in rural areas.

Existing models do provide a good working framework, but the absence of objective validation limits their widespread application. Some models were developed without validation. Others were validated by comparing model predictions to independent assessments by bicyclists despite the fact that bicyclists have been found to misjudge the safety of routes (Garder, 1994). Objective data is needed for both model development and validation.

1.1 Problem Statement and Research Objective

The desire of public officials to encourage bicycling as a viable transportation mode choice. is tempered by the recognition that bicyclists experience both higher rates of accidents and increase injury severity levels. A problem exists in that there are no route selection methods available for officials to use in reducing the accident risk for bicyclists. Methods are needed to identify both the factors that affect route safety and their relative weight. With such a tool, officials could not only select safe bicycle routes, they could determine which capital improvements would offer the greatest improvement and return on investment in terms of improving bicycle safety.

The goal of this research effort is to develop a multivariate logistic model for use in rating the safety of bicycle routes based on their physical characteristics. The model will identify variables to be considered and provide an understanding of the differences in injury severity/ accident outcomes between urban and rural condition, commuter and recreational bicyclists and child and adult bicyclists. The model will be fit using objective injury severity data and will be practical and easy to apply, requiring only data that is readily available. Overall, the model will provide a rational approach and will meet accepted statistical performance measures.

1.2 Technical Approach

The development of a multivariate logistic safety rating model involves the application of mathematical techniques to a series of data points to seek a relationship between a Y, the response variable, and the Xs, the independent variables, to establish a formula where Y is a function of X. The logical choice of bicycle crash rates for the response variable is complicated in the U.S. by the difficulty in obtaining true exposure based rates, i.e. crashes per vehicle miles traveled. Exposure rates are preferable because they are route specific and as continuous variables, they may be used as response variables in standard Ordinary Least Squares (OLS) model building techniques.

Total crash events is an inferior choice for a model's response variable because it is not independent of bicyclist route choice (Epperson, 1994). Not surprisingly, due to the very presence of the bicyclists, popular bicycle routes will report higher crash numbers than other less traveled and potentially dangerous routes. Crash injury severity is a valid choice for the model's response variable in that it is reported for each crash and it is independent of route volume.

Given injury severity as the response variable, the choice of explanatory variables must be addressed. To identify these variables requires an understanding of both the causes of bicycle crashes and the factors that influence injury severity. The research has been conducted to explore the impacts of such operational factors as speed and volume, and physical factors such as lane width and grade. In addition, environmental factors such as weather, lighting and roadway conditions often act as confounders in predicting crashes for bicycles as well as motor vehicles (Shankar, 1995).

Injury severity is a categorical response variable, which precludes the use of OLS methods for model development. Ordinary Least Squares solutions cannot be constrained to combine the model parameters in such a fashion to generate only responses that are integers within the specified range. OLS solutions require that the dependent variable be continuous and able to assume any value between $-\infty$ and ∞.

Nonlinear transformations relax these restrictions. While there are a number of possible transformations, the logistic form has been historically popular because it is well developed and mathematically easy to derive. It is formalized as follows:

$$Y = \frac{1}{\left(1 + e^{-Z}\right)} \tag{1.1}$$

$$\text{Where} \quad Z = \alpha + \sum \beta_k X_k \tag{1.2}$$

In the logistic formulation, Y, the probability that the outcome of a crash will produce an injury with a given level of severity is derived from the logit Z. Z, as described in Equation 1.2, is a linear function of X, the set of physical and operational explanatory variables as well as temporal and personal confounders. Using historical crash data and standard statistical software packages, a logistic model will be developed to select parameter estimates to maximize the model's predictive power.

1.3 Organization

This dissertation consists of seven chapters. This research will be presented in the following manner. Chapter 1 consists of the introduction, problem statement, technical approach and dissertation organization.

After this initial introduction, Chapter 2 submits a full literature review to answer the questions: 1. Why is a bicycle route safety rating model needed? 2. Why are existing models inadequate? and 3. What is the nature of bicycle crashes with special attention to the factors that cause crashes and factors which exacerbate injuries?

Chapter 3 presents an understanding of categorical models, explaining why the ordered logistic form is most appropriate for an injury severity model. Maximum likelihood equations for standard dichotomous and polytomous forms will be presented.

Chapter 4 focuses on the data including an explanation of how the data were obtained. In instances where data were normalized or aggregated for use in the model, an explanation of the methods applied is presented. Statistical distributions and cross tabulations with injury severities are presented for nineteen explanatory variables and confounders in addition to Injury Severity, the response variable.

Chapter 5 presents the bicycle route safety prediction model. It will explain the modeling techniques used to build the model. Statistical measurements of its goodness of fit and predictive properties will be discussed in detail as well as an understanding of the model's factors including the relative magnitude of these factors and their direction.

Chapter 6 applies the model to a route selection problem: the Jersey City Bicycle Plan as prepared by the Rutgers Transportation Policy Institute (2000). A series of route alternatives will be identified. Using data collected for the individual route alternatives, the model will generate safety ratings. Based on these ratings, a comparison will be made as to the merits of the candidate routes.

Chapter 7 concludes with a review of the model and a generalized assessment of the multivariate logistic technique using injury severity to predict route safety. Future research needs will be considered. Most importantly, recommendations will be made as to the manner in which this safety prediction factor should be included in currently accepted bicycle suitability models.

CHAPTER 2

LITERATURE REVIEW

2.1 Purpose

The purpose of the literature search is threefold. Before embarking on a lengthy research project it is necessary to affirm that: 1. A problem exists; 2. There is a need to solve it; and 3. That it has not yet been solved. The problem under consideration is the prediction of bicycle route safety. Section 2.2 answers the question of whether bicycling is a dangerous mode of transportation which requires improved safety measures. Section 2.3 addresses the benefits of bicycling including the reasons why society should encourage this activity. Section 2.4 reviews the previous attempts to solve this problem, i.e. the development of reliable models which identify safe bicycle routes. This review of existing models will also explain why further model development work is needed.

Once the first phase of the literature search has provided sufficient justification for the research, the second phase will serve to amass all of the tools available to solve the bicycle route selection problem. Section 2.5 examines the nature of the accidents in an attempt to understand the causal factors. This section also explores the interrelationships between the accident types and the severity of the injury suffered by the victim. Section 2.6 discusses physical and operational contributing factors. Temporal confounders will be discussed in Section 2.7. Operator confounders will be discussed in Section 2.8.

Additional reviewed literature discusses the available data for both its breadth and limitations. Also, a review has been conducted of the theoretical modeling tools available to the modeler. In the interest of continuity, this material will be presented in Chapter 4, Data Collection and Chapter 3, Theoretical Approach, respectively.

2.2 Bicycle Safety

According to National Highway Traffic Safety Association (NHTSA, 2000), over 47,000 bicyclists have died in traffic crashes since 1932, the first year since bicycle crash records were recorded. At that time, bicycle fatalities represented 1.3% of all accidents for a total of 350 fatalities. Today, the bicycle accident rate has nearly doubled. In 2000, bicycle accidents resulted in 51,000 injuries. The number of fatalities totaled 690 representing 2% of all fatalities, although bicycle trips accounted for only 0.9% of all trips.

There is no one explanation for the fact that bicyclists are over-represented in both injuries and fatalities. The most obvious cause for the increase in serious injury is the vulnerability of the bicyclist. Without the metal shell of a vehicle, the unprotected flesh of the bicyclist is infinitely more prone to harm as a result of a crash. For this reason, helmet use has been found to reduce injury by as much as 70% (Rivara, et al., 1996). Unfortunately, helmet use is far from universal. In their study of bicyclists in Arizona, Cynecki, et al. (1993) viewed helmets on only 15% of the 480 observed bicyclists.

Although the lack of personal protection explains the increased probability of an injury in motor vehicle / bicycle crashes, it does not explain the increased likelihood of a crash event. One reason for the increased likelihood of a motor vehicle / bicycle crash is that bicycles are inherently less stable than motor vehicles. Poor pavement quality, catch basin grates, physical obstructions, pedestrians and other bicyclists can all be sources of falls. In fact, Forester (1983) found that 44% of all bicycle accidents resulted from falls. This is large when compared to the 18% of accidents which involve motor vehicles.

Also, contributing to the prevalence of bicycle crashes is the fact that bicyclists are not as visible to motorists. Bicycle crashes are particularly prevalent at dusk, during the hours between 4:00 and 8:00 pm (NHTSA , 2000). Motorists are known to scan the roadways, searching for potential conflicts with other motorists while ignoring the presence of bicyclists. Sumala, et al. (1996) when videotaping motorists at two sight obstructed intersections in Helsinki found that 100% of drivers turning left looked left to avoid conflicts with motorists. Only 7% of right turning motorists, however, looked right, probably because only bicyclists not motorists could obstruct their path on the right. This finding demonstrates that not only are bicyclists smaller and therefore not as visible as motor vehicles, typical motorist behavior causes them not to anticipate or look for bicyclists.

In addition, a sizeable number of bicycle / motor vehicle crashes arise from operator error, both bicyclist and motorist. In a study of bicycle crashes for a six-year period in Hawaii, 83.5% of these crashes were found to be due to motorist error and 16.5 % were due to bicyclist error (Kim, et al. 1996). Motorists were subject to cause crashes by speeding, failing to yield and following too closely. Bicyclists, especially young bicyclists without drivers licenses, disregarded intersection controls, crossed centerlines, traveled in the wrong direction and made improper turns before an intersection.

Irrespective of the involvement of operator error, physical characteristics of the route such as speed and volume exacerbate these failures of judgement. Considering the magnitude of these injuries and fatalities, decision makers have sought tools to identify the sources of risk and evaluate countermeasures. These efforts have been hampered by a lack of exposure data which would allow analyses based on accident rates. Alternative methods are needed for safety assessment because the level of harm to the cycling public is too great to ignore.

2.3 Bicycle Mode Choice - Public Policy

The environmental and personal benefits of bicycling are great, encompassing improved health, reduction in space needed for parking and travel, and most importantly, reduction in exhaust from fossil fuels. The reduction in fuel use and consequently, the reduction in air pollution was estimated by Komanoff, et al. (1993) through an assessment of the bicycle vehicle miles traveled and a per mile estimate of emissions and fuel consumption for the travel if the trip had been made with a motor vehicle. Using FHWA's *National Personal Transportation Study* of 1990, Komanoff developed estimates of total bicycle trips. He estimated the total mileage for bicycle and pedestrian trips ranged between 26.3 and 65.4 billion miles per year. Based on the number of trips displaced and the nature of these trips, the authors estimated that non-motor vehicle trips annually displaced between 1.2% and 5% of passenger vehicle emission of carbon dioxide (CO_2), nitrous oxide (NOX), carbon monoxide (CO) and volatile organic compounds (VOC).

In evaluating the high rate of emissions, Komanoff stressed that the vehicle trips replaced by bicycles are typically of a short duration, high polluting type. Short trips pollute more than long trips because cold engines at start up emit CO and VOC at higher rates than vehicles on long highway trips. Then, even after the engine is turned off, vehicles continue to emit VOCs.

In recognition of these environmental benefits, the Transportation Efficiency Act for the Twenty-first Century (TEA21) provided flexible funding to increase investment in bicycling infrastructure in the hopes of increasing bicycling. Some optimistic projections anticipated as much as a 15% reduction of harmful emission by the year 2000 if walking and bicycling trips had met projected levels.

Beyond the air pollution benefits, bicycling and walking have other environmental benefits. Non-motorized vehicles require less road space because their reduced size and speed require less headway. Reduction in demand for new roadways reduces loss of open space, conversion of farm lands, loss of permeable land in drainage basins and promotes more concentrated land use. Bicycle use reduces noise pollution. The amount of land relegated to parking is also greatly reduced for bicycles.

In consideration of all of these environmental benefits, it is no surprise that public policy has undertaken programs to increase bicycle ridership. Yet, as stated earlier, the bicycle accident rate is higher than the motor vehicle accident rate. If the public is to be encouraged to switch from automobiles to bicycles, public officials must utilize their resources to identify and address the sources of risk.

2.4 Bicycle Route Evaluation Models

The need to assess the suitability and safety of bicycle routes has long been recognized by many researchers and government officials. In the past two decades a number of models have been developed to rate existing bicycle routes or to improve or establish new bicycle routes. Some models base their ratings on the safety of the route. Others include the quality and convenience of the bicycle experience to establish a suitability rating. Still, others seek a level-of-service measurement analogous to the widely accepted *Highway Capacity Manual* analyses for roadway operators.

Models vary greatly in the methods used to validate them. Validation criteria have included gross accident occurrences, expert opinions, and accident severities. Some bicycle route selection models did not undergo validation. Validation is important because without validation, there is no means to evaluate the accuracy of the model's predictions. The following section presents a historical discussion of the most significant bicycle route evaluation models.

2.4.1 Bicycle Safety Index Rating

In 1987, Davis, a pioneer in the development of bicycle safety models, developed the Bicycle Safety Index Rating (BSIR). The BSIR attempts to rate the relative safety of comparative bicycle safety routes. A BSIR is determined by assigning a Road Segment Index (RSI) to the individual road segments and an Intersection Evaluation Index (IEI) to the intersections. The road segment rating, that ranges from 0 (excellent) to 6 (poor), is computed as:

$$RSI = \left[\frac{AD}{L \times 2{,}500}\right] \div \left[\frac{S}{56}\right] \div \left[(4.25 - W) \times 1.635\right] + \sum[PF] + \sum[LF] \tag{2.1}$$

Where the RSI is a function of the speed limit (S), width of the outside traffic lane (W), pavement condition (PF), a location factor (LF) which reflected geometric and operational hazards, and the average daily traffic volume per lane (AD) averaged over the segment length (L). The intersection evaluation index (IEI) is computed as:

$$IEI = \left[\frac{VC \div VR}{10{,}000}\right] \div \left[\frac{VR \times 2}{VC \div VR}\right] + \sum[GF] + \sum[SF] \tag{2.2}$$

This index is a function of the intersection cross street volume (VC) divided by the main route volume (VR), geometric factors (GF) and signalization factors (SF).

Neither objective nor subjective criteria were used to validate the rating indices calculated by Equations 2.1 and 2.2. Davis (1994) did apply these ratings to seven routes in Chattanooga, Tennessee. He found the rating system to be oversensitive to pavement and location factors which overwhelmed contributions of speed, lane width or volume. The intersection factor reflected the impact of individual intersections but disregarded the frequency of the intersections.

In response to the Chattanooga experience, Davis (1995) modified his BSIR by dropping the intersection component. He used the revised formula to rate eight routes in Atlanta, Georgia. After comparing these ratings to the perceptions of twenty-nine cyclists, he concluded that the calculated ratings differed greatly from the bicyclists' perceptions because the bicyclists' sensitivity to volume and speed overshadowed all of the other factors.

2.4.2 Roadway Condition Index

Broward County, Florida adopted a bicycle suitability index titled the Roadway Condition Index (RCI). The RCI is essentially the Davis' BSIR model also modified by dropping the intersection assessment component from the model. Using this rating system, Broward County assessed all major county streets and highways. The results of these assessments were color coded on a Broward County map as shown in Figure 2.1

As an alternative to subjective bicyclist scores, Epperson (1994) used actual bicycle accident rates as validation criteria. The roadway network of Hollywood, Florida was evaluated by city planners using a modified RCI. As in the Broward County RCI, the intersection evaluation index was dropped. Other modifications included adjusting the location and pavement factors so that they made less of a contribution and multiplying the lane width term by the speed limit to place a greater weight on narrow road segments with high vehicle speeds.

To determine the bicycle accident rate, each motor vehicle / bicycle accident which occurred within a twenty-month period between 1990 and 1991 was plotted on a map of Hollywood, Florida. The accidents were weighted according to a severity scale of 1 (no injury) to 5 (fatality). Accident rates were obtained for each roadway section by dividing the weighted total by the roadway length resulting in a weighted number of accidents per mile.

The problem with this approach is that it relied on the gross number of accidents, which is independent of bicycle exposure. Unless the rate adjusts the gross number of accidents by the volume of bicycles, sections of roadway with high bicycle usage will naturally reflect higher bicycle accidents. This result leads to the faulty conclusion that a route preferred by bicyclists is more dangerous than a route which bicyclists avoid. Furthermore, the weighting of the injury severity assumed that this value is continuous and possesses a ratio relationship. Not surprisingly, the correlation between the predicted RCI and the computed accident rate was low (approximately eighteen percent).

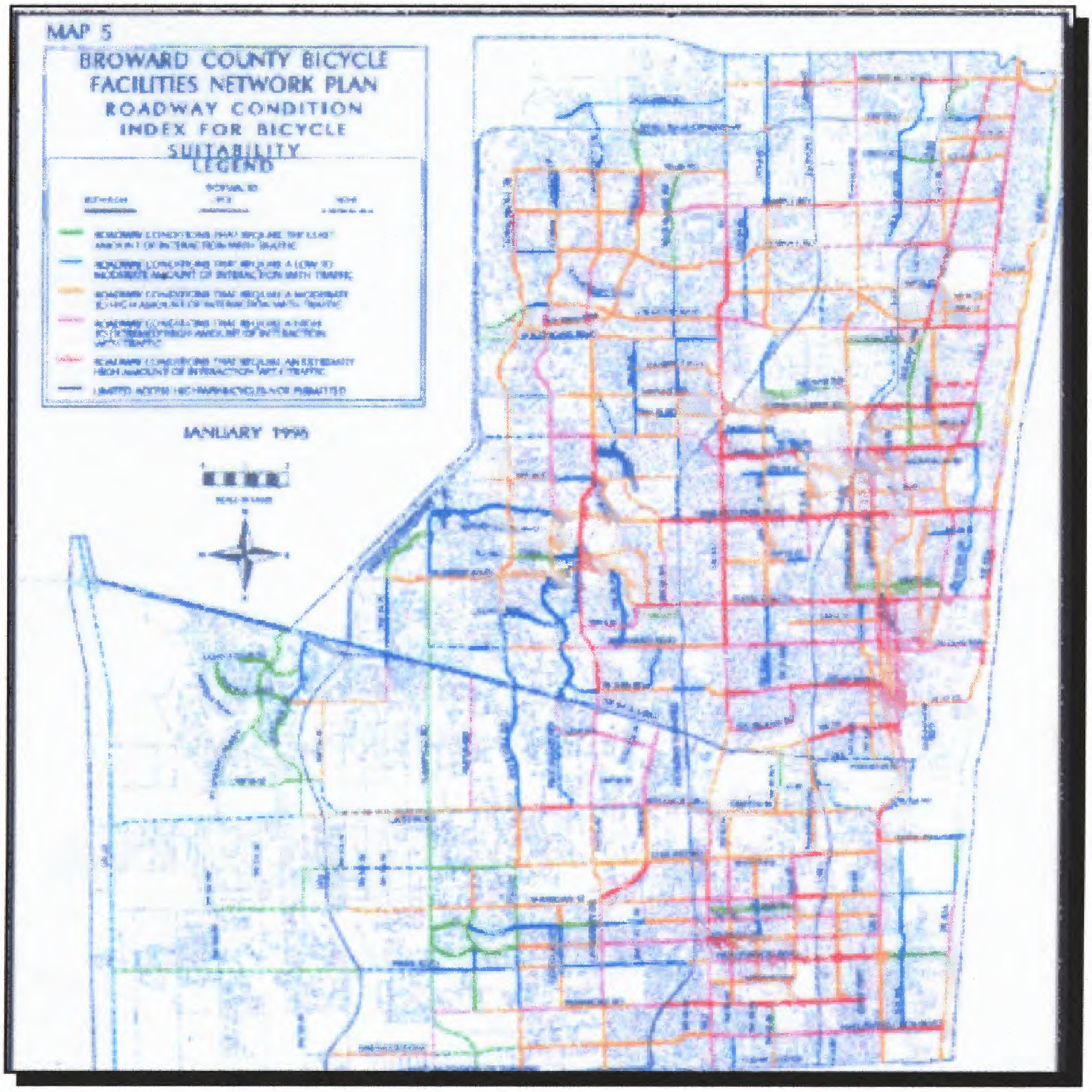


Figure 2.1 Broward County Bicycle Facilities Network Plan.
Source: Shawn M. Turner, C. Scott Shafer and William P. Stewart, Bicycle Suitability Criteria: Literature Review and State-of-the Practice Survey, Texas Transportation Institute, July, 1997.

A similar application of the modified BSIR was made by Eddy (1996) to map small urban areas in Oregon and Washington State. In Table 2.1, Eddy defined five Bicycle Level-of-Service categories.

Table 2.1 Bicycle Level of Service

LOS	Description
Superior (< 3.00)	Conducive to bicycle use. Minor improvements, if any, needed.
Good (3.00 - 3.99)	Accommodates most cyclists. Minor improvements may improve to superior rating.
Fair (4.00 - 4.99)	Useable by many cyclists but poses hazards. Improvements such as shoulder or lanes, may be needed.
Poor (5.00 - 6.99)	Useable by some cyclists but poses significant hazards. Improvements, such as shoulders or lanes, probably needed.
Very poor (>6.99)	Substandard conditions combined with heavy traffic create significant hazards. Should be improved.

Source: Nils Eddy, Developing Suitability of Roadways for Bicycle Use, Pro Bike/Pro Walk 96 Resource Book, Proceeding of the Ninth International Conference on Bicycle and Pedestrian Programs Resource Book, Bicycle Federation of America and Pedestrian Federation of America, 1996.

Eddy recommended using these ratings to estimate the effect of improvements, evaluate bicycle facility networks, provide maps to cyclists, and make comparisons between different urban areas based on the quality of their bicycle facilities. These ratings place a heavy reliance on the underlying Davis BSIR model. Considering Davis' earlier findings of the disparity between the bicyclists' scores and BSIR ratings, caution should be exercised before implementing any of these recommendations.

2.4.3 Bicycle Stress Level

Sorton and Walsh (1994) criticized these earlier bicycle route selection models because of their reliance on ADT instead of peak hour volume, lack of rational basis for validation, and failure to make distinction between rural and urban routes. The researchers also theorized that the experience level of the bicyclist was a key factor in determining a route's suitability. They classified bicyclists into four categories: 1. Child - under the age of ten which should only cycle with adult supervision; 2. Youth - secondary school student with some street riding; 3. Casual - recreational cyclists and/or those making discretionary trips who place a high priority on low congestion and safe environment; and 4. Experienced - bicyclists who commute and tour, preferring direct and convenient routes.

Stress Level was stratified from Level 1 which would be suitable for even the most casual cyclist to Level 4 where no one would wish to bicycle under any conditions. Three factors determined the Stress Level. Stress Levels were a function of peak hour traffic volume, curb lane width and speed limit. The three factors were averaged to establish a single stress level.

An expert judgement model approach was used by showing sixty-one bicyclists, the experts, videotapes of twenty-three roadway segments. The model was developed by fitting the experts' ratings (response variables) to the routes' characteristics (explanatory variables). The bicyclists were asked to rate themselves as to their level of experience and rate the roadway segments as to the perceived stress levels. Upon examination, the researchers discovered that most bicyclists overestimated their experience level. The bicyclists' sensitivity to the stress factors became less pronounced as the bicyclist's level of experience increased. However, the sensitivity of the bicyclists' perceptions to the stress factors was far less than expected. Sorton's experience demonstrates the difficulties in relying on the perceptions of bicyclists for the development of expert judgement models.

2.4.4 Interaction Hazard Score

To obtain bicycle route selection ratings which would better reflect the preferences of the cycling public, researchers have used expert judgement models (Landis, 1998; Harkey, 1999; Jones, 2003 and Noel, 2003) whereby models are fit based on the judgements or ratings of the experts, in this case the bicyclist themselves. Using the perceptions of bicyclists as the response variable, the model directly fit the variables defined in Equation 2.3 using standard regression techniques. The bicyclists' subjective ratings used in these expert judgement models were obtained by a variety of methods. Some researchers had bicyclist raters perform field tests, rating roadway sections after actually riding on them. Other researchers collected bicyclist ratings after showing them video taped or simulated roadway sections. The Highway Safety Research Center at the University of North Carolina is using virtual reality to simulate roadway sections for bicyclists to rate.

Landis (1997) developed an Interaction Hazard Score (IHS) using 150 bicyclists who varied in age, gender, experience levels and geographic origin. He considered using simulated riding conditions, but opted for real urban traffic and roadway conditions to better represent all vehicle and operator response factors.

Each of the test sections rated by the bicyclists varied in the amount of traffic volume in the outside lane, the traffic speed, the mix of vehicle types, interactions with driveways and intersecting streets, pavement condition and degree of separation between the traffic stream and the bicyclist. Coefficients for the variables were established using standard multiple linear regression. The resulting model, with a $R^2 = 0.73$ is stated as:

$$\begin{aligned} BLOS = {} & 0.589\ln(VOL_{15}/L) + 0.826\ln[SPD_P(1+\%HV)] \\ & + 0.019\ln(COM_{15} * NCA) + 6.406PC_5 - 0.005(W_e)^2 - 1.579 \end{aligned} \tag{2.3}$$

Where BLOS, the Bicycle Level of Service (the perceived hazard of the shared-roadway environment) is a function of the volume of directional traffic in 15 minute time period (VOL_{15}), total number of through lanes (L), posted speed limit (SPD_P), percentage heavy vehicle (%HV), trip generation intensity of the land use adjoining the road segment with a stratification of 15 (COM_{15}), effective frequency per mile of noncontrolled vehicular access (NCA), FHWA's five point pavement condition rating (PC_5), and the effective width of the shoulder lane (W_e) calculated as the sum of the curb lane plus any paved section beyond the lane stripe minus any encroachments.

This model predicts the bicyclists' perception of the route's safety, not the route's safety. Most models avoid claims of safety prediction by designating their indices as bicycle route suitability, compatibility or stress level.

2.4.5 Bicycle Compatibility Index

Harkey (1998) also used an expert judgement model to develop the Bicycle Compatibility Index (BCI), which has become FHWA's standard approach for bicycle route selection. Model response variables were obtained using twenty-four participants to rate 13 locations by viewing 40-second video clips which resulted in 312 (24 x 13) data points. Video clips of roadway sections were used instead of real-time road riding because there were no risks to the bicyclists, specific variables could be presented to the bicyclists in a controlled environment, the bicyclists could be exposed to a greater number of operational and geometric conditions and the same conditions could be evaluated by bicyclists from different cities. The BCI model was developed using standard multiple linear regression techniques. Harkey reported a correlation coefficient of $R^2 = 0.89$ for the following model:

$$\begin{aligned} BCI = 3.67 - 0.966BL - 0.410BLW - 0.498CLW + 0.002CLV \\ + 0.0004OLV + 0.22SPE + 0.506PKG - 0.264AREA + AF \end{aligned} \tag{2.4}$$

The BCI includes the presence of a bicycle lane or paved shoulder (BL), bicycle lane or pave shoulder width (BLW), curb lane width (CLW), curb lane volume (CLV), other lane volume (OLV), the 85th percentile speed (SPD), the presence of parking (PKG), the type of roadside development (AREA), and an adjustment factor that reflects hourly curb lane volume, parking time limit and peak hour right turn volume (AF).

2.4.6 Accident Severity

Accident prediction models based on injury severity are an objective alternative to expert judgement models. Injury severity is assigned by the police officer completing the police accident report. The police officer classifies the level of injury into one of the following categories: 0 - no injury, 1 - minor injury, 2 - injury, 3 - incapacitating injury and 4 - fatality. In recognition of the non-continuous, non-ratio quality of injury severity, categorical regression techniques such as logit and probit models are used to determine the weights of the coefficients for the physical and operational parameters. Accident severity models have been applied to predict accidents for motor vehicles (Vogt, 1996) and motorcycle accidents (Shankar, 1996), among others. These models have been found to be reliable based on their statistical goodness-of-fit, robustness and stability of injury severity expression coefficients (Saccomanno, 1996).

Klop, et al. (1999) applied a probit model to develop an injury severity model for bicycle accidents on rural roadway segments in North Carolina. Bicycle / motor vehicle crash data were obtained from FHWA's Highway Safety Information System (HSIS) for the period between 1990-1993 (N=1,025). Using this method, they developed an accident severity model based on horizontal curves, grades, Average Daily Traffic (ADT), speed limit, intersections, right shoulder width, darkness, rain and fog.

The goodness-of-fit for this probit model was measured using a modified ρ^2 as calculated by Equation 2.5.

$$\rho^2 = 1 - \left[\frac{\ln L_b}{\ln L_0}\right] \tag{2.5}$$

Where,

L_b = log– likelihood at convergence

L_0 = log– likelihood at 0

The fit of the overall model was low (p^2 = 0.024), but improved slightly for only rural conditions (p^2 = 0.026). Further improvements may be possible by including additional parameters, confounders and interactions, improving data and exploring non-linear effects.

2.5 Contributing Factors

The accurate prediction of the safety of a bicycle route requires an understanding of the causes of bicycle accidents and the factors which affect the severity of the injury suffered. This understanding gained from experience and the literature review is essential to the development of the data collection program. To maintain focus and avoid expending excessive effort, data should only be collected for those variables that may impact either the frequency or injury severity of motor vehicle / bicycle accidents.

Forester, in his book *Bicycle Transportation (*1983), examined accident rates of various groups of bicyclists. He based his analysis on Planek, Klecker and Driessen's survey of elementary school children (1975), Schupak and Driessen's survey of college cyclists (1976), Kaplan's survey of League of American Wheelmen (1984) and S.M. Watkin's study of cyclists of the British Touring Club (1984). The accident experience of these groups is shown in Table 2.2. These studies are important in that they include all bicycle accidents, not just bicycle/motor vehicle accidents. In general they found that accident rates decreased with experience and that females had a 60% higher accident rate than males. As the individual ages and obtains a motor vehicle license, he learns safe operating procedures for road use. Accident rates drop further with cycle club members who are accustomed to riding in heavy traffic and learn through organized cycling experience. Still, the accident rates for even experienced cyclists are ten times the rate of motor vehicles.

Table 2.2 General Accident Rates

Type of Cyclist	Miles per Year	Accident per million miles
Elementary School	580	720
College	600	500
League of American Wheelman	2,400	113
Cyclists' Touring Club	2,000	66

Source: Forester, J., *Bicycle Transportation*, MIT Press, Cambridge, Massachusetts, 1983.

Bicycle collisions with moving vehicles are neither responsible for the highest percentages of total accidents nor serious accidents. Nonetheless, serious injuries are most prominent in crashes that involve motor vehicles. Table 2.3 shows that serious injuries are over-represented in bicycle motor vehicle crashes with only 18% of the accidents producing 26% of the serious injuries. Falls, the largest source of single vehicle accidents, result from either cyclist error, road surface faults or obstructions.

Table 2.3 Accident Types and Frequencies

Type	% of All Accidents	% of Serious Accidents
Fall	44	38
Collision w/vehicle	18	26
Collision w/other bicycle	17	13
Collision w/dog	8	10
Collision w/parked car	4	2
Bicycle failure	3	3
Collision w/pedestrian	1	1
Other	5	7

Source: Forester, J., *Bicycle Transportation*, MIT Press, Cambridge, Massachusetts, 1983.

As an outgrowth of the NHTSA's earlier effort to measure and classify bicycle accidents, Cross and Fisher (1977) developed a bicycle crash typology defined by crash descriptors, location descriptors, bicyclist characteristics, intersection details, driver contributing factors, motorist contributing factors, bicyclist contributing factors, environmental contributing factors and fault of driver, bicyclist or both. Hunter, et al. (1995)

applied this typology to the NHTSA's General Estimating System's (GES) database which consisted of three thousand bicycle motor vehicle cases drawn from the states of California, Florida, Maryland, Minnesota, North Carolina and Utah.

Table 2.4 Motor Vehicle / Pedacycle Crash Types

Accident Type	Percentage
Parallel Path Crashes	35.5%
Motorist turning/merging into bicyclist's path	12.2%
Motorist overtaking	8.6%
Bicyclist turning/merging into motorist's path	7.3%
Crossing Path Crashes	57.5%
Motorist failing to yield when crossing bicyclist's path	21.7%
Bicyclist failing to yield at intersection	16.8%
Bicyclist failed to yield midblock	11.8%
Specific Circumstances	6%

Source: William W. Hunter, Wayne E. Pein and Jane C. Stutts, Bicycle-Motor Vehicle Crash Types: The Early 1990's, Transportation Record 1502, Transportation Research Board, National Research Council, Washington, D.C., 1995.

These accidents were distributed as summarized in Table 2.4. In reviewing these crash distributions, it is appropriate to ask: Did any characteristic of the roadway exacerbate the condition? Did a crosswalk located too far from the intersection increase right turn crashes? Did the absence of a shoulder increase over-taking crashes? With such a high percentage of accidents, as reported in Table 2.4, resulting from crossed paths (57.5%) the question arises as to the role signalization might play in the frequency and injury severity produced by such accidents. The interactions between crash type and injury severity need to be explored in the context of the physical and operational characteristics of the crash location.

2.5.1 Bicyclist Left Turn in Front of Traffic

The following diagrams provided by FHWA (1990) illustrate the seven most common categories of bicycle accidents. By far, the most dangerous type of accident is the "Bicyclist Left Turn in Front of Traffic " as shown in imoves to the road's center line preparing to execute a left turn in advance of the intersection The bicyclist is hit from behind by an overtaking motor vehicle. Factors that may contribute to this type of crash include high vehicular speed, high roadway volume or no control at the intersection. This type of accident is responsible for only 4.3% of all crashes. However, 28% of this accident typology results in serious and fatal injuries.

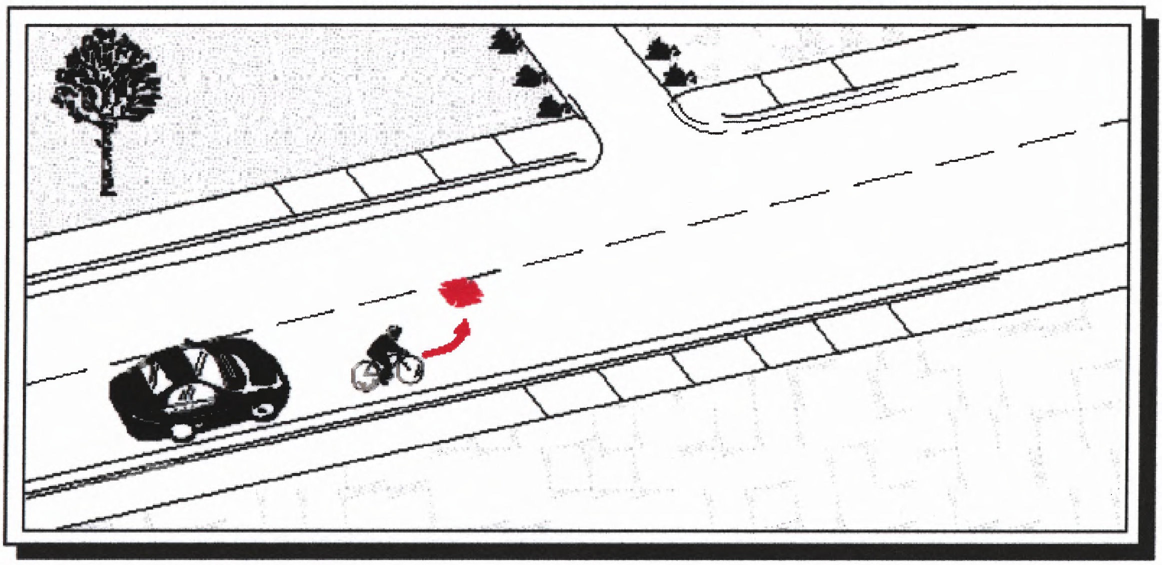


Figure 2.2 Bicyclist Left Turn in Front of Traffic.
Source: FHWA Bicycle Crash Types: A 1990's Information Guide, FHWA-RD-96-163, 1997.

2.5.2 Motorist Left Turn in Front of Bicyclist

The second most dangerous bike-car crash occurs when a bicyclist traveling through an intersection is hit by a vehicle traveling in the opposing direction turning left in front of the bicyclist. Factors which may contribute to this type of crash include poor signal timing either providing inadequate clearance for the cyclist or for the left turning vehicle. This type of crash, as shown in Figure 2.3, results in only 5.9% of all motor vehicle / bicycle crashes of which 24% result in serious or fatal injuries.

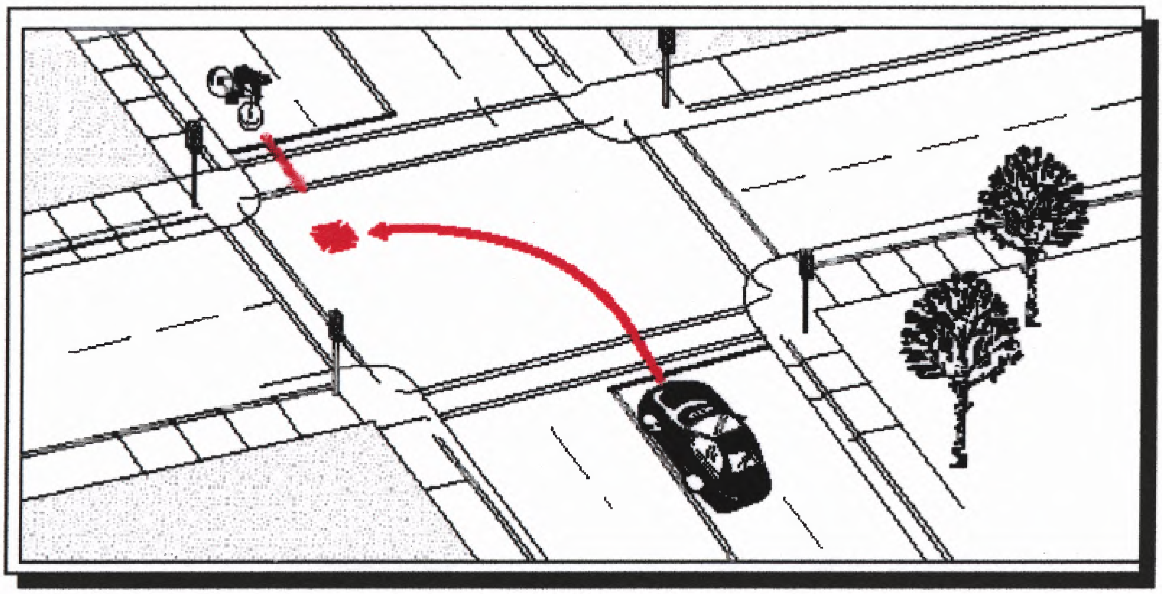


Figure 2.3 Motorist Left Turn in Front of Bicyclist.
Source: FHWA Bicycle Crash Types: A 1990's Information Guide, FHWA-RD-96-163, 1997.

2.5.3 Riding Out of Residential Driveways

"Riding Out of Residential Driveways," as shown in Figure 2.4, also produces 24% serious or fatal injuries. This type of accident is most common among children. Physical factors which can contribute to the risk of this type of crash would include obstructions in both the motorist or bicyclist's line of sight.

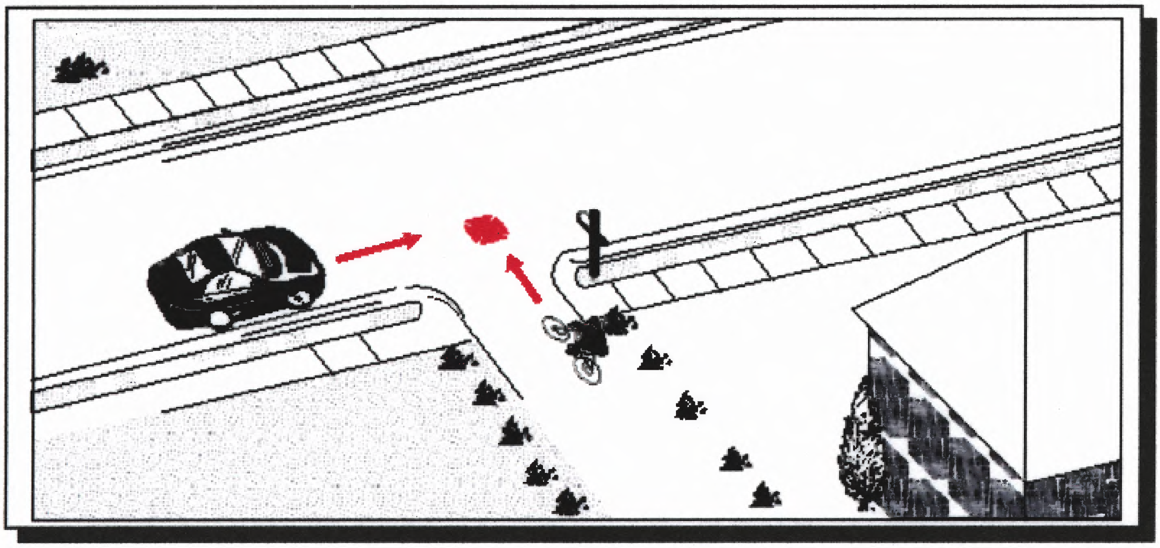


Figure 2.4 Riding Out At Residential Driveway.

Source: FHWA Bicycle Crash Types: A 1990's Information Guide, FHWA-RD-96-163, 1997.

2.5.4 Riding Out at Stop Sign

"Riding Out at Stop Sign" accidents, as shown in Figure 2.5, produces 23% of serious or fatal injuries. In this accident either the bicycle does not sufficiently yield to the motorist or the bicyclists failed to obey stop signs because they did not perceive cross traffic.

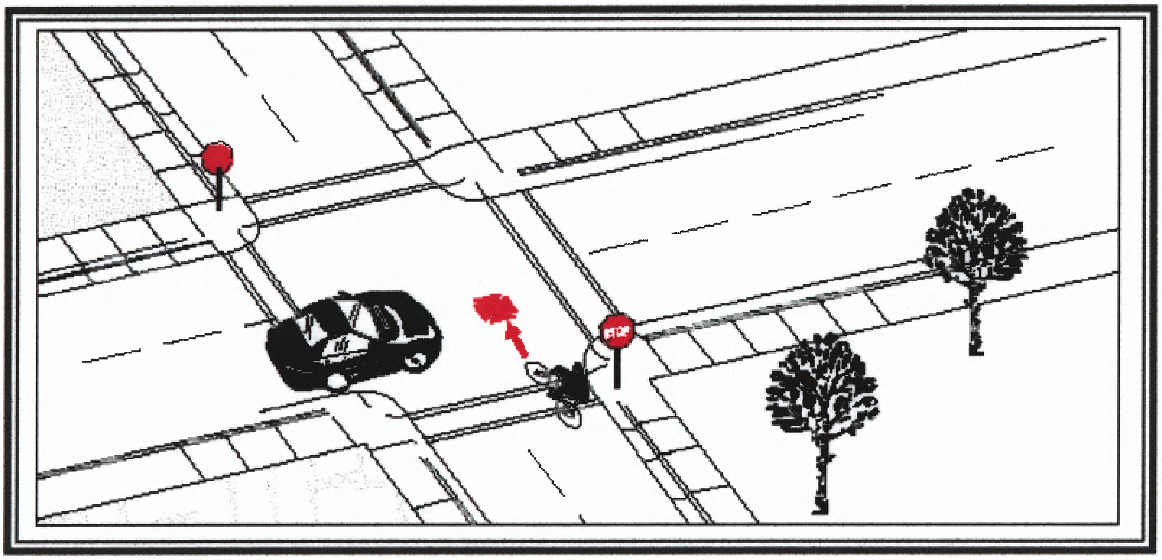


Figure 2.5 Riding Out At Stop Sign

Source: FHWA Bicycle Crash Types: A 1990's Information Guide, FHWA-RD-96-163, 1997.

2.5.5 Motorist Making Right Turns

"Motorist Making Right Turns" accidents, as shown in Figure 2.6, at signalized intersections are also dangerous to bicyclists. These crashes are known to seriously injure or kill 11% of the bicyclists involved in this type of accident. In this type of crash, the bicyclist may be hit by the motorist traveling in the same or perpendicular direction.

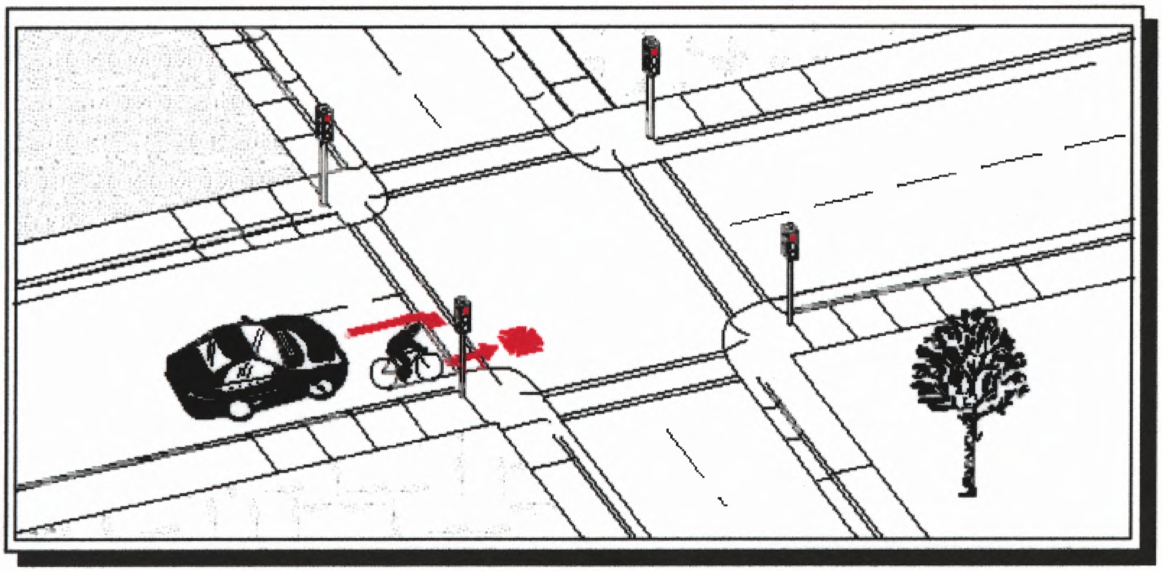


Figure 2.6 Motorist Making Right Turn.
Source: FHWA Bicycle Crash Types: A 1990's Information Guide, FHWA-RD-96-163, 1997.

2.5.6 Driving Out at a Stop Sign

"Driving Out at a Stop Sign" accidents as depicted in Figure 2.7, occurs when the motorist fails to yield to the bicyclist. This type of crash results in serious or fatal injuries in 10% of this type the crashes. As in the previous type, visual obstructions may be a factor. Alcohol consumption by the vehicle driver may also be a factor.

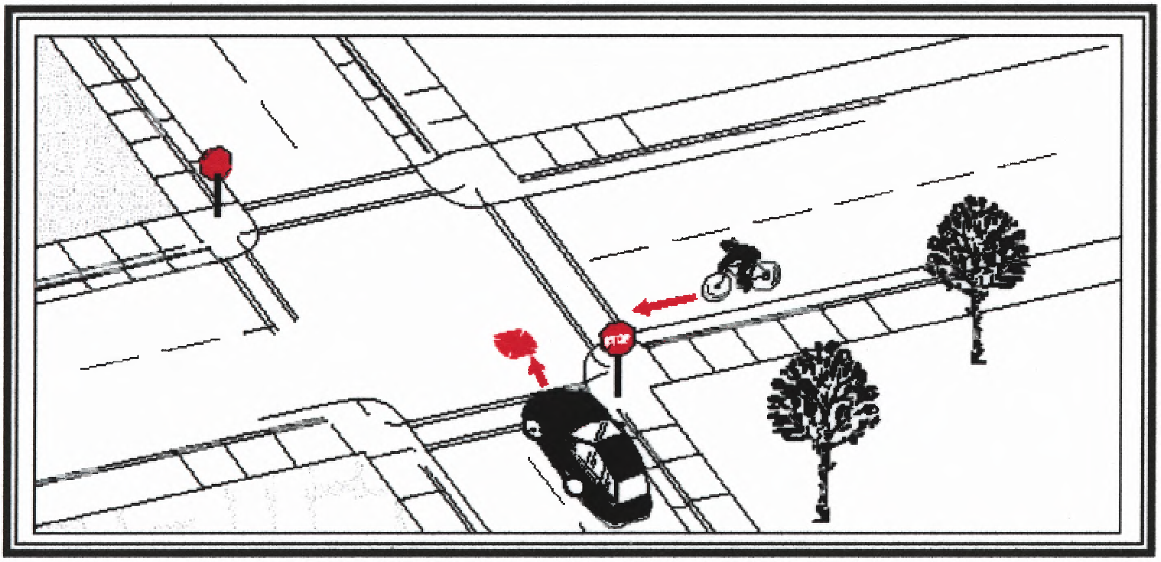


Figure 2.7 Drive Out at Stop Sign.

Source: FHWA Crash Types: A 1990's Information Guide, FHWA-RD-96-163, 1997.

2.5.7 Drive Out at Midblock

The final type of bicycle crash is "Drive Out at Midblock" as depicted in Figure 2.8. In this case a bicyclist traveling either on the street or sidewalk is struck by a vehicle driving out of a driveway at midblock. This type of accident results in less than 7% serious or fatal injuries, probably due to the lower speed of the motorist.

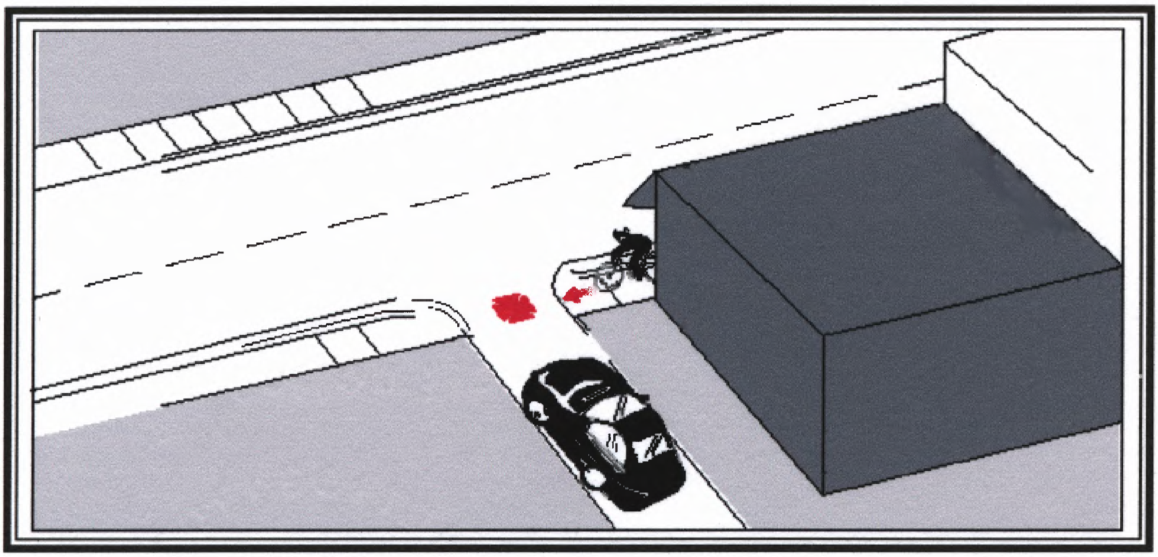


Figure 2.8 Drive Out at Midblock

Source: FHWA Bicycle Crash Types: A 1990's Information Guide, FHWA-RD-96-163, 1997.

2.5.8 Typology Review

A review of these seven accident typologies which have been found to produce the most severe accident injuries yields a number of clues as to the reasons why some accident produce higher injury severity than others. In all of the typologies, the bicyclist is hit by the motorist as opposed to situations where the motorist is struck by the bicyclist. They all occur when the bicyclist and motorist cross paths. The conflict is typically not anticipated because of failures to signal, to obey traffic controls or to observe the other vehicle. These considerations stress the need to consider to presence of signalization, lane width, lighting conditions and age as explanatory variables.

2.6 Physical and Operational Route Characteristics

Bicyclists have been found to take geometric and operational factors into consideration when choosing their routes (Aultman-Hall, 1998). In a study of bicycle commuters in Ontario, Canada, GIS was used to compare shortest path routes with actual routes. A large portion of the bicyclists surveyed preferred the shortest path route. A total of 14.6% preferred the absolute shortest path and 37.5% were within 0.1 m of the shortest path. Table 2.5 reports the difference in the actual route used with the shortest route. The difference in the values of specific factors where bicyclists deviated from the shortest available path to an alternate path provides insight into the importance of these factors to bicyclists and the possibility that the bicyclists recognized the impacts of these factors on accident severity.

The means of each of these factors are calculated by summing the variable over both the actual and shortest routes, then averaging it for all reported trips. Those factors which are deemed significant to a 0.05 level are indicated on Table 2.5 with an asterisk to the left of the variable name. In the subsequent discussion, the first term of the pair refers the variable mean from the preferred path and the second term refers to the variable mean from the shortest path.

Table 2.5 Variable Means for Actual and Shortest-Path Routes

	Variable	Actual Route	Shortest Path Route
	Turns	6.0	5.7
*	Turns per kilometer	1.8	1.9
*	Signals	3.9	3.4
	Signals per kilometer	1.0	1.0
*	Major Signals	1.0	0.9
*	Turns at Signals	1.2	0.7
*	Turns at Major Signals	0.3	0.2
*	Proportion of movements between a Major and Minor Road with a signal	0.09	0.06
*	Proportion of movement from a Minor/Path to a Minor/Path across an arterial with a signal	0.03	0.02
	Proportion of Route on Arterial Roads	0.5	0.4
*	Proportion of Route on Collector Roads	0.14	0.11
	Proportion of Route on Local Roads	0.3	0.4
*	Proportion of Route Off-road	0.05	0.07
*	Travel on Grades (km)	0.09	0.12
*	Travel with > 2 bus routes (km)	0.29	0.35
	Road Bridges	0.5	0.6
*	Level Railway Crossings	0.61	0.55

* Variable significant at o 0.05 level or better

Source: Aultman-Hall, Lisa, Fred L. Hall, and Brian B. Baetz, Analysis of Bicycle Commuter Routes Using Geographic Information Systems, Implications for Bicycle Planning, *Transportation Record 1578*, Transportation Research Board, National Research Council, Washington, D.C. 1998.

For example, to calculate the mean of Travel on Grades (km), the total lengths of sections with steep grades in kilometers were measured for both the preferred and the shortest routes. The mean length of steeply graded section in the paths the bicyclists actually rode was 0.09 km. This was 0.03 km lower than the 0.12 km of steeply graded sections that would have been found on the shortest section. In examining the relative means of these factors reported in Table 2.5, it can be deduced that bicyclists prefer flatter grades (0.09, 0.12) because the shortest routes have on average 25% more steeply graded section than the preferred routes. Table 2.5 reports that bicyclists avoid turns, the mean of the number of turns on the actual route was 1.8 versus 1.9 on the shortest path. This preference for straight sections reverses when a traffic signal is presence. The mean number of turns at traffic signals for the preferred paths is 1.2, almost twice as high as the 0.7 mean number of turns at traffic signals for the shortest path. Clearly, bicyclists are traveling out of there way to make turns at traffic signals. Overall, they liked signals (3.9, 3.4). They disliked bus routes (0.29, 0.35), road bridges (0.5, 0.6) and railroad tracks (0.61, 0.55).

The preceding demonstration of bicyclists' preference for routes, which either include or avoid certain physical factors, justifies the need to explore these factors. An effort will be made to collect the data to determine whether and to what extent these factors affect injury severity.

The following provides a discussion of the candidate explanatory variables as examined in the literature. The findings from previous research will be used to assess the potential of these factors to contribute to or mitigated the severity of an injury resulting from a bicycle / motor vehicle accident.

2.6.1 Facility Width

The width of the curb lane, the lane in which the bicycle operates is critical because it defines the interrelationship between the bicyclist and the adjacent traffic. Curb lane width in urban areas is calculated as the distance from lane marking to curb minus any obstruction. Obstructions may include parking, gutters without bicycle safe grates, bus stops, rumble strips, etc. According to the Highway Capacity Manual (2000), curb lanes that are wider than 14 feet allow the bicyclists to share the lane without compromising either the motorists' or the bicyclists' operations. Harkey (1998) found that on average, motorists positioned themselves between 5.9 feet and 6.4 feet away from the bicyclist when passing, although they will accept a slightly smaller separation if a striped bicycle lane is present. When the curb lane width is less than 11 feet, the bicyclist becomes equivalent to a passenger car because he forces the motorist to move into the adjoining lane to pass. Narrow streets may in fact be safer for bicyclists as they do not have sufficient width to allow the "Car-overtaking Bicyclist" accidents that Cross and Fisher (1977) found to produce the highest fatality rate. A narrow width or the inclusion of a bike lane may also act to calm traffic and reduce operating speeds.

2.6.2 Grade and Curve

The presence of horizontal and vertical curves increase the likelihood of an accident in two ways. For the motorist, both horizontal and vertical curves reduce his sight distance, (AASHTO, 2000). In the presence of high posted speeds, the motorist may find himself bearing down on a bicyclist without advance warning. For the bicyclists, vertical curves are especially dangerous. Downhill stretches increase speeds to unsafe levels. Sudden braking, particularly of the front wheel, can cause fly overs. Uphill, the bicyclist must strain to maintain momentum, possibly dismounting, swerving or peddling while standing. This additional effort can also distract the bicyclist from paying attention to the actions of the motorists passing him. As evidenced by Aultman-Hall (1998), bicyclists will travel longer distances to avoid grades.

2.6.3 Road Division

A number of options exist for separating two way traffic that protect drivers traveling in opposing directions from head on collisions. The simplest and least expensive is a double yellow line. In areas where accident experience is high, especially due to head-on collisions, travel directions may be separated by a grassy median if a generous right-of-way exists, or a Jersey barrier if not. The presence or absence of road division may have a direct impact on the severity of a bicyclist's accident. Medians can serve as safety refuges for bicyclists when making turning movements. However, care must be taken in including both the road division and highway classification variables into the model to be developed as the two variables may be correlated. Most state highways which operate at higher speeds provide some road division. Most local and county roads do not.

2.6.4 Pavement

Pavement quality can impact the severity of an accident in a number of ways. Poor pavement surfaces may cause bicyclists to fall. Collisions can occur when bicyclists and/or motorists swerve to avoid potholes. The frequency of roadway paving can be used as an estimate of the roadway's condition. It has the potential of predicting both pavement quality and other civil improvement such as bicycle-safe inlet grate, good drainage, lane markings and striping. All of these factors should improve bicycle safety. Conversely, there is also the chance that recent paving may be necessitated by high traffic volumes, particularly heavy vehicle volumes. Furthermore, poor roadway surfaces may have the unanticipated benefit of actually discouraging motorists from speeding.

2.6.5 Highway

The road system type as provided in NJDOT accident reports, is determined by the entity which has jurisdiction over the roads. The roadway system can be classified as an interstate, state highway, state/interstate authority, county, municipality and other categories. Many operating characteristics which may vary with road classifications such as operating speed, frequency of access points, restrictions on non-motorized vehicles may affect the severity of an accident.

The results of Wessels' (1996) review of six years of bicycle accident data from Washington State are shown in Figure 2.9. The figure shows that the total number of accidents is higher on city streets, yet, the severity is much greater on County and State Roads.

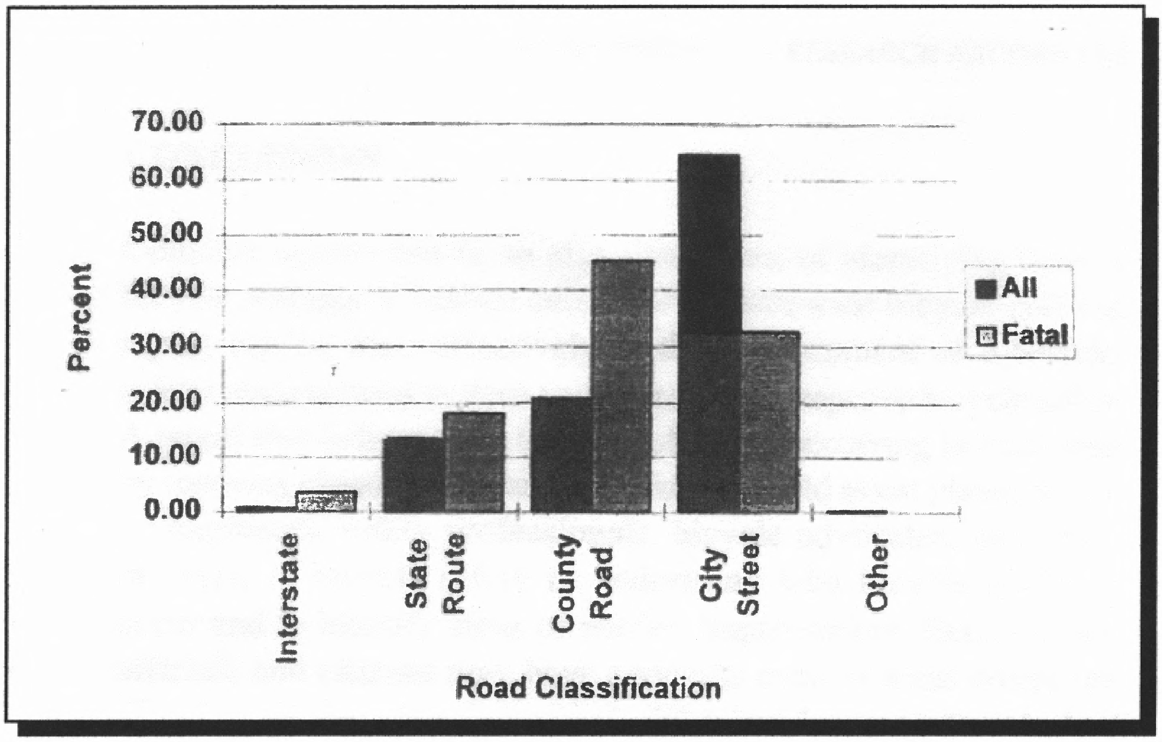


Figure 2.9 All Collisions Versus Fatal Collisions by Highway Classification.
Source: Wessels, Ralph L., Bicycle Collisions in Washington State: A Six-Year Perspective, 1988-1993, *Transportation Record 1538*, Transportation Research Board, National Research Council, Washington, D.C. 1996.

2.6.6 Speed

High motor vehicle operating speeds and consequently high momentum are frequently associated with higher injury severities for bicyclists . High speed limits and consequently high operating speeds increase a vehicle's safe stopping distance (AASHTO, 2000) and is thus more likely to cause serious accidents. Furthermore, the increased momentum of a faster vehicle produces more severe injuries. Garder (1994) analyzed four years of bicycle accidents from

1988-1991 in Maine. Of the twelve fatal accidents, nine involved motor vehicles of which six took place on roads with speed limits higher than 40 mph. Garder calculated a "most likely ratio" as the ratio between the number of fatal bicycle accidents divided by the total bicycle accidents for a given speed limit. As indicated by Figure 2.10, the likelihood of a fatality rises sharply with speed.

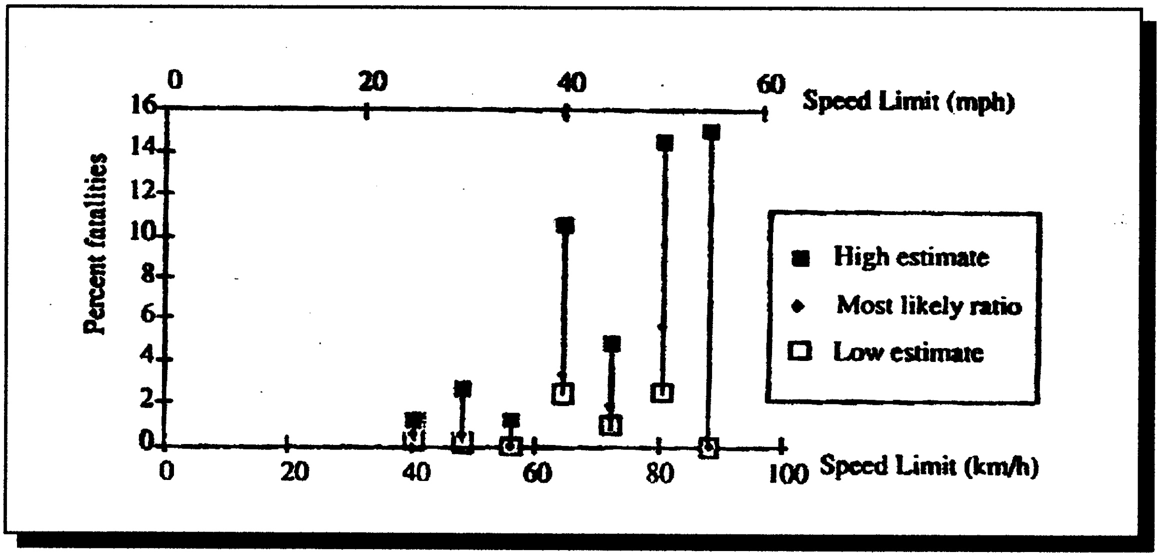


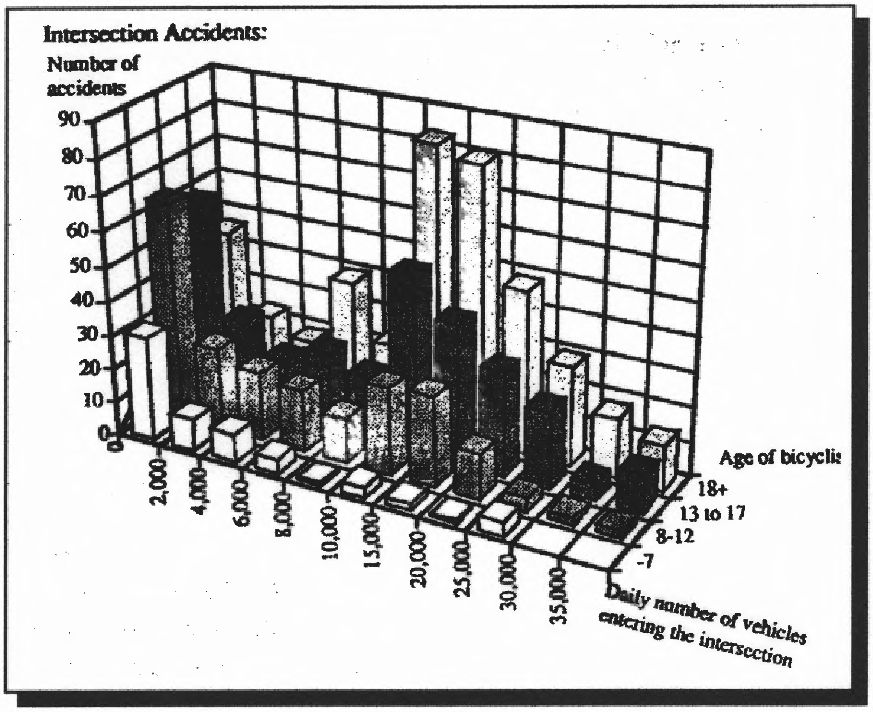
Figure 2.10 Most Likely Ratio of Fatal Accidents for Speed Limits.

Source: Garder, Per, Bicycle Accidents in Maine: An Analysis, *Transportation Record 1438*, Transportation Research Board, Natinal Research Council, Washington, D.C. 1994.

2.6.7 Volume

Motor vehicle volume (ADT) has historically been included in many bicycle route suitability models (Turner, 1997). Sorton (1994) wrote that excessive traffic volume, greater than 450 vphpl is stressful for the bicyclist because the bicyclist must shift his focus from the roadway to the passing car. The bicyclist must exercise caution to prevent any sudden swerving. Traffic volume is especially problematic for bicyclists in the presence of narrow curb lanes on two lane roads as the bicyclist does not have the option to shift onto a shoulder to avert impact. If the traffic volume is dominated by a single flow direction, the motorist can move into the lane for opposing traffic, however, if traffic volume is high and bi-directional, motorists can find themselves stuck behind a bicyclist. The resulting frustration may lead the motorist to take chances and exhibit unsafe behavior.

Garder's (1994) study of Maine bicyclists did not show a higher accident rate on higher volume roads. Figure 2.11 shows higher rates at low volume roads for non-intersection accidents. At intersections, however, roads with moderate volume (15,000 ADT) produce the greatest number of accidents. European studies (Linderholm, 1992 and Brundell-Freij, 1990) revealed that risk, measured as bicycle accidents per mile ridden, actually decreased with increased ADT. This paradox may exist because bicyclists may become more careful when they ride on high-volume roads. In congested urban areas a relationship may exist between traffic volume and operating speed.



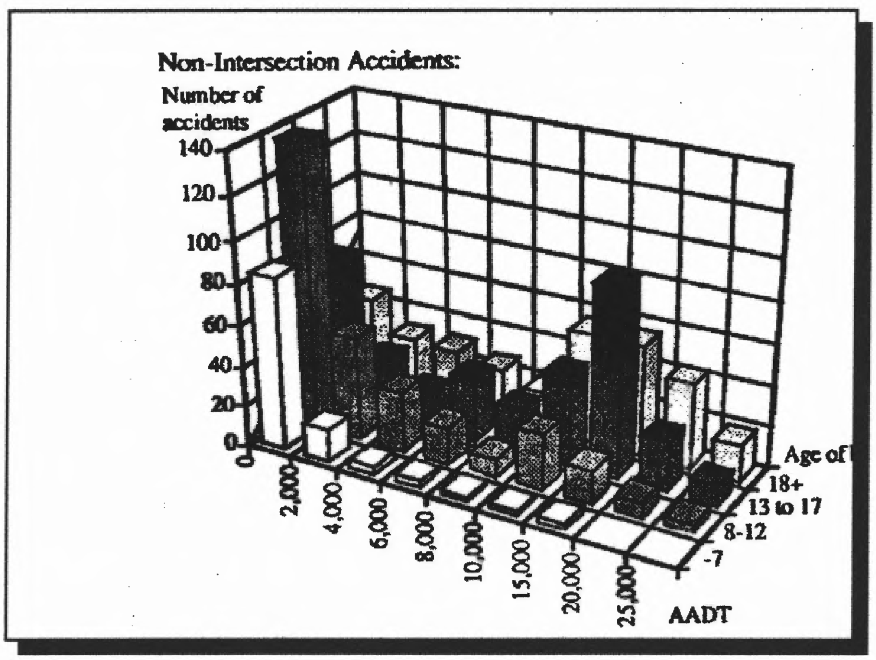


Figure 2.11 Bicycle Accidents in Maine by Age and Vehicle Volume.
Source: Per Garder, Bicycle Accidents in Maine: An Analysis, *Transportation Record 1438*, Transportation Research Board, National Research Council, Washington, D.C. 1994.

2.6.8 Bus

The presence of buses on a route may pose visibility problems for the bicyclist because buses are larger than passenger vehicles and they reduce the lateral space available for through traffic. Buses stopping in traffic force riders to cross the bicyclist's path as they enter or depart the bus. Bus stops in advance of an intersection can obscure the presence of a bicyclist from a right turning motorist. Pedestrians attempting to "catch" the bus, may execute dangerous and erratic maneuvers. Furthermore, since most bus stops are at intersections, bicyclists passing on the right may collide with turning buses. Motorists, frustrated by the slow travel speed of buses may undertake risky passing maneuvers and collide with bicyclists hidden by the bus. There is also the possibility that bus routes are primarily located on through arterials, which probably operate at higher speeds than do the short local streets which are not serviced by buses. Conversely, as in the cases of a number of the variables discussed earlier, the presence of buses may ultimately slow traffic down and thus reduce the severity of the accident. Aultman-Hall (1998) found that commuting bicyclists diverted to longer routes to avoid bus routes. The mean travel length on route sections containing bus routes was 0.29 km for the preferred route versus 0.35 km for the shortest path route.

2.6.9 Parking

Parking maneuvers affect bicyclists in a number of ways. Motor vehicles moving in and out of parking spaces may either collide with bicyclists or may obscure the vision of both the bicyclists and other vehicles in the traffic flow. To avoid the parking vehicle, bicyclists as well as other vehicles may swerve to avoid the parkers. Conversely, numerous parking maneuvers may serve to slow down vehicle speeds and thereby reduce the severity of an accident.

Bicycle accidents with parked cars come into two forms. In the fist case, a motorist opens a door into the bicyclist's path. In the second case, an unobservant bicyclist rides into the rear of a parked car. These types of accidents have caused injuries and even fatalities. At a minimum, parked cars function as a transverse obstruction and effectively narrow the width of the roadway. Cross and Fisher (1977) attributed 8% of non-motorist accidents to parked cars.

2.6.10 Signalization

For motorists, traffic signals have the potential to reduce certain types of accidents such as broadsides, while increasing others such as rear end collision. In the same manner, signalization could be a benefit or a deficit to a bicycle route. By their basic function, signals slow the speed of vehicles approaching and departing the intersection. Yet aggressive drivers may actually speed up on an amber in their hurry to "make the light." Bicyclists, are also guilty of red light running, operating on the questionable assumption that police officers will not issue traffic violations to bicyclists. Traffic signals do provide gaps for safe crossing of heavy volume roads.

Controlled intersections, provided they are properly designed, offer bicyclists many benefits. They establish clear right-of-way between opposing directions, provide gaps for bicyclists proceeding through high volume intersections and facilitate turning movements. Motor vehicle drivers are more likely to see the bicyclist because they are more likely to be attentive at intersections.

Aultman-Hall (1998) found commuting bicyclists actually seek routes which included signalized intersections, increasing their travel path, especially if they needed to make turns. The actual paths chosen contained 1 signal versus 0.9 signals in the shortest path route. This preference for signals was even more pronounced at turns where the bicyclist selected paths that included 1.2 signals at turns as opposed to 0.7 signals at turns for the shortest path option.

2.7 Temporal Factors

Conditions which are temporal in nature such as weather and lighting also impact the likelihood and type of the accident as well as the severity of the injury. Similar to operator characteristics, temporal factors are included in the model as confounders lest these effects cloud the impact of the true variables such as the route's physical and operational characteristics.

A confounder is a pseudo variable that may affect the outcome of an event although it is not a decision variable. For example, wet roads may increase the severity of an accident, yet there is no intent to create a model that selects routes for a specific weather condition. Instead, if weather were found to be significant the model would be fit using weather as a confounder in order to capture the variance that results from weather. During implementation, confounder values are set at one level e.g. weather is dry to allow for unbiased comparisons of bicycle routes under identical weather conditions.

2.7.1 Lighting

Most bicycle accidents occur in the daylight. Garder (1994) found 83% of the accidents in his Maine study occurring during daylight conditions. Of the remaining 17%, half occurred during dawn or dusk, an additional 40% occurred on streets where lighting was present. Accidents where the motorist simply did not see the bicyclist tend to be the most serious. Forester (1983) estimates that ten percent of urban car-bike collisions occur during darkness and that the rate of car-overtaking-bike collisions is thirty times higher at night than during the daytime. This accident typology, however, is largely a problem of rural, unlit roads.

The USDOT Crash Outcome Date Evaluation System (CODES) Project has been funded to correlate police report injury data with hospital records. By comparing the total bicycle accidents to the percentage to that of severe outcomes, inferences can be made as to whether a factor is likely to increase or reduce the injury severity of the outcome.

For example, confirming earlier statements that speed increases severity, CODES data in Table 2.6 shows that while only 10% of all accidents occurred on roads posted with speed limits greater than or equal to 45 mph, these accidents were responsible for 22% of all serious injuries and deaths. No strong relationship exists for time of day. Although the morning seems to be somewhat safer (26% accidents versus 21% serious injuries) than afternoon or nighttime accidents, the difference is not significantly pronounced to make definite conclusions.

Hunter (1995) did find injury severity linked to lighting conditions. Almost 80% of the bicycle-motor vehicle crashes occurred during daylight conditions. Serious and fatal injuries to the bicyclist were heavily over-represented during conditions of darkness with no streetlights.

Table 2.6 Bicycle Crash Factors and Outcomes, 1995-1996 Maine CODES Results

Police Reported Crash Information		Total Bicyclists	Hospitalized or Killed
		713 (100%)	68 (100%)
Crash Location	Urban	568 (80%)	39 (57%)
Crash Location	Rural	145 (20%)	29 (43%)
Road Location	Straight or Curved Road	234 (33%)	32 (47%)
Road Location	Intersections	362 (51%)	27 (40%)
Road Location	Driveway	117 (16%)	9 (13%)
Posted Speed Limit	25 MPH or Less	407 (57%)	24 (35%)
Posted Speed Limit	30-40 MPH	212 (30%)	24 (35%)
Posted Speed Limit	45 MPH and Over	69 (10%)	15 (22%)
Time of Crash	7-12 AM	183 (26%)	14 (21%)
Time of Crash	1-6 PM	413 (58%)	41 (60%)
Time of Crash	7 PM-6AM	117 (16%)	13 (19%)

Source: Bicycle-Motor Vehicle Crashes in Maine, Results of the Maine CODES (Crash Outcome Data Evaluation System) Project, Transportation Research Board Annual Meeting, Washington, D.C., January, 2002.

2.7.2 Weather

Weather indicated in the accident report is determined by the presence of precipitation and visibility, i.e. rainy, snowy or snowy. Bicyclists typically avoid poor weather both because of physical discomfort and because of the impairment to their bicycle's operating performance. Wet brakes are slow to stop bicycle tires. Wet roads cause skids and consequently falls (Forester, 1983). Spray from passing vehicles obscures visibility. In their hurry to get out of the rain, bicyclists may take risks such as red light running in order to arrive at their destination as soon as possible.

Bad weather affects a bicycle's handling ability. Wet road surfaces lengthen the braking distance for both motor vehicles and bicycles. Precipitation reduces visibility. In their haste to seek shelter, bicyclists may take more risks during bad weather. Hunter, et al. (1995)

found fatal injuries over-represented during winter months. Fortunately, bicyclists clearly recognize the danger of riding in icy and snowy conditions. Garder (1994) reported only 0.6% of accidents occurring under those conditions. However, 6% of accidents in the Maine study did occur during wet weather. As with lighting, weather will be explored as a potential confounder for the accident injury model.

2.8 Bicycle Operator Characteristics

A number of factors pertaining to the bicycle operator have been found to be significant to the likelihood and severity of a serious bicycle crash. Although such factors as age, gender, alcohol use and helmet use are not specific to the bicycle route chosen, they must be included in the model as controls or confounders in the same manner as temporal factors.

2.8.1 Age

A number of reasons exists as to why a victim's age may affect the outcome of an accident. A child's lower body mass may be less able to resist the force impact. Also, the child's height places his head and vital organs at the level of the vehicle where an adult would be struck below the thigh. The National Center for Statistics and Analysis reported that in 2000, the fatality rate for children between the ages of 5 and 15 was twice the rate of older bicyclists. Kim (1996) found that while 35% of all riders were over 40, this group experienced only 9% of the reported accidents. The CODES data in Figure 2.12 shows children of ages 5-9 are more likely to die as a bicyclist or pedestrian than while riding in a motorized vehicle.

Age may also reduce the severity rate. Children get in numerous accidents. To a child, the bicycle represents entertainment in addition to transportation. They use their bicycles as toys performing "wheelies", etc. Such behavior causes frequent falls, collisions with fixed objects and other accidents less likely to cause severe injury. Children under 16 have no experience operating motor vehicles and are not acquainted with traffic law and performance limitations of motor vehicles such as braking distance. Because of these reasons, children may exhibit risky behaviors that make them more vulnerable to accidents than mature bicyclists who possess motor vehicle licenses.

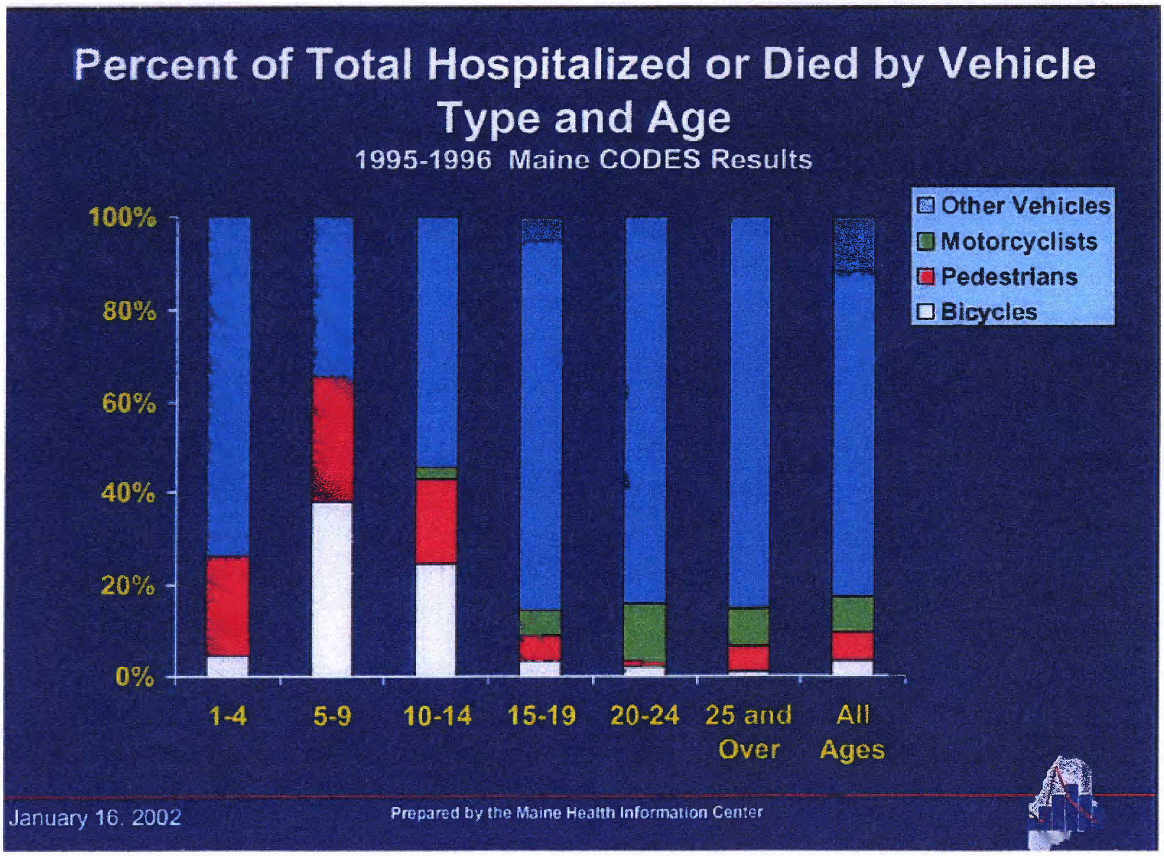


Figure 2.12 Percent of Total Hospitalized or Died by Vehicle Type by Age 1995-1996 Maine CODES Results.
Source: Bicycle-Motor Vehicle Crashes in Maine, Results of the Maine CODES (Crash Outcome Data Evaluation System) Project, Transportation Research Board Annual Meeting, Washington, D.C., January, 2002.

The age of the bicyclist also affects the type of accidents which are likely to occur as shown in Figure 2.13 (Cross and Fisher, 1977). This relationship affects the type of crash and the injury severity, as discussed in Section 2.5. Young children are the most likely to crash riding out of residential driveways, on sidewalks and over curbs. As the cyclist ages and he becomes skilled in handling the bicycle, he begins to cycle in the roadway. Risk-taking and bicyclist error such as wrong way cycling and running stop signs, dominate the crashes which involve older children and teenager. These types of crashes are less frequent amongst adult bicyclists who have learned the rules of the road from driving motor vehicles. Motorist error is a dominant cause of accidents involving adult riders.

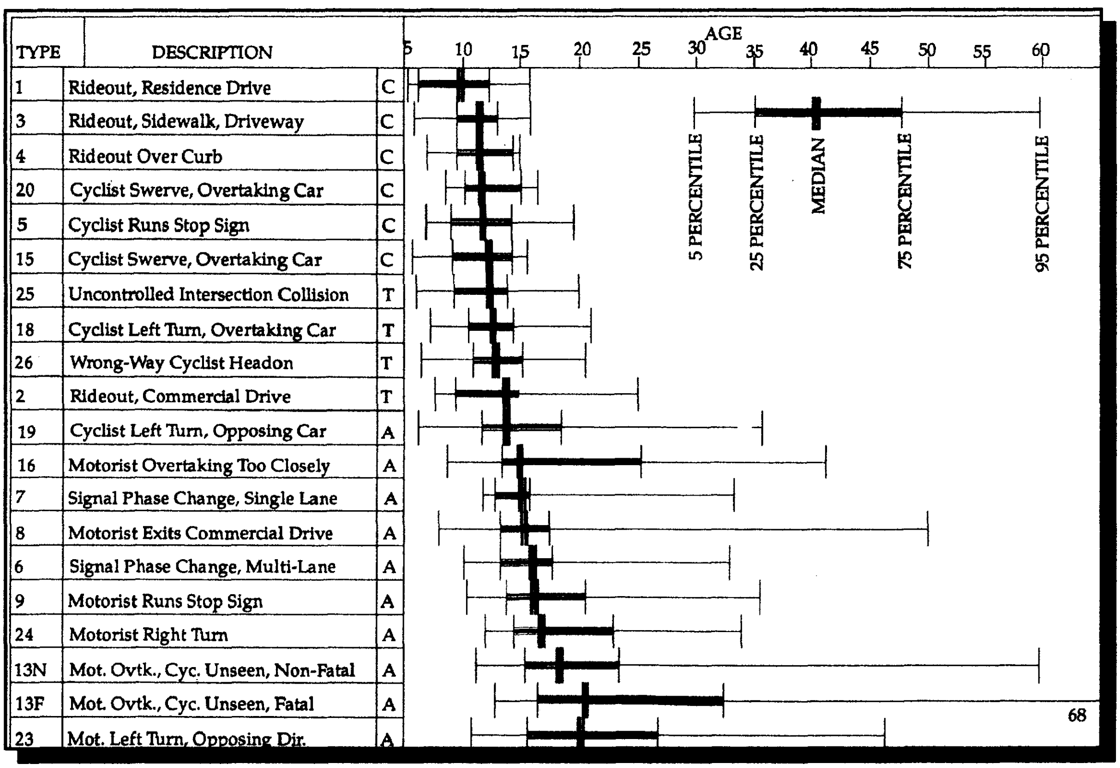


Figure 2.13 Car-Bike Collisions, Arranged in Order of Increasing Median Age of Cyclist. Source: Cross, Kenneth D and Gary Fisher, A study of Bicycle/Motor Vehicle Accidents: Identification of Problem Types and Countermeasure Approaches; National Highway Traffic Administration, 1977.

Wachtel (1994) found that adult bicyclists were 1.8 times more likely to have an accident with a vehicle than children. He theorized that this increased risk may be due to greater exposure of bicyclists to motor vehicle errors. Rodgers (1995) also found adult bicyclists to be over represented in fatalities.

2.8.2 Alcohol Use

NHTSA (2000) reported that alcohol was involved - either for the bicyclist or motorist in one-third of all bicycle fatalities. Excessive blood alcohol levels were present in 26% of the bicyclist fatalities. Alcohol impairs the abilities of bicyclists in much the same way as motor vehicle drivers. Perception is dulled, reaction time is slowed and the potential for driver error is increased. These results combine to increase the likelihood of a crash irrespective of the location.

Drunk drivers are a special threat to nighttime bicyclists operating with insufficient illumination, especially in car overtaking bike accidents (Forrester, 1983). This type of accident is especially dangerous as it is the only class whose proportion of fatal collisions significantly exceed its proportion of nonfatal collisions. The involvement of alcohol is evidenced by the disproportionate amount of car overtaking accidents which occur on Friday and Saturday evenings.

2.8.3 Helmet Use

Wearing a bicycle helmet reduces the risk of head and brain injury among cyclists by about 70% (Rivara, 1996). The use of this safety equipment is reported on the police accident report. Since its effect on the severity of the injury is so overwhelming, helmet use should be included as a confounder in the model.

2.8.4 Gender

Females experience lower fatality and serious accident rates than their male counterparts. Rodgers (1995) found that while just under half of all cyclists were female, only 15% of bicycle fatalities were women. In Kim's (1996) study, over 78% of the injuries occurred to male cyclists. In Wachtel's (1994) exposure based study, he found that among all cyclists, male cyclists were 1.2 times as likely to be injured as female cyclists. This behavior is even more pronounced among children where boys are 1.7 times as likely to have accidents as girls. This difference is probably due to greater risk taking behavior amongst boys and not due to any increased resistance to injury by females. This distinction is important in that while gender would probably be a factor in a frequency-based accident prediction model, it probably will not bias the results of an injury severity model.

2.9 Summary

The preceding review illustrates the voluminous effort that has been undertaken to understand the nature of bicycling during the past three decades. The benefits of bicycling have been universally accepted. Despite the broad-based effort to quantify the nature of the dangers associated with bicycle riding, there is still no reliable means to use this information to increase the safety of the bicycling public. The challenge will be to use the knowledge gained from the myriad operational and geometric factors that affect bicycle safety and combine them with the temporal and operator confounders to create a workable Bicycle Route Safety Rating model.

CHAPTER 3

THEORETICAL APPROACH

3.1 Methodology Choice

Predicting the safety of a bicycle route based on its physical characteristics is more difficult than most accident modeling efforts. A direct approach would be to fit the accident rate of a specific location with operational and physical factors specific to that location. However, for bicycle accidents, that approach is not possible because, unlike other modes of transport, bicycle accident rates are not easily obtainable.

Conventional accident rates are either population-based, i.e. total accidents divided by an area's population, or exposure based - total accidents divided by traffic volume, miles traveled, or other measures of exposure. The former population-based rates are unsuitable for this modeling effort because such calculations are not route specific. For example, a population based accident rate for Jersey City, would result in the same rate for the New Jersey Turnpike Extension, an interstate, as the adjacent Liberty State Park two lane, local access road.

Bicycle accident rates based on motor vehicle volumes are route specific but are not based on the bicyclists' exposure. Reliance on motor vehicle volume based accident rates will result in the false conclusion that a highway interstate with high motor vehicle volume and little bicycle accident experience was safer than a local collector with low motor vehicle volume and some observed bicycle accidents. The absence of bicycles on the highway interstate guarantees that the number of bicycle accidents will be lower. Accident frequency, another commonly used measure of safety, is also impractical because the rarity of bicycle accidents requires extremely long analysis periods.

Since bicycle accident rates are not available to serve as the model's response variable, an alternative indicator must be selected. To be acceptable this variable must be: 1. practical - the index must be provided or be easily obtained from available data; 2. reliable - the index must be rational and widely accepted by traffic engineers; and 3. the index must be an objective standard independent of any bias on the part of the researcher.

These requirements can be met by employing accident severity as the model's dependent variable. Injury severity can be assigned a categorical index dependent on the injury suffered by the bicyclist. The underlying rationale is that the severity of the accident can then be extrapolated to the safety of the route, a fatality is more likely to occur on a dangerous route than a safe route.

Severity, however, is not a continuous variable. Only integer values are possible. Accident severity does not have a ratio relationship, which means an accident resulting in a moderate injury (Injury severity = 2) is not ½ as fortunate as a victim suffering a fatality (Injury severity = 4). The ratio may actually be 1 to 100, or 1 to 1000.

Although ordinary least squares (OLS) regression is permissible for categorical explanatory variables with continuous response variables, a categorical response variable violates the Gauss-Markov conditions for linear regression in that a categorical variable is not normally distributed. Non-linear transformations such as logit are used in lieu of OLS for categorical prediction models. Prior to embarking on a discussion of categorical methods which is the selected methodology, the fundamentals of OLS will be reviewed for comparison purposes.

3.2 Linear Regression Modeling

Various mathematical techniques have been developed to provide researchers with the tools to derive models from sets of experimental data or observations. Linear regression, one of the most commonly used methods, seeks to define the best fitting relationship between a number of predictors, X (explanatory variables) and outcome, Y (response variable), by minimizing the square of the error between the predicted outcome and the observed outcome (Vining, 1998). The simplest form of this model assumes a linear relationship between a single independent variable X:

$$Y = \alpha + \beta X + \varepsilon \tag{3.1}$$

Where:

Y - observed value of the response variable,
X - explanatory variable
α - intercept
β - slope or weight of the explanatory variable
ε - error term

Given n data points, the intercept and the slope of the independent variables can be derived using the ordinary least square method in the following manner. Equation 3.2 is constructed as the prediction of Y as a function of the intercept and slope.

$$\hat{Y}_i = \alpha + \beta X_i \tag{3.2}$$

Where $\hat{Y}_i$ is the predicted value of the i_{th} response

Using Equation 3.3, the error term ε or residual is calculated as the difference between the observed response and the predicted response.

$$\begin{aligned} \varepsilon_i &= Y_i - \hat{Y}_i \\ SS_{res} &= \sum \left(Y_i - \hat{Y}\right)^2 \\ &= \sum \left[Y_i - \left(\alpha - \beta X_i\right)\right]^2 \end{aligned} \tag{3.3}$$

Taking the partial derivatives of Equation 3.3 and setting them equal to 0 to find the minimum point, yields the following equations which define the model's coefficients:

$$\begin{aligned} \alpha &= \overline{Y} - \beta \overline{X} \\ \beta &= \frac{S_{xy}}{S_{xx}} \\ S_{xy} &= \sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) \\ S_{xx} &= \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \end{aligned} \tag{3.4}$$

Where:

$\overline{Y}$ and $\overline{X}$ are the mean of Y and X respectively

This methodology can be expanded to multivariate conditions in situations where there is more than one independent variable. In the case of k predictors, the generalized form of Equation 3.1 is:

$$Y = \alpha + \sum_{i=1}^{k} \beta_{\kappa} X_k + \varepsilon \tag{3.5}$$

To apply the OLS method, the data must meet the three Gauss - Markov assumptions:

- All relevant and no irrelevant X's are included in the model;
- The residual ε is homoscedastic which requires that ε is normally distributed with a constant variance σ^2 across all observations; and

- The residual ε is serially independent in that it is not correlated with any other independent variable.

Provided these assumptions are met, the OLS method produces Y predictors which are "Blue" - the best, linear, unbiased estimators (Aldrich, 1984).

3.3 Categorical Models

Categorical models are models whose dependent variables are non-continuous. They may be dichotomous having only two states, i.e. yes or no, Patient is dead $Y = 0$ or Patient is alive $Y=1$. These models also may be polytomous or have more than two states. For examples, injury severity is a polytomous variables with four levels: driver is not injured ($Y = 1$), driver is mildly injured ($Y = 2$), driver is seriously injured ($Y = 3$), and driver is dead ($Y = 4$).

Although OLS solutions can be obtained with categorical explanatory variables, OLS is not an option when the response variable is categorical. Not only does categorical data violate the requirement of homeoscedasticity, and thus forfeiting the promise that the OLS solution would produce the best linear unbiased estimate, the OLS solutions cannot be constrained to combine the Xs in such a fashion to generate only Ys that are integers of the level allowed by the target response variable. OLS solutions require that the dependent variable be continuous and able to assume any value between $-\infty$ and ∞.

One solution to this problem is to define Y as the expected value or probability that $Y = y$ and in the dichotomous case, Y can assume only two values: 0 or 1, then:

$$\begin{aligned} E[Y] &= P(Y = y) \\ &= 0 * P(Y= 0) + 1 * P(Y_i = 1) \\ &= P(Y = 1) \end{aligned} \tag{3.6}$$

As shown in Figure 3.1, a number of mathematical functions have been used to transform categorical variables to meet the requirement that probability distributions are continuous and range from 0 to 1.

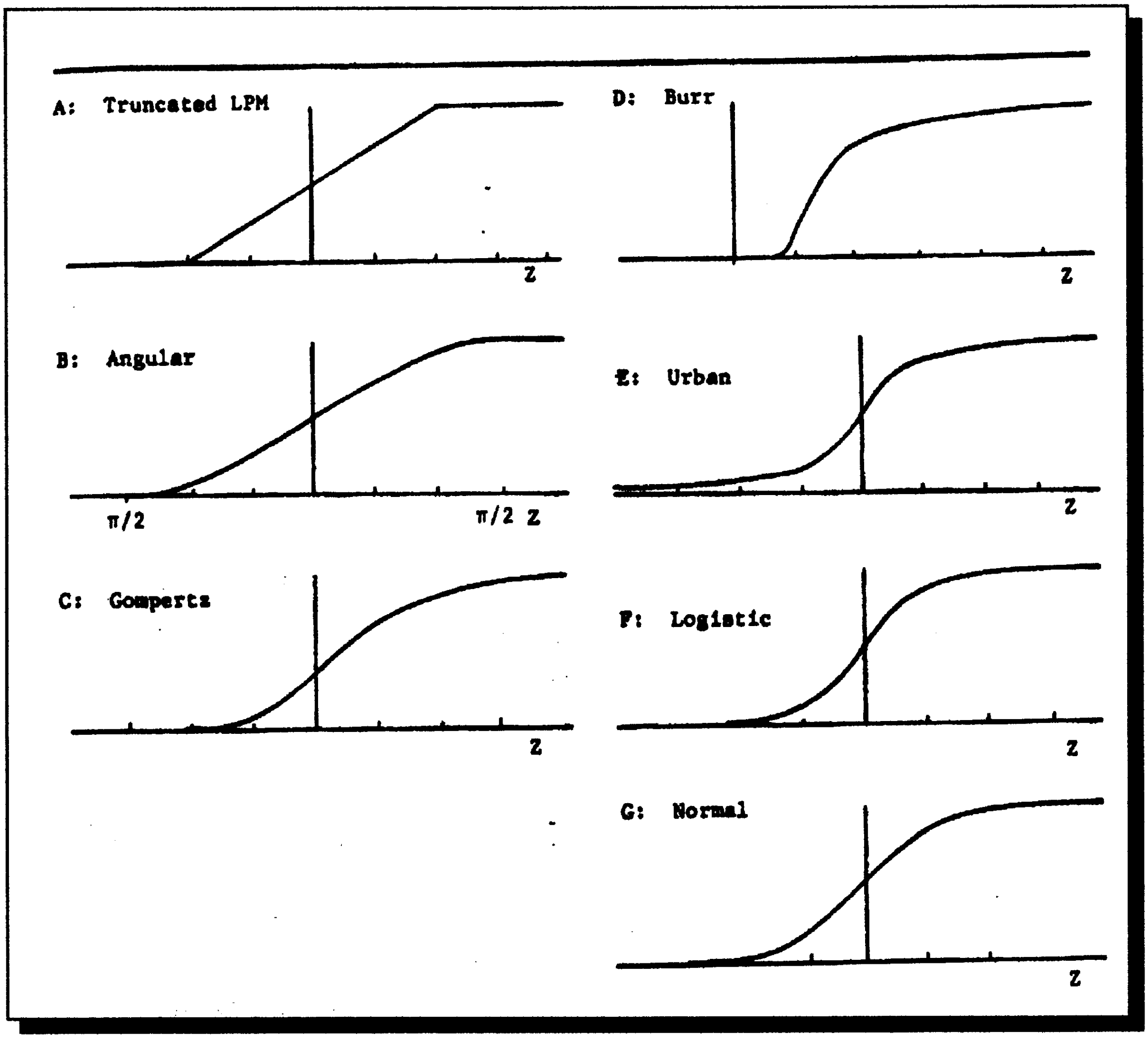


Figure 3.1 Graphs of alternative specifications.
Source: Aldrich, John H. and Forrest D. Nelson, *Linear Probability, Logit, and Probit Models*, Series; Quantitative Applications in the Social Sciences, Sage Publications, Inc., Newbury Park, California, 1984.

These transformations differ in their symmetry about the origin, their constraints on Z and the thickness of their tails. The logistic curve is the transformation used for logit models as defined by Equation 3.7. The transformation which is based on the assumption of a randomly distributed error term, is given as follows.

$$F(Z) = \frac{e^Z}{1+e^Z} \tag{3.7}$$

The standard normal distribution function curve is used for the probit model. It is based on the assumption of a normally distributed interaction between response and explanatory variables. The logistic curve and normal curves are so similar that with few exceptions they produce nearly identical results as defined by Equation 3.8.

$$F(Z) = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right) du \equiv \Phi(Z) \tag{3.8}$$

The Gompertz Curve is used for log log models. Unlike the logit and probit transformations which are symmetrical around p=0.50, the complementary log-log transformation as defined by Equation 3.9 is asymmetrical which is closely related to continuous time models for the occurrence of events (Allison, 2001).

$$\text{Log}[-\text{Log}(1\text{-}Y)] = Z \tag{3.9}$$

Of these three categorical model choices, the logistic approach is probably best suited for the bicycle routes safety rating model. Considering the unexplained elements that cannot be predicted based on the explanatory variables included in the model, there is no guarantee that these unexplained elements are normally distributed (Train, 2002). The unexplained element may be due to differences in police officers' assessments of the injury and accuracy in recording the specific attributes of the location, grade, surface condition, etc. The injury severity may be affected by the strength of the individual, the resiliency of his bones and how he fell. Although there are many variables such as operating speed or volume that may be normally distributed, there is no injury that does not have some random element associated with it. Moreover, because these models are a function of the probability of the event and not its direct outcome, there is no error term to evaluate the form of its distribution. Therefore, since there is no strong reason to use an alternative model and because the logit model is attractive for many practical reasons, such as the availability and flexibility of software, logit was chosen as the categorical method to develop the bicycle route safety rating model.

3.4 Logit

The logistic form as shown in Figure 3.2 has been historically popular because it is well developed and mathematically easy to define and interpret. Irrespective of the value of Z, the logistic f(z) ranges between 0 and 1 which corresponds to probability functions. Another reason for its popularity is the S-shape of its curve which defines upper bound and lower bound thresholds for a response range. This shape corresponds with many commonly found probability functions which produce flat slopes with relatively small responses at extreme levels and steep slopes with large responses when the independent variable is close to the origin.

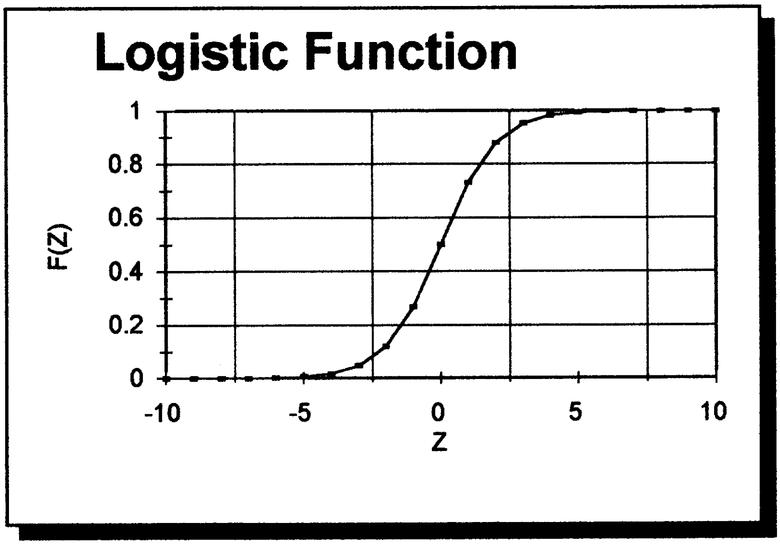


Figure 3.2 Logistic curve.

The response variable Y can then be transformed using a logistic function of the form:

$$Y = \frac{1}{\left(1 + e^{-Z}\right)} \tag{3.10}$$

Where:

$$Z = \alpha + \beta X_k \tag{3.11}$$

Substituting Equation 3.10 into Equation 3.11 yields:

$$Y = \frac{1}{1 + e^{-\left(\alpha + \beta_k X_k\right)}} \tag{3.12}$$

The ratio of the likelihood of an event occurring to the likelihood of not occurring is known as the odds ratio. The odds ratio relationship using a logistic transformation is defined as follows:

$$\left[\frac{Y}{1-Y}\right]=\left[\frac{\dfrac{1}{1+e^{-Z}}}{1-\dfrac{1}{1+e^{-Z}}}\right]=e^{Z} \tag{3.13}$$

Because exponential and fractional denominators are mathematically difficult to manipulate and fit observation data to, Equation 3.13 is transformed using natural logarithms to develop a linear relation. The resulting expression, Equation 3.14, is termed the log odds ratio or logit of Y:

$$Z=\ln\left[\frac{Y}{1-Y}\right] \tag{3.14}$$

In a familiar context, at a race track a horse might be given odds of 3-1 implying that he has one chance of winning versus three chances of losing. In the event the horse won, a \$2 bet would pay \$8, three times the money bet plus the return of the initial bet. Thus the lower the odds ratio, the greater the likelihood of an event occurring and the lower the payout. While a high odds ratios is akin to a long shot justifying a higher return on investment. The logit form is generally chosen because it linearizes a function, Equation 3.14, that would otherwise be difficult, if not impossible to fit. The fact that the logit has a common language interpretation is a bonus in that it converts a complex mathematical function into easy to conceive, familiar terminology.

To obtain a logistic model from a set of observations, the dependent variable P(Y), or for simplicity in notation, Y, is first transformed into Z using Equation 3.14. After a model is obtained using the transformed variable Z to fit the set of independent variables X using Equation 3.11, Z can be used to predict an outcome Y by reversing the transformation using Equation 3.15 as given below:

$$Y = \frac{e^Z}{1+e^Z} \tag{3.15}$$

3.5 Maximum Likelihood Estimation (MLE) Method

Unlike continuous functions, the variables α and β for a transformed categorical function cannot be obtained by fitting a line through a sample of data points by minimizing the square of the error. Instead, the MLE method is employed by fitting the logit model by maximizing the likelihood that the predicted probability of the event matches the observed probability (Hosmer and Lemeshow, 2000). Mathematically, this likelihood equation expressed for a sample of n pairs of data points is:

$$\ell(\beta) = \prod_{i=1}^{n} P(X_i)^{Y_i} \left[1 - P(X_i)\right]^{1-Y_i} \tag{3.16}$$

As with the initial logistic transformation, it is easier to work with the sum of a series instead of a product. Again, this is accomplished through the application of natural logarithms to Equation 3.16 and restating it as follows:

$$\ln\left[\ell(\beta)\right] = \sum_{i=1}^{n} \{Y_i \ln P(X_i) + (1 - Y_i)\ln\left[1 - P(X_i)\right]\} \tag{3.17}$$

Partial derivatives of the log likelihood function as given in Equation 3.17, are taken with respect to the β coefficients. The resulting equations shown below are set to 0 to obtain the values of the β coefficients which will maximize Equation 3.17 as shown below:

$$\sum \left[Y_i - P(X_i)\right] = 0 \tag{3.18}$$

$$\sum X_i\left[Y_i - P(X_i)\right] = 0 \tag{3.19}$$

Because these equations are non-linear with respect to X, they cannot be solved using standard linear algebra. Computer programs have been developed to solve them using computational algorithms. These equations, if successfully solved, produce solutions that converge after a number of iterations. The optimum β_k will achieve the maximum estimation of the observed probabilities.

3.6 Polytomous Models

The previous section presented the Maximum Likelihood Equation (MLE) for the case of a dichotomous or Bernoulli variable, i.e. Yes or No, 0 or 1, dependent variable. A further distinction must be made between nominal and ordinal polytomous variables. Nominal variables such as transportation choice, i.e. bus, car, train, walk, have no fixed order or magnitude. Ordinal polytomous variables, such as in the current case where injury levels are expressed in terms of increased severity, do have a relationship between each other. Unlike nominal variables, cumulative probabilities may be derived for ordinal variables. For example a cumulative probability can be established for a levels of severity less than or equal to a Level 3. Later in this section, the distinction between ordinal and nominal variables will be useful in deriving the MLE equations for polytomous, ordinal response variables.

As discussed earlier, the response variable based on the severity index, has more than two levels. Such categorical models are termed polytomous with J different levels of Y_i. The simplest approach to this problem is to set each j level as a separate dichotomous nonlinear probability model. For each level of j, the question simplifies to what is the probability that the observation, based on the dependent variables will assume that value of j. The following equation defines the probability that Y is equal to a specific level, in this case j = 1. Similar equations could be created for other levels of j = 2 and 3 as follows:

$$P(Y_i = j) = P(Y_{ij} = 1) = f\left(\sum^k \beta_{jk} X_{jk}\right) \tag{3.20}$$

The MLE method could be used to solve this series of j equations to obtain k coefficients for each level of j. However, there is no guarantee that if the probabilities for each level were summed, the total would equal 1. An alternative approach that does satisfy this requirement is similar to the log odds ratio used for the dichotomous problem. In this case, one of the dependent variable levels, say J, is used for the denominator and each probability is transformed into the logit model. For j = 1, 2, 3, . . ., J, there are a total of J-1 ratios of the form:

$$\frac{P(Y = j)}{P(Y = J)} = \exp(Z_j) \tag{3.21}$$

Multiplying both sides of the equation by P(Y=J) yields the following expression for P(Y=j):

$$P(Y = j) = \exp(Z_j) \cdot P(Y = J) \tag{3.22}$$

A value for P(Y=J) can be obtained by requiring that the summation of the probabilities of each of the response variables must equal 1.0. While this is never a problem with ordinal variables which imply a variable for every level of response, it can pose a problem for nominal response variables. For example, should a class be queried as to their favorite ice cream flavor, Vanilla, Chocolate or Strawberry, a student who preferred Coffee would be omitted from the count. This problem is corrected by the addition of a final category of "None of the Above" thereby capturing all of the respondents in the count. Thus, Equation 3.23 can be constructed as a summation of all probabilities of all J possible responses as follows:

$$P(Y=1)+P(Y=2)+\ldots+P(Y=J-1)+P(Y=J)=1 \tag{3.23}$$

Substituting Equation 3.22 into 3.23 yields:

$$\begin{aligned}\exp(Z_1)P(Y=J)+\exp(Z_2)P(Y=J)+..\\ +\exp(Z_{J-1})P(Y=J)+P(Y=J)=1\end{aligned} \tag{3.24}$$

Collecting the terms P(Y=J) in Equation 3.24 yields:

$$P(Y=J)\cdot\left[\exp(Z_1)+\exp(Z_2)+\ldots+\exp(Z_{J-1})+1\right]=1 \tag{3.25}$$

Summing these terms and rearranging produces an expression for P(Y=J) as in the following expression:

$$P(Y=J)=\frac{1}{\left[\sum_{j=1}^{J-1}\exp(Z_j)+1\right]} \tag{3.26}$$

When j is ordinal, there is a fixed order of the polytomous response variables and a threshold μ_j associated with moving from one level of severity to the next. For convenience μ_1 is set to 0, where μ_j is equal to the cumulative probability for each response variable level which adheres to the following relationship:

$$0 < \mu_2 < \mu_3 < \ldots < \mu_{j-1} < 1 \tag{3.27}$$

It is accepted that the curve for each level of response variable is identical but offset by a different intercept. The odds ratio for an ordinal model addresses this cumulative relationship in the following manner:

$$\log\left[\frac{P(y \le j|X)}{1 - P(y \le j|X)}\right] = \mu_j - \sum_{k=1}^{K} \beta_k X_k \tag{3.28}$$

$$P(y \le j) = P(y^* \le \mu_j) = \frac{e^{\mu_j - \sum_{k=1}^{K} \beta_k X_k}}{1 + e^{\mu_j - \sum_{k=1}^{K} \beta_k X_k}} \tag{3.29}$$

Equation 3.29 will estimate the cumulative probability for a given response level which includes the probability of all the lower events. To compute the probability for a specific response level, it is necessary to subtract the cumulative probability of all lower levels. Thus Equation 3.30 estimates the probability that event j will occur as the difference between the cumulative probabilities of j and j -1 or:

$$P(y = j) = P(y \le j) - P(y \le j - 1) \tag{3.30}$$

Figure 3.3, excerpted from the *JMP Start Statistics* © (1996) depicts this ordinal relationship with each curve for a given event level identical in shape to all other event levels, but shifted along the x axis.

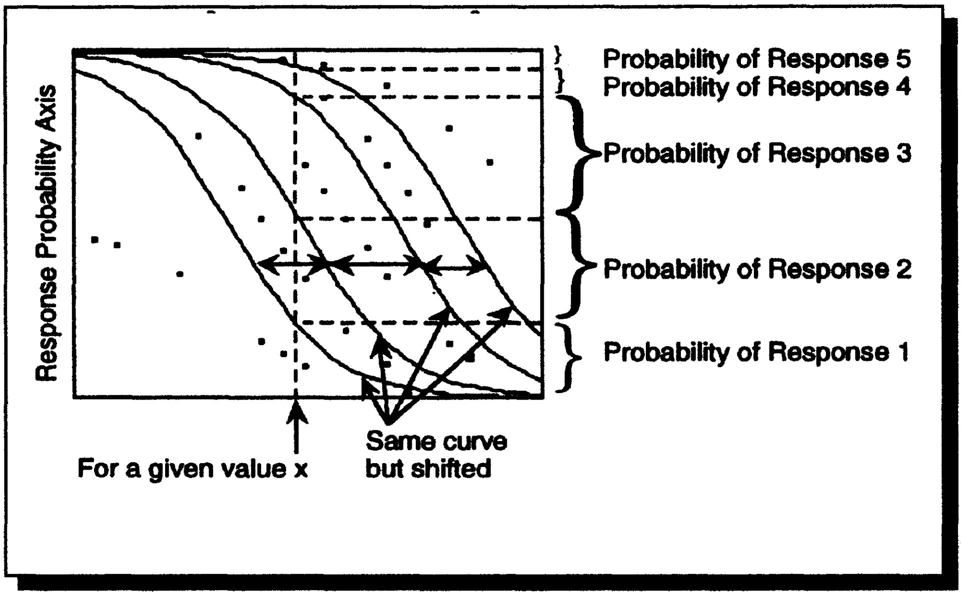


Figure 3.3 Ordinal Logistic Regression Cumulative Probability Plot.
Source: Sall, John and Ann Lehman, SAS Institute, *JMP Start Statistics*, Duxbury Press, 1996.

The algebraic expressions for the shifting of these cumulative equations and the computation of the delta probabilities are expressed by Equations 3.31 through 3.34 as follows:

$$\mathrm{Prob}(y=1)=f\left(-\sum_{k=1}^{K}\beta_k X_k\right) \tag{3.31}$$

$$\mathrm{Prob}(y=2)=f\left(\mu_2-\sum_{k=1}^{K}\beta_k X_k\right)-f\left(-\sum_{k=1}^{K}\beta_k X_k\right) \tag{3.32}$$

$$\mathrm{Prob}(y=3)=f\left(\mu_3-\sum_{k=1}^{K}\beta_k X_k\right)-f\left(\mu_2-\sum_{k=1}^{K}\beta_k X_k\right) \tag{3.33}$$

$$\mathrm{Prob}(y=J)=1-f\left(\mu_{J-1}-\sum_{k=1}^{K}\beta_k X_k\right) \tag{3.34}$$

3.7 Statistical Inference

As with all models, a categorical model must be assessed in terms of its statistical quality. Simply expressed, how good is the model? How well does the model predict the observed responses? Does the model maximize the number of hits, the events correctly predicted, and minimize the number of misses, the number of incorrect predictions? Is the model efficient? Were the right factors chosen? Are important factors missing? The extra time and cost to accumulate the data for a large number of independent variables may not be justified if a comparable level of accuracy can be attained with a smaller subset.

3.7.1 Level of Confidence

First, it is important to establish an appropriate level of confidence. The establishment of an appropriate confidence level is a factor of both the model's input: the quality of the data upon which the predictions will be made and the need for accuracy. Certainly the level of impurities in infant formula or the expected failure rate of space shuttle o-rings require a higher level of confidence than the predicted pounds of turkeys Americans will consume next Thanksgiving. Although, the life and safety of the bicyclist is of great importance, it is the quality of the data which necessitates a lower confidence interval. Traffic volume measurements vary widely over proximity to the accident site and date taken. Even uniform pavement management systems, if they are available in a community, do not produce ratings precise to the hundredth of a decimal point. Given the random nature of the bicyclist's behavior and the quality of the data itself (police reports taken at the accident scene under stressful conditions), a 99% level of confidence is unwarranted. Moreover, restricting the explanatory variables to meet standards that are too rigid, will force the elimination of factors that might have made a contribution to the overall model.

On the other hand, setting the confidence level too low, less than 90%, would not produce a model with sufficiently reliable rating predictions on which to make route choice or investment decision. As a result of these considerations, the model developed in this research will have a confidence interval set at 90% with a p value of 0.1 for a one-tailed test. For comparison purposes, a second model will be presented that meets a 95% confidence limit.

Care must be exercised when ranking bicycle routes or making investment decision based on estimates obtained from a model developed with this moderate levels of precision. One route cannot be chosen over another because of the difference of a hundredth in predicted severity level. The precision of the severity index will be established to correspond to a given confidence level to provide users with the criteria to make choices.

3.7.2 Whole Model Test

The first step in evaluating a model is to assess its goodness-of-fit. In other words, based on the sample data, how well does this model predict the observed responses? Does the model make better predictions than a set of random occurrences, thereby justifying the rejection of the null hypothesis? If the data's explanatory variables, Xs, were plotted against the data's responses, Ys, and found to be perfectly horizontal, i.e. there is no better explanation of the response than the mean of the data, the null hypothesis which implies that that the model's explanatory variable have no (null) effect on the prediction of the response variable.

Had the model been a linear regression obtained using the OLS Method, the null hypothesis would have been evaluated by comparing the F statistic with the ratio of the model's mean square to the error mean square. This is the joint hypothesis that all the coefficients except the intercept are equal to 0. Since the MLE method is used instead of the OLS to obtain the variable coefficients for a logit model, the goodness-of-fit is evaluated using an alternative test statistic. The loglikelihood ratio G^2, a test statistic that is compared to a chi-square distribution, is computed as:

$$G^2 = -2\ln\left[\frac{(\text{likelihood of the fitted model})}{(\text{likelihood of the constrained model})}\right] \tag{3.35}$$

The likelihood of the fitted model is obtained by using all of the variables multiplied by each of the fitted coefficients. The constrained model sets each of the variables equal to 0 and calculates the likelihood as the result of the intercept alone, in essence a mean of the data. The ratio between these two numbers is referred in the literature as the likelihood ratio. When this ratio is very large such that twice the natural log of its value multiplied by -2 is significantly greater than the χ^2 for K-1 levels of freedom, the model can be accepted as explaining a statistically greater part of the data's variance than a simple mean. As discussed in Section 3.6.1, a p value of 0.1 would indicate a meaningful model for the purposes of predicting accident severity.

3.7.3 Coefficient of Determination

Evaluations of linear regression models place great importance on the value of the coefficient of determination, R^2, the square of the model's variance divided by the square of the total variance. Simply put, R^2 represents the portion of the data variance which is explained by the variables. A high R^2 demonstrates that the model is explaining a large part of the variance. The following expression (Vining, 1998) defines R^2 as the ratio of the error that is explained by the model divided by the total error:

$$R^2 = \frac{SS_{reg}}{SS_{total}} = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2} \tag{3.36}$$

The same computation cannot be made for categorical models. Categorical models are developed not by minimizing variance, but by maximizing the probability or likelihood of achieving a correct response. A corresponding measure which may be used in lieu of R^2 (Allison, 2001) might be constructed for categorical models as follows:

$$R^2 = \frac{-\text{loglikelihood for Model}}{-\text{log likelihood for Constrained Model}} \tag{3.37}$$

In categorical models, high R^2s are rare . The ratios are very sensitive to the magnitude of the intercept. The lower the intercept, the lower the ratio. For most purposes, the R^2 for categorical models is not used.

3.7.4 Coefficient Estimates

After passing the whole model test and lack of fit test affirming that the model contains all of the variables and interactions required to produce meaningful estimates, the next question that must be addressed is whether extraneous variables have also been included. A Wald chi-square test should be performed on each coefficient to determine if it is significant and what would be its corresponding confidence interval. The test statistic should be calculated as follows:

$$t = \left(\frac{\hat{\beta}}{s_{\hat{\beta}}}\right)^2 \tag{3.38}$$

Where:

$\hat{\beta}$ is the estimated coefficient
$s_{\hat{\beta}}$ is the standard error of the coefficent

The Wald statistic is compared to a χ^2. If p <0.1 then the null hypothesis, which states that the model's predictions based on that variable are no better than random observations, could not be rejected. Coefficients which meet the Wald χ^2 test will be retained in the model, otherwise, the variable is eliminated.

Confidence intervals for categorical models may be computed as in linear regression models and inspected to verify that the range does not include the origin, i.e. horizontal or 0 slope. Flat slopes imply a lack of significance between the relationship between the response variables and the explanatory variable.

3.8 Logistic Model Interpretation

After a model has met the goodness-of-fit criteria and all insignificant variables have been eliminated, the model can be interpreted by performing the following: 1. Examining the coefficients' signs; 2. Determining the marginal effects of the model's variables; and 3. Predicting the probability of the response variable given a set of explanatory variables. A detailed discussion of these three model interpretation steps follows.

3.8.1 Coefficient Sign

Through examination of the sign of the coefficient, much can be learned about the impact of a variable to the response of the model. In linear regression, a positive coefficient implies a positive relationship. For example, a positive coefficient for an explanatory variable in an accident prediction model such as traffic volume results in increasing relationship between accident severity and volume, while a negative coefficient results in a decreasing relationship.

Linear probability models such as the logistic model cannot have negative values. The expected severity Y, as defined in Equation 3.10, ranges between 0 and 1. As Z becomes increasingly negative, the denominator of this equation increases forcing Y to 0. A positive Z reduces the denominator and increases the probability of the function as Y approaches 1. In this manner, coefficients can be evaluated based on their effect on Z.

3.8.2 Marginal Effects

Typically, one examines the magnitude of the coefficients of a model to determine the marginal effect of that particular explanatory variable. However, unlike the coefficients of a linear regression model, the β_k coefficients of a logistic model do not represent a constant effect. Given that the predicted probability of a logistic model is based on the logit transformation and that the very slope of its S shaped curve is a function of Z, the marginal effect of a variable can only be assessed by taking the partial derivative of the logistic model as follows:

$$\begin{aligned}\frac{dP(Y=1)}{dX_k} &= \frac{\exp(Z)}{1+\exp(Z)} \cdot \frac{1}{1+\exp(Z)} \beta_k \\ &= [Y(1-Y)]\beta_k\end{aligned} \tag{3.39}$$

Thus the impact on Y of a change in an explanatory variable X_K is a function of its log odds ratio and coefficient $\beta_{K,}$ and is not completely determined by β_K. Thus the impact of a single variable on the response variable varies with the probability of Y. The variable may have a greater or lesser impact at different levels of severity.

3.8.3 Predicted Response Variable Probability

The logit model generates a set of coefficients for each of the polytomous levels of the response variable. Applying these formulae to a given set of explanatory variables will produce a probability for each j level where the levels for accident severity include 1, 2 or 3. Since the sum of the probabilities for all possible outcomes must equal 1.0, it is important that a given set of k explanatory values produce j equations that when summed together also equal 1.0. It is also important that the formulae make reasonable predictions across the full cumulative range of probabilities given a fixed set of explanatory variables. For example, it would be difficult to accept a model that responded to the steady increase of one or more of its factors by vacillating between increasing and decreasing levels of severity. For example, as speed limit increased by 10 mph, we would expect to see severity levels increase from 1 to 3. If instead, such a change in speed limit resulted in a predicted severity level of 3 at 30 mph, a decrease to a severity level of 1 at 40 mph and an increase to a severity level of 3 at 50 mph, this erratic behavior would call to question the entire validity of the model. This problem is avoided by using the parallel curves shifted by the intercepts generated by the ordinal logit procedure.

3.9 Summary

After establishing the difference between continuous and categorical response variables, this chapter has explained why categorical models cannot be developed using standard OLS methods. The maximum likelihood equation (MLE) method was explored as an alternate approach to fit the model based on a logistic transformation of a probability distribution model. Then, the dichotomous solution was expanded to a polytomous form to accommodate the three levels of severity. Chi-square tests were presented as a means to evaluate the significance of the goodness-of-fit for the whole model using the loglikelihood ratio and the significance of the beta estimates using the Wald statistic. Finally, coefficient signs and estimates were used to interpret the model and the variable's marginal effects. With these statistical tools, it will be possible to build a model using the data collected in the following chapter.

In concluding this discussion of the technical approach, it is necessary to stress the need to exercise caution in applying this or any mathematically derived model. Model building is both a science and an art. It demands an intuitive understanding of the underlying nature of the variables, how they will promote certain accident typologies, how they may interact with other variables and ultimately, in what way they will impact the safety of the route. The variables considered in this model building effort are real characteristics at real locations where real people were injured. These variables are not abstract numbers. Notwithstanding the sophisticated techniques employed to develop the model and irrespective of the quality of the statistical measures, any model can and should be rejected if its application does not produce rational results.

CHAPTER 4

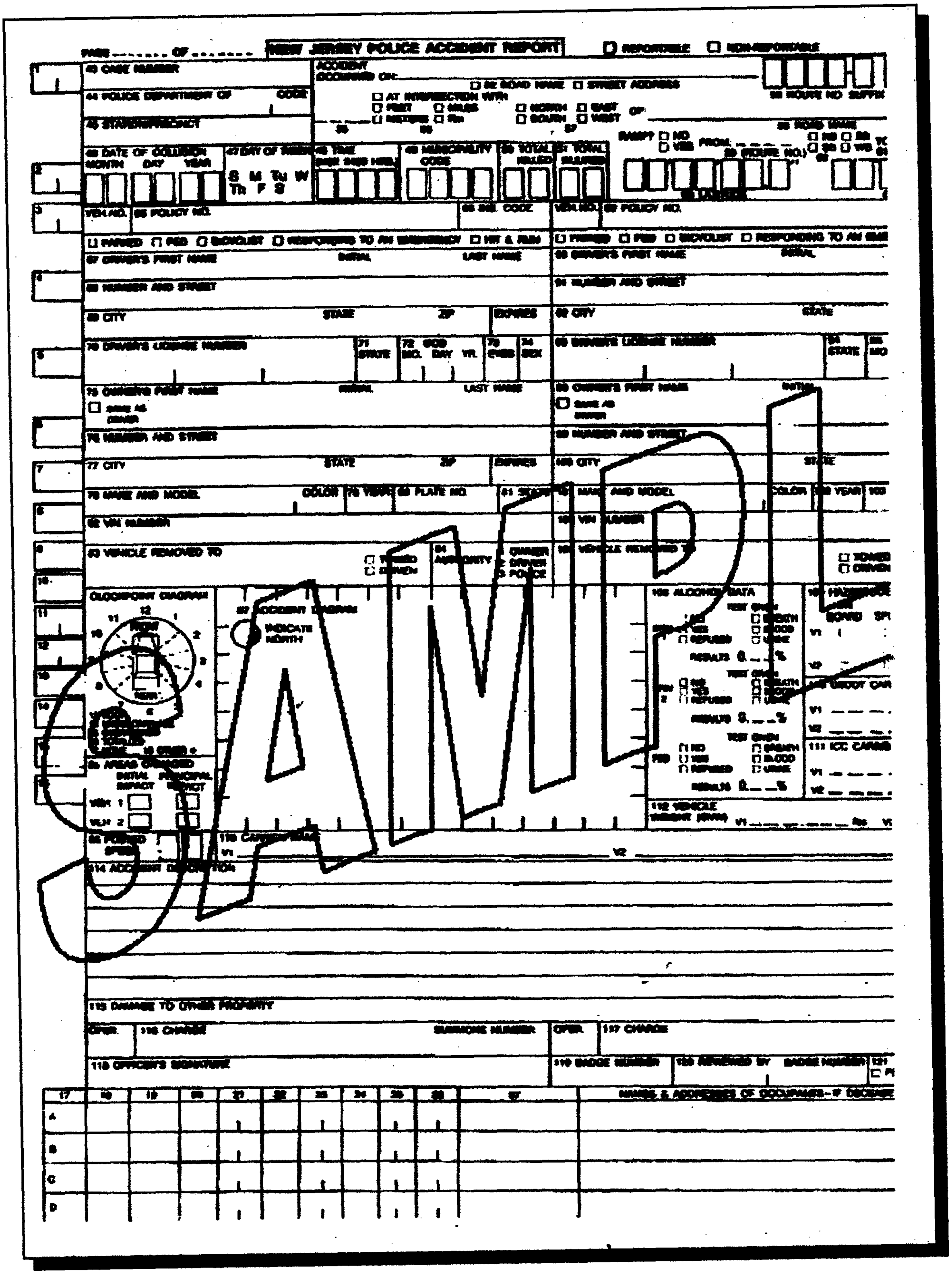
DATA COLLECTION

4.1 Data Acquisition

The bicycle route safety rating model was developed using Jersey City, New Jersey as the study area. This selection was based on a number of factors. Its population of 225,000 is sufficiently large to be classified as a Metropolitan Planning area. The municipality has identified bicycling as a viable travel mode choice and intends to include it in their current master planning effort. Finally, and most importantly, the Jersey City Engineering Division and Police Department were generous in sharing their data resources.

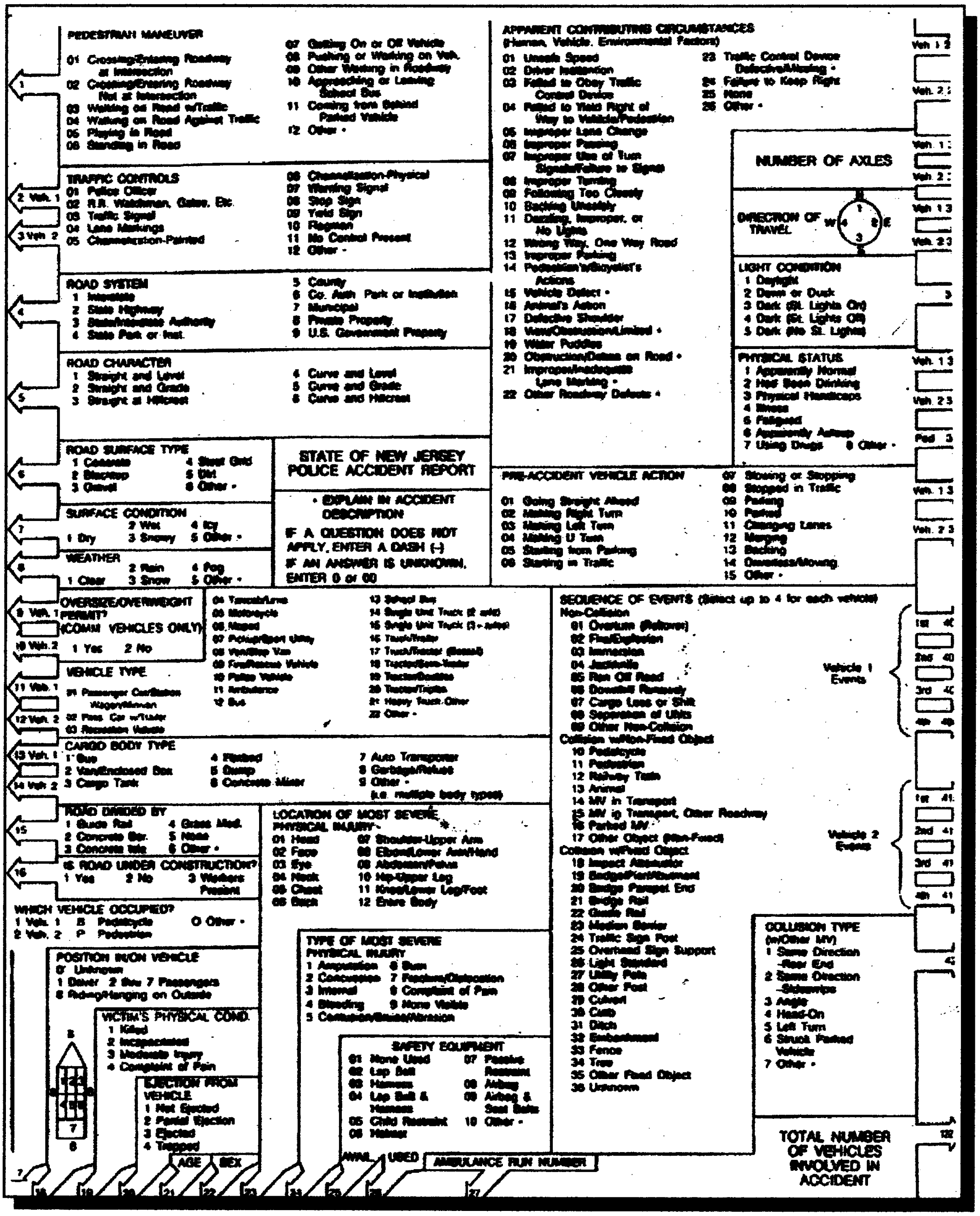
The New Jersey Department of Transportation (NJDOT) accident database was used as the source of accident records. This database does not contain all accidents, but is limited to bicycle accidents which involve motor vehicles. This limitation initially appears to be serious in that the majority of bicycle crashes do not involve motor vehicles. Bicycle / Motor vehicle collisions are typically less than 20% of all crashes (Forester, 1983). However, Tinsworth (1993) found that while most of the 500,000 bicycle-related emergency room visits per year do not involve motor vehicles, 90% of fatalities do involve motor vehicles. Thus, a study based on the State's motor vehicle accident will produce the most serious injuries.

The NJDOT database publishes its crash data on the Internet on an annual, county by county basis for the period 1997-2000. Each county file contains over 20,000 events with the information derived from standard police accident forms, see Figure 4.1.


NEW JERSEY POLICE ACCIDENT REPORT

SAMPLE

Figure 4.1 New Jersey State Police Report Form.


STATE OF NEW JERSEY
POLICE ACCIDENT REPORT

PEDESTRIAN MANEUVER
01 Crossing/Entering Roadway at Intersection
02 Crossing/Entering Roadway Not at Intersection
03 Walking on Road w/Traffic
04 Walking on Road Against Traffic
05 Playing in Road
06 Standing in Road
07 Getting On or Off Vehicle
08 Pushing or Working on Veh.
09 Other Working in Roadway
10 Approaching or Leaving School Bus
11 Coming from Behind Parked Vehicle
12 Other *

TRAFFIC CONTROLS
01 Police Officer
02 R.R. Watchman, Gates, Etc.
03 Traffic Signal
04 Lane Markings
05 Channelization-Painted
06 Channelization-Physical
07 Warning Signal
08 Stop Sign
09 Yield Sign
10 Flagman
11 No Control Present
12 Other *

ROAD SYSTEM
1 Interstate
2 State Highway
3 State/Interstate Authority
4 State Park or Inst.
5 County
6 Co. Auth. Park or Institution
7 Municipal
8 Private Property
9 U.S. Government Property

ROAD CHARACTER
1 Straight and Level
2 Straight and Grade
3 Straight at Hillcrest
4 Curve and Level
5 Curve and Grade
6 Curve and Hillcrest

ROAD SURFACE TYPE
1 Concrete
2 Blacktop
3 Gravel
4 Steel Grid
5 Dirt
6 Other *

SURFACE CONDITION
1 Dry
2 Wet
3 Snowy
4 Icy
5 Other *

WEATHER
1 Clear
2 Rain
3 Snow
4 Fog
5 Other *

* EXPLAIN IN ACCIDENT DESCRIPTION
IF A QUESTION DOES NOT APPLY, ENTER A DASH (-)
IF AN ANSWER IS UNKNOWN, ENTER 0 or 00

OVERSIZE/OVERWEIGHT PERMIT? (COMM VEHICLES ONLY)
1 Yes 2 No

VEHICLE TYPE
01 Passenger Car/Station Wagon/Minivan
02 Pass. Car w/Trailer
03 Recreation Vehicle
04 Taxicab/Limo
05 Motorcycle
06 Moped
07 Pickup/Sport Utility
08 Van/Step Van
09 Fire/Rescue Vehicle
10 Police Vehicle
11 Ambulance
12 Bus
13 School Bus
14 Single Unit Truck (2 axle)
15 Single Unit Truck (3+ axle)
16 Truck/Trailer
17 Truck/Tractor (Bobtail)
18 Tractor/Semi-Trailer
19 Tractor/Doubles
20 Tractor/Triples
21 Heavy Truck-Other
22 Other *

CARGO BODY TYPE
1 Bus
2 Van/Enclosed Box
3 Cargo Tank
4 Flatbed
5 Dump
6 Concrete Mixer
7 Auto Transporter
8 Garbage/Refuse
9 Other *
(i.e. multiple body types)

ROAD DIVIDED BY
1 Guide Rail
2 Concrete Bar.
3 Concrete Isle
4 Grass Med.
5 None
6 Other *

IS ROAD UNDER CONSTRUCTION?
1 Yes 2 No 3 Workers Present

WHICH VEHICLE OCCUPIED?
1 Veh. 1
2 Veh. 2
B Pedalcycle
P Pedestrian
O Other *

POSITION IN/ON VEHICLE
0 Unknown
1 Driver
2 thru 7 Passengers
8 Riding/Hanging on Outside

VICTIM'S PHYSICAL COND.
1 Killed
2 Incapacitated
3 Moderate Injury
4 Complaint of Pain

EJECTION FROM VEHICLE
1 Not Ejected
2 Partial Ejection
3 Ejected
4 Trapped

AGE SEX

LOCATION OF MOST SEVERE PHYSICAL INJURY
01 Head
02 Face
03 Eye
04 Neck
05 Chest
06 Back
07 Shoulder-Upper Arm
08 Elbow/Lower Arm/Hand
09 Abdomen/Pelvis
10 Hip-Upper Leg
11 Knee/Lower Leg/Foot
12 Entire Body

TYPE OF MOST SEVERE PHYSICAL INJURY
1 Amputation
2 Concussion
3 Internal
4 Bleeding
5 Contusion/Bruise/Abrasion
6 Burn
7 Fracture/Dislocation
8 Complaint of Pain
9 None Visible

SAFETY EQUIPMENT
01 None Used
02 Lap Belt
03 Harness
04 Lap Belt & Harness
05 Child Restraint
06 Helmet
07 Passive Restraint
08 Airbag
09 Airbag & Seat Belts
10 Other *

AVAIL. USED AMBULANCE RUN NUMBER

APPARENT CONTRIBUTING CIRCUMSTANCES (Human, Vehicle, Environmental Factors)
01 Unsafe Speed
02 Driver Inattention
03 Failed to Obey Traffic Control Device
04 Failed to Yield Right of Way to Vehicle/Pedestrian
05 Improper Lane Change
06 Improper Passing
07 Improper Use of Turn Signal/Failure to Signal
08 Improper Turning
09 Following Too Closely
10 Backing Unsafely
11 Dazzling, Improper, or No Lights
12 Wrong Way, One Way Road
13 Improper Parking
14 Pedestrian's/Bicyclist's Actions
15 Vehicle Defect *
16 Animal's Action
17 Defective Shoulder
18 View/Obstruction/Limited *
19 Water Puddles
20 Obstruction/Debris on Road *
21 Improper/Inadequate Lane Marking *
22 Other Roadway Defects *
23 Traffic Control Device Defective/Missing *
24 Failure to Keep Right
25 None
26 Other *

NUMBER OF AXLES

DIRECTION OF TRAVEL

LIGHT CONDITION
1 Daylight
2 Dawn or Dusk
3 Dark (St. Lights On)
4 Dark (St. Lights Off)
5 Dark (No St. Lights)

PHYSICAL STATUS
1 Apparently Normal
2 Had Been Drinking
3 Physical Handicaps
4 Illness
5 Fatigued
6 Apparently Asleep
7 Using Drugs
8 Other *

PRE-ACCIDENT VEHICLE ACTION
01 Going Straight Ahead
02 Making Right Turn
03 Making Left Turn
04 Making U Turn
05 Starting from Parking
06 Starting in Traffic
07 Slowing or Stopping
08 Stopped in Traffic
09 Parking
10 Parked
11 Changing Lanes
12 Merging
13 Backing
14 Driverless/Moving
15 Other *

SEQUENCE OF EVENTS (Select up to 4 for each vehicle)
Non-Collision
01 Overturn (Rollover)
02 Fire/Explosion
03 Immersion
04 Jackknife
05 Ran Off Road
06 Downhill Runaway
07 Cargo Loss or Shift
08 Separation of Units
09 Other Non-Collision
Collision w/Non-Fixed Object
10 Pedalcycle
11 Pedestrian
12 Railway Train
13 Animal
14 MV in Transport
15 MV in Transport, Other Roadway
16 Parked MV
17 Other Object (Non-Fixed)
Collision w/Fixed Object
18 Impact Attenuator
19 Bridge/Pier/Abutment
20 Bridge Parapet End
21 Bridge Rail
22 Guide Rail
23 Median Barrier
24 Traffic Sign Post
25 Overhead Sign Support
26 Light Standard
27 Utility Pole
28 Other Post
29 Culvert
30 Curb
31 Ditch
32 Embankment
33 Fence
34 Tree
35 Other Fixed Object
36 Unknown

Vehicle 1 Events
Vehicle 2 Events
1st 2nd 3rd 4th

COLLISION TYPE (w/Other MV)
1 Same Direction -Rear End
2 Same Direction -Sideswipe
3 Angle
4 Head-On
5 Left Turn
6 Struck Parked Vehicle
7 Other *

TOTAL NUMBER OF VEHICLES INVOLVED IN ACCIDENT

Figure 4.2 New Jersey Police Accident Form Instructions.

The crash data contain over forty fields of information about conditions at the time of the accident, including existing roadway conditions as well as information about the driver and damage that occurred. Figure 4.2 contains the instructions for the accident report which serves as the metadata for the files which explains the descriptions of the different values the variables may take.

After first downloading these large county files for each of the four available years, the text files were imported to Excel as comma delimited files. Since the text files did not include column labels, a template was created using the metadata as a guide and pasted onto the first row of each year's accident. Next, Jersey City bicycle accidents were extracted from the files. Bicycle related accidents were identified from the vehicle occupant field where bicyclists are denoted as "B." The separate data files for each of the four years were combined into one large data file for model building. Standardized case numbers were created to keep each record unique.

For the four-year period 1997-2000, there were 97,310 crashes in Hudson County of which 36,623 or 37.6% motor vehicle accidents were reported in Jersey City. Of these records 328 or 0.9% were bicycle accidents. Thirteen of these accidents were eliminated due to insufficient data leaving a remaining 314 bicycle / motor vehicle accidents for study.

The raw data obtained from the NJDOT database has a number of limitations which must be addressed before model building can proceed. Some fields were either blank or assigned implausible values. Many fields had categories that are extraneous to this study, e.g. driver's license number. Other categories, such as alcohol testing, had data that were too

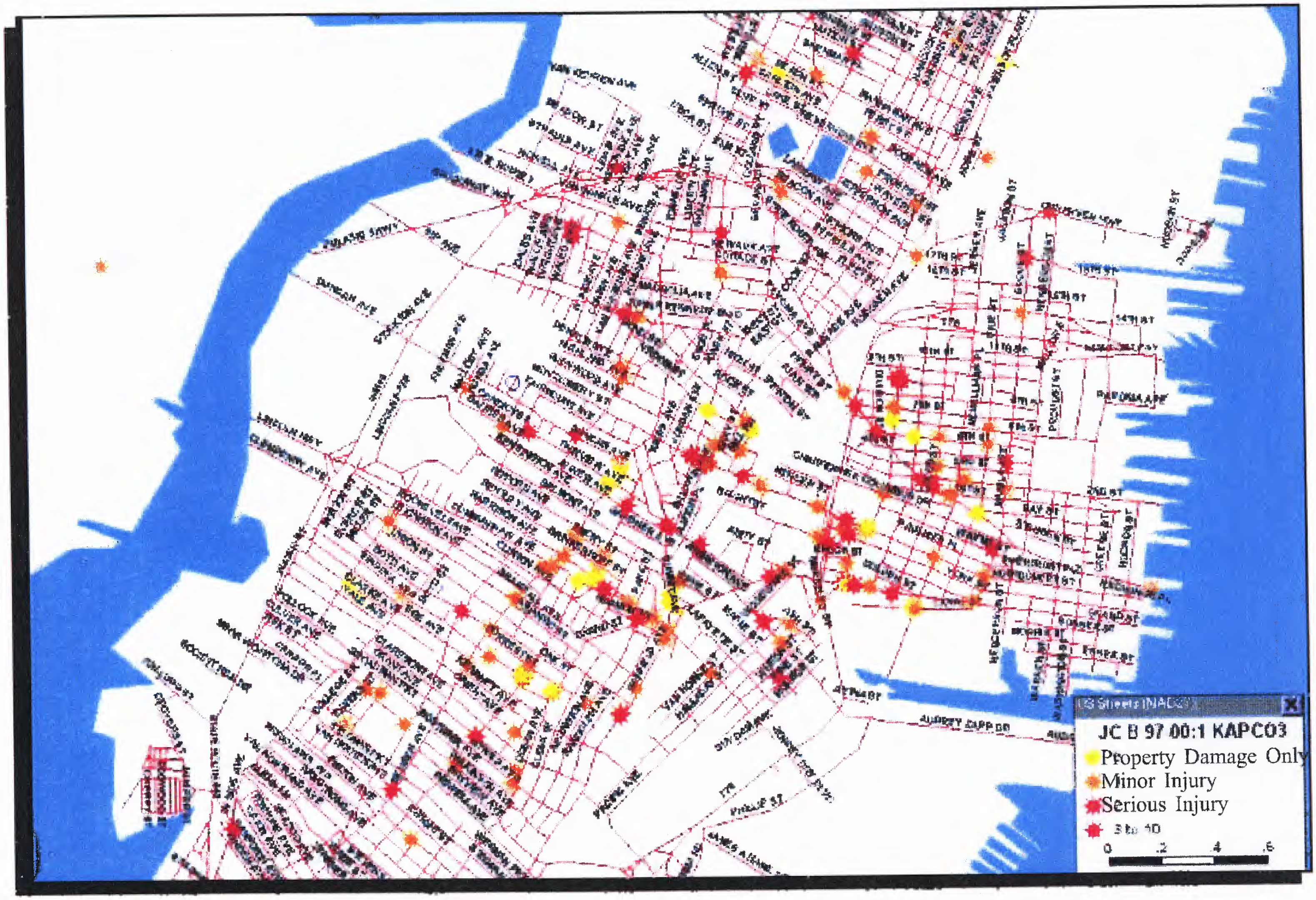


Figure 4.3 Jersey City bicycle accidents (1997-2000)

sparse to be useable. Not all desired study variables such as traffic volume and lane width were provided by the database. Because of these problems, the NJDOT database was enhanced with local information sources. Over 100 original police reports were obtained to search for missing data. Using the location and cross streets provided in the NJDOT database, crash locations were mapped on a Jersey City street map, Figure 4.3. Then data were obtained from local municipal codes, zoning maps, transit maps as well as Jersey City's internal records. In the following section describing the data, an explanation will be provided as to how the data were obtained and if necessary "cleaned up."

4.2 Data Analysis Techniques

In the following data analysis sections, the data is summarized and analyzed using data plots generated by JMP © software, a SAS © product. These plots provide a concise and visual summary of the data collected. Plots for categorical data are limited to frequency and cumulative probability distributions. Plots for continuous variables also provide quantiles, moments including mean and standard deviations and "Box and Whiskers" plots that visually display ranges and frequency concentrations.

Plots are provided of the interaction between injury severity and each of the explanatory variables. Categorical data is provided as a mosaic plot. The X-axis indicates the level of the categorical variable: yes (+1) or no (-1) if the variable is dichotomous and level equal 1, 2, 3 . . . n for polytomous variables with n levels. The frequency of the response variable is plotted for each level of injury severity, the explanatory variable. The right Y-axis shows the level of the response variable. The left Y-axis shows the cumulative probability distribution. Interaction diagrams are also provided for continuous explanatory variables. In lieu of mosaic plots, as cartesian coordinates are plotted for each of the continuous variables' data points. The curve is a plot of the logit of X for n-1 levels of Y, the cumulative probability distributions as a function of X.

Examination of the interaction plot for both categorical and continuous variables provide a visual display of the nature (or absence) of the relationship between the individual explanatory and response variables. Both categorical and continuous plots report chi-square tests of the intercepts and beta coefficients. Based on the results of these tests, a decision will be made as to whether the interaction is sufficiently strong to include the variable during the model building phase. Variables that do not meet the specified confidence levels when combined with the other retained model variables will ultimately be dropped.

4.3 Accident Severity

The NJDOT Accident data base provides an injury severity rating for each of the vehicle occupants including the bicyclists. The severity of the injury ranged from "No Injury," "Complaint of Pain," "Moderate Injury," "Incapacitated," to "Killed." Of the 314 bicycle accident records extracted from the NJDOT accident data base for Jersey City for the four-year period 1997 - 2000, no fatalities or incapacitating injuries were reported.

Injury severity reported in the NJDOT data base is derived from standard police reports. However, police officers do not have complete information and are not as skilled in assessing injury severity as hospital emergency department staff (Stutts, 1990). It is possible that a person removed from the scene by an ambulance may be released hours later after being treated for abrasions, while another person may walk away from the accident with a head injury and later die. There is also a great deal of variability between police officers' assessments of victims' physical conditions. The USDOT Crash Outcome Data Evaluation (CODES) could improve the accuracy of the reported accident severity by correlating motor vehicle accident report data with emergency room accident data. New Jersey, however, is not one of twenty-three designated CODES states.

Police accident reports provide greater detail than what is contained in accident data base records. To investigate the reliability of the reported accident severity, 149 original police reports were obtained from the Jersey City Police records by using the case number provided by the NJDOT database. The descriptive text from these accident reports provided an understanding of the sequence of events, the typology of the accident and the nature of the injuries. Comparisons of the NJDOT database records with the original police records confirmed that there was great variability in the individual policemen's application of the severity level. One policeman classified the injury severity of a victim who refused medical attention as serious while another policeman reported the injury severity of a child with a bleeding head who was taken from the scene by an ambulance as minor.

Table 4.1 Type of Most Severe Physical Injury

NJDOT Injury Class	Injury Severity Index	Description
0	1	Not known
1	3	Amputation
2	3	Concussion
3	3	Internal
4	3	Bleeding
5	3	Contusion/Bruise/Abrasion
6	3	Burn
7	3	Fracture/ Dislocation
8	2	Complaint of Pain
9	1	None visible.

Source: NJDOT accident database

To create a uniform and objective severity index, a new index was generated directly from the reported injury. The injury classifications are shown in Table 4.1. Supported by the police report detailed writeups, injury severity was derived from reported injury in the following manner. All injuries which were unknown or not visible were deemed to be property damage only (PDO) or Level 1. Those injuries that were not visible, although the victim complained of pain were deemed minor, Level 2. All other observed injuries were deemed serious, Level 3. Had there been fatalities, they would have been classified as Level 4. Since no fatalities were reported for bicycle accidents during this period, Level 4 was eliminated. Figure 4.4 shows that of all bicycle accident reported, 19% were PDO, 47% were Minor Injury and 34% were Serious Injury.

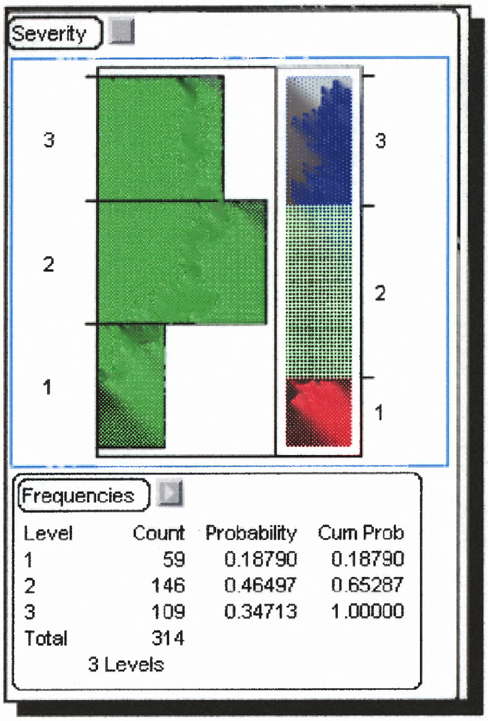


Figure 4.4 Severity Level Distribution.

4.4 Physical Factors

4.4.1 Lane Width

Lane width was not provided by the NJDOT database. Lane width was calculated by subtracting the amount of roadway dedicated to parking from the total width and dividing the difference by the number of through lanes. Parking lanes were assumed to be eight feet wide.

$$\text{Lane Width} = \frac{[\text{Carriage way} - 8 * (\text{Number Parking Lanes})]}{\text{Number of Travel Lanes}} \qquad (4.1)$$

The carriageway width was obtained from Jersey City historical records. This required looking up each roadway and locating the closest intersection to the accident. The number of lanes was determined by field observations. The number of parking lanes were obtained by looking up the section of roadway in the Jersey City Municipal Code Schedule 3, parking regulations.

Figure 4.5 shows the median lane width is 12 feet. Lane widths ranged between a minimum of 7' to a maximum of 30'. The constricted width of 7' occurred at a few locations where parking exists on both sides of an already narrow street two street. In reality, such narrow streets accommodate two-way traffic by using the numerous loading zones and driveways that are interspersed between the parking lanes. Passage of opposing traffic, especially for wide vehicles is accomplished by one vehicle pulling into vacant curb space to let the other pass.

The interaction plot as shown in Figure 4.6 shows a strong relationship between lane width and injury severity. When the street width is 7, there is an approximately 70% chance that the injury severity is a PDO or Minor Injury. As the width increases to 30, the probability of a PDO or Minor Injury drops below 40%. Width will be retained in the model building stage.

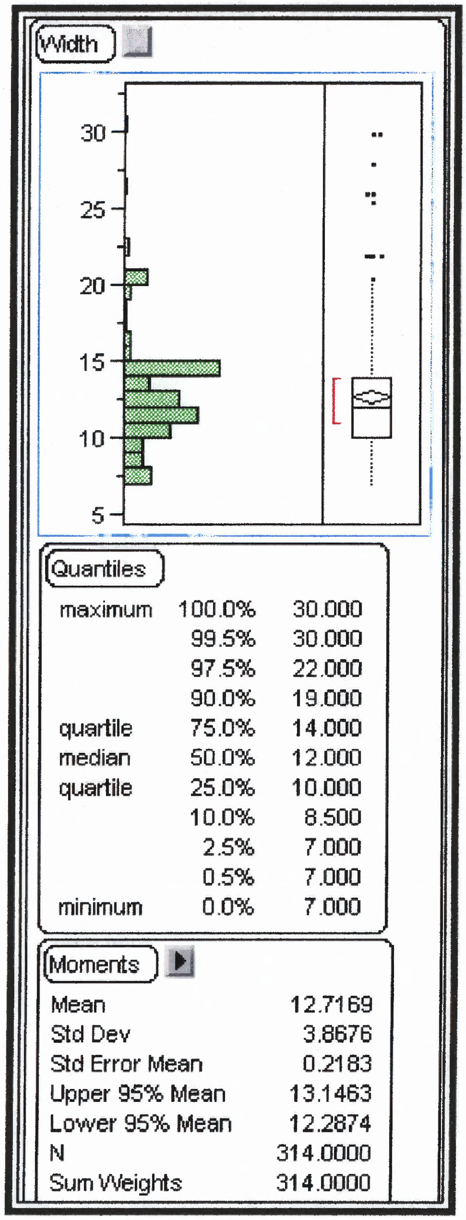


Figure 4.5 Lane Width Distribution.

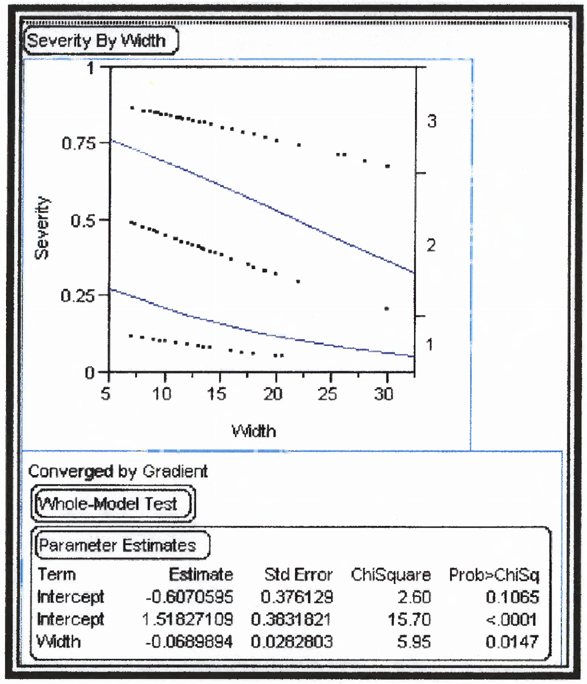


Figure 4.6 Width Severity Interaction .

4.4.2 Grade and Curve

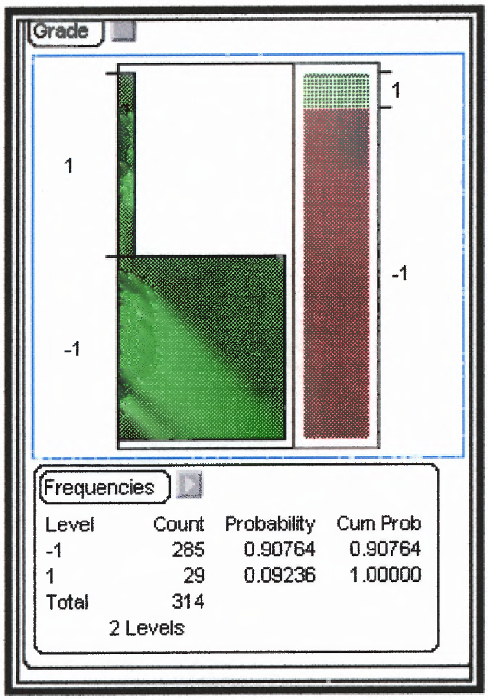
NJDOT provided grade information and curve information for most accidents as their Road Character variable. Missing road character data were obtained by field observations. The various classifications of grades for Road Character were aggregated into the dichotomous variables Grade and Curve as shown in Table 4.2. Figures 4.7 and 4.9 show that of all the accident locations, 9% were on grade, 81% were on level ground, 4% took place on a curve, 96% on straight away.

Table 4.2 Grade and Curve Variable Aggregation

NJDOT	Grade	Curve	Description
1	-1	-1	Straight and Level
2	1	-1	Straight and Grade
3	-1	-1	Straight at Hillcrest
4	-1	1	Curve and Level
5	1	1	Curve and Grade
6	-1	1	Curve at Hillcrest

Source: NJDOT accident database

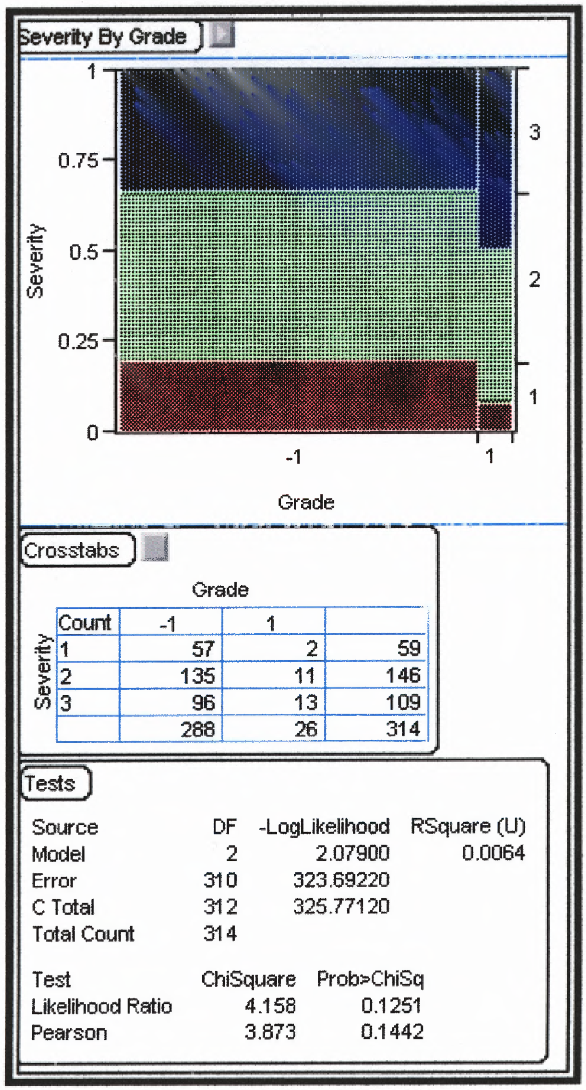
A comparison of the mosaic plots as shown in Figures 4.8 and 4.10 for Grade and Curve shows a visible increase in the highest level of severity with grade, but little or no change with curve. The middle severity level, Level 2 - Complaint of pain, is reduced for both variables, but the reduction is far more pronounced for grade. A comparison of the p-values confirms this fact. The variable Curve's p-value of 0.8986 is so high that any significance is doubtful. The variable Curve will not be considered as a potential model variable. Grade's p-value is not low enough to guarantee entry into the model, but it will be retained for possible interaction with other variables.



Level	Count	Probability	Cum Prob
-1	285	0.90764	0.90764
1	29	0.09236	1.00000
Total	314		

2 Levels

Figure 4.7 Grade Distribution.



	Count	Grade -1	Grade 1	
Severity	1	57	2	59
	2	135	11	146
	3	96	13	109
		288	26	314

Tests

Source	DF	-LogLikelihood	RSquare (U)
Model	2	2.07900	0.0064
Error	310	323.69220	
C Total	312	325.77120	
Total Count	314		

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	4.158	0.1251
Pearson	3.873	0.1442

Figure 4.8 Grade Severity Mosaic.

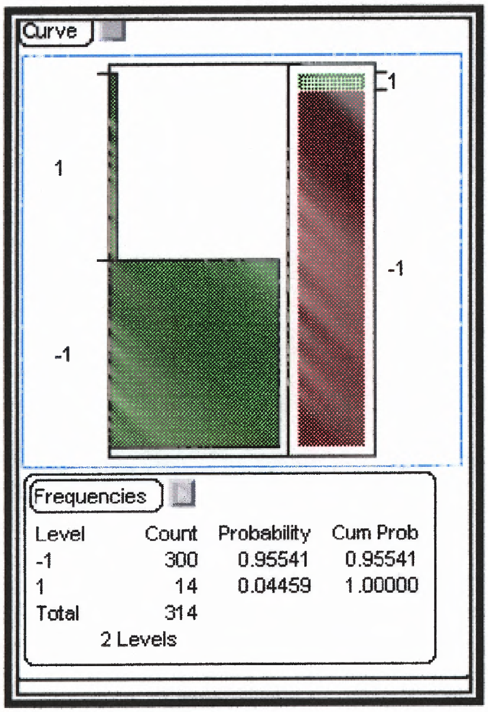


Figure 4.9 Curve Distribution.

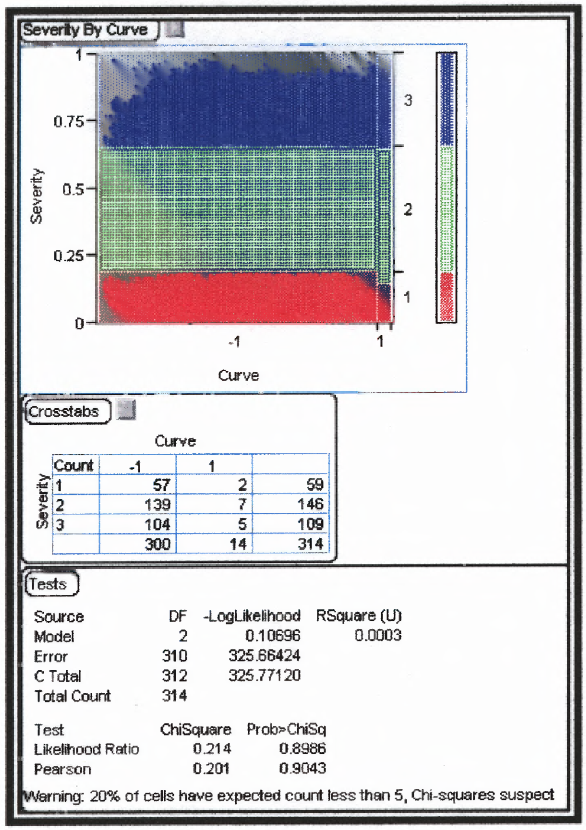


Figure 4.10 Curve Severity Mosaic.

4.3.3 Road Division

Road Division data were obtained from the NJDOT database. The data were aggregated as shown in Table 4.3.

Table 4.3 Road Division Aggregation

NJDOT	Road _Div	Description
1	1	Guide Rail
2	1	Concrete Bar
3	1	Concrete Isle
4	1	Grass Median
5	-1	None
6	-1	Other

Source: NJDOT accident database

There were 15 records with the variable Road_Divided_By designated as "Other" as well as additional 15 records with no designations. Field inspections determined whether medians existed at these locations. Of the 314 accidents, only 10 or 3.1% occurred on roads with medians.

Road_Div appears to have no significance with a large p-value of 0.9655. This lack of significance may result from the fact that in Jersey City with narrow pre-automobile roads, few median median separations exist. In another, non-urban area, Road_Div may be significant. Road_Div will not be considered as a model variable.

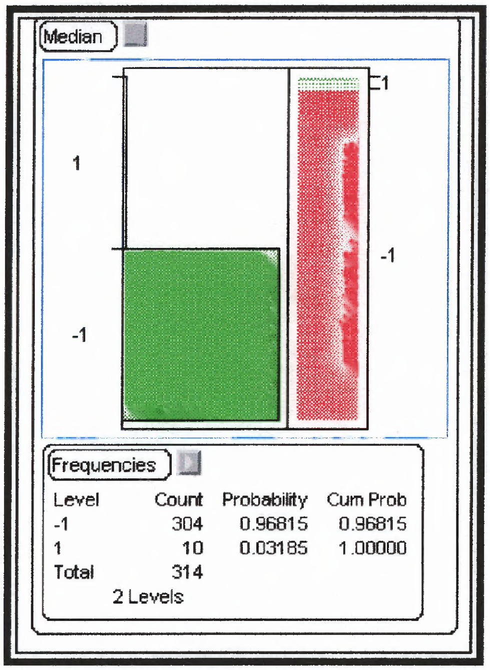


Figure 4.11 Road_Div Distribution.

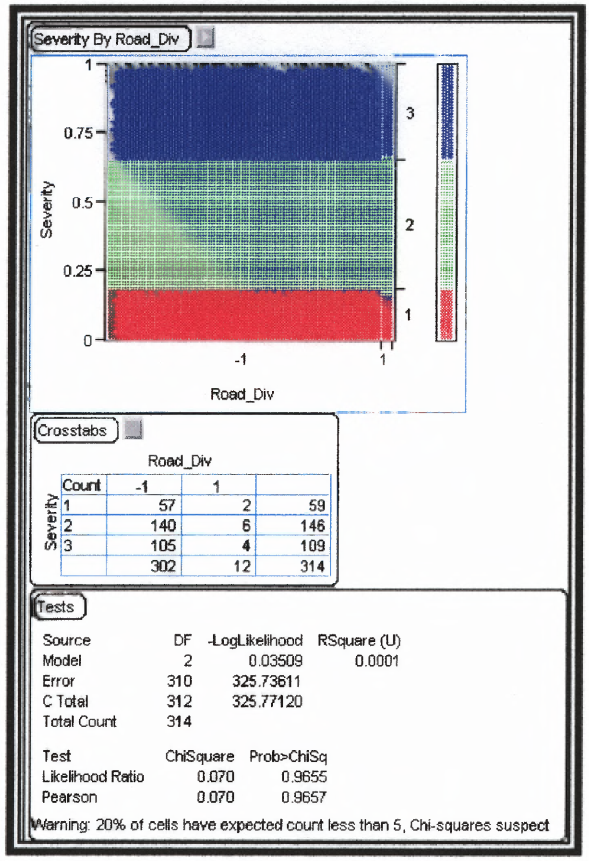


Figure 4.12 Road_Div Severity Mosaic.

4.4.4 Pavement

The pavement variable answers the question: Was the accident location paved within the ten years prior to the accident? In the absence of a pavement management system with a standardized pavement rating system, the assumption is made that a newer pavement is in better condition. Newer pavement is less likely to have utility patches, pot holes, ruts or cracking. There are many other reasons for poor quality pavement besides age, such as traffic volume, particularly heavy vehicle volume, quality of installation, underlying base, amount of utility cuts, etc. Nevertheless, a section of roadway paved within the past ten years is much less likely to exhibit the degradation present in older pavement sections.

Pavement age is also a marker for bicyclist safe grates. Stream flow catch basin grates with longitudinal spacing were originally installed in the 1960's and 1970's to improve drainage. The front wheels of bicycles, however, can easily become caught between the bars of these grates. Newer grates with a closely spaced grid do not cause this problem. The switch to bicycle safe grates has been a gradual process. In Jersey City, all roads constructed within the past ten years have replaced existing grates with bicycle safe grates. Therefore, pavement age reflects both pavement quality and the presence of bicycle safe grates.

The capital improvement records of the State of New Jersey, Hudson County and Jersey City were investigated to determine the date of the most recent paving. Care was taken that the paving occurred within ten years of the date of the accident, not ten years prior to the date of this study. The accidents in the database occurred over a four year period. Paving which took place after the accident would have no impact on the outcome. Jersey City maintains a map of their ten-year pavement program.

The distribution between recently and older pavement sections is shown in Figure

4.13. Roughly half or 52.5% of the roads have been paved within ten years of the accident. Pavement is a very significant variable with a p-value of 0.0678. The mosaic plot shows little impact of pavement condition on the difference between serious and minor injury, but a pronounced difference between an injury and a property damage only accident.

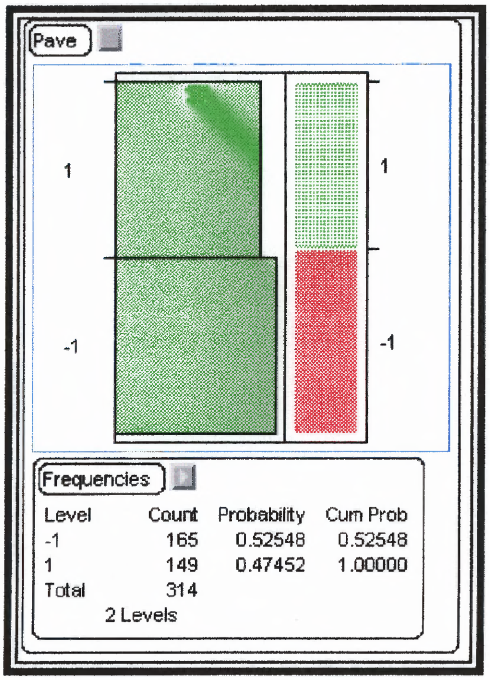


Fig
ure

4.13 Pavement Distribution

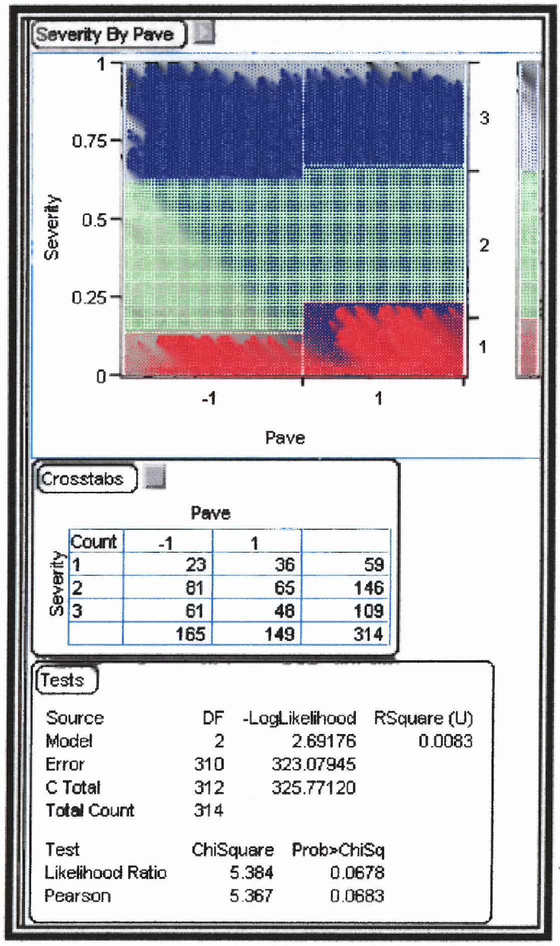


Figure 4.14 Pavement Severity Mosaic

4.4.5 Highway

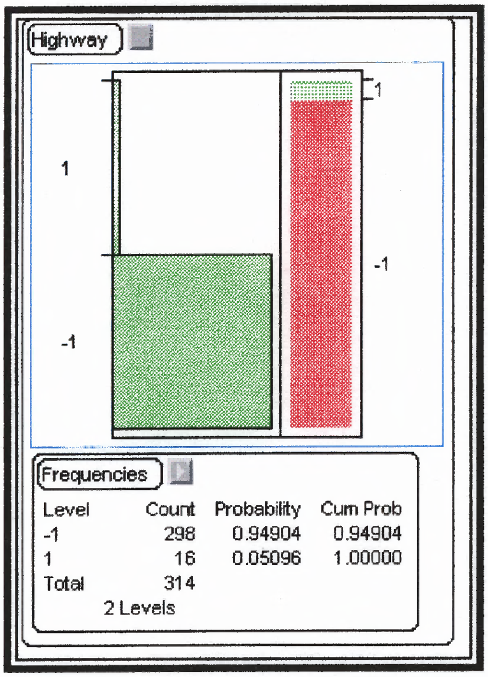
The NJDOT accident database designates the Road System for each of the locations where accidents occurred. This variable was aggregated into the dichotomous variable referred to as Highway as shown in Table 4.4. The variable simply answers the question, "Is the road a State highway or not?" Of all accident locations, 5% were highways, 95% were local or county roads as shown in Figure 4.15.

Table 4.4 Highway Variable Aggregation

NJDOT	Highway	Description
1	1	Interstate
2	1	State Highway
3	1	State/Interstate Authority
4	1	State Park or Institution
5	-1	County
6	-1	County Authority Park or Institution
7	-1	Municipal
8	-1	Private Property
9	-1	U.S. Government Property

Source: NJDOT accident database

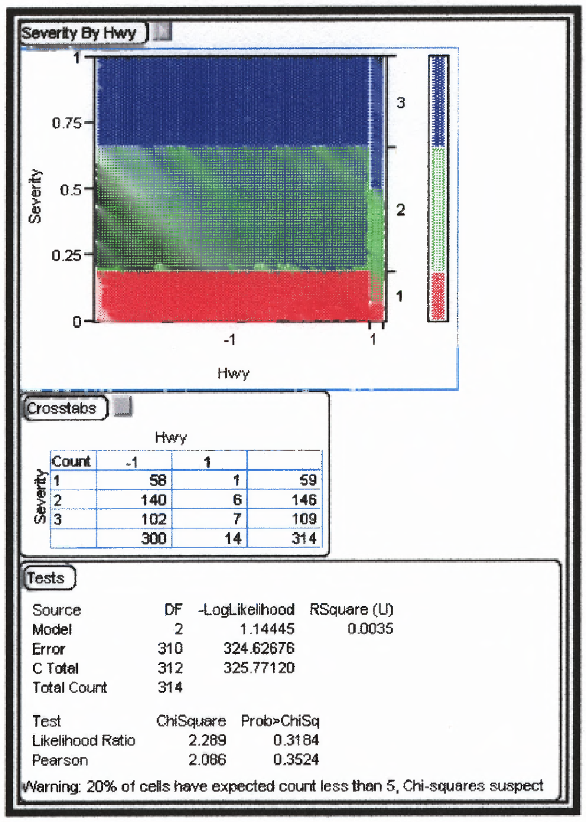
The highway mosaic plot depicted in Figure 4.16 shows a marked increase in severe accidents on State highways. Because of limited records of bicycle accidents on highways, the chi-square test may be suspect. However, the Highway variable will be entered into the model to determine if it interacts with any other variables despite its relatively high p-value of 0.3164.



Level	Count	Probability	Cum Prob
-1	298	0.94904	0.94904
1	16	0.05096	1.00000
Total	314		

2 Levels

Figure 4.15 Highway Distribution.



Count	-1	1	
1	58	1	59
2	140	6	146
3	102	7	109
	300	14	314

Tests

Source	DF	-LogLikelihood	RSquare (U)
Model	2	1.14445	0.0035
Error	310	324.62676	
C Total	312	325.77120	
Total Count	314		

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	2.289	0.3184
Pearson	2.086	0.3524

Warning: 20% of cells have expected count less than 5, Chi-squares suspect

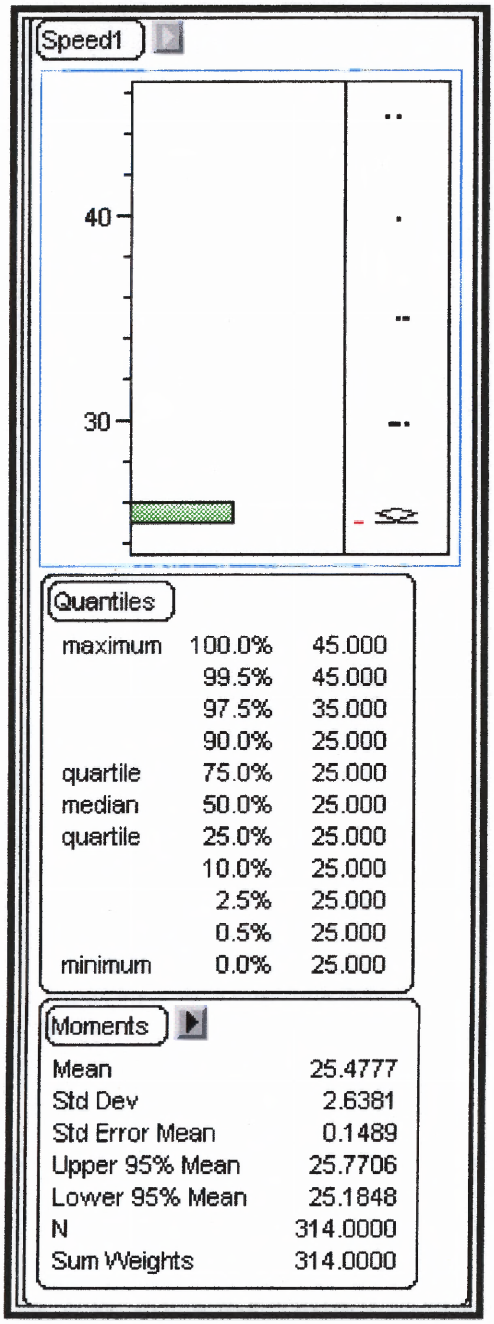
Figure 4.16 Highway Severity Mosaic .

4.5 Operational Factors

4.5.1 Speed

Posted speed was obtained directly from the NJDOT database. Missing data were obtained either from NJDOT straight line diagrams or the Jersey City Municipal Code. The density distribution shown in Figure 4.17 reveals that there is little distribution in speed limit. Over 90% of the streets are posted at 25 miles per hour with the exception of Garfield Avenue and State Highways Route 1, 139, 169 and 440. Given this absence of variance in the data, it is not surprising to see its mosaic plot, as shown in Figure 4.18, to be completely flat. With a p-value of 0.9290, it is inconceivable that speed will be entered into the model. Speed will not be considered as a model variable.

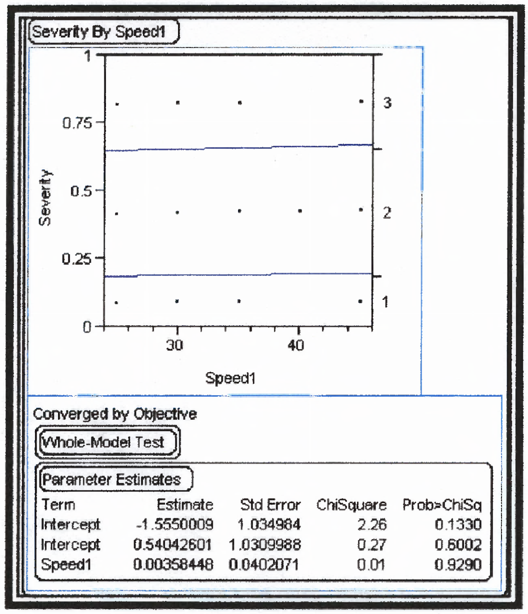
In truth, posted speed and operating speed differ greatly in congested urban areas. Operating speeds of 10-15 mph are far more common than 25 mph. This is confirmed in the morning and evening peak periods by Figures 4.19 and 4.20 which show the measured operating speeds on Jersey City roads, (Voorhees, 1979). Few operating speed studies were available in Jersey City's files. Unfortunately, the number of these studies were too small to analyze. If funds and time were not a constraint, a speed study to determine operating speeds may reveal that the most significant factor in urban bicycle accident injury severity is operating speed.



maximum	100.0%	45.000
	99.5%	45.000
	97.5%	35.000
	90.0%	25.000
quartile	75.0%	25.000
median	50.0%	25.000
quartile	25.0%	25.000
	10.0%	25.000
	2.5%	25.000
	0.5%	25.000
minimum	0.0%	25.000

Mean	25.4777
Std Dev	2.6381
Std Error Mean	0.1489
Upper 95% Mean	25.7706
Lower 95% Mean	25.1848
N	314.0000
Sum Weights	314.0000

Figure 4.17 Speed Distribution.



Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-1.5550009	1.034984	2.26	0.1330
Intercept	0.54042601	1.0309988	0.27	0.6002
Speed1	0.00358448	0.0402071	0.01	0.9290

Figure 4.18 Speed Severity Interaction.

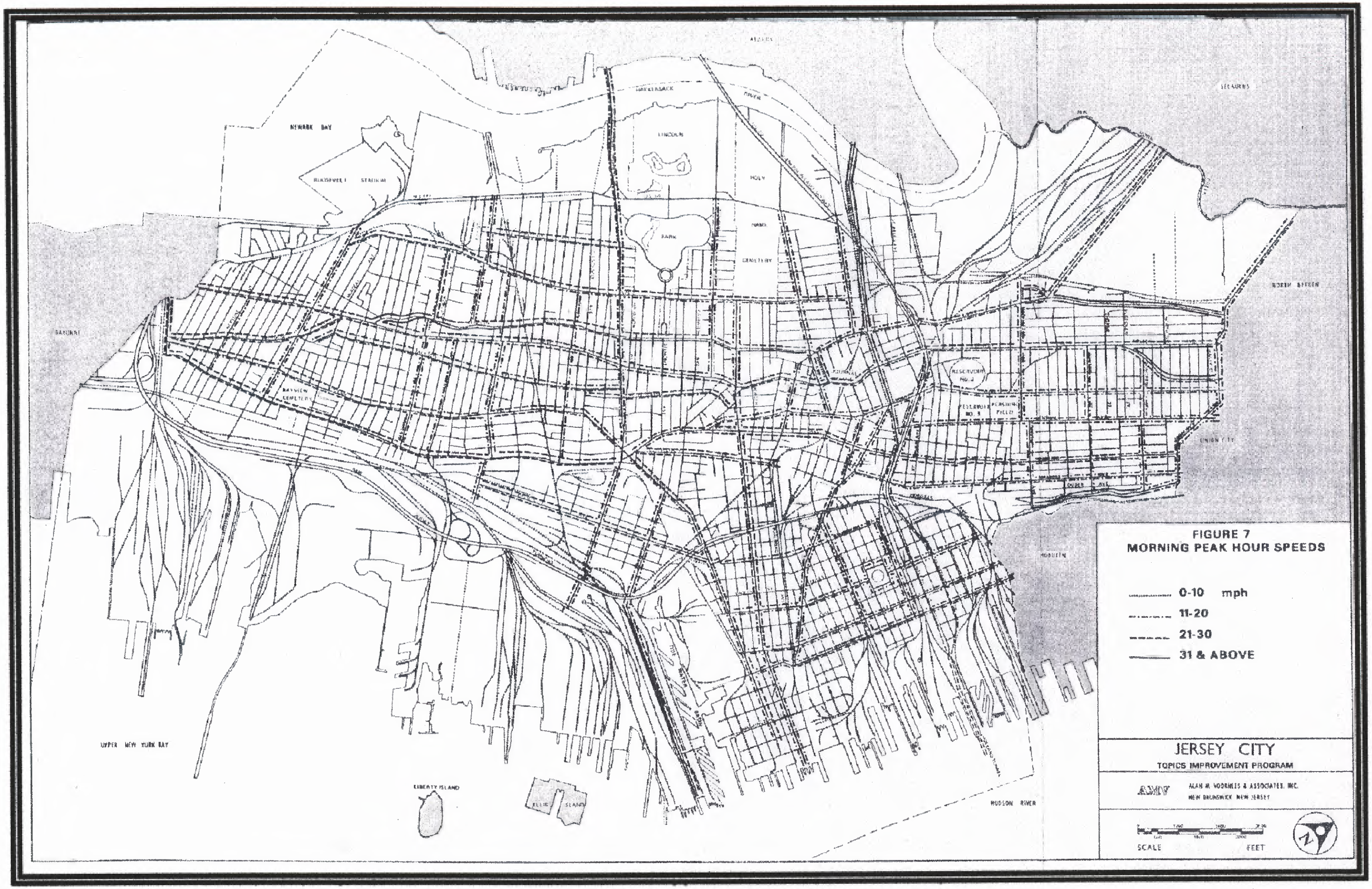


Figure 4.19 Morning Peak Hour Speeds
Source: Voorhees, Alan M. & Associates, Jersey City Topics, Prepared for New Jersey Department of Transportation in Cooperation with United States Department of Transportation Federal Highway Administration and City of Jersey City, New Jersey, 1973.

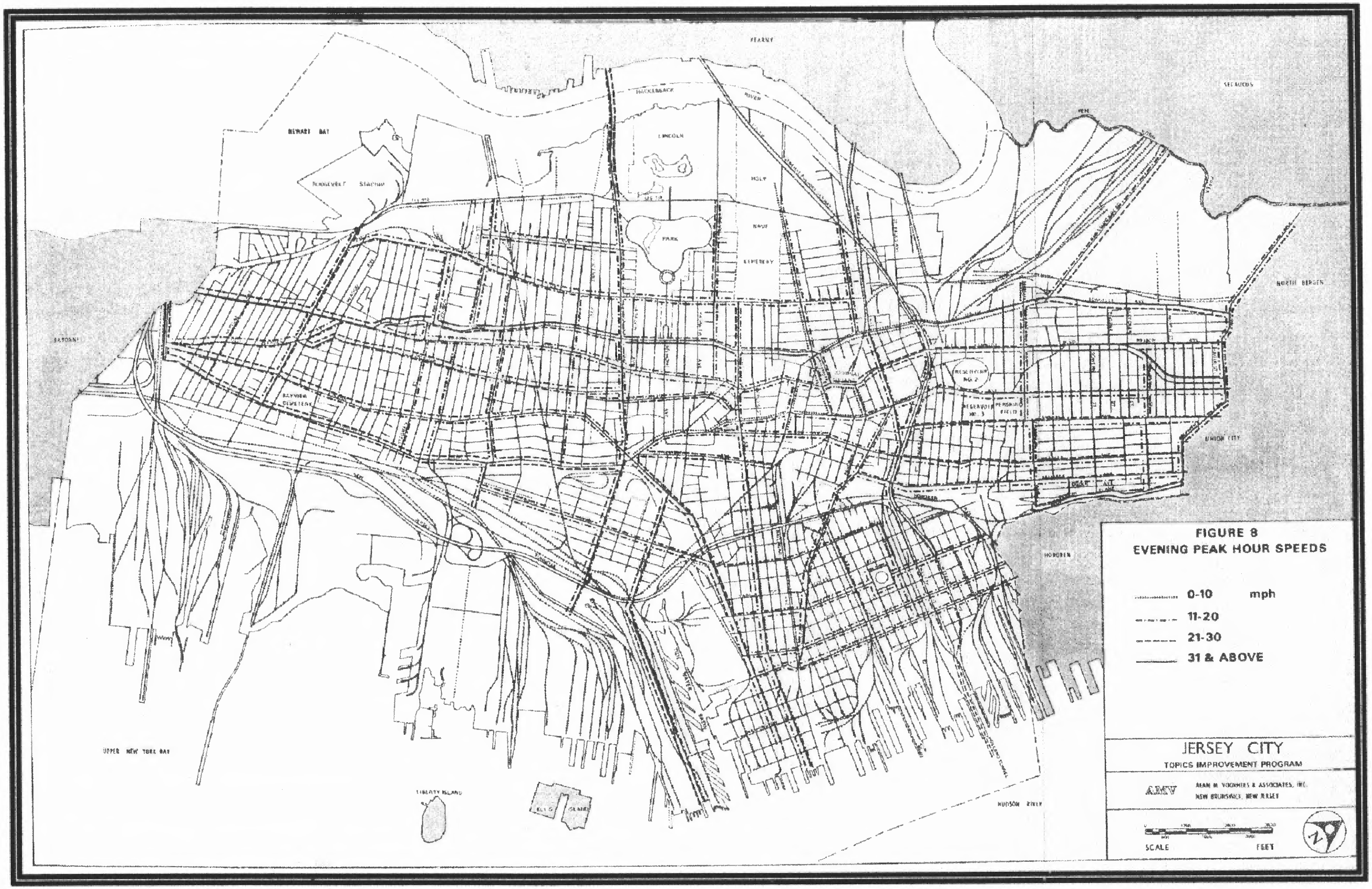


Figure 4.20 Evening Peak Hour Speeds
Source: Voorhees, Alan M. & Associates, Jersey City Topics, Prepared for New Jersey Department of Transportation in Cooperation with United States Department of Transportation Federal Highway Administration and City of Jersey City, New Jersey, 1973.

4.5.2 Lane Volume

The NJDOT database does not provide the ADT at the accident location. It was therefore necessary to obtain the roadway volumes from Jersey City's Division of Traffic files. Historically, Jersey City measures traffic volume locations at intersections when requests are made to install new or upgrade existing traffic signals. On occasion, volume counts may be taken at the request of an elected official, a resident, or as a part of a study of traffic impacts, proposed detours, etc. Traffic counts were not maintained electronically. Manual retrieval was required using a cumbersome and time consuming card catalog method. The indices to these files are listed on index cards filed in a card catalog by the major street of the intersection. As a byproduct of this research, Jersey City was provided with an electronic database of traffic volume.

The task to obtain the volume for a given accident location required examining all of the available traffic volume records. If no record existed at the specific location, volumes would be sought for proximate locations taking into consideration the direction of traffic flow. For example, if a count existed north of the accident location and another count existed south of the location, the southbound traffic volume from the northern location would be added to the northbound traffic volume from the southern location to compute a total ADT for both directions. Cross street volumes were not included in the ADT because it would produce unreasonably skewed volumes for mid-block crashes which had no intersecting street.

Also not considered, was the directionality or distribution of the percentage of traffic that flows in each direction. In rural areas where significant passing occurs, volumes with heavy one directional flows impact the bicyclist less because vehicles are able to pass the bicyclist by moving into the opposing lane. Urban streets are either one way or striped with a double yellow line that prohibit passing in the opposing lane. Therefore, although it would affect a bicyclist if the majority of traffic was moving in the direction opposite of his travel, the effect will not be as pronounced as in a rural area.

If no traffic volumes were available along the specific road from a proximate location, traffic volumes were substituted from roads that were similar in land use, width, length and circulation. In the rare cases where no reasonable volume counts were available, a minimal value of 250 ADT was applied. It is assumed that these locations that had absolutely no volume count history were generally little traveled roadways that never warranted a signal or a traffic investigation.

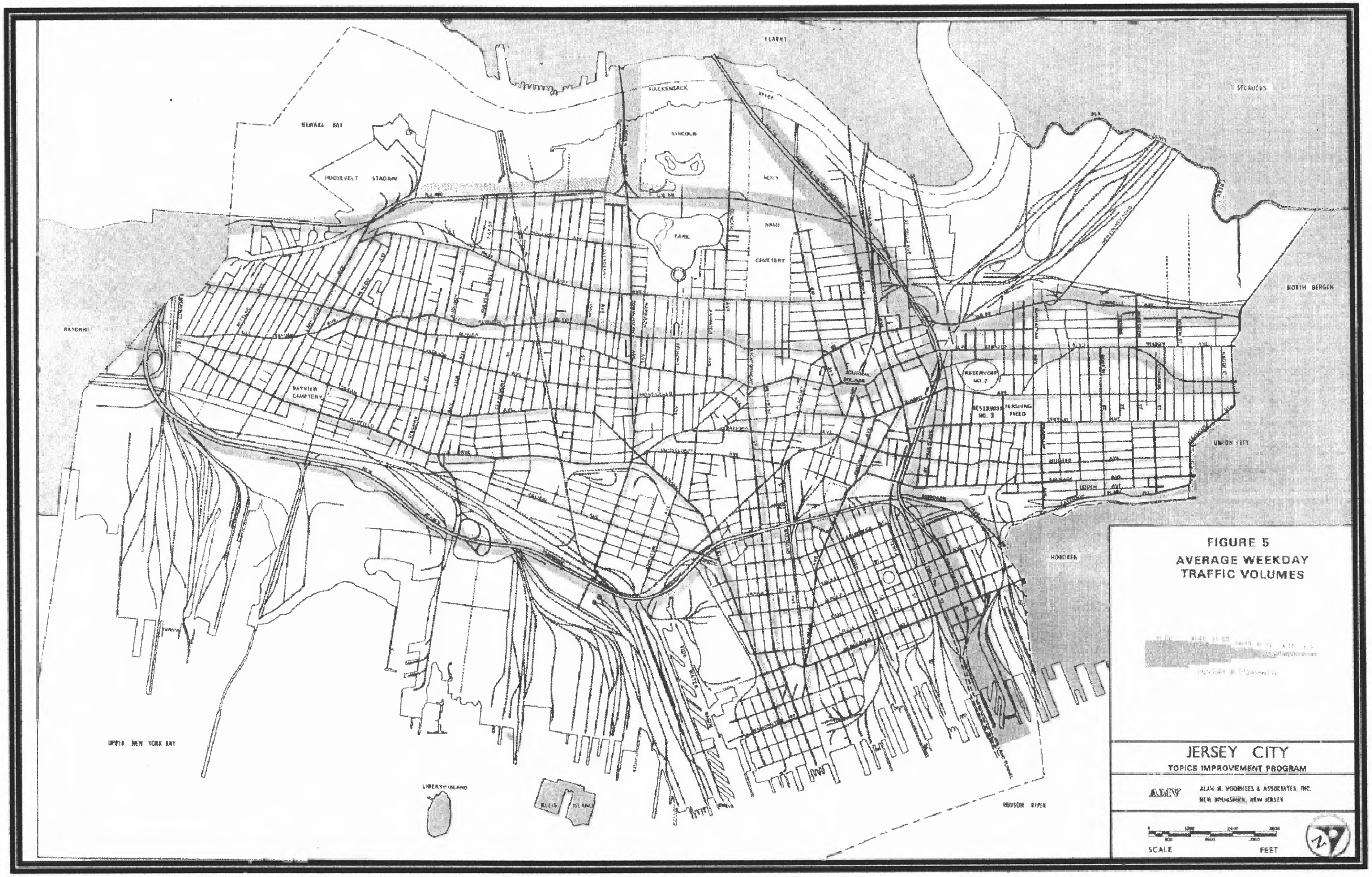


Figure 4.21 Average Weekday Traffic Volume
Source: Voorhees, Alan M. & Associates, Jersey City Topics, Prepared for New Jersey Department of Transportation in cooperation with United States Department of Transportation Federal Highway Administration and City of Jersey City, New Jersey, 1973.

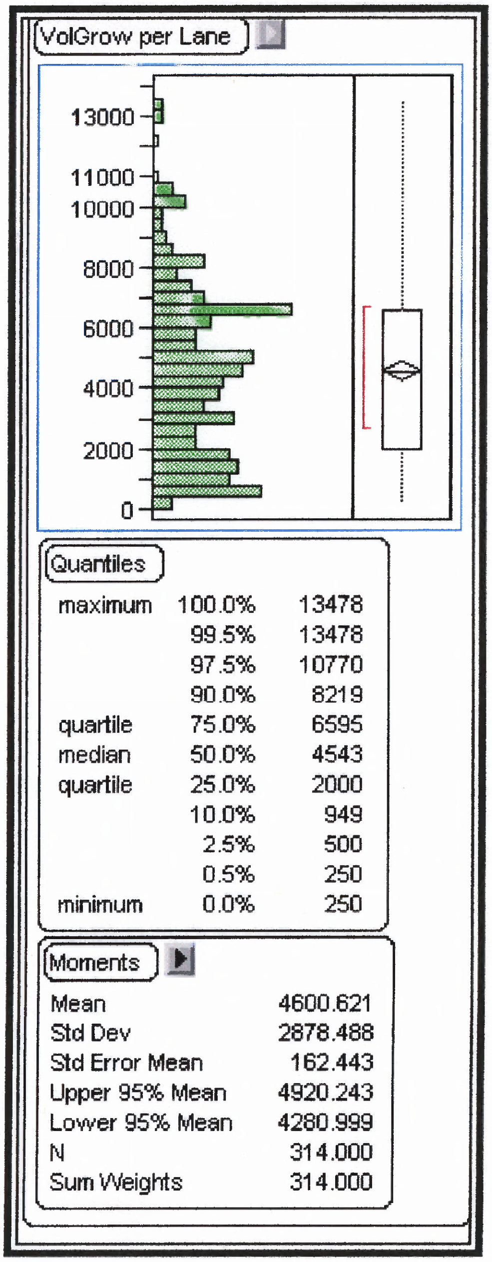


Figure 4.22 Lane Volume Distribution.

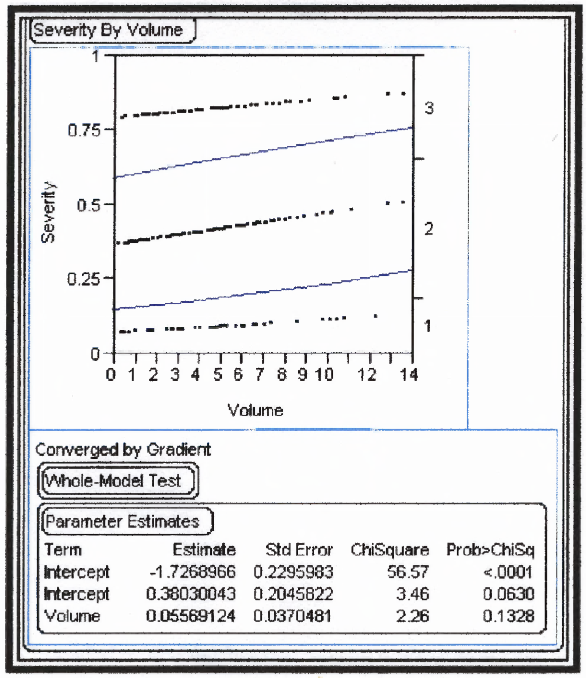


Figure 4.23 Volume Severity Interaction.

Jersey City has been able to generate computerized traffic counts using automated traffic recorders (ATR's) since the early 1980's. Prior to this time, ATR tapes were transcribed manually. Because of the concerns that a thirty-year old traffic count may have questionable validity today and because of the potential error introduced by manual transcription and summations, an arbitrary cut off date for inclusion of traffic volume data were set at 1980. All traffic volume studies used in this analysis were taken after 1980 and modified with a growth factor of one percent per annum from the year the count was taken. Considering that according to the U.S. Census (2000), Jersey City's population grew over 7% in that period, a growth factor or one percent would be conservative.

For analysis purposes, traffic volume was divided by the number of lanes to create a lane volume because the bicyclist's friction with motor vehicles is determined by the intensity of traffic immediately adjacent to his path of travel. Jersey City has no records of the number of lanes for each of its roads. Obtaining this information required field investigations.

Volume per lane ranges from a minimum of 250 vehicles per day (ADT) to a maximum of 13,478 ADT with a mean of 4600 ADT. The variable volume has a reverse effect on injury severity. As volume increases, severity decreases. The significance level of volume at 0.1328 does not meet the criteria for inclusion in the model. However, it will be retained for the model building effort at this time because of potential interactions with other potential variables.

4.5.3 Bus Routes

A determination of whether a bicycle accident was located on a bus route was undertaken by identifying the location of the crash on a New Jersey Transit map for Hudson County. Those crashes which occurred on a bus route were designated as +1. Those crashes which did not occur on a bus route were designated as -1. The designation of 1 and -1 are convenient modeling labels because they facilitate implementation of the model whose coefficients are set at $+\beta$ for the presence of the condition, in this case a bus route, and $-\beta$ in the absence of the condition. A distribution of the accidents on bus routes is shown in Figure 4.24

Slightly more than half, or 57% of all recorded bicycle crashes occurred on bus routes. The mosaic plot shown in Figure 4.25 does show a slight increase in property damage only accidents and a slight decrease in serious injury accidents on bus routes. With a p-value of 0.2506 there is a slight chance that the Bus variable may combine with another variable to meet the significance limit. For this reason, the Bus variable will be retained to test its significance when combined with other variables during the model building stage.

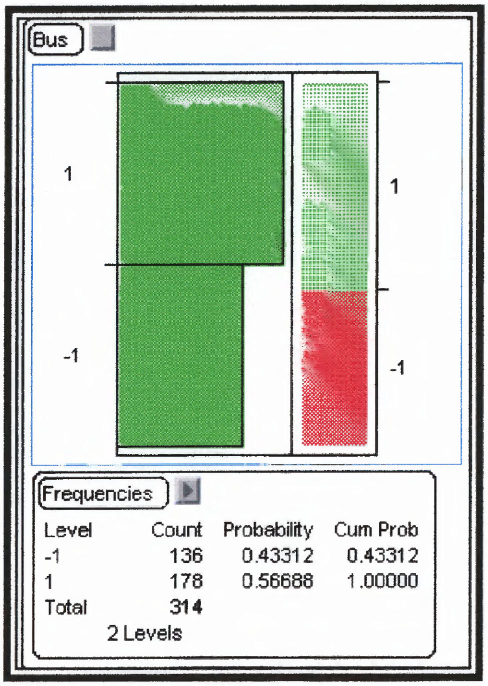


Figure 4.24 Bus Distribution.

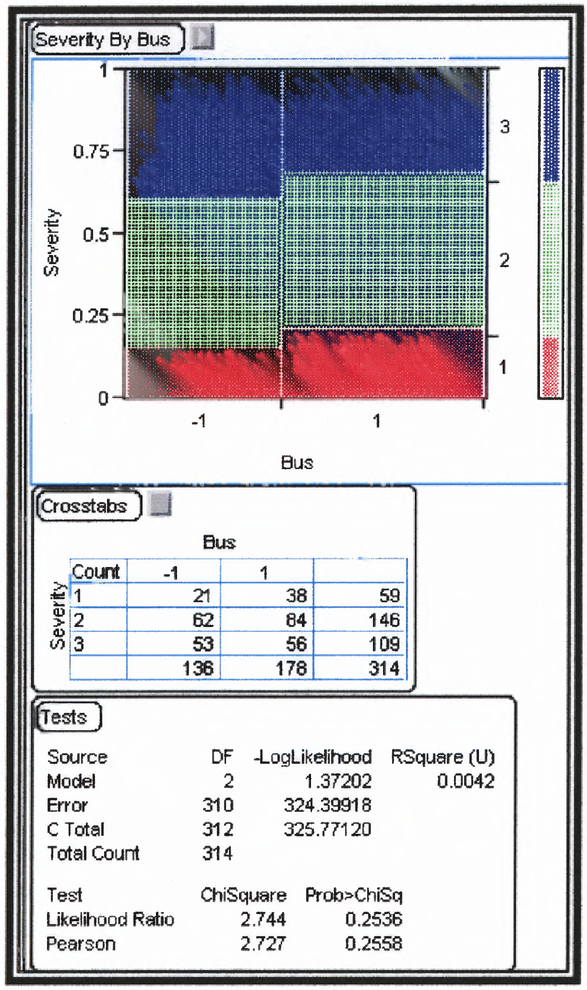


Figure 4.25 Bus Route Severity Mosaic.

4.5.4 Truck Routes

Jersey City Municipal Code Schedule 7 designates the sections of roads that are truck routes. The determination of whether an accident occurred on a truck route is determined by pinpointing the accident location on a truck route map. The Jersey City Truck route map is shown in Figure 4.26. If the accident did not occur on a truck route it is scored -1, if it did occur on a bus route, it is scored +1.

Figure 4.27 shows that the majority of accidents, over 72%, occurred on truck routes. As shown in Figure 4.28, the significance of the variable Truck is doubtful with its high p-value of 0.4887, although there is a marked increase in serious injury accidents on truck routes. An attempt will be made to investigate the Truck variable's interaction with other variables by retaining it for testing during the model building stage.

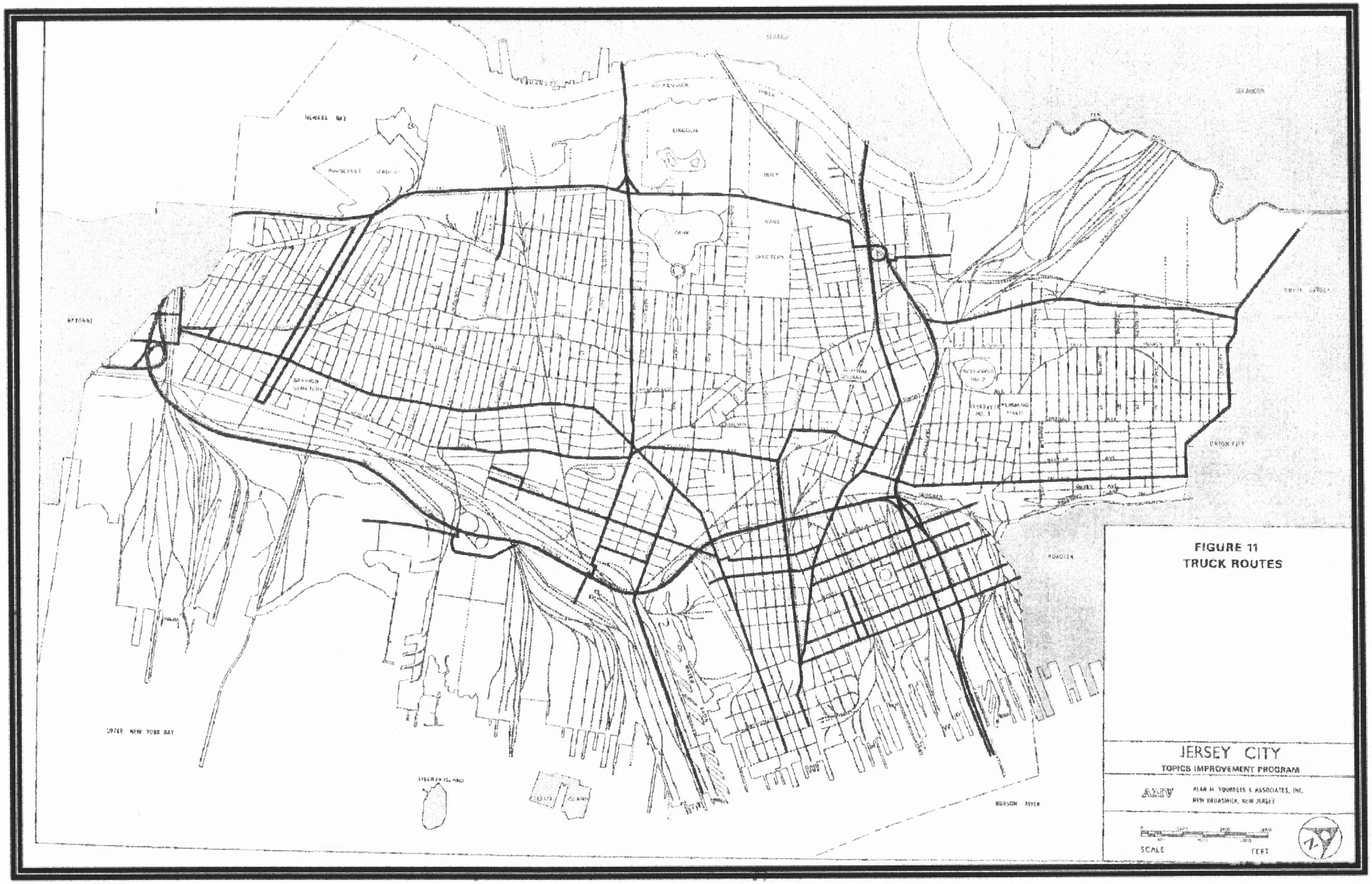
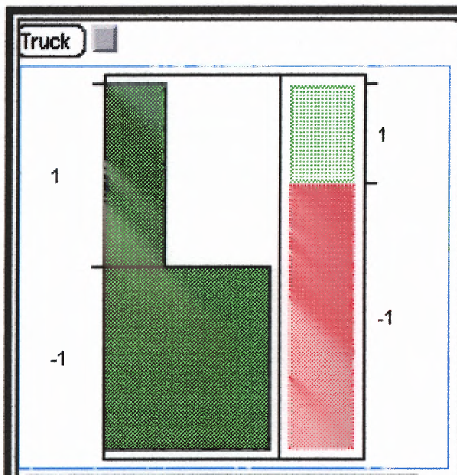


Figure 4.26 Truck Route
Source: Voorhees, Alan M. & Associates, Jersey City Topics, Prepared for New Jersey Department of Transportation in Cooperation with United States Department of Transportation Federal Highway Administration and City of Jersey City, New Jersey, 1973.

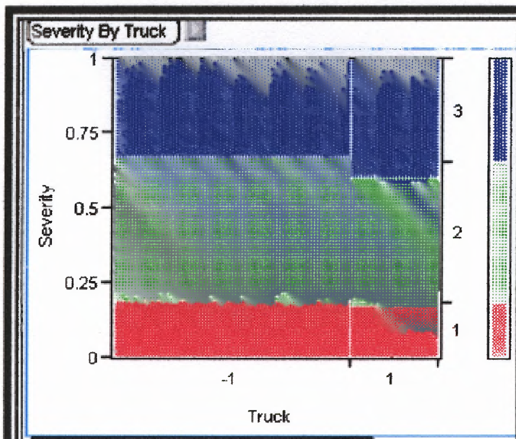


Frequencies

Level	Count	Probability	Cum Prob
-1	229	0.72930	0.72930
1	85	0.27070	1.00000
Total	314		

2 Levels

Figure 4.27 Truck Distribution.



Crosstabs

Severity	Count	Truck -1	Truck 1	
	1	44	15	59
	2	110	36	146
	3	75	34	109
		229	85	314

Tests

Source	DF	-LogLikelihood	RSquare (U)
Model	2	0.71599	0.0022
Error	310	325.05521	
C Total	312	325.77120	
Total Count	314		

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	1.432	0.4887
Pearson	1.450	0.4844

Figure 4.28 Truck Severity Mosaic.

4.5.5 One-Way

A determination of whether the accident occurred at the intersection with at least one leg being limited to one-way travel was determined by looking up the road segment in the Jersey City municipal code Schedule 1 as mapped on Figure 4.29.

Figure 4.30 show that 64% of all accidents had a one-way street on at least one leg of the intersection. Figure 4.31 shows a definite relationship between the variable One-Way and injury severity. With a p-value of 0.0968, the variable One-Way will probably be retained in the model.

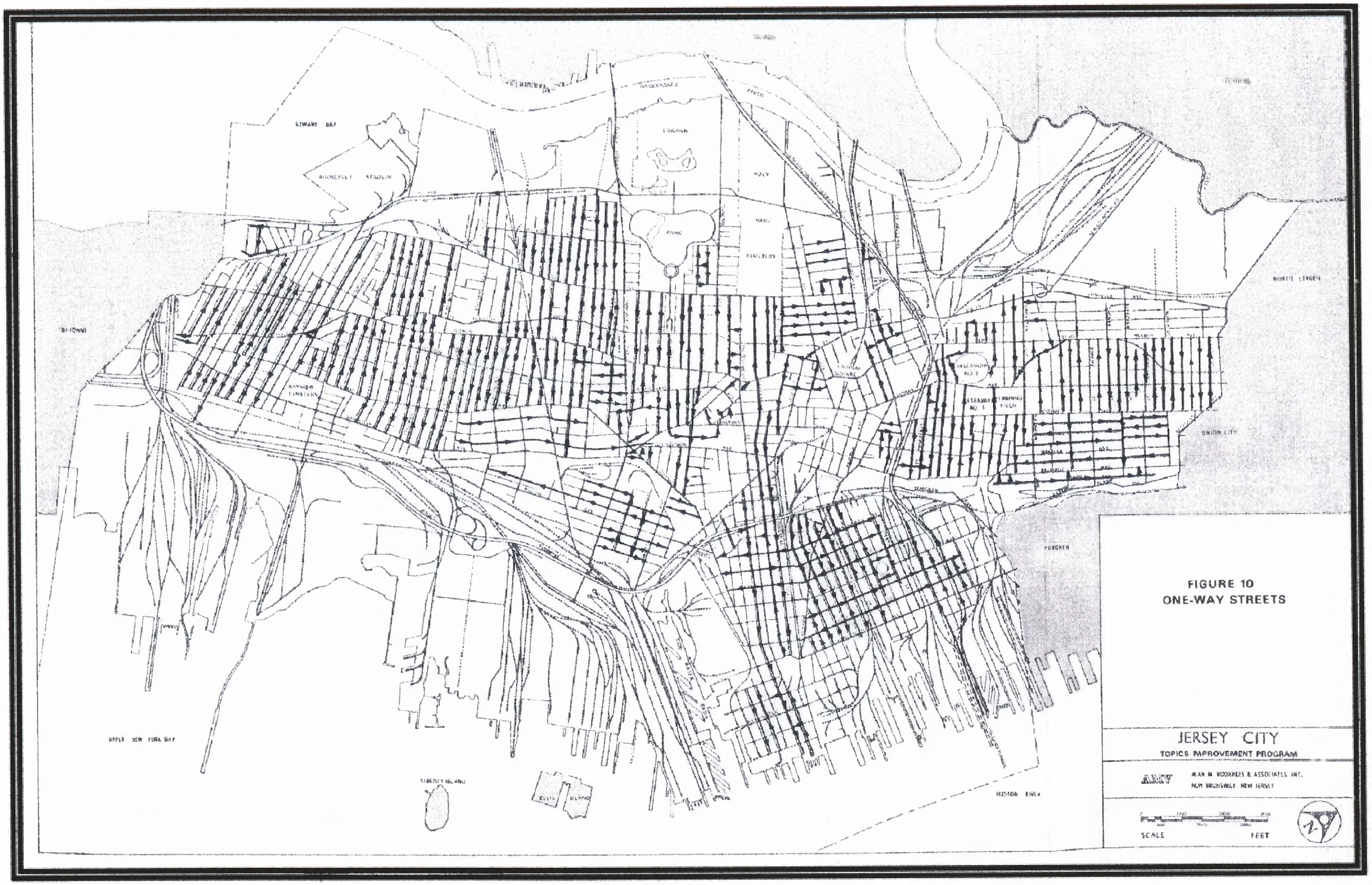
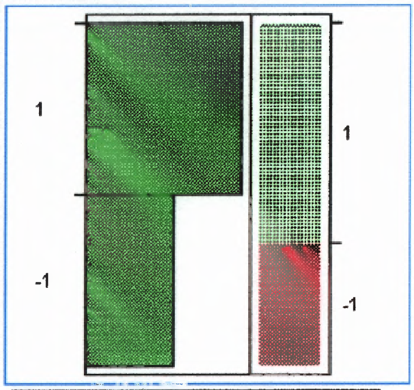


Figure 4.29 One Way Streets
Source: Voorhees, Alan M. & Associates, Jersey City Topics, Prepared for New Jersey Department of Transportation in Cooperation with United States Department of Transportation Federal Highway Administration and City of Jersey City, New Jersey, 1973.

One Way



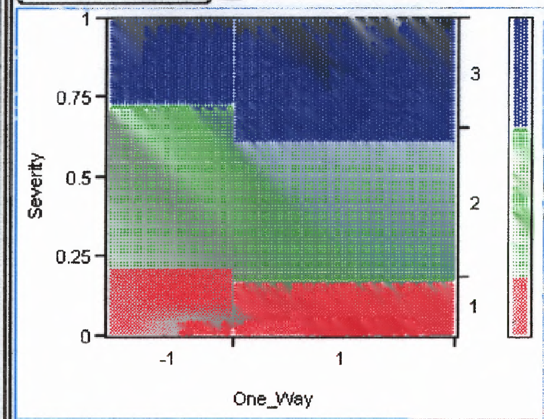
Frequencies

Level	Count	Probability	Cum Prob
-1	114	0.36306	0.36306
1	200	0.63694	1.00000
Total	314		

2 Levels

Figure 4.30 One Distribution.

Severity By One_Way



Crosstabs

One_Way

Severity	Count	-1	1	
	1	25	34	59
	2	58	88	146
	3	31	78	109
		114	200	314

Tests

Source	DF	-LogLikelihood	RSquare (U)
Model	2	2.33749	0.0072
Error	310	323.43372	
C Total	312	325.77120	
Total Count	314		

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	4.675	0.0966
Pearson	4.594	0.1006

Figure 4.31 One Way Severity Mosaic.

4.5.6 Parking

The Jersey Municipal Code Schedule 3 sets forth all parking restrictions enacted by municipal ordinance. Those roads with parking restrictions on either side of the street were set as -1. Those locations without parking restrictions were set as 1. Parking restrictions limited to street sweeping were not considered. The plot in Figure 4.32 illustrates that very few locations, little more than 10% of all accidents had parking restrictions.

Although the parking variable with a p-value of 0.1852, as shown in Figure 4.33, does not meet the significance limit. However, the plot does show that there is a visible increase in serious accidents in locations without parking. As with other variable with marginal significance, it will be re-examined during the model building state as it may interact significantly with other variables.

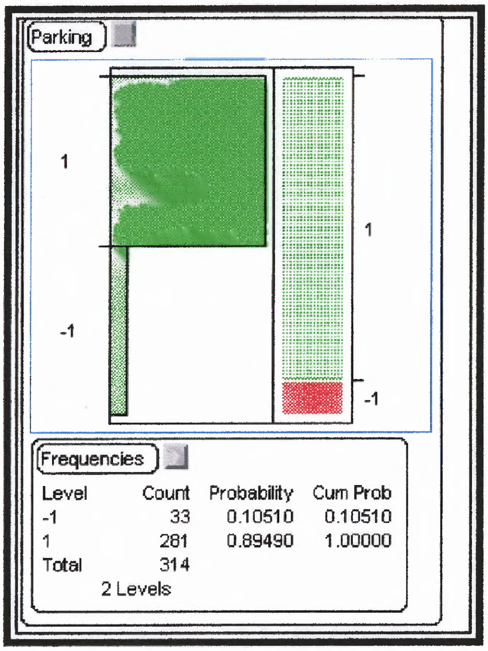


Figure 4.32 Parking Distribution.

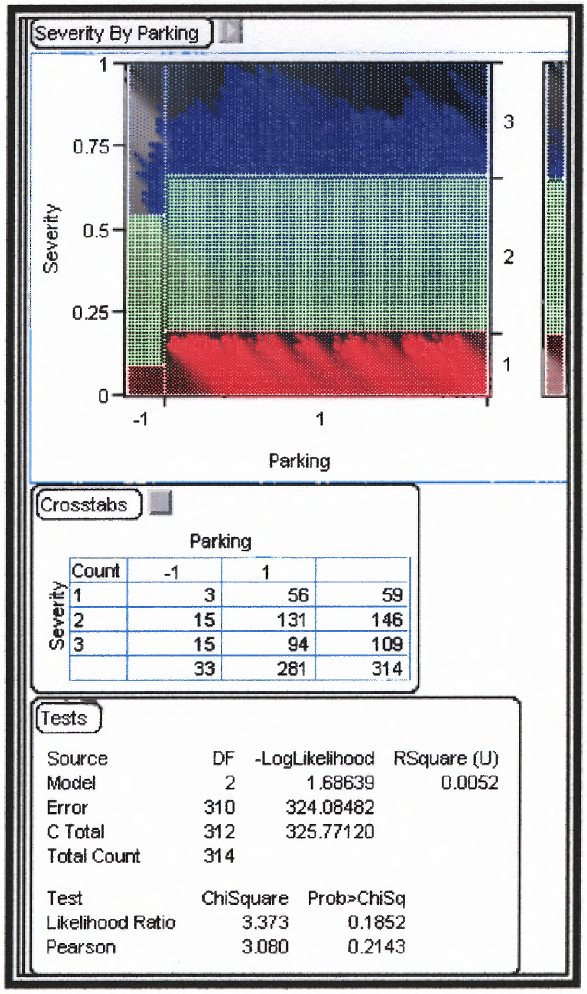


Figure 4.33 Parking Severity Mosaic.

4.5.7 Signalization

Jersey City has traffic signals located at 244 city streets which are mapped on Figure 4.34. In addition, Hudson County maintains traffic signals on the following county roads: Kennedy Boulevard, Patterson Plank Road and Secaucus Road. The State has traffic signals on Route 1, Route 139, Route 169 and Route 440. Each of these signals were located either on the map provided in Figure 4.34 or on the updated schedule. To determine whether a given accident location was at a signalized location, the location was looked up on the map. If the location was at a signalized intersection, the Signal variable was coded 1. If it was not, Signal was coded -1.

As shown on the Signal distribution in Figure 4.35, exactly one half of all intersections were signalized. The mosaic plat in Figure 4.36 shows that the variable Signal has a disappointingly low significance with a p-value of 0.7556 and little variation amongst severity levels. It will not be considered as a model variable.

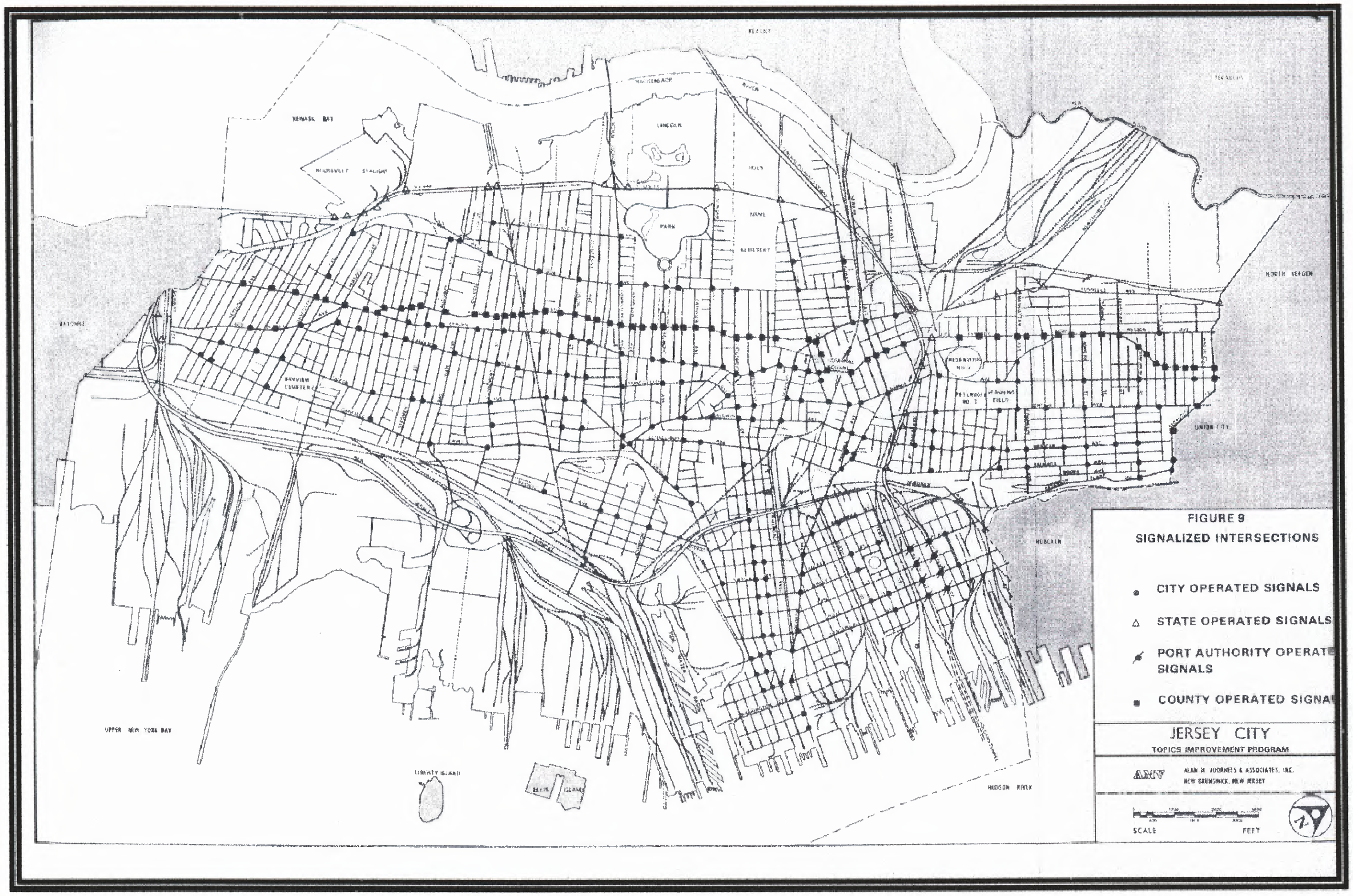
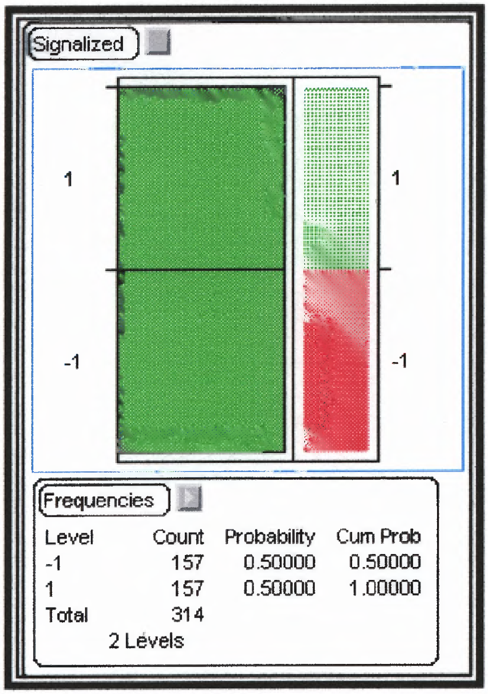


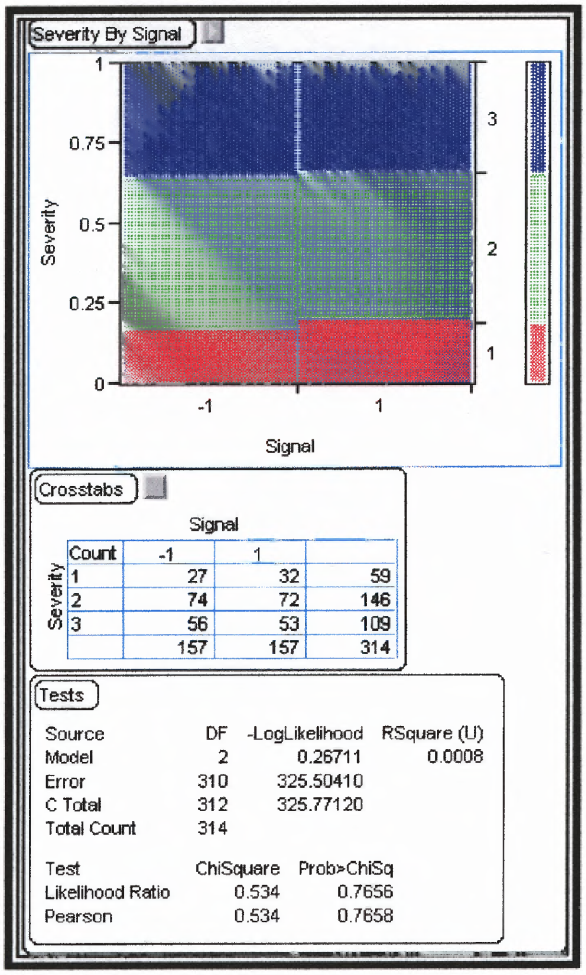
Figure 4.34 Signalized Intersections
Source: Voorhees, Alan M. & Associates, Jersey City Topics, Prepared for New Jersey Department of Transportation in Cooperation with United States Department of Transportation Federal Highway Administration and City of Jersey City, New Jersey, 1973.



Level	Count	Probability	Cum Prob
-1	157	0.50000	0.50000
1	157	0.50000	1.00000
Total	314		

2 Levels

Figure 4.35 Signal Distribution.



Severity / Count	Signal -1	Signal 1	
1	27	32	59
2	74	72	146
3	56	53	109
	157	157	314

Tests

Source	DF	-LogLikelihood	RSquare (U)
Model	2	0.26711	0.0008
Error	310	325.50410	
C Total	312	325.77120	
Total Count	314		

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	0.534	0.7656
Pearson	0.534	0.7658

Figure 4.36 Signal Severity Mosaic.

4.6 Socioeconomic Factors

4.6.1 Density

Density is defined as the population per square mile. High density areas might create more street activity and numerous low impact accidents as high density areas may also slow down traffic.

Data were available on the U.S. Census Bureau web site for each of the census tracts. As shown in Figure 4.37, density ranged from a minimum of 310 persons per square mile to a maximum of 60,245 persons per square mile with a mean of 29,698 persons per square mile. The interaction plot shown in Figure 4.38 suggests that injury severity increases with population density. Although, the curve is fairly flat and a high p-value of 0.4652, it may have the potential to interact with other factors. As with other variables in the range of significance, it will be reexamined during the model building stage.

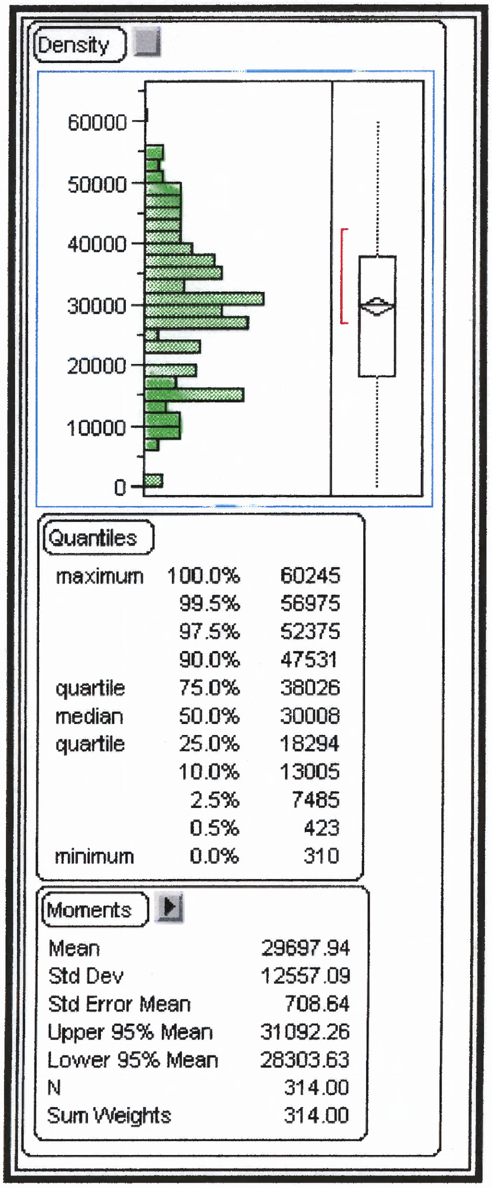


Figure 4.37 Density Distribution.

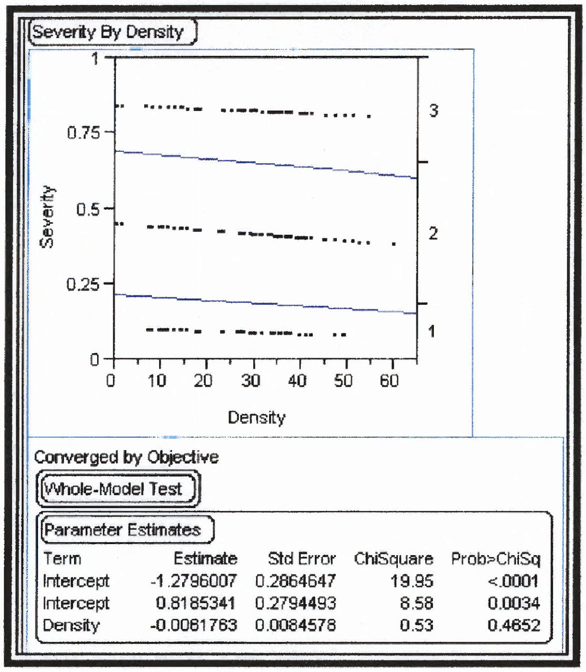


Figure 4.38 Density Severity Interaction.

4.6.2 Income

To investigate the significance of income to accident severity, average income was obtained from the US 2000 Census data. The census tracts were looked up for each accident using a web site look up facility. Average income levels for each census tract were then applied to the accident based on its census tract number. Figure 4.39 shows income levels ranged from a minimum of $6,846 per person to a maximum of $67,435 per person with a mean of $17,937.

The mosaic plot shown in Figure 4.40 shows a highly significant relationship between injury severity and income, with a p-value of 0.0173. Examining the shape of the curve reveals that injury severity increases with increasing per capita income. One plausible explanation is that high income drivers who are preoccupied with cell phone and other electronic equipment may be responsible for the serious accidents which result when the motorist does not observe the bicyclist. Alternatively, high income victims may be more likely to understand the financial implications of an accident any may also receive more sympathetic treatment from the police. There is also the possibility that a few high income outlier points may be responsible for the strong interaction. Nonetheless, Income will be included in the model building stage.

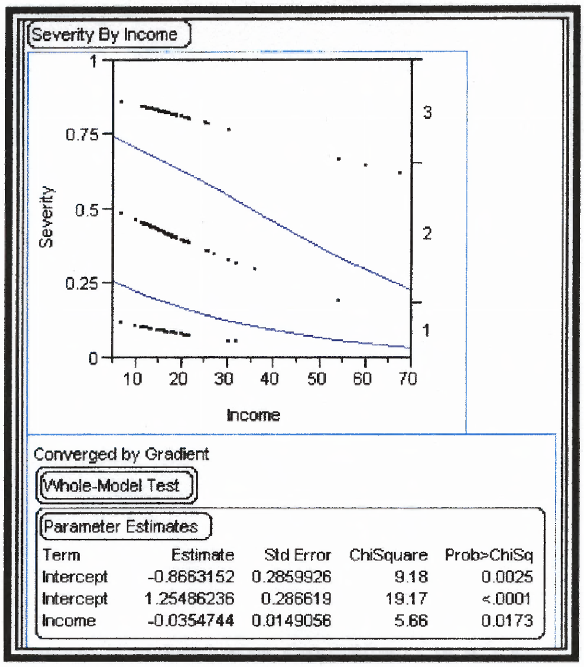


Figure 4.40 Income Severity Interaction.

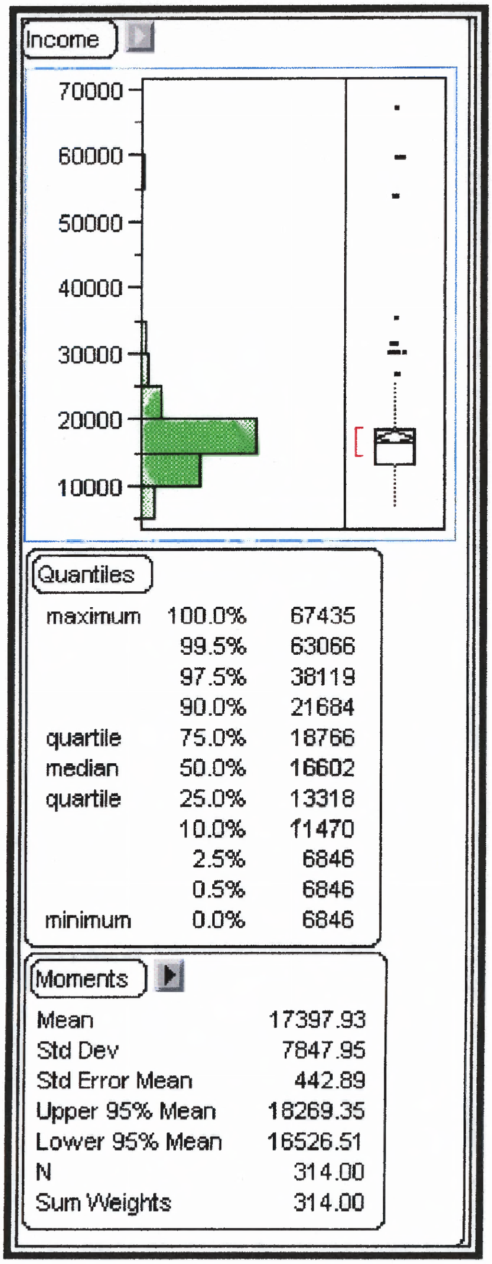


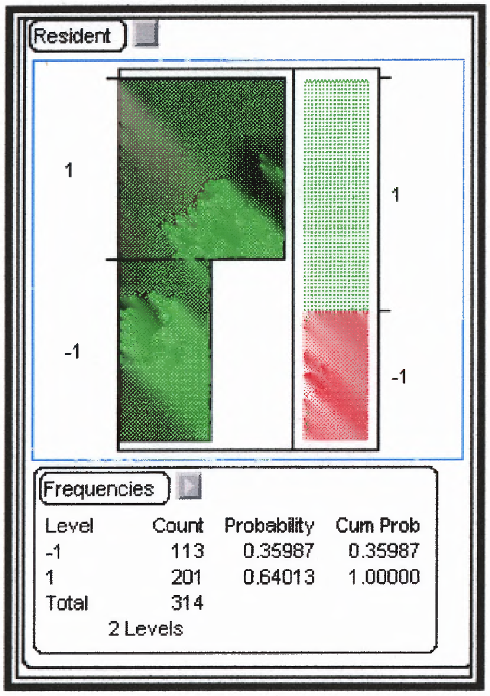
Figure 4.39 Income Distribution.

4.6.3 Land Use

Land use may affect the severity of an accident. Commercial areas tend to have high volumes, slow moving traffic, frequent parking maneuvers, frequent transit stops and active truck unloading zones. Residential areas have lower traffic volumes, with a larger percentage of children bicyclists who use their bicycles for recreation as opposed to accessing working or shopping destinations.

Jersey City has mapped its land use on a zoning map. Land use for each of the records was determined by looking up the accident location on the zoning map. Those accidents which occurred in areas zoned for some level of residential use were coded 1. Industrial and commercial uses were coded -1. A distribution of residential use is shown in Figure 4.41. As shown, slightly over a third or 35% of the accidents occurred in nonresidential zones.

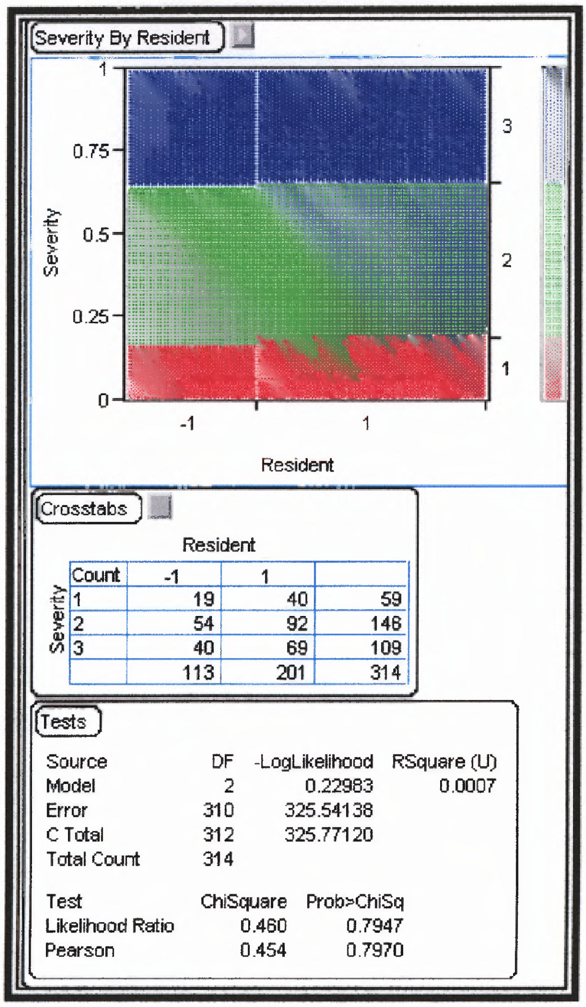
The mosaic plot shown in Figure 4.42 shows injury severity to be fairly constant irrespective of the zoning. This may be a result of the fact that much of Jersey City was built prior to the creation of zoning regulations. The present land use may differ greatly from the formal zoning map. In any case, Resident with a p-value of 0.7970 cannot be considered a candidate for inclusion into the model.



Level	Count	Probability	Cum Prob
-1	113	0.35987	0.35987
1	201	0.64013	1.00000
Total	314		

2 Levels

Figure 4.41 Resident Distribution.



	Count	Resident -1	Resident 1	
Severity	1	19	40	59
	2	54	92	146
	3	40	69	109
		113	201	314

Tests

Source	DF	-LogLikelihood	RSquare (U)
Model	2	0.22983	0.0007
Error	310	325.54138	
C Total	312	325.77120	
Total Count	314		

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	0.460	0.7947
Pearson	0.454	0.7970

Figure 4.42 Resident Severity Mosaic.

4.7 Temporal Factors

4.7.1 Weather

Although weather is not a true model parameter, it is being considered for inclusion in the model as a confounder. It is important to examine the effects of weather on the outcome of the accident because if all other factors are equal, yet the accident severity varied due to the weather, it would be incorrect to attribute all of the variance to non temporal factors. If weather is ultimately included in the final model, it could be held constant to determine which are the safest roads in dry conditions and then at a second level, in wet conditions. Weather is derived from the NJDOT's accident database Weather field shown in Table 4.5.

Table 4.5 Weather Aggregation

NJDOT	Weather	Description
1	1	Dry
2	2	Rain
3	2	Snow
4	2	Fog
5	2	Other

Source: NJDOT accident database

Out of 314 records, only 4 had blanks Weather fields. These were assumed to be dry. Only one accident occurred during a snowy condition. A distribution of the weather conditions is shown in Figure 4.43. Over 90% of accidents occurred in dry conditions. Not surprisingly, this confirms that few bicyclists enjoy riding in the rain.

Although the mosaic plot, as shown in Figure 4.44, does indicate a visible reduction of injury accidents during good weather, the effect is not pronounced enough to produce a high significance. With a p-value of 0.6034, the variable cannot be entered into the model.

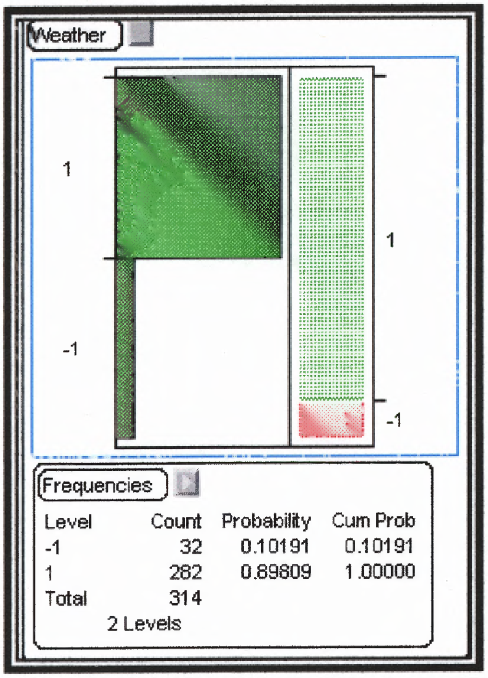


Figure 4.43 Weather Distribution.

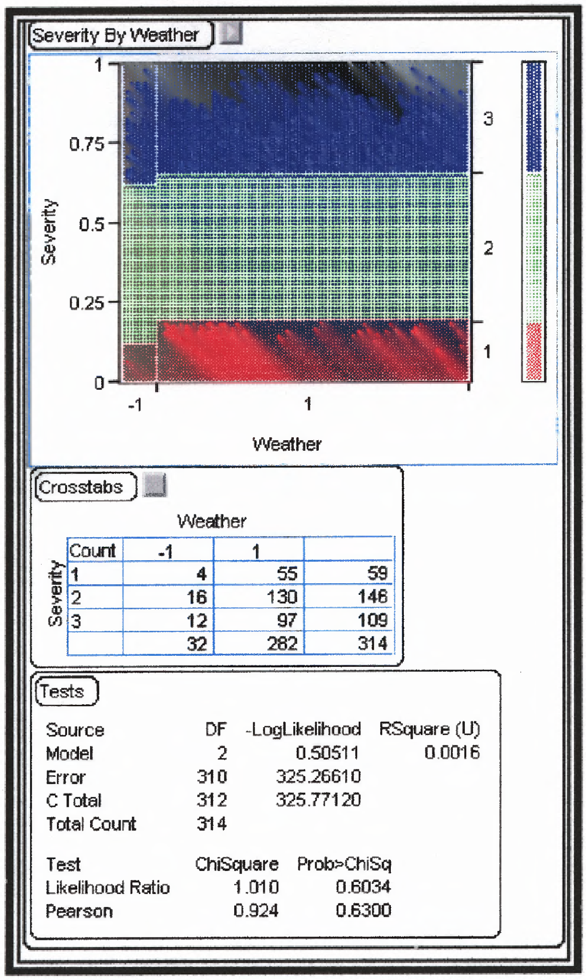


Figure 4.44 Weather Severity Interaction.

4.7.2 Daylight

Daylight was derived from the Lighting Condition variable in the NJDOT Accident database as shown in Table 4.6. If the accident occurred during full daylight, the variable was coded 1. At all other times, irrespective of the lighting condition, it was coded as -1. Six accidents had a blank value for daylight, but this value was imputed from time of day and month of year.

Table 4.6 Daylight Aggregation

NJDOT	Daylight	Description
1	1	Daylight
2	2	Dawn or Dusk
3	2	Dark (Street lights on)
4	2	Dark (Street lights off)
5	2	Dark (No Street lights)

Source: NJDOT accident database

The distribution of daylight accidents is shown in Figure 4.45. Out of 314 accidents, roughly one third, or 34.5% occurred during non-daylight hours. Surprisingly, the mosaic plot in Figure 4.46 shows an actual increase in serious accidents during daylight hours. Possibly, nighttime riders are more experienced and less likely to be recreational riders. Daylight, with a borderline p-value of 0.1074 has a definite chance of ultimately becoming a model variable.

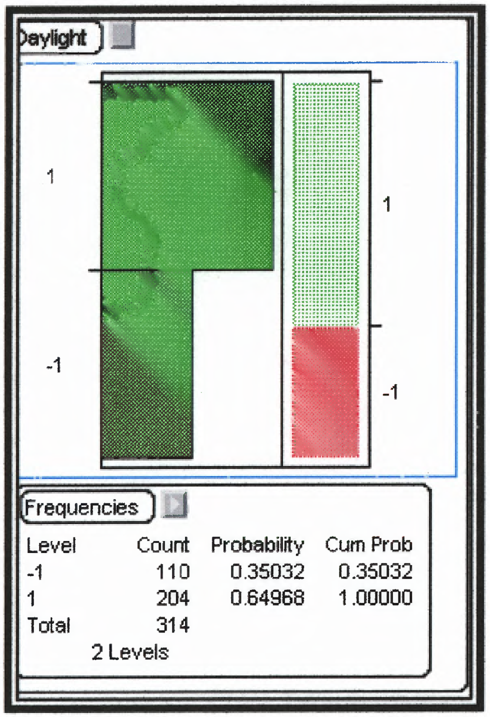


Figure 4.45 Daylight Density.

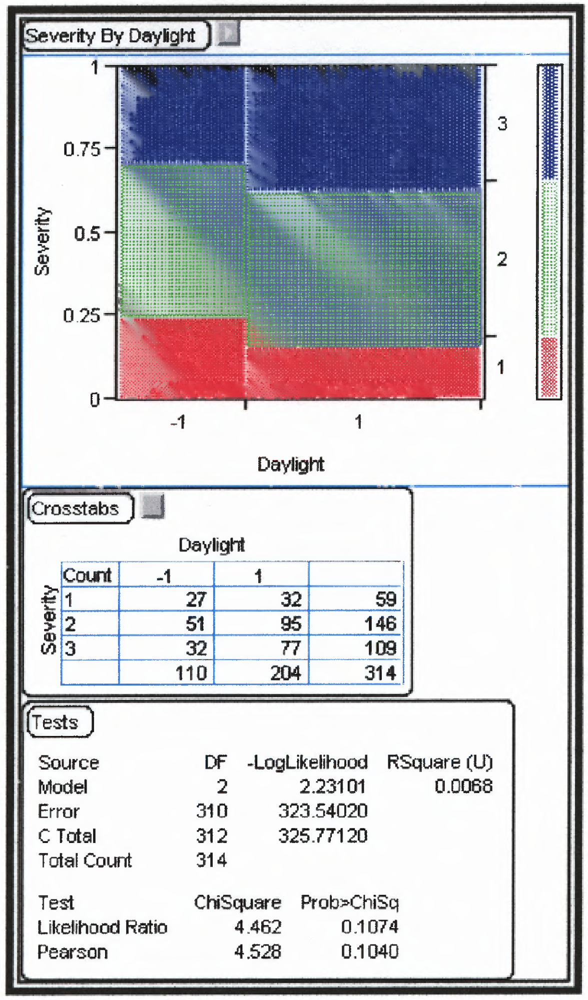


Figure 4.46 Daylight Severity Mosaic.

4.8 Operator Factors

4.8.1 Age

Age is not truly a continuous variable although it may have a wide spectrum of values between 0 (baby on bike seat) and 100, it need not be integer. Using Age as a continuous variable implies a ratio relationship that does not exist. The severity of an accident for a 30 year old person is not expected to be less than the accident of a 35 year old person. The true question contracts to "Was the victim a child or not?" The question of when the child is an adult must be defined differently than the age of legal majority. A seventeen year old, 6' tall, 165 pound youth would not be expected to suffer a greater injury than a twenty-five year old 6' tall, 165 pound adult. It would be better to establish some criteria for height or weight instead of age, but that information is not provided in police reports.

To determine the most appropriate age breakpoint for classification of Child, a series of unilateral models were fit with severity and age as a categorical variable with breakpoints of increasing age, similar to Stewart's (1996) CART method. To start, a unilateral model was fit and a p-value obtained for the dichotomous variable, "Is the victim's age less than or equal to 7 years? A second curve fitting was next done for the question, "Is the victim's age less than 8 years?" which yielded a p-value of 0.8559. These cumulative age variables were tested for 14 unilateral models ending with, "Is the victim's age less than or equal to 20?" The results are shown in Table 4.7. For each age breakpoint, a p-value was obtained. From a comparison of the p-values, fifteen years old was established as the breakpoint as it had the highest significance with a p-value of 0.1150. Victims younger than 16 were classified as children, Child =1. All others were classified as not children, or Child = -1.

Table 4.7 Age Breakpoint Determination

Age	P-Value	
7	0.7140	
8	0.8559	
9	0.4085	
10	0.7428	
11	0.1994	
12	0.3804	
13	0.1524	
14	0.1446	
15	0.1150	Break Point
16	0.6484	
17	0.8027	
18	0.9302	
20	0.9380	

The NJDOT accident database provided the age of the victim in several fields. In some records, the victim's age was provided in the occupant injury fields. On other records, the date of birth was provided as the vehicle operator. The age data had to be extracted from each of these fields after first determining which vehicle was driven by a bicyclist or which occupant was the bicyclist. A number of missing fields had to be obtained by obtaining the original accident report. Ten of the 314 accident records had no Age information. The remaining 9 records were imputed in the following manner. Since 105 of the 305 records or 34.4% were classified as children, the unknown 9 records were designated, 3 children, 6 adult. This Child distribution is shown in Figure 4.48.

The mosaic plot shown in Figure 4.49 shows a pronounced reduction in severe accidents which involve children. This observation may be anticipated because children frequently get into accidents while playing in their residential neighborhoods. These accidents are not as severe as those which occur while traveling in traffic.

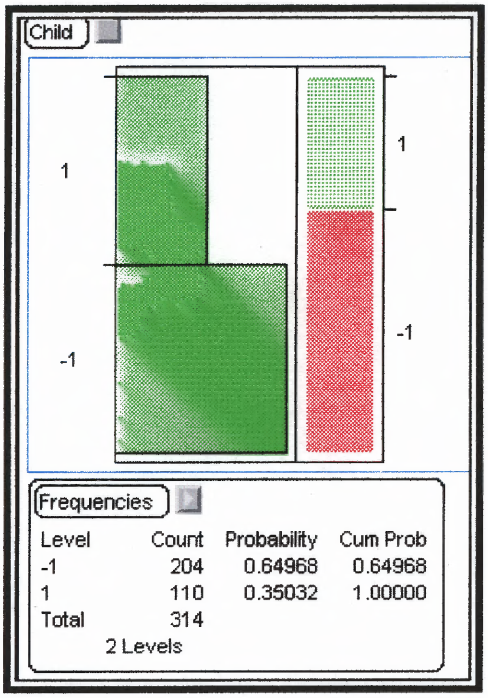


Figure 4.48 Child Distribution.

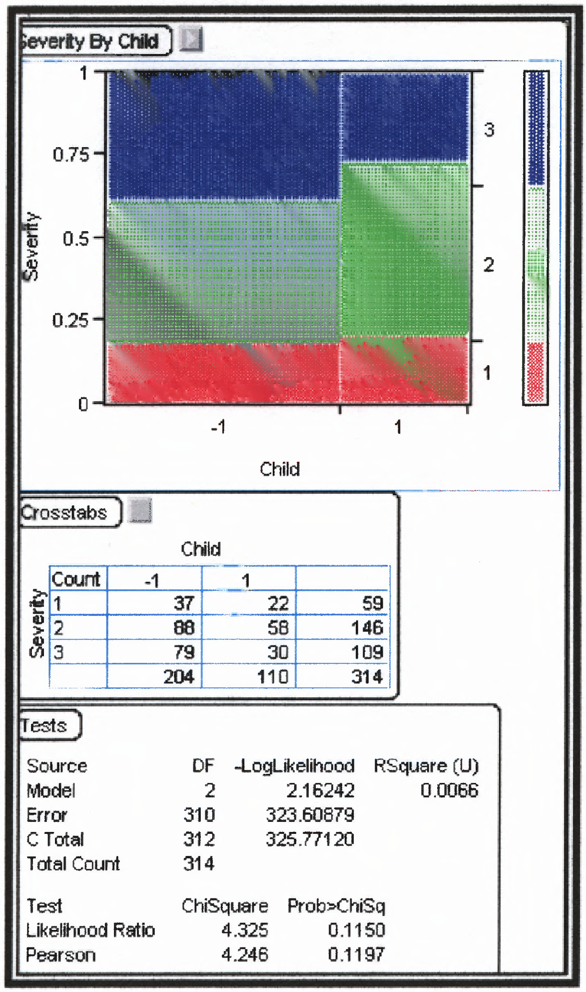


Figure 4.49 Child Severity Mosaic.

4.8 Data Summary

The data collection effort examined fifteen true variables (Width, Pavement, One_Way, Grade, Volume, Parking, Bus, Highway, Density, Truck, Signal, Resident, Curve, Speed, and Road_Div) and four confounders (Income, Child, Daylight and Weather). Of these variables, only four (Width, Pavement, One-Way and Income) met the prescribed significance level of 0.10. Another twelve variables (Grade, Volume, Parking, Bus, Hwy, Density, Truck, Child and Daylight) did not meet the significance level, but will be included in the model building effort, in the event that they interact positively without variables which increases their significance. Six variables (Signal, Resident, Curve, Speed Road_Div and Weather) were so insignificant with p-values greater than 0.50, that they were eliminated from further consideration. Several variables that were discussed in the Literature Review Section, such as helmet use, alcohol involvement and gender were not included in this section due to scarcity of data.

CHAPTER 5

MODEL DEVELOPMENT

5.1 Selection Criteria

Logistic software packages are available to quickly estimate the best coefficients for a model's intercept and parameters which will maximize the model's likelihood of predicting the observed data. The algorithms within the software perform this optimization only on a set of user provided variables. It is the user's responsibility to select the variables which should be included in the model. This task of determining which variables should be included in the model and which should be omitted is both the most crucial and the most arduous task.

There are numerous models that can be developed from a given set of dependent variables. If the model is limited to linear combinations to determine the logistic function, then:

$$Z = \alpha + \sum_{k=1}^{n} \beta X_k \tag{5.1}$$

Where:

Z = logit
α = intercept
β = variable coefficients
X_k = expanatory variables

There are n! possible combinations of X_k within the model. For example, if n =3, all of the following 3! (six) models could be considered:

$$\begin{aligned}
Z &= \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \\
Z &= \alpha + \beta_1 X_1 + \beta_2 X_2 \\
Z &= \alpha + \beta_1 X_1 + \beta_3 X_3 \\
Z &= \alpha + \beta_2 X_2 + \beta_3 X_3 \\
Z &= \alpha + \beta_1 X_1 \\
Z &= \alpha + \beta_2 X_2 \\
Z &= \alpha + \beta_3 X_3
\end{aligned} \tag{5.2}$$

For the bicycle safety model, with nineteen variables under consideration, 19! or 3.56×10^{14} combinations could be considered. If higher order terms and interactions are included, the number of viable models become practically infinite.

Obviously, a systematic method is needed to rationally identify the most likely combinations lest the task becomes unmanageable. Before proceeding, it is essential to define the precise criteria for choosing one model as the best model. One must resist the initial temptation to seek the model which produces the maximum likelihood value as defined in Equation 3.17 without regard to the significance of the individual independent variables. This approach would be incorrect as even the most insignificant variables and even correlated variables can increase the likelihood of the predictions. The resulting model would be overfit to the data set from which it was derived. Its capacity to make valid predictions for other data sets would be questionable.

Thus, the criterion definition for the "Best Model" is refined to seeking that model which maximizes the likelihood function while using only variables which meet a specified level of significance. Due to the inexact nature of human injuries and severity outcomes, a 90% confidence level should be appropriate. However, to satisfy reviewers who seek a more stringent criterion, a second model will be identified meeting a 95% confidence level.

A final criteria refinement of parsimony using the least amount of variables will not be the goal of this effort. Typically, researchers are motivated to obtain models that require the least amount of data because smaller data sets are easier and less expensive to acquire and maintain when using the model for application purposes. In this modeling effort, however, the goal was to actually maintain as many variables as possible in order to evaluate the magnitude and direction of their impact on the safety of a route. It was also observed that whenever a variable was added to the model, the log likelihood value was reduced. Thus, the best model for research purposes is the model that includes as many significant variables as possible and maximizes the log likelihood, although it may not be the most parsimonious model that would be better suited for general practice. In summary, the criteria for developing three models for discussions are as follows:

Model 0: Complete

Model fit using all nineteen explanatory variables in the data base regardless of their significance.

Model 1:Best 90%

Maximizes Likelihood equation using maximum number of variables whose intercept and parameters meet a 90% confidence level (Wald Chi Square tests $\leq$0.10).

Model 2: Best 95%

Maximizes Likelihood equation using maximum number of variables whose intercept and parameters meet a 95% confidence level (Wald Chi Square tests $\leq$0.05).

5.2 Variable Selection

The output for Model 0 is shown in Appendix A. It was built using all of the nineteen variables. The whole model meets the goodness-of-fit test with its log-likelihood test of its intercept and covariates equal to 0.0005, well below the target limit of 0.1. An examination of its statistics for the individual beta tests, however, reveals that the p- value of many of the variables grossly exceed the acceptable limits. Clearly, a parking variable with a p-value of

0.9973 cannot be considered a significant variable and does not belong in the model. It is tempting to delete all variables with p-values exceeding 0.10. In this haste good variables may be unfairly eliminated due to unfavorable interactions with the insignificant variables.

One popular approach, is entitled Forward Selection. In this selection method, variables are fit unilaterally with the response variable to determine their initial significance. Then, one at a time, in order from most significant (lowest p-value) to least significance (highest p-value), the variables are added to the model. If their addition improves the model's overall goodness of fit and individual variables continue to meet significance limits, the variable is retained in the model. If not, the variable is removed. When a new variable is retained in the model, the previously rejected variables are given another chance of inclusion to determine if through interaction with the recently added variable, they now meet the significance levels. The process continues until all variables are either included or rejected. Unfortunately, the Forward Selection method can produce models that fail to include variables which are not significant on their own, but are significant when combined with other factors. Because these borderline factors are never entered into the model at the same time, this interaction is never observed.

An alternative selection approach is entitled Backward Elimination. In this approach, the full model is initially fit with all variables under consideration. Then, each of the variables' p-values is examined. The variable with the highest p-value is dropped and the model is refit. This process continues until all remaining variables meet the confidence level. Using this method, Model 1 included in Appendix B, was developed including nine variables (Width, Volume, Density, One-way, Grade, Pave, Hwy, Truck and Daylight) and is included in Appendix B. An in depth discussion of the significance of the model and these variables was given previously in Sections 4.3 through 4.8.

If instead of a 90% confidence level, a 95% level was sought, the entire process would be repeated using a p-value of 0.05 to allow entry of an explanatory variable into the model. Using this criteria, a model was obtained by eliminating all variables with the exception of Width, Volume, One-Way, Truck and Daylight. The SAS output for Model 2 is included in the Appendix C.

5.3 SAS Statistical Measures

With the objectives of the model building process thus defined, it is now possible to use statistical software to obtain candidate models from which to select the final model. While there are many good statistical packages for doing logit regression, SAS is highly regarded due to the efficiency and stability of its algorithms and quality of its documentation (Allison, 2001). Before comparing the results of different trials, it is beneficial to discuss the statistical measures provided by the SAS logistic procedure output. This is accomplished by providing a brief explanation of the SAS output for Model 1 contained in Appendix B.

5.3.1 Model Information

The first section confirms the basic analysis facts: the name of the data file is Bike Data; the name of the dependent variable is Severity which has three increasing levels; the number of observations which is 314 crashes; the model which is cumulative logit; and the Optimization Technique, Fisher's Scoring which uses iteratively reweighted least squares to solve the maximum likelihood equation.

The cumulative probability relationships for a categorical response variable containing three levels is defined as follows:

$$P(1) = P(Y \le 1),$$
$$P(2) = P(Y \le 2) - P(1) \quad (5.3)$$
$$P(2) = 1 - P(2) - P(1)$$

5.3.2 Response Profile and Class Level Information

The Response Profile section of the output then provides a distribution of the response variable, Severity's levels: Level 1 (PDO has 59 accidents), Level 2 (minor injuries has 146) accidents and Level 3 (serious injuries has109 accidents). It reconfirms that the software will proceed under the assumption that they are cumulative from the lower level of 1 to the higher level of 3. The class levels are defined for the six categorical variables: One-way, Grade, Pave, Hwy, Truck and Daylight.

5.3.3 Convergence Status

The Convergence Status section verifies that the optimization technique was able to produce a solution which converged within the allotted number of iterations.

5.3.4 Proportional Odds Test

The Score Test for the Proportional Odds assumption verifies whether the ordinal restrictions are valid. The model was built assuming that the log of the curve for each level of Severity was parallel but shifted by a constant intercept, α. If the proportional odds test fails, this assumption must be rejected and separate curves fit for each level of Severity. High p-values are desirable signifying that there is no reason to reject the form of the ordinal model. The reported value of 0.5062 is good.

5.3.5 Model Fit Statistics

The Model Fit statistics for Model 1 report the negative of twice the log of the maximum likelihood equation for both the intercept only, “Constrained Condition,” and the intercept and covariates, “Covariates Model.” Note, since the log-likelihood is actually a negative number, by multiplying it by -2, the maximum likelihood is actually represented by the lower number. The constrained condition equation is essentially the log of the averaged odds ratios of the severity which is equal to 651.542. It assumes that the inclusion of the covariates will have no effect on the predicted outcomes. The Covariates Model calculates the maximum likelihood using the fitted model which is equal to 615.735.

Other model fit statistics are also provided. The Akaike's information criterion (AIC) is calculated as:

$$\begin{aligned} \text{AIC} &= -2 \text{ x log-likelihood} + 2\text{k} \qquad (5.4) \\ &= 615.738 + 2{*}11 = 637.738. \end{aligned}$$

Where k is the number of independent variables and intercepts

The motivation of the AIC is to penalize the likelihood (increasing -2 log-likelihood value) as the number of variables are added to the model. The Schwartz Criterion (SC) produces an even harsher penalty with a value calculated as

$$SC = -2 \times \text{log-likelihood} + k \log n \tag{5.5}$$

$$= 615.738 + 11*\log(314) = 678.982.$$

Where n is the sample size

These measures are only valuable when the goal is to seek the most parsimonious solution.

5.3.6 Global Null Hypotheses

The Testing Global Null Hypothesis: Beta=0 section uses the log-likelihood of the intercept only equation to compare it with the fitted equation using a chi-square table for the given degrees of freedom. The difference between these two values is the chi-square value reported in the Likelihood Ratio. It determines whether the probability of such an improvement could occur randomly or in common terminology, whether the model is better than nothing. For the model to meet a 90% confidence level, the p-value must be less than or equal to 0.10. The likelihood ratio of Model 1 has a p-value of less than 0.0001 as shown in Table 5.1.

Again, SAS performs two additional tests to supply options for other researcher's preferences. The Score statistic is a function (a quadratic form) of the first and second derivatives of the log-likelihood function under the null hypothesis. The reported Score and Wald statistics are 0.0001. Generally, there is little difference between the two, but in small samples with extreme data patterns, Jennings (1986) has demonstrated that the likelihood chi-square ratio is superior.

5.3.7 Maximum Likelihood Estimates

Using the model fitting algorithms, SAS determines the estimates of an alpha for each of the intercepts and a beta for each of the explanatory variables along with their standard errors. The Wald chi-squares are computed by dividing the beta estimates by their standard error and squaring the results comparing to a chi-square table to determine their p-values which may be interpreted as the probability that their contribution to the model is random. Depending on the desired confidence limits the p-value must be below 0.10 (90%) or below 0.05 (95%). The significance of each of the model variables are discussed in depth in Section 5.4.

5.3.8 Odds Ratio Estimates

The coefficients of a logistic model are difficult to interpret. In linear regression, the coefficients can be examined for their magnitude and slope. For example, a slope of -5 implies that a unit increase in the variable would reduce the value of the overall equation five times the value of the variable. Due to the logistic transformation which was used to fit the model, this relationship with the estimated coefficients is not valid.

Instead, it is possible to interpret the impacts to the odds ratio by exponentiating these coefficients. The odds ratios reported in this section are the exponent of each of the coefficient estimates for the curve fitted for each of the severity level curves. These parallel curves are interpreted as the change in the ratio of a probability of a given severity level, for example, the shift from a PDO accident to an injury accident based on a unit change in the variable. Thus, an odds ratio less than 1.0, such as Grade with 0.466 implies that the odds of a PDO only accident of (1-0.466) or 54.4% lower than those accidents which do not occur on road sections with grade. Conversely, an accident which occurs on a section of roadway which was paved in the last ten years is (1.550-1) or 55.5% more likely to produce a PDO accident than one which occurred on a road section that had not been maintained. The confidence intervals provided are determined by the defined by the defined level of confidence set at 90% for Model 0 and Model 1 and 95% for Model 2.

5.3.9 Predicted Probabilities and Observed Responses

The model's ability to predict the observed outcome is determined by a different set of criterion than the one used to determine how well the data is fit. In linear regression, the coefficient of determination R^2 is used. Logistic models cannot be evaluated using R^2 because this value is based on the ratio between the explained error and the total error. Since logistic functions predict probabilities, not precise values, there are no error terms to evaluate.

SAS provides four measures to evaluate the model's predictive power. These measures are derived by evaluating all of the data points as a series of pairs. The total number of N pairs that exist for the bicycle accident data set with 314 records is:

$$\text{Pairs} = (314*313) / 2 = 49{,}141 \tag{5.6}$$

Those pairs with identical severity levels are eliminated leaving a total of P = 30,959 pairs in this case. The question is then asked whether the model would properly predict a higher value for the record with the higher severity level and a lower value for the crash with the lower severity level. If the answer is yes, the pair is deemed Concordant. The total percent of Concordant pairs is C or 64.6%. If the answer is no, the pair is deemed Discordant. The total percent of Discordant pairs is D or 34.5%. Pairs which have equal ratings or tied are labeled T. The total percent of tied pairs is 0.8%. The measure of concordance(C), discordance (D) and ties (T) are used to calculate the four measures of the model's predictive power. These measures entitled the Sommers'D, Gamma, Tau-a and c are useful when making comparisons between candidate models. They are defined below as:

$$\text{Sommers' D} = \frac{C-D}{C+D+T} = \frac{(0.647-.345)P}{(0.647+0.345+0.008)P} = 0.302 \tag{5.7}$$

$$\text{Gamma} = \frac{C-D}{C+D} = \frac{(0.647-0.345)P}{(0.647+0.345)P} = 0.305 \tag{5.8}$$

$$\text{Tau}-a = \frac{C-D}{N} = \frac{(.647-.345)P}{N} = \frac{.302*30{,}959}{49{,}144} = 0.190 \tag{5.9}$$

$$c = 0.5*(1+\text{Sommers' D}) = 0.5*(1+.302) = 0.651 \tag{5.10}$$

A preferable measure of a model's predictive ability is the generalized R^2 value developed by Cox and Snell (1989) which is constructed as:

$$R^2 = 1-\exp\left(-\frac{L^2}{n}\right) \tag{5.11}$$

where n is the sample size and L is the likelihood ratio chi-square which is -2 times the difference in the log likelihood of the fitted constrained (intercept only) model and the fitted unconstrained (intercept plus covariates) model. Substituting these values into Equation 5.12, produces:

$$R^2 = 1-\exp\left(-\frac{32.8354^2}{314}\right) = 1-\exp(-3.8054) = 0.9677 \tag{5.12}$$

The use of this generalized R^2 is recommended because it is based on log-likelihood, the quantity that is being maximized, it never diminishes when variables are added to the model, and its calculated values are usually quite similar to the R^2 obtained from fitting a linear probability model by ordinary least squares. Using this measure of predictive power, the model fit in Equation 5.13 with a R^2 of 0.9677 would be expected to have a strong predictive power.

5.4 Model Interpretation

Given the results of the model building effort, the logit can be constructed for the beta estimates. These estimates, p-values for the individual variables as well as the overall model are presented in Table 5.1 for all the models. Returning to Model 1 out of Appendix B, the section entitled Analysis of Maximum Likelihood Estimates provides the estimated beta for each of the explanatory variables. The logit equation derived from Model 1 is:

$$\begin{aligned} Z = {} & -0.0728\text{Width} + 0.0861\text{Volume} - 0.020\text{Density} - 0.3126\text{One\_Way} \\ & - 0.3817\text{Grade} + 0.2191\text{Pave} - 0.5174\text{Hwy} - .3965\text{Truck} - .2744\text{Daylight} \end{aligned} \tag{5.13}$$

If this model were linear, instead of logistic, it would be possible to interpret the magnitude and direction of the impact of an individual variable. For example, if this model were the result of a linear regression, a unit change in width would reduce the response variable by 7.28%. This is a logistic model and Z is not severity. From Z, the probability that Y is less than or equal to a given level of Severity can be predicted using the logistic transformation of:

$$P(Y \le y) = \frac{1}{1 + e^{-Z}} \tag{5.14}$$

Because the coefficient estimates cannot be used to directly interpreted the effects of the explanatory variables on the model, SAS provides Odds Ratio Estimates. The odds ratio is defined as the ratio of the probability of an event occurring to the probability of that event not occurring.

$$\text{O.R.} = \frac{P(Y=1)}{1-P(Y=1)} \tag{5.15}$$

Thus a variable with an odds ratio of less than 1.0 can be understood to be a section of roadway along which that variable has (O.R.-1.0)% less chance of having a PDO, Severity Level 1, accident than an accident without that condition. On the other hand, a variable with a O.R. greater than 1.0 implies that a section of roadway possessing that particular condition has a (O.R.-1.0)% greater chance of having a PDO (Severity Level 1) accident. An O.R. equal to 1.0 implies that the variable has no impact on the severity level. While none of the explanatory variables included in Model 1 have O.R. equal to 0, three variables (Pave, Highway and Grade) do contain the origin in their confidence intervals. The O.R.s will be examined for each of model's explanatory variables.

Table 5.1 Logit Model Coefficient Results

Explanatory Variable	Estimate(p-value)		
	Model 0: Full	Model 1: 90%	Model 2:95%
Intercept 1	-1.5351(0.5332)	-1.2498(0.0395)	-1.0556(0.0104)
Intercept 2	0.7999(0.7453)	1.0346(0.0877)	1.1093 (0.0071)
Speed	0.0259(0.7297)		
Width	-0.0788(0.0238)	-0.0728(0.0129)	-0.0611(0.0316)
Volume	0.1212(0.0588)	0.0861(0.0482)	
Income	-0.0328(0.0540)		
Density	-0.0217(0.0270)	-0.0200(0.0359)	
One_Way	-0.2669(0.0476)	-0.3126(0.0100)	-0.2345(0.0387)
Road_Div	0.1183(0.7700)		
Grade	-0.3078(0.1617)	-0.3817(0.0680)	-0.4568(0.0245)
Curve	-0.0871(0.7605)		
Pave	0.2318(0.0483)	0.2191(0.0568)	
Hwy	-0.6017(0.0903)	-0.5174(0.0568)	
Parking	0.0084(0.9973)		
Bus	-0.0371(0.8229)		
Truck	-0.3768(0.0152)	-0.3965(0.0076)	
Signal	0.0292(0.8094)		
Resident	0.1031(0.4132)		
Weather	0.2139(0.2653		
Daylight	-0.2956(0.0127)	-0.2744(0.0176)	
Child	0.1597(0.1810)		
LogLikelihood Ratio	0.0005	0.0001	0.0027
-2 Log L	605.829	615.738	637.374
Proportional Odds	0.0856	0.5062	0.8535

* The p-value is contained in the parentheses.

5.5 Explanatory Variables

5.5.1 Width

The odds ratio of Width is 0.930. Thus a unit or 1' increase in the width of a curb lane will be 0.930-1.0 or 7% less likely to have a PDO accident. With a p-value of 0.0129 for its chi-square test, Width is one of the more significant variables in this model. At first the direction of Width's impact may appear counterintuitive. It may be anticipated that wider streets are safer for bicyclists, yet wider streets may also produce higher operating speed. Wide lanes allow cars to pass bicyclists without changing lanes, thus creating an environment for the most fatal bicycle accident typology. Without designated bike lanes, wide roads in Jersey City have produced more serious accidents. Width ranges in the database from a minimum of 7' to a maximum of 30'.

5.5.2 Volume

The odds ratio for Volume is 1.090. Thus a unit of 1000 increase in ADT of Volume will produce (1.090-1) or 9% increase in the probability of producing a PDO accident. With a p-value of 0.0482, Volume would be included in a 95% model. As in Width, the direction of the Volume effect is counter-intuitive. As volume increases the severity decreases. Again, the underlying cause may be the operating speed. As volume increases, speed decreases. Volume (1,000's) ranges from 0.250 to 13.478 ADT.

5.5.3 Density

The odds ratio for Density is 0.980. Thus a unit increase of 1000 persons per square mile increases the probability that an accident will be a PDO accident by 2%. Increased population density would generate more car bike interactions than a sparsely populated area.

5.5.4 One-Way

The odds ratio for One-Way streets is 0.535. Thus a One-Way street is (0.535-1) or 46.5.1% less likely to have PDO accidents. It was anticipated that One-Way street having fewer points of conflict would produce less severe accidents than two way streets. This false sense of security may encourage both bicyclists and motorists to be less attentive. Bike riders may be more likely to ride on the wrong side, i.e. left side of the street, and be less cautious when turning left than on two-way streets. Also, one-way streets tend to possess wider lanes and operate at higher speed than their two way counterparts. With a p-value of 0.0100, One-Way streets is a highly significant explanatory variable.

5.5.5 Grade

The odds ratio for Grade is 0.466. Thus a roadway section with a Grade is (0.466-1) or 52.4% less likely to have a PDO accident than a flat roadway. This effect was anticipated in that steep roads produce higher bicyclist speeds on the downslope and maneuvering and line of sight problems on the upslope. However, with a p-value of 0.0680, Grade would be dropped from a 95% confidence model.

5.5.6 Pave

The odds ratio for Pave is 1.550. Thus a roadway section which was paved in the past ten years is (1.550-1) or 55% more likely to have a PDO accident than a roadway section with an older pavement surface. As anticipated, smoother riding surfaces and/or the presence of bicycle safe grates create safer bicycling conditions for bicyclists. The significance of the Pave variable is also borderline with a p-value of 0.0568 which would just miss being included in a 95% confidence model.

5.5.7 Highway

The odds ratio for Highway is 0.355. Thus, a roadway section which is on a State highway is (1-0.355) or 64.5% less likely to produce a PDO accident than an accident which occurred on a county or local road. Not surprisingly, bicyclists operating on State Route 1 or 440, are more likely to being seriously injured than a bicyclist on Grove Street, a local street. State highways are designed for and consequently posted for higher speeds. Thus, motorist on state highways operate at higher speed and do not anticipate the presence of bicyclists. The resulting high speed collisions cause more serious injuries. Highway's p-value of 0.0565 would also be considered a borderline variable at a 95% level.

5.5.8 Truck

The odds ratio for Truck is 0.452. Thus, a truck route is (1-0.452) or 54.8% less likely to produce a PDO accident than a non-truck route. Certainly, any bicycle accident with a truck would produce serious injuries. Even non-truck accidents may increase the likelihood of serious injuries as trucks reduce visibility for all vehicles. Trucks also make wide right hand turns and may not see a bicycle passing on the right in their blind spot. Based on these findings, it would be inadvisable to locate bike routes on truck routes. The truck variable is very significant with a p-value of 0.0076.

5.5.9 Daylight

The final explanatory variable Daylight is not a function of the location, but it is included as a confounder. Its effect on the injury outcome can be held as a control while calculating predictions for accident severity. The odds ratio for Daylight is 0.578. Thus, accidents occurring during the day are (0.578-1.0) or 42.4% less likely to have a serious accident than a nighttime accident. This is surprising because it would be anticipated that the better visibility during the day would produce less severe accidents. Since Jersey City is an urban area with street lights generally present at all locations, it is possible that there are less children and less recreational bicyclists during non-daylight hours. Certainly, the causes for this observation should be explored further. Daylight is a very significant variable with a p-value of 0.0176.

5.6 Predicted Probabilities

For comparison purposes and to understand the application of the model results, consider two roadway sections. Section 1 possesses all of the most favorable conditions to make it the safest section for bicyclists. The values of its explanatory values are: Width = 7', Volume = 13.478 (000) ADT and Density = 0.310 (000) population per square mile. Its roadway section is not a one-way street. It does not have a grade, has been paved during the last ten years and is not a state highway or a truck route. The accident severity is determined for evening conditions. Under these conditions the probability of is calculated as follows:

$$\begin{aligned} Z_1 &= -0.0728*7+0.0861*13.478-0.0020*0.310-0.3126*(-1)-0.3817*(-1) \\ &\quad +0.2181*(1)-0.5174*(-1)-0.3965*(-1)-0.2744*(-1) = 4.2465 \end{aligned} \tag{5.16}$$

$$Z_2 = -1.2498+4.2465 = 2.9967 \qquad Z_3 = 1.0346+4.2465 = 5.2811$$

$$P(Y \le 1) = \frac{1}{1+e^{-(2.9967)}} = 95.2\% \quad P(Y=1) = P(Y \le 1) = 95.2\%$$

$$P(Y \le 2) = \frac{1}{1+e^{-(5.2811)}} = 99.5\% \quad P(Y=2) = P(Y \le 2) - P(Y \le 1) = 99.5\% - 95.2\% = 4.3\%$$

$$P(Y \le 3) = 1.0 \qquad P(Y=3) = P(Y \le 3) - P(Y \le 2) = 1 - 99.5\% = 0.5\%$$

These calculations state that an accident which occurred on a roadway section that possessed all of the most favorable attributes, the resulting injury would have a 95.2% probability of being a PDO, a 4.3% chance of being a minor injury and a 0.5% chance of being a serious injury. On the other extreme, hypothetical scenario was envisioned where all of the factors were as negative as possible, then the predicted severity distribution would be a 0.4%

probability of being a PDO, a 3.0% chance of being a minor injury and a 96.6% chance of being a serious injury.

These probabilities show that this hypothetical scenario was created only for discussion purposes. Its existence is extremely impossible because there are no roadway sections that meet all of these conditions. State highways simply do not exist with 30' lanes experiencing only 250 ADT that are also truck routes on one-way streets.

Now that these two extreme conditions have been calculated, the question remains as to how well would the model would predict injury severity levels on independent data. Unfortunately, the data set was not large enough to retain a portion for validation. For information purposes only, the model was applied to the conditions of each of the 314 accident records will produce a probability for each of the severity levels. The severity level with the highest probability is designated as the predicted value. The predicted values are summed over each of the observed levels to produce the following cross tabulation, shown in Table 5.2.

Table 5.2 Cross Tabulations

Observed Severity Injury Levels	Predicted Severity Injury Levels: 1	2	3	Total
1	5	49	5	59
2	1	114	31	146
3	1	72	36	109
Total	7	235	72	314

Using this approach, the model overestimates the severity of PDO injuries and underestimates the magnitude of serious injuries. Only 8.5% of Level 1 and 33% of Level 3 injuries were accurately predicted. Level 2 predictions are better with an overall accuracy rate of 78.1%. Accuracy over all levels is 49.4%. While 49.4% may not sound impressive, a comparison with the one-third chance of randomly predicting the injury severity demonstrates that the model has improved predictions by an additional 16.1%.

Table 5.2 oversimplifies the model's capabilities. By selecting a predicted value from a set of probabilities from each of the severity levels, much of the predictive power is lost. Imagine two hypothetical predicted probability distributions that both predict a minor injury, but possess the following probability distributions for the [P(Y=Level 1), P(Y=Level 2), P(Y=Level 3)] of (0.49, 0.51, 0.0) and (0.0, 0.51, 0.49). Clearly the first prediction is a safer street with a lower expected value of injury severity than the second scenario. The shortcoming of relying on cross tabulations is that once the prediction is made, the information contained in the probabilities distributions is lost.

Another measure of the model's performance is its Log-Likelihood. The log of the predicted probabilities of the observed levels are summed over each of the 314 observations. If, for example, the observed level was Level 1 and the model made a perfect prediction, the probability of a Level 1 would be 100% and its log likelihood would be Ln(1) or 0. If each of the predictions for the 314 accidents were perfect, the sum of the log likelihoods would be 314*0 or 0. Conversely, if each of the predictions were completely wrong, they would have a probability of 0. Logs of 0 are not possible, but as the probability approaches 0, its log approaches negative infinity. Summing these infinitely low numbers over all records produces an infinitely low number. Thus, as the log likelihood decreases and approaches zero, the more effective is the model. The model's prediction over the 314 records produced a total log-likelihood of 615.738/-2 or 307.86 with an average of -.9805, or an average predicted probability of $e^{-.9805}$ or 0.3751, not much better than a random probability of 0.3333.

Probably, the most effective measure that incorporates the full spectrum of the predicted probabilities is the expected value. Constructing the expected value as

$$E[Y] = \sum P(Y = y) * Y \tag{5.17}$$

Returning to the two extreme cases discussed at the beginning of this section, the expected value for the most favorable conditions is computed in Equation 5.18. Thus the expected severity for an injury which occurred on a roadway section possessing all of the most favorable conditions would be SI = 1.333 or approximately Level 1 (PDO.)

$$E[Y] = 0.7073 * 1 + 0.2519 * 2 + 0.0408 * 3 = 1.333 \tag{5.18}$$

Conversely, if all of the conditions for a roadway section happened to be the most unfavorable, a parallel computation would result in an expected value of 2.997, an almost absolute certainly that the injury would be of the highest level. The expected value for the most unfavorable conditions is computed in Equation 5.19.

$$E[Y] = 0.0003 * 1 + 0.0027 * 2 + 0.9970 * 3 = 2.997 \tag{5.19}$$

If a similar calculation is performed on each of the accident records, each of the expected values could be calculated. Averaging the expected values over each of the Severity levels produces expected values for Level 1, Level 2 and Level 3 of 2.041, 2.205 and 2.465 respectively. Clearly, the model is able to distinguish between each level.

Ultimately, the goal is to use the model for comparing one prospective route to another. For this reason, the expected value is the best measure for making these public policy decisions. This expected value method will be used in Chapter 6 to compare sections of bicycle routes in Jersey City.

CHAPTER 6

MODEL APPLICATION

6.1 Jersey City Transportation Profile

As stated in the data section, Jersey City located in Hudson County, New Jersey was chosen for the study location because of its urban character, the author's familiarity with the street network and the availability of data. Aside from these practical considerations, there are important reasons why Jersey City is particularly suited for bicycling.

Hudson County is compact and densely populated. According to the 2000 census, 608,975 people live within a total land mass of 46.69 square miles creating a population density of 13,043 people per square mile, which is nearly 13 times the state average of 1,134.4 persons per square mile. This is the highest population density in the state of New Jersey, already the most densely populated state in the nation. This density creates many employment opportunities within "bikeable" distances. In fact, over 47% of Hudson County residents work within the county. There are many accessible retail destinations such as the Newport Mall. There are also excellent recreational attractions from the Hudson Waterfront walkway with its views of Manhattan, to Liberty State Park and access to Ellis Island.

Jersey City's superior transit system with the PATH subway, the Hudson Bergen Light Rail System and numerous bus line already produces an admirable modal split that approaches a 33% transit share in the downtown area. As development continues, even with good modal splits, additional vehicles are still added to the roadway network. Diverting a portion of these trips is desirable. Bicycle trips could replace auto trips to transit Park'n'rides. Coupled with transit, bicycling can serve as an intermodal link to jobs in Manhattan, Newark and suburban New Jersey.

In addition to all of the aforementioned reasons to bicycle, there are equally strong disincentives to drive. Jersey City's streets are congested. Trans-Hudson commuter traffic clogs its interstates and state highways as each day over 80,000 vehicles use the Holland Tunnel. Every work day roadways in Jersey City including the New Jersey Turnpike, New Jersey Route 1&9 Truck, the Pulaski Skiway and Tonnele Circles exceed capacity. Trans-Hudson traffic destined for New York spills onto the City's street network already congested with vehicles traveling to local destinations. Normal operating speeds of less than 10 miles per hour are common. Given the operating conditions of Jersey City streets, a bicyclist can complete his trip in less time than a car.

Even when the vehicle is not moving, Jersey City motorists face obstacles. Many of the residents of pre-automobile neighborhood developments have no access to off-street parking. Parking garages and lots are expensive, charging as much as $20 per hour in the waterfront commercial districts. The little on-street parking remaining after reductions for bus stops, fire hydrants, loading zones and handicap parking is frequently metered or restricted by the parking permit program. High accident and auto theft rates and, as a result, high auto insurance rates, make auto ownership in Jersey City an expensive proposition. In 1998, CNN named Jersey City as the auto-theft capital of the nation with 1 out of every 36 vehicles stolen.

Not to be overlooked is the most important factor in the bicycle mode split equation - the population from which the potential bicyclists are drawn. Jersey City has a high percent

(38.5%) of immigrants. An immigrant, coming from countries where bicycling is a dominant mode of transportation, with some encouragement is likely to continue riding in his new home. Also, stricter driver license regulations for foreigners have increased the difficulty of obtaining driver licenses.

As demonstrated in the description above, Jersey City is an attractive place to bicycle. Car operation is difficult and costly. The population has a large concentration of potential bicyclists. Then, why aren't more people bicycling? The reasons stated nationally for not bicycling include no secure place to leave bicycles, too dirty and too dangerous (Goldsmith, 1992). Jersey City could do much to change both the perception and the reality of bicycling risks. Through the adoption of a bicycle master plan and through its implementation with signage and striping, bicyclists could ride, confident in knowing that their local officials had provided the safest route for their trip.

6.2 Jersey City Bicycle Plan

The Transportation Policy Institute, a unit of the Alan M. Voorhees Transportation Center within the Edward J. Bloustein School of Planning and Public Policy at Rutgers University, in conjunction with the Jersey City Division of Planning developed a Jersey City Bicycle Master Plan (Rutgers, 2000). The goal of the plan was to develop a comprehensive bicycle program in Jersey City by providing guidelines for bicycle lanes, shared use lanes, bicycle parking and bicycle friendly community development. The plan provided a network of recommended bicycle routes for Jersey City to incorporate, over time, into its existing road system.

The planners defined the benefits of bicycling. For the bicyclist, they contended that the bicycle was fast in maneuvering through traffic jams and required little time to locate parking spaces. It was convenient as most destinations within the city could be reached in less than twenty minutes from the city center at Journal Square. It was healthy exercise. It was inexpensive to operated since the cost of acquisition and maintenance was much lower than that for an automobile and it does not need insurance, registration and fuel.

The report also noted that increased bicycle usage benefits non-cyclists. Diverting automobile trips to bicycle trips reduced congestion, air pollution and wear on roads. Allocations of land for highway expansion and parking garages also could be reduced. Bicycle usage provided access to jobs thereby reducing unemployment.

The planners acknowledged that although Jersey City was an ideal location for functional and recreational bicycle travel, roadway safety issues and lack of secure bicycle storage areas discourage its use. Their goal was to:

1. Encourage bicycle use in the city
2. Reduce conflicts between bicycles, motor vehicles and pedestrians.
3. Encourage bicycle tourism

To accomplish these goals, the Plan made the following recommendations:

A. Implement a well-connected network of bicycle lanes and share use lanes that connect and bring the city's diverse neighborhoods together.

B. Install safe and visible bicycle parking facilities at all major destinations.

C. Apply for state grants explicitly available for bicycle projects.

Three criteria were used to develop the proposed bicycle route network. The first criterion focused on making the routes “destination focused,” i.e. providing access to major destinations such as PATH stations, shopping districts, parks and schools. Second, the route selected must be able to accommodate the bicycles without reduction in parking and roadway capacity. Finally, the plan sought to service the entire city with no origin more than one-quarter mile from a bicycle route. The plan did not address the safety of the routes.

A map of the recommended citywide bicycle Network is shown in Figure 6.1. As depicted, numerous routes have been designated. Many of these routes are redundant. Because of constrained funding, it is advisable to identify which routes will be the safest and what investments are needed to improve the route safety. As a demonstration of the decision making capabilities of the bicycle routes safety model developed in this research, the plan for the Jersey City Heights section of the city will be examined.

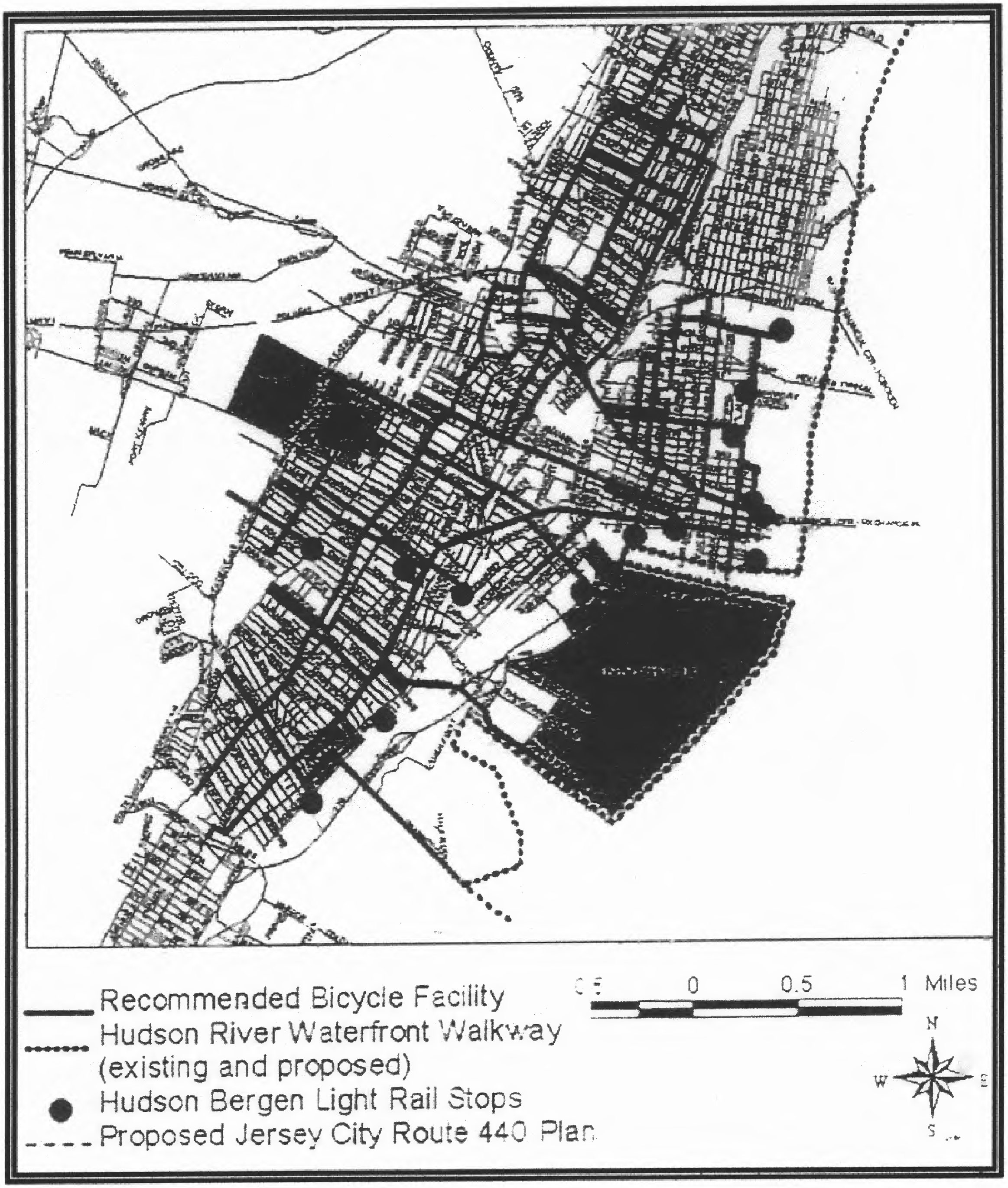


Figure 6.1 Recommended Bicycle Route Citywide.
Source: Rutgers Transportation Policy Institute, *Jersey City Bicycle Plan, Report prepared for the City of Jersey City*,, New Brunswick, New Jersey, April, 2000.

6.3 Jersey City Heights Bicycle Route Comparison

The Heights, formerly, Hudson City, was the last city to be amalgamated into the present-day Jersey City. Although it has no PATH service, it is serviced by many bus routes and ultimately the 8^{th} Street Hudson Bergen Light Rail Transit (HBLRT) station will be connected to the Heights with an elevator. Four north-south streets are recommended for inclusion by the plan: Paterson Plank Road, Palisades Avenue, Central Avenue and Kennedy Boulevard. A map of the Recommended Bicycle Network for the Heights is shown in Figure 6.2. Field observations were conducted during a non-peak period, Sunday afternoon, April 13, 2003. Each of the comparison sections were bicycled. Finally, these observations were compared with ratings obtained from the safety rating model. As shown, all four roads provide access between Franklin Street and Congress Street. The objective is this exercise is to determine which one of these streets is the safest choice for a bicycle route.

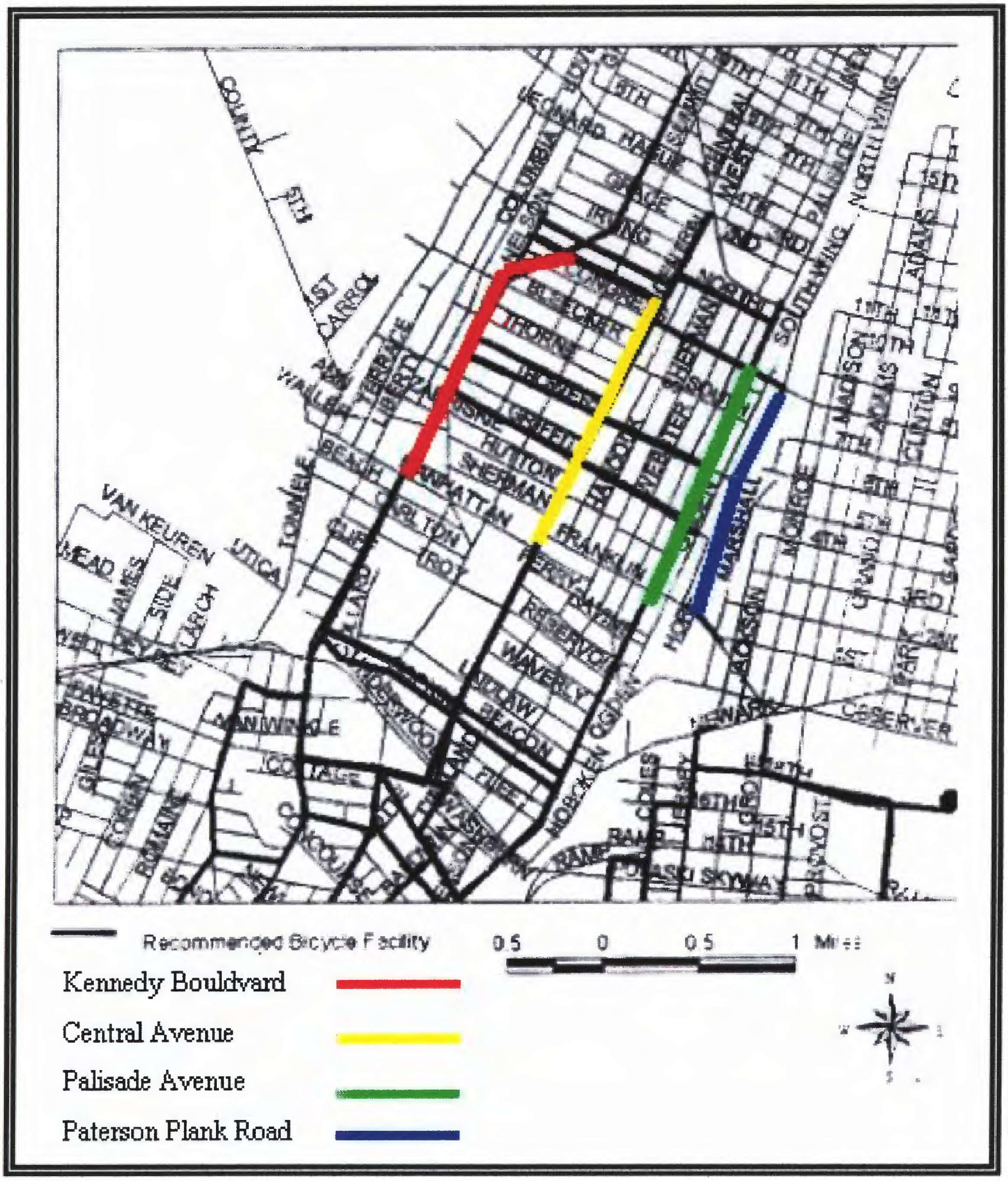


Figure 6.2 Recommended Bicycle Network - The Heights.
Source: Rutgers Transportation Policy Institute, *Jersey City Bicycle Plan, Report prepared for the City of Jersey City*,, New Brunswick, New Jersey, April, 2000.

6.3.1 Kennedy Boulevard

Figure 6.3 Kennedy Boulevard.

Kennedy Boulevard, also known as County Route 501, is a two-way, four-lane arterial beginning at the Jersey City /Bayonne border at milepost 26.7 and extending to milepost 32.7 at the Jersey City West New York border. As shown in Figure 6.3, it is primarily residential, but also includes retail and small commercial offices. It is zoned R-1, single family residential. Although its speed limit is 25 mph, most vehicles travel between 30 and 35 mph. On-street parking is allowed on both sides of the street. Most intersections are signalized.

The road is relatively flat with a fairly good surface, free from potholes and rutting. The No. D55/WS, H1, H2 and H5 buses stop on Kennedy Boulevard. Kennedy Boulevard is not a truck route.

Bicycling on Kennedy Boulevard was not a comfortable experience. Despite the option of the second lane, cars frequently remained in the curb lane to pass. Passing was done at high speed and with tight lateral clearances between the motor vehicle and bicyclist. The general perception was that the motorists were "in a hurry" and had no time to be courteous to bicyclists. Frequent traffic signals which were synchronized for motorists forced bicyclists to stop frequently. There was a strong temptation to "run the light" especially at intersections without cars visible at the cross streets.

The consensus was that Kennedy Boulevard could be a safe bicycling route if it were supported by a driver education program. Informative signage encouraging drivers to be courteous to bicyclists and to shift left when passing would greatly contribute to bicycle safety.

6.3.2 Central Avenue

Figure 6.4 Central Avenue.

Central Avenue, as shown on Figure 6.4, is a busy two-way, two-lane retail arterial. Many shops, small offices and restaurants give it a typical urban bustle. Parking is primarily on-street metered with numerous parking maneuvers per hour. Traffic signals are frequent. The grade is relatively flat. The street surface was excellent as it had been newly resurfaced. It is zoned Neighborhood Commercial. The No. 87 and privately operated Central Avenue bus lines serve Central Avenue. Its operating travel speed hovered between 10 and 15 mph allowing the bicyclists to keep up with and even pass moving traffic. Central Avenue is not a truck route.

The field observation confirmed that the street was narrow, at times forcing cars to wait for gaps in opposing traffic before passing the bicyclists by shifting close to the double yellow centerline. The narrow street also discouraged the bicyclist from easing forward between parked cars and queued cars at red lights. Numerous parking maneuvers required constant attention to avoid collisions with opening doors. Yet, the high level of activity on the avenue seemed to force both the motorist and bicyclist to be more attentive. Central Avenue drivers seemed resigned to the numerous delays from parked cars, buses and pedestrians. No one appeared to try to "make good time." Surprisingly, the level of activity made the overall experience of bicycling on Central Avenue both exciting and enjoyable.

6.3.3 Palisade Avenue

Figure 6.5 Palisade Avenue.

Palisade Avenue, as shown on Figure 6.5, is a two-way, two-lane arterial. Its uses are primarily residential, although some retail and small offices are present. Parking is primarily on-street metered with moderate level of parking maneuvers per hour. Traffic signals are less frequent than Kennedy Boulevard and Central Avenue. The grade is relatively flat. Sections of roadway were in poor condition. It is zoned R-2 multi-family residential.

The No. D99S, 67, 68, 84, 86, and 87 buses serve Palisade Avenue. Palisade Avenue is not a truck route. Palisade Avenue is not a truck route. Its operating travel speed ranged between 20 and 25 mph.

Similar to Central Avenue, field observations confirmed Palisade Avenue to be very narrow. Passing motorists frequently waited for gaps in opposing traffic before squeezing next to the centerline. However, with the absence of Central Avenues marketplace ambiance, the Palisade Avenue motorists seemed less patient. On occasion, the presence of potholes forced the bicyclists to swerve into traffic at inopportune times.

6.3.4 Paterson Plank Road

Figure 6.6 Paterson Plank Road.

Paterson Plank Road Avenue, as shown in Figure 6.6, is a two-way, two-lane retail arterial. Its uses are primarily industrial with limited access points. Parking is not allowed. The only traffic signal is at the intersection of Paterson Plank Road with Congress Street. The grade is steep in several locations. The road is currently being reconstructed. It is zoned R1 and R3 multi-family mid-rise residential. Busses 85, 87 and 89 serve Paterson Plank Road. Paterson Plank Road is a truck route. Its operating travel speed during non-peak hours ranged between 30 and 35 mph.

Bicycling south on Paterson Plank Road was an exhilarating and, at times, frightening experience. The Manhattan skyline created a dramatic vista to the east and the cliff of the Palisades on the west, made this a exceptionally scenic route. Traveling downhill, speeds exceeding the 25mph speed limit were reached. It was a pleasurable experience, but potentially dangerous, especially in wet weather conditions. The many blind driveways posed an additional hazard. The wide carriageway would lend itself easily to the installation of bicycle lanes.

6.4 Numerical Analysis

The safety rating model was applied to the sections of Kennedy Boulevard, Central Avenue, Palisade Avenue and Paterson Plank Roads discussed previously in this chapter. Each of these sections which were bounded by Congress Street on the north and either Franklin Street or Manhattan Avenue on the south, are approximately 0.60 miles in length.

Prior to entering the data into the model, the following simplifications were made to reduce the number of calculations. First, since the expected value will be the measure for comparison, a direct equation for the routes' expected accident severity will be derived in the following manner:

$$E[Y] = \sum_{y} y * p(y) \quad \text{(Vining,1998)}, \tag{6.1}$$

Then it follows that E[Y] for a three level categorical variable is:

$$E[Y] = 1*P(Y=1) + 2*P(Y=2) + 3*P(Y=3)$$
$$P(Y=1) = P(Y \le 1)$$
$$P(Y=2) = P(Y \le 2) - P(Y \le 1)$$
$$P(Y=3) = 1 - P(Y \le 2)$$

Denoting $P(Y \le y)$ as Y

$$E[Y] = 1*Y_1 + 2*(Y_2 - Y_1) + 3*(1 - Y_2)$$
$$E[Y] = 3 - (Y_1 + Y_2) \tag{6.2}$$
$$\text{Where: } Y_1 = \frac{1}{1+e^{-(\alpha_1+Z)}} \text{ and } Y_2 = \frac{1}{1+e^{-(\alpha_2+Z)}}$$

Next, the equation for the logit and the intercepts may be simplified in the following manner from equation 5.13:

$$Z = \alpha_k - 0.0728\,\text{Width} + 0.0861\,\text{Volume} - 0.020\,\text{Density} - 0.3126\,\text{One-Way} \tag{6.3}$$
$$- 0.3817\,\text{Grade} + 0.2191\,\text{Pave} - 0.5174\,\text{Hwy} - 0.3965\,\text{Truck} - 0.2744\,\text{Daylight}$$

Where: α_1= -1.2498 and α_2=1.0346

Two of these variables (Highway and Daylight) are in practice, unlikely to be treated as variables and are more appropriately set as constants. Highway is not a true variable. It helped to explain a number of severe accidents which occurred on State highways and thus improve the predictive power of the model. However, highways will never be seriously considered as bicycle routes. Therefore the highway variable may be fixed as -1 i.e. no.

Daylight should be treated as a model confounder. It is not specific to the location, but was found to be significant in fitting the model to the data. It may be fixed at +1 (daylight) for comparison purposes. Using these assumptions, the logit model may be reduced to:

$$Z = -0.0728 \text{ Width} + 0.0861 \text{ Volume} -0.0020 \text{ Density} -0.3126 \text{ One-Way}$$
$$- 0.3817 \text{ Grade} + 0.2191 \text{ Pave} - 0.5174 (-1) - 0.3965 \text{ Truck} -0.2744 (1)$$
$$= 0.2430 -0.0728 \text{ Width} + 0.0861 \text{Volume} - 0.0020 \text{ Density} -0.3126 \text{ One-Way}$$
$$-0.3817 \text{ Grade} +0.2191 \text{ Pave} -0.3965 \text{ Truck} \qquad (6.4)$$

The constant term +0.2430 can be removed from the calculation of Z by using it to modify the two model intercepts α_1 and α_2 in the following manner:

$$\alpha_1 = -1.2498 + 0.2430 = -1.0068 \qquad \alpha_2 = 1.0346 + 0.2430 = 1.2776 \qquad (6.5)$$

Table 6.1 incorporates these simplifications in order to efficiently assign an expected levels of severity for each of the considered roadway segments.

Table 6.1 Jersey City Route Comparison

	Width	Volume	Density	OneWay	Grade	Pave	Truck	Expected Value
Coefficient	-0.0728	0.0861	-0.0200	-.3126	-.3817	0.2191	-0.3965	
Kennedy Blvd.	11	6.7	42.8	-1	-1	1	-1	1.87
Central Ave.	12	6.7	50.7	-1	-1	1	-1	1.95
Palisade Ave.	12	5.6	31.4	-1	-1	-1	-1	2.00
Paterson Plank Rd.	15	4.6	31.4	-1	1	-1	1	2.63

Examining the expected values for all four segments, it is clear that Paterson Plank Road being a truck route with a steep grade and wide lane width would be a poor choice for a bicycle route. While the three remaining roads have relatively low indices, Kennedy Boulevard would be the preferred choice by a narrow margin.

Before a final recommendation is made, particularly if other factors favor a specific route, the planner may consider "What if?" scenarios which may mitigate a selected route. In the evaluation considered above, one might consider what if both Palisade Avenue and Paterson Plank Road were paved. Both roads have received grant funding and paving plans are currently in design. Another consideration would be to reduce the lane width of Paterson Plank Road by installing a 4' bike lane. Incorporating these three mitigations would produce Table 6.2.

Table 6.2 Mitigated Jersey City Heights Route Comparison

	Width	Volume	Density	OneWay	Grade	Pave	Truck	Expected Value
Coefficient	-0.0728	0.0861	-0.0200	-0.3126	-0.3817	0.2191	-0.3965	
Kennedy Blvd.	11	6.7	42.8	-1	-1	1	-1	1.87
Central Ave.	12	6.7	50.7	-1	-1	1	-1	1.95
Palisade Ave.	12	5.6	31.4	-1	-1	1	-1	1.84
Paterson Plank Rd.	11	4.6	31.4	-1	1	1	1	2.41

Now paving Palisade Avenue improves its index and is preferable to Central Avenue and Kennedy Boulevard. Paterson Plank Road, due to its grade and lower volume, still is the least preferable choice of the four roads. Based on this analysis, Palisade Avenue would be the preferred as the safest route. The other routes may be recommended for convenience or aesthetics, but a mapping of the area would designate Palisade Avenue as the safest route.

This section has demonstrated how the bicycle route safety rating model can be applied to rank alternative road sections. Route ratings can be developed for routes consisting of many linked non-homogenous sections by computing a weighted average rating based on segment length.

CHAPTER 7

CONCLUSIONS AND FUTURE RESEARCH

7.1 Summary

The objectives of this dissertation were to develop a practical model to rate the safety of bicycle routes. The rating would provide bicycle route planners with a tool for the comparison of alternate routes. It would also allow planners to formulate capital programs to improve bicycle routes by enabling them to conduct cost benefit analyses on selected factors which impact bicycle safety. The model developed used the following multivariate ordinal logistic transformation to define the bicycle route safety rating (RS) as:

$$RS = 3 - (Y_2 + Y_1) \tag{7.1}$$

$$\text{For } k = 1,2 \Rightarrow Y_k = \frac{1}{1+e^{-Z_k}}; \quad \text{For } k = 3, Y_k = 1$$

The accident data for bicycle accidents in Jersey City, New Jersey for the period (1997-2000) was used to fit the logit (Z) and the intercepts (α) for the cumulative probability distributions:

Where:

$$Z_k = \alpha_k - 0.0728\ \text{Width} + 0.0861\text{Volume} - 0.0020\ \text{Density} - 0.3126\ \text{One-way} - 0.3817\ \text{Grade} + 0.2191\ \text{Pave} - 0.3965\ \text{Truck} \tag{7.2}$$

Y_k = Cumulative probability of an injury severity level k accident

k = 1 (Property damage only), 2 (minor injury) or 3 (serious injury)

α_1, α_2 = -1.0068, 1.2776, intercepts shifting the logit to severity level k+1

Width = Traveled lane width (feet)
Volume = Motor vehicles per day (1,000 ADT)
Density = Population density (persons per square mile)
One-way = Indicates whether street is one-way or two-way
Grade = Indicates presence of a vertical grade, either upgrade or downgrade
Pave = Indicates whether road has been paved in past 10 years
Truck = Indicates presence of a truck route

Each of the model's variable coefficients is statistically significant at a 90% confidence level. The sign and the magnitude of the model's variable coefficients provide route planners and bicyclists with a new understanding of factors which increase or reduce the safety of chosen routes. Some of the recommendations appear to confirm the obvious. For example, new pavement that included the installation of bicycle safe grates are desirable factors, steep grades and truck routes are dangerous. Other recommendations provide new insights into the realities of bicycling in urban areas. As an example, contrary to common belief, neither increased lane width nor reduced traffic volume increases bicycle safety. This research has demonstrated that in fact, reduced lane width and increased traffic volume reduces the injury severity, possibly by calming traffic.

An important conclusion to be drawn from this understanding of the factors which affect route safety is the acceptance that bicyclists need not be relegated to deserted, circuitous routes in a city's outskirts. Bicyclists may safely choose direct routes through the city center. Bicyclists need not be segregated from the traffic stream. Through wise route choice and public education, bicycles could safely become an important mode of urban transportation.

7.2 General Applicability

The model developed in this research has demonstrated its usefulness in predicting the safety of proposed bicycle routes in Jersey City. Can this model be applied to other cities? Can it be applied to other regions of the country? Can it be applied to rural areas? These questions can only be answered through the study of other accident databases. The model's predictive properties can be evaluated by comparing its success in predicting the observed injury severities.

The answer to the questions of the model's general applicability may be no. The model was developed by fitting it to four years of Jersey City data (314 accidents) which determined both the selection of the variables and the values of their coefficients. To achieve a model which could be applied to other locations across the state or across the nation, a larger database is recommended.

Nonetheless, the development of this safety prediction model for Jersey City has demonstrated a valid method for creating objective safety prediction models. Using this method, customized models could be developed for other communities. Statewide or national databases could produce a model with broader applications. Better results may be obtained by segmenting the data and producing separate models for rural, suburban and urban jurisdictions.

Despite the possibility that the model may be limited to either Jersey City or other mature, urban areas, the goal to obtain an objective means to predict the safety of a bicycle route has been accomplished. Prior to this research effort, such a method did not exist. Current practitioners of bicycle route planning either totally ignore route safety or erroneously use traffic volumes as a measure. Concern for the safety of the bicycling public indicates that this problem should be addressed. Further research and development is advised.

7.3 Future Research

Data limitations constrained the model development in a number of significant ways. Future research should address these issues to improve the accuracy of the model. Specific improvements needed are discussed below:

7.3.1 Injury Reporting

The severity index was established based on the NJDOT accident database field for "most serious injury" which had been completed by the police. The reliability of these assessments is limited since the assessments are made without the benefit of a medical examination. The CODES program is working to corroborate police and hospital records. At this time, New Jersey has not received CODES funding. Future modeling efforts should use medical assessments, not police reports to establish the model's response variable.

7.3.2 Speed

Speed was found to be an insignificant variable during the model fitting stage. As suggested, this finding may be a result of the fact that operating speed and posted speed may vary greatly, especially in congested areas. The significance of a number of explanatory variables such as lane width and traffic volume may be linked to their interaction with operating speed. In fact, there is the possibility that operating speed could be the most significant predictor of a bicycle route's safety. Technology exists to take field measurements of operating speed. The outcome of developing an accurate and simplified model would justify the cost and time to undertake such an effort.

7.3.3 Pavement Quality

Pavement age is used by the model to predict route safety. A better indicator would be pavement quality. At this time, however, Jersey City has not adopted a uniform paving rating system. A true pavement rating functioning as a continuous variable may have better predictive properties that the categorical "Pavement Age less than 10 years." It is recommended to repeat this model fitting method using a city with a current pavement management system. Unilateral model fitting will evaluate the significance of pavement rating on injury severity.

7.3.4 Volume

The data used for volume was spread over a twenty year period. A growth factor of 1.0% was applied uniformly. However, different portions of the city have experienced different rates of growth. The waterfront's conversion from industrial to mixed use commercial zoning has created more intense growth than that experience by more established neighborhoods. A traffic volume collection program along proposed routes would provide more uniform data. Actually bicycle volumes from which actual bicycle accident rates could be obtained would be especially valuable enabling an independent validation of the model developed in this research.

7.3.5 Intersection Data

Signalization of intersections did not prove to be a significant variable. The investigation was limited to the simplistic question of whether the intersection was signalized or not. Additional factors may be relevant when considering the impacts of signals: Are turning bays present? Is the signal semi-actuated? Does the timing have pedestrian phases?

A closer examination of the interrelationship between injury severity and intersection signalization may provide a better understanding on how the operation of signals impact bicycle safety.

7.3.6 Bicycle Facilities

As of this date, none of the bicycle facilities that were recommended in the Rutgers Bicycle Plan had been implemented. The presence of facilities such as signage and lane markings may offer bicyclists real benefits. An important variable in the model might be whether such facilities are present. A before and after study of bicycle facilities would provide an answer to this question. The variable indicating the presence of these facilities would then be examined for significance. Again the logistic method must be used because a linear regression based on total accidents could produce misleading results. Total accidents may actually increase as more bicyclists are attracted to the official improved route. The logistic modeling techniques used in this research would assess whether the accidents were more or less severe than accidents which occurred in facilities without such improvements.

7.4 Graphical Interface Systems

In Chapter 6, a demonstration was provided of the technique to apply the model to chose a single roadway segment. In practice, numerous roadway segments must be combined to produce a complete route. The model generated expected severity value could be computed for each candidate link, then combined together using a weighted average based on link length. Given numerous alternatives, the task could quickly become onerous. The effort could be automated by using a GIS interface. A data base of the key explanatory variables could be linked to the GIS map. Planners could compare candidate routes between desired origins and destinations by stringing together links to complete alternate paths and selecting the path with the lowest weighted average.

7.5 Conclusion

The entire effort pursued in this research has been to develop a tool to assess the safety of a bicycle route. Certainly, there are other considerations for bicyclists in route selection. Scenery and rest stops may be of great concern for a recreational bicyclist. However, trip duration is generally most important to a commuting cyclist. The safety ratings from the model developed in this research can be combined with these other stated preferences to determine the route that best fulfils the bicyclists' stated objectives. Truly unsafe routes could be either mitigated or eliminated from consideration. Public officials would be negligent if they were to totally ignore the safety of the route in deference to these other stated objectives. Encouraging a bicyclist to chose a route that is slightly longer or slightly less scenic is worth the extra effort. Ultimately, the goal of the bicyclist is to reach his destination without the assistance of an ambulance.

APPENDIX A

MODEL 0, FULL MODEL

This appendix contains the SAS software output for Model 0, the Full Model. This model was fit using all nineteen study variable irrespective of its significance.

The LOGISTIC Procedure

Model Information

Data Set	WORK._TMP_0	
Response Variable	Severity	Severity
Number of Response Levels	3	
Number of Observations	314	
Model	cumulative logit	
Optimization Technique	Fisher's scoring	

Response Profile

Ordered Value	Severity	Total Frequency
1	1	59
2	2	146
3	3	109

Probabilities modeled are cumulated over the lower Ordered Values.

Class Level Information

Class	Value	Design Variables 1
One_Way	1	1
	-1	-1
Road_Div	1	1
	-1	-1
Grade_	1	1
	-1	-1
Curve	1	1
	-1	-1

Pave	1	1
	-1	-1
Hwy	1	1
	-1	-1
Parking	1	1
	-1	-1
Bus	1	1
	-1	-1
Truck	1	1
	-1	-1

The LOGISTIC Procedure

Class Level Information

Class	Value	Design Variables 1
Signal	1	1
	-1	-1
Resident	1	1
	-1	-1
Weather	1	1
	-1	-1
Daylight	1	1
	-1	-1
Child	1	1
	-1	-1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
12.6500	19	0.8560

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	655.542	647.829
SC	663.041	726.566
-2 Log L	651.542	605.829

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	45.7134	19	0.0005
Score	41.7639	19	0.0019
Wald	40.5546	19	0.0028

Model 0: Full Model 12:50 Tuesday, May 20

The LOGISTIC Procedure

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Speed	1	0.1194	0.7297
Width	1	5.1121	0.0238
Volume	1	3.5708	0.0588
Income	1	3.7134	0.0540
Density	1	4.8940	0.0270
One_Way	1	3.9243	0.0476
Road_Div	1	0.0855	0.7700

Grade_	1	1.9584	0.1617
Curve	1	0.0929	0.7605
Pave	1	3.9007	0.0483
Hwy	1	2.8696	0.0903
Parking	1	0.0000	0.9973
Bus	1	0.0501	0.8229
Truck	1	5.8911	0.0152
Signal	1	0.0581	0.8094
Resident	1	0.6697	0.4132
Weather	1	1.2407	0.2653
Daylight	1	6.2147	0.0127
Child	1	1.7894	0.1810

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	-1.5351	2.4635	0.3883	0.5332
Intercept	2	1	0.7999	2.4627	0.1055	0.7453
Speed		1	0.0259	0.0749	0.1194	0.7297
Width		1	-0.0788	0.0348	5.1121	0.0238
Volume		1	0.1212	0.0641	3.5708	0.0588
Income		1	-0.0328	0.0170	3.7134	0.0540
Density		1	-0.0217	0.00982	4.8940	0.0270
One_Way	1	1	-0.2669	0.1347	3.9243	0.0476
Road_Div	1	1	0.1183	0.4045	0.0855	0.7700
Grade_	1	1	-0.3078	0.2199	1.9584	0.1617
Curve	1	1	-0.0871	0.2858	0.0929	0.7605
Pave	1	1	0.2318	0.1174	3.9007	0.0483
Hwy	1	1	-0.6017	0.3552	2.8696	0.0903
Parking	1	1	0.000841	0.2473	0.0000	0.9973
Bus	1	1	-0.0371	0.1657	0.0501	0.8229
Truck	1	1	-0.3768	0.1552	5.8911	0.0152
Signal	1	1	0.0292	0.1212	0.0581	0.8094
Resident	1	1	0.1031	0.1260	0.6697	0.4132
Weather	1	1	0.2139	0.1920	1.2407	0.2653
Daylight	1	1	-0.2956	0.1186	6.2147	0.0127
Child	1	1	0.1597	0.1194	1.7894	0.1810

Model 0: Full Model 12:50 Tuesday, May 20

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Speed	1.026	0.886	1.188
Width	0.924	0.863	0.990
Volume	1.129	0.996	1.280
Income	0.968	0.936	1.001
Density	0.979	0.960	0.998
One_Way 1 vs -1	0.586	0.346	0.994
Road_Div 1 vs -1	1.267	0.259	6.186
Grade_ 1 vs -1	0.540	0.228	1.280
Curve 1 vs -1	0.840	0.274	2.575
Pave 1 vs -1	1.590	1.004	2.518
Hwy 1 vs -1	0.300	0.075	1.208
Parking 1 vs -1	1.002	0.380	2.641
Bus 1 vs -1	0.929	0.485	1.778
Truck 1 vs -1	0.471	0.256	0.865
Signal 1 vs -1	1.060	0.659	1.705
Resident 1 vs -1	1.229	0.750	2.014
Weather 1 vs -1	1.534	0.723	3.256
Daylight 1 vs -1	0.554	0.348	0.881
Child 1 vs -1	1.376	0.862	2.197

Association of Predicted Probabilities and Observed Responses

Percent Concordant	67.2	Somers' D	0.348
Percent Discordant	32.3	Gamma	0.350
Percent Tied	0.5	Tau-a	0.219
Pairs	30959	c	0.674

APPENDIX B

MODEL 1, BEST 90%

This appendix contains the SAS software output for Model 1, the Best 90 Model. This model was fit using only the nine study variable which are significant to a 90% confidence level.

Model 1: Best 90% Confidence Level
12:50 Tuesday, May 20, 200

The LOGISTIC Procedure

Model Information

Data Set	WORK._TMP_0	
Response Variable	Severity	Severity
Number of Response Levels	3	
Number of Observations	314	
Model	cumulative logit	
Optimization Technique	Fisher's scoring	

Response Profile

Ordered Value	Severity	Total Frequency
1	1	59
2	2	146
3	3	109

Probabilities modeled are cumulated over the lower Ordered Values.

Class Level Information

Class	Value	Design Variables 1
One_Way	1	1
	-1	-1
Grade_	1	1
	-1	-1
Pave	1	1
	-1	-1
Hwy	1	1

	-1	-1
Truck	1	1
	-1	-1
Daylight	1	1
	-1	-1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model 1: Best 90% Confidence Level
12:50 Tuesday, May 20, 200

The LOGISTIC Procedure

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
8.2796	9	0.5062

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	655.542	637.738
SC	663.041	678.982
-2 Log L	651.542	615.738

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	35.8040	9	<.0001
Score	33.1810	9	0.0001
Wald	32.8354	9	0.0001

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Width	1	6.1889	0.0129
Volume	1	3.9032	0.0482
Density	1	4.4021	0.0359
One_Way	1	6.6437	0.0100
Grade_	1	3.3295	0.0680
Pave	1	3.6296	0.0568
Hwy	1	3.6298	0.0568
Truck	1	7.1173	0.0076
Daylight	1	5.6330	0.0176

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	-1.2498	0.6070	4.2396	0.0395
Intercept	2	1	1.0346	0.6059	2.9156	0.0877
Width		1	-0.0728	0.0293	6.1889	0.0129
Volume		1	0.0861	0.0436	3.9032	0.0482
Density		1	-0.0200	0.00952	4.4021	0.0359
One_Way	1	1	-0.3126	0.1213	6.6437	0.0100
Grade_	1	1	-0.3817	0.2092	3.3295	0.0680

Model 1: Best 90% Confidence Level

12:50 Tuesday, May 20, 200

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Pave	1	1	0.2191	0.1150	3.6296	0.0568
Hwy	1	1	-0.5174	0.2716	3.6298	0.0568
Truck	1	1	-0.3965	0.1486	7.1173	0.0076
Daylight	1	1	-0.2744	0.1156	5.6330	0.0176

Odds Ratio Estimates

Effect			Point Estimate	90% Wald Confidence Limits	
Width			0.930	0.886	0.976
Volume			1.090	1.015	1.171
Density			0.980	0.965	0.996
One_Way	1	vs -1	0.535	0.359	0.798
Grade_	1	vs -1	0.466	0.234	0.928
Pave	1	vs -1	1.550	1.062	2.262
Hwy	1	vs -1	0.355	0.145	0.868
Truck	1	vs -1	0.452	0.277	0.738
Daylight	1	vs -1	0.578	0.395	0.845

Association of Predicted Probabilities and Observed Responses

Percent Concordant	64.7	Somers' D	0.302
Percent Discordant	34.5	Gamma	0.305
Percent Tied	0.8	Tau-a	0.190
Pairs	30959	c	0.651

APPENDIX C

MODEL 2, BEST 95%

This appendix contains the SAS software output for Model 2, the Best 95 Model. This model was fit using only the three study variable which are significant to a 95% confidence level.

Model 2: Best 95% Confidence Level
12:50 Tuesday, May 20, 200

The LOGISTIC Procedure

Model Information

Data Set	WORK._TMP_0	
Response Variable	Severity	Severity
Number of Response Levels	3	
Number of Observations	314	
Model	cumulative logit	
Optimization Technique	Fisher's scoring	

Response Profile

Ordered Value	Severity	Total Frequency
1	1	59
2	2	146
3	3	109

Probabilities modeled are cumulated over the lower Ordered Values.

Class Level Information

Class	Value	Design Variables 1
One_Way	1	1
	-1	-1
Grade_	1	1
	-1	-1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
0.7830	3	0.8535

Model 2: Best 95% Confidence Level

12:50 Tuesday, May 20, 200

The LOGISTIC Procedure

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	655.542	647.374
SC	663.041	666.121
-2 Log L	651.542	637.374

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	14.1682	3	0.0027
Score	13.7243	3	0.0033
Wald	13.3454	3	0.0039

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Width	1	4.6190	0.0316
One_Way	1	4.2726	0.0387
Grade_	1	5.0578	0.0245

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	-1.0556	0.4119	6.5692	0.0104
Intercept	2	1	1.1093	0.4118	7.2570	0.0071
Width		1	-0.0611	0.0284	4.6190	0.0316
One_Way	1	1	-0.2345	0.1134	4.2726	0.0387
Grade_	1	1	-0.4568	0.2031	5.0578	0.0245

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Width	0.941	0.890	0.995
One_Way 1 vs -1	0.626	0.401	0.976
Grade_ 1 vs -1	0.401	0.181	0.889

Model 2: Best 95% Confidence Level

12:50 Tuesday, May 20, 200

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses

Percent Concordant	55.1	Somers' D	0.168
Percent Discordant	38.3	Gamma	0.179
Percent Tied	6.6	Tau-a	0.106
Pairs	30959	c	0.584

REFERENCES

Agresti, Alan, *An Introduction to Categorical Data Analysis*, John Wiley & Sons, New York, 1996.

American Association of State Highway and Transportation Officials, *A Policy on Geometric Design of Highways and Streets,* Federal Highway Administration, U.S. Department of Transportation, 2001.

American Association of State Highway and Transportation Officials *Manual on Uniform Traffic Control Device (MUTCD): Millennium Edition,* Federal Highway Administration, U.S. Department of Transportation, 2000.

Aldrich, John H. and Forrest D. Nelson, *Linear Probability, Logit, and Probit Models*, Series; Quantitative Applications in the Social Sciences, Sage Publications, Inc., Newbury Park, California, 1984.

Allen, D., N. Rouphail, J. Hummer and J. Milazzo, II, Operational Analysis of Uninterrupted Bicycle Facilities, *Transportation Record 1636* Transportation Research Board, National Research Council, Washington, D.C. 1998.

Allison, Paul D., *Logistic Regression Using the SAS System*, SAS Institute and Wiley, Cary, North Carolina, 2001.

Antonakos, Cathy L. Environmental and Travel Preferences of Cyclists, *Transportation Record 1438*, Transportation Research Board, National Research Council, Washington, D.C. 1994.

Ashley, C. and C. Bannister, Cycling to Work from Wards in a Metropolitan Area, 1. Factors Influencing Cycling to Work, *Traffic Engineering and Control,* Vol. 30, No. 6, 1989.

Aultman-Hall, Lisa, Fred L. Hall, and Brian B. Baetz, Analysis of Bicycle Commuter Routes Using Geographic Information Systems, Implications for Bicycle Planning, *Transportation Record 1578*, Transportation Research Board, National Research Council, Washington, D.C. 1998.

Aultman-Hall, Lisa and Fred Hall, Research Design Insights from a Survey of Urban Bicycle Commuters, *Transportation Record 1636*, Transportation Research Board, National Research Council, Washington, D.C. 1998.

Aultman-Hall, Lisa, Sidewalk Bicycling Safety Issues, *Transportation Record 1578*, Transportation Research Board, National Research Council, Washington, D.C. 1997.

Baltes, Michael R., Factors Influencing Nondiscretionary Work Trips by Bicycle Determined from 1990 U.S. Census Metropolitan Statistical Area, *Transportation Record 1538*, Transportation Research Board, National Research Council, Washington, D.C. 1996.

Barancik, J. L., & Fife, D. Discrepancies in Vehicular Crash Injury Reporting: Northeastern Ohio Trauma Study IV, *Accident Analysis and Prevention, 17(2), 147-152, 1985.*

Ben-Akiva, M., and S.R. Lerman, *Discrete Choice Analysis: Theory and Application in Travel Demand,* Cambridge, Massachusetts, The MIT Press, 1985.

Betz, Joe, Jim Dustrude and Jill Walker, Intelligent Bicycle Routing in the United States, *Transportation Record 1405*, Transportation Research Board, National Research Council, Washington, D.C. 1993.

Bovy, P. and M. Bradley, Route Choice Analyzed with Stated-Preference Approaches, *Transportation Record 1037* Transportation Research Board, National Research Council, Washington, D.C. 1985.

Brundell-Freij, K. and L. Ekman, Safety and Exposure - Aspects of the Relationship with Special Attention to Vulnerable Road-Users, Department of Traffic Planning and Engineering, Lund Institute of Technology, Lund, Sweden, 1990.

Chen, Wan-Hui and Paul Jovanis, Method for Identifying Factors Contributing to Driver-Injury Severity in Traffic Crashes, *Transportation Record 1717*, Transportation Research Board, National Research Council, Washington, D.C., 2000.

Clark, Stephen D., and Matthew W. Page, Cycling and Urban Traffic Management and Control Systems, *Transportation Record 1705*, Transportation Research Board, National Research Council, Washington, D.C. 1992.

Clarke, Andy, Bicycle-Friendly Cities: Key Ingredients for Success, *Transportation Record 1705*, Transportation Research Board, National Research Council, Washington, D.C. 2000.

Cooke, R.M., 1991, *Experts in Uncertainty, Opinion and Subjective Probability in Science*, Oxford University Press, New York.

Cox, D.R. and E.I. Snell, *Analysis of Binary Data,* Second Edition, Chapman & Hall, London, England, 1989.

Cross, Kenneth D and Gary Fisher, A Study of Bicycle/Motor Vehicle Accidents: Identification of Problem Types and Countermeasure Approaches; National Highway Traffic Administration, 1977.

Cynecki, Michael J., Grace Perry, and George Frangos, Study of Bicyclist Characteristics in Phoenix, Arizona, *Transportation Record 1405*, Transportation Research Board, National Research Council, Washington, D.C. 1993.

Data, Jeffrey W., *Bicycle Safety Evaluation.* Auburn University, City of Chattanooga, and Chattanooga-Hamilton County Regional Planning Commission, Chattanooga, TN, 1987.

Cho, Yoon-Ho and Dong-hyn Lee, An Evaluation of Bikeway by Discriminant Analysis and Analytic Hierarchy Process, *Transportation Record Preprint*, Transportation Research Board, National Research Council, Washington, D.C. 2002.

Davis, Jeffrey W., Bicycle Test Route Evaluation for Urban Road Conditions. In *Transportation Congress: Civil Engineers–Key to the World Infrastructure.* Proceedings of the 1995 Conference, American Society of Civil Engineers, San Diego, CA, 1995, pp. 1063-1076.

Dixon, Linda B., Bicycle and Pedestrian Level-of-Service Performance Measures and Standards for Congestion Management Systems, *Transportation Record 1538*, Transportation Research Board, National Research Council, Washington, D.C. 1996.

Duncan, C.S. and A. Khattak, and F.M. Council, Applying the Ordered Probit Mode to Injury Severity in Truck-Passenger Car Rear-End Collisions, *Transportation Record 1635*, Transportation Research Board, National Research Council, Washington, D.C. 1996.

Eck, Ronald W. and E. Darin Simpson, Using Medical Records in Non-Motor Vehicle Pedestrian Accident Identification and Countermeasure Development, *Transportation Record 1538*, Transportation Research Board, National Research Council, Washington, D.C. 1996.

Eddy, Nils. Developing Suitability of Roadways for Bicycle Use, *Pro Bike/Pro Walk 96 Resource Book.* Proceeding of the Ninth International Conference on Bicycle and Pedestrian Programs Resource Book, Bicycle Federation of America and Pedestrian Federation of America, 1996, pp. 310-314.

El Taller Collaborative in association with Frederic R. Harris and New Jersey Department of Transportation, *Jersey City-Route 400 Pedestrian and Bicycle Local* Planning Assistance, Concept alternatives for Key Pedestrian and Bicycle Issues, January, 2000.

Epperson, Bruce, Demographic and Economic Characteristics of Bicyclists Involved in Bicycle-Motor Vehicle Accidents, *Transportation Record 1502*, Transportation Research Board, National Research Council, Washington, D.C. 1995.

Epperson, Bruce, Evaluating Suitability of Roadways for Bicycle Use: Toward a Cycling Level-of-Service Standard, *Transportation Record 1438*, Transportation Research Board, National Research Council, Washington, D.C. 1994.

Federal Highway Administration, *Bicycle Crash Types: A 1990's Information Guide, FHWA-RD-96-163*, 1997.

FHWA, Motor Vehicle Accident Severity, *FHWA Internal Publication HHS-12*, July 6, 1992.

Forester, J., *Bicycle Transportation*, MIT Press, Cambridge, Massachusetts, 1983.

Garder, Per, Bicycle Accidents in Maine: An Analysis, *Transportation Record 1438*, Transportation Research Board, National Research Council, Washington, D.C. 1994.

Garder, Per and Lars Leden and Urho Pulkkinen, Measuring the Safety Effect of Raised Bicycle Crossing Using a New Research Methodology, *Transportation Record 1636*, Transportation Research Board, National Research Council, Washington, D.C. 1998.

Goldsmith, S., *Reasons Why Bicycling and Walking Are Not Being Used More Extensively As Travel Modes,* Publication No. FHWA-PD-92-041, Federal Highway Administration National Bicycle and Walking Study: Case Study #1, 1992, http://www.bikefed.org/PDF/CASE1.PDF.

Harkey, D.L. and J. R. Stewart, Evaluation of Shared-Use Facilities for Bicycles and Motor Vehicles, *Transportation Record 1578*, Transportation Research Board, National Research Council, Washington, D.C. 1997.

Harkey, David L., Donald W. Reinfurt, and M. Knuiman,. Development of the Bicycle Compatibility Index, *Transportation Record 1636*, Transportation Research Board, National Research Council, Washington, D.C. 1998.

Hosmer, David W. and Stanley Lemeshow, *Applied Logistic Regression*, Second Edition, John Wiley & Sons, Inc., New York, New York, 2000.

Howard, C. and E. Burns, Cycling to Work in Phoenix: Route Choice, Travel Behavior, and Commuter Characteristics, *Transportation Record 1773* Transportation Research Board, National Research Council, Washington, D.C. 2001.

Hunter, William W., Wayne E. Pein, and Jane C. Stutts, Bicycle-Motor Vehicle Crash Types: The Early 1990's, *Transportation Record 1502* Transportation Research Board, National Research Council, Washington, D.C. 1995.

Hunter, William W., J. Richard Stewart, and Jane C. Stutts, Study of Bicycle Lanes Versus Wide Curb Lanes, *Transportation Record 1674*, Transportation Research Board, National Research Council, Washington, D.C. 1999.

Hunter, William W., David Harkey, J. Richard Stewart, and Mia L. Birk, Evaluation of Blue Bike-Lane Treatment in Portland, Oregon, *Transportation Record 1705*, Transportation Research Board, National Research Council, Washington, D.C. 2000.

Hunter, William W., Evaluation of a Combined Bicycle Lane/Right-Turn Lane in Eugene, Oregon, *Transportation Record Preprint*, Transportation Research Board, National Research Council, Washington, D.C. 2002.

Hunter, William W., Evaluation of Innovative Bike-Box Application in Eugene, Oregon, *Transportation Record 1705*, Transportation Research Board, National Research Council, Washington, D.C. 2000.

Hyodo, Tetsuro, Norikazu Suzuki, and Katsumi Takahashi, Modeling of Bicycle Route and Destination Choice Behavior for Bicycle Road Network Plan, *Transportation Record 1705*, Transportation Research Board, National Research Council, Washington, D.C. 2000.

Institute of Transportation Engineers Pedestrian and Bicycle Task Force, *ITE Journal*, September 2000.

Jones, Elizabeth G., and Carlson, Development of a Bicycle Compatibility Index for Rural Roads in Nebraska, *Transportation Research Board* , Transportation Research Board, National Research Council, Washington, D.C. 2003.

Kaiser, S. H., Urban Intersections that Work: A Definition for Level of Service, *Transportation Record 1438*, Transportation Research Board, National Research Council, Washington, D.C. 1994.

Khan, A.M. and A. Bacchus, Bicycle Use of Highway Shoulders, *Transportation Record 1502*, Transportation Research Board, National Research Council, Washington, D.C. 1995.

Kim, Karl, Crash- and Injury-Outcome Multipliers, *Transportation Record 1717*, Transportation Research Board, National Research Council, Washington, D.C. 2000.

Kim, Karl and Lei Li, Modeling Fault Among Bicyclists and Drivers Involved in Collisions in Hawaii, 1986-1991, *Transportation Record 1538*, Transportation Research Board, National Research Council, Washington, D.C. 1996.

King, Michael, Urban Bicycle Facility Selection Guides, Transportation Research Board, National Research Council, Washington, D.C., 2003.

Kleinbaum, David G., *Logistic Regression,* Springer-Verlag New York, Inc., New York, New York, 1994.

Klop, Jeremeny R. and Asad J. Khattak, Factors Influencing Bicycle Crash Severity on Two-Lane, Undivided Roadway in North Carolina, *Transportation Record 1674,* Transportation Research Board, National Research Council, Washington, D.C., October, 1999.

Komanoff, Charles, Cora Roelofs, Jon Orcutt, and Brian Ketcham, Environmental Benefits of Bicycling and Walking in the United States, *Transportation Record 1405*, Transportation Research Board, National Research Council, Washington, D.C. 1993.

Landis, Bruce W, Bicycle Interaction Hazard Score: A Theoretical Model, *Transportation Record 1438*, Transportation Research Board, National Research Council, Washington, D.C. 1994.

Landis, Bruce W., Venkat R. Vattikuti, Ottenberg, et. al., Intersection Level of Service: The Bicycle Through Movement, *Transportation* Research Board, National Research Council, Washington, D.C. 2003.

Landis, Bruce W., Venkat R. Vattikuti, and Michael T. Brannick, Real-Time Human Perceptions Toward a Bicycle Level of Service, *Transportation Record 1578*, Transportation Research Board, National Research Council, Washington, D.C. 1997.

Leden, Lars, Per Garder, Urho Pulkkinen, An Expert Judgement Model Applied to Estimating the Safety Effect of a Bicycle Facility, *Accident Analysis and Prevention,* Vol 32, pp 589-599.

Liao, Tim Futing, *Interpreting Probability Models Logit, Probit, and Other Generalized Linear Models* Sage Publications, Inc., Thousand Oaks, California, 1994.

Linderholm, L., Traffic Safety Evaluation of Engineering Measures - Development of a Method and Its Application to How Physical Layouts Influence Bicyclists at Signalized Intersections, Department of Traffic Planning and Engineering, Lund Institute of Technology, Lund, Sweden, 1992.

Lord, Dominique and Bhagwant N. Persaud, Accident Prediction Models With and Without Trend, *Transportation Record 1717*, Transportation Research Board, National Research Council, Washington, D.C. 2000.

Macbeth, Andrew G., Bicycle Lanes in Toronto, *ITE Journal, April, 1999, pp 38 - 46..*

McDonald, Charlene and Elizabeth K. Burns, Cycling to Work in Phoenix: Route Choice, Travel Behavior and Commuter Characteristics, *Transportation Record Preprint*, Transportation Research Board, National Research Council, Washington, D.C. 2001.

McShane, William R., Roger P. Roess, Elena S. Prassas, *Traffic Engineering*, Prentice Hall, Upper Saddle River, New Jersey, 1998, p. 1990.

Moritz, William E., Adult Bicyclists in the United States, Characteristics and Riding Experience in 1996, *Transportation Record 1636*, Transportation Research Board, National Research Council, Washington, D.C. 1998.

New Jersey Department of Transportation, *Bicycle Compatible Roadways and Bikeway Planning and Design Guidelines,* Trenton, New Jersey, 1996.

New Jersey Department of Transportation, *Bicycle Touring Route Locations Study - Phase 1, High Point to Cape May*, Trenton, New Jersey, 2000.

Noel, Nathalie, C. Leclerc, M. Lee-Gosselin, CRC Index: Compatibility of Roads for Cyclists in Rural and Urban Fringe Areas, *Transportation Record,* Transportation Research Board, National Research Council, Washington, D.C. 2003.

O'Dea, J., *NCHRP Synthesis of Highway Practice 192: Accident Data Quality, TRB,* National Research Council, Washington, D.C., 1994, pp. 50-55.

O'Donnell, C. and D. Connor, Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice, *Accident Analysis and Prevention,* Vol. 28, No. 6, 1996, pp 739-753.

Raad, Tamin, Use of the Bicycle for the Journey to Work: An Analysis of the 1996 Census Results for the Greater Vancouver Regional District, *Transportation Record Preprint*, Transportation Research Board, National Research Council, Washington, D.C. 2002.

Replogle, Michael, *Integrating Pedestrian and Bicycle Factors into Regional Transportation Planning Models*, Environmental Defense Fund, 1995.

Rivara, F.P., D.C. Thompson, and R.S. Thompson, 1996. Circumstances and Severity of Bicycle Injuries, Seattle, Washington: Snell Memorial Foundation, Harborview Injury Prevention and Research Center.

Rodgers, Gregory B., Bicyclist Deaths and Fatality Risk Patterns, *Accident Analysis and Prevention*, Volume 27, No. 2, pp. 215-223, 1995.

Rutgers Transportation Policy Institute, *Jersey City Bicycle Plan, Report prepared for the City of Jersey City,*, New Brunswick, New Jersey, April, 2000.

Saccomanno, F.F., S. A. Nassar, and J.H. Shortreed, Reliability of Statistical Road Accident Injury Severity Models, *Transportation Record 1542* Transportation Research Board, National Research Council, Washington, D.C. 1996.

Saccomanno, E.F., K. C. Chong, and S.A. Nassar, Geographic Information System Platform for Road Accident Risk Modeling, *Transportation Record 1581*, Transportation Research Board, National Research Council, Washington, D.C. 1997.

Sall, John and Ann Lehman, SAS Institute, *JMP Start Statistics*, Duxbury Press, 1996.

Shafizadeh, K. and D. Niemeir, Bicycle Journey-To-Work: Travel Behavior Characteristics and Spatial Attributes, *Transportation Record 1578* Transportation Research Board, National Research Council, Washington, D.C. 19975.

Shankar, Venkataraman, Fred Mannering, and Woodrow Barfield, Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies, *Accident Analysis and Prevention,* Vol. 27, No. 3, pp. 371-389, 1995.

Shankar, Venkataraman and Fred Mannering, An Exploratory Multinomial Logit Analysis of Single-Vehicle Motorcycle Accident Severity, *Journal of Safety Research,* Vol. 27, No. 3, pp. 183-194, 1996.

Sharples, Rosemary, The Use of Main Roads by Utility Cyclists in Urban Areas, *Traffic Engineering and Control*, January, 1999, pp. 18-22.

Sorton, Alex, and Thomas Walsh, Bicycle Stress Level as a Tool To Evaluate Urban and Suburban Bicycle Compatibility, *Transportation Record 1438*, Transportation Research Board, National Research Council, Washington, D.C. 1994.

Stewart, Richard, Applications of Classification and Regression Tree Methods in Roadway Safety Studies, *Transportation Record No.1542*, National Academy Press, Washington, D.C.1996.

Stinson, Monique A. and Chandra R. Bhat, An Analysis of Commuter Route Choice Using a State Preference Study, *Transportation Record*, Transportation Research Board, National Research Council, Washington, D.C. 2003.

Stutts, Jane C., Study of Bicycle Lanes Versus Wide Curb Lanes, *Transportation Record 1674*, Transportation Research Board, National Research Council, Washington, D.C. 1999.

Stutts, J.C., J.E. Williamson, T.Whitely, and F.C. Sheldon, Bicycle Accidents and Injuries: A Pilot Study Comparing Hospital and Police Reported Data, *Accident Analysis and Prevention,* 22(1), 1990, 67-78.

Summala, Heikki, Eero Pasanen, Mikko Rasanen and Jukka Sievanen, Bicycle Accidents and Drivers' Visual Search at Left and Right Turns, *Accident Analysis and Prevention* Vol. 28, No. 2, pp. 147-153, 1996.

Taylor, Dean B., Analysis of Traffic Clearance Interval Requirements for Bicycle-Automobile Mixed Traffic, *Transportation Record 1405*, Transportation Research Board, National Research Council, Washington, D.C. 1993.

Taylor, Dean B., and Hani S. Mahmassani, Coordinating Traffic Signals for Bicycle Progression, *Transportation Record 1705*, Transportation Research Board, National Research Council, Washington, D.C. 2000.

Taylor, Dean and W. Jeffery Davis, Research in Bicycle Traffic Science, Traffic Operations, and Facility Design, *Transportation Record 1674*, Transportation Research Board, National Research Council, Washington, D.C. 1999.

Thom, R. and A. Clayton, Accident Data Requirements for Improving Cycling Safety, *Transportation Record 1405*, Transportation Research Board, National Research Council, Washington, D.C. 1993.

Thom, R. and A. Clayton, Low Cost Opportunities for Making Cities Bicycle-Friendly Based on a Case Study Analysis of Cyclist Behavior and Accidents, *Transportation Record 1372*, Transportation Research Board, National Research Council, Washington, D.C. 1992.

Train, Kenneth, *Discrete Choice Methods with Simulation,* Cambridge University Press, 2002.

Turner, Shawn, and C. Scott Shafer and William Stewart, Bicycle Suitability Criteria: Literature Review and State-of-the Practice Survey, *Research Report 3988-1, Texas Transportation Institute*, 1997.

National Biking and Walking Study Five Year Status Report, U.S. Department of Transportation, Federal Highway Authority, Washington, D.C., 1999.

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

U.S. Department of Transportation, *Traffic Safety Facts 2000 Pedacyclists,* National Highway Traffic Safety Administration, Washington, D.C. 2000.

U.S. Department of Transportation, *Transportation Statistics Annual Report*, Washington, D.C., 1999.

Vining, G. Geoffrey, *Statistical Methods for Engineers*, Thomson Publishing Inc., 1998.

Vogt, Andrew, and Joe Bared, Accident Models for Two-Lane Rural Segments and Intersection, *Transportation Record 1635*, Transportation Research Board, National Research Council, Washington, D.C. 1996.

Voorhees, Alan M. & Associates, Jersey City Topics, prepared for New Jersey Department of Transportation in Cooperation with United States Department of Transportation Federal Highway Administration and City of Jersey City, New Jersey, 1973.

Wachtel, Alan and Diana Lewiston, Risk Factors for Bicycle-Motor Vehicle Collisions at Intersections, *ITE Journal*, September, 1994, pp. 30-35.

Wang, Yinhai and Nancy L. Nihan, Estimating the Risk of Collisions Between Bicycles and Automobiles at Signalized Intersections, *Transportation Record Prepring*, Transportation Research Board, National Research Council, Washington, D.C. 2002.

Wessels, Ralph L., Bicycle Collisions in Washington State: A Six-Year Perspective, 1988-1993, *Transportation Record 1538*, Transportation Research Board, National Research Council, Washington, D.C. 1996.

Zeeger, C.V., *NCHRP Synthesis of Highway Practice 91; Highway Accident Analysis Systems,* Transportation Research Board, National Research Council, Washing, D.C., 1982.