This is to certify that the

dissertation entitled

EVALUATION OF SAFETY AT FREEWAY INTERCHANGES

presented by

Nakmoon Sung

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Civil Engineering

William C. Taylor
Major professor

Date 30 OCT. 2000

EVALUATION OF SAFETY AT FREEWAY INTERCHANGES

By

Nakmoon Sung

A DISSERTATION

Submitted to
Michigan State University
In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Civil and Environmental Engineering

2000

ABSTRACT

EVALUATION OF SAFETY AT FREEWAY INTERCHANGES

By

Nakmoon Sung

This research focused on several issues that arise when the Negative Binomial distribution rather than the Poisson distribution(which have been the commonly accepted assumption in analyzing traffic accidents), is found to better fit the accident data.

On the basis of the Negative Binomial distribution, the framework of the rate quality control method was redefined as a basis for the identification of hazardous sites. This produced conceptually more reasonable results than the existing approaches such as the Poisson distribution based rate quality control method, or the Bayes approach.

However, it is sometimes not efficient for traffic engineers to apply this approach since the parameters of the Negative Binomial distribution can not be easily estimated. Therefore, a Normal approximation method to overcome this issue was developed. The Normal approximation method identified the same hazardous sites from a list of two common interchange types found on several freeways in Michigan.

Although the rate quality control method based on the Negative Binomial distribution is an effective technique for the identification of hazardous sites, it has two limitations. First, the selection of reference sites is a matter of judgement.

Second, a sufficient number of reference sites with similar characteristics are not always available to assure statistical accuracy. As an alternative, a prediction model method was developed. This method produced results similar to those from the rate quality control method. By using the prediction model method, the conceptual and practical problems associated with the identification of hazardous sites can be reduced. The Generalized Linear Model concept was used to calibrate the accident prediction models.

To Haein and Jinmo

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my major professor, Dr. William C. Taylor for his invaluable wisdom and kindness during the journey of this research.

I would like to express my gratitude to the other members of the committee, Dr. Thomas L. Maleck, Dr. Virginia P. Sisiopiku and Dr. Vincent Melfi for their suggestions and interest in this work.

A particular note of thanks is extended to Dr. Robert Maki and Mr. Robert Rios of the Michigan Department of Transportation for helping me use several databases, and Mr. Oleg V. Makhnin, Mr. Linyuan Li and Mr. Jan Hannig of the Statistical Consultant Service(SCS) of the Department of Statistics and Probability for advice in understanding various statistical concepts.

Finally, I would like to express my thanks to my wife, Byungcheon, for her love and patience, and my parents for their encouragement through the years.

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES xi

CHAPTER 1
INTRODUCTION 1
1.1 Background and problem identification 1
1.2 Proposed research objectives 4
1.3 Structure of this dissertation 5

CHAPTER 2
THE PROBABILITY DISTRIBUTION OF TRAFFIC CRASHES AT FREEWAY INTERCHANGES 6
2.1 General 6
2.2 Concept of the Poisson distribution and Negative Binomial distribution 8
2.3 Phenomena of over-dispersion over time 9
    2.3.1 Analysis by crash types 9
    2.3.2 Analysis by annual crash frequency per interchange (Diamond interchanges) 13
    2.3.3 The results 16

CHAPTER 3
A TRAFFIC CRASH PREDICTION MODEL FOR FREEWAY INTERCHANGES 17
3.1 General 17
3.2 Dependent variable description 20
    3.2.1 Classification by interchange type 21
    3.2.2 Crash data summary 25
3.3 Independent variable description 33
3.4 Preliminary analyses of variables 36
    3.4.1 Correlation analysis 36
    3.4.2 Analysis of variance (ANOVA) 38
3.5 Model structure 42
3.6 Model calibration and analysis 49
    3.6.1 Link functions for the Generalized linear model (GLIM) approach 50
    3.6.2 Assessing the goodness of fit of the model 52
    3.6.3 The procedures used in parameter calibration 55
    3.6.4 Results of the model calibration 58
    3.6.5 Pearson Residuals 61
3.7 A comparison of model calibration results according to the error structure (Normal versus Negative Binomial assumption) 63
3.8 Sensitivity analysis of the accident prediction model 65

CHAPTER 4
THE DEVELOPMENT OF A METHOD TO IDENTIFY HAZARDOUS SITES BASED ON THE NEGATIVE BINOMIAL DISTRIBUTION 68
4.1 General 68
4.2 A review of the statistical methods for identifying hazardous sites 69
4.2.1 The rate quality control method 69
4.2.2 The Bayes approach 74
4.2.3 The problems resulting from the use of the Poisson distribution 77
4.3 The identification of hazardous sites based on the Negative Binomial distribution 79
4.3.1 Concept 79
4.3.2 Estimation of the parameters 81
4.3.3 Application and validation of the Negative Binomial approach 87

CHAPTER 5
A SIMPLIFIED APPROACH FOR OVER-DISPERSION (NORMAL APPROXIMATION METHOD) 93
5.1 General 93
5.2 Concept 93
5.3 Application and validation of the Normal approximation method 95
5.3.1 Estimation of parameters 96
5.3.2 Validation of the Normal approximation approach 96
5.4 An examination of the assumptions 100
5.4.1 Goodness of fit of the Normal distribution 100
5.4.2 Large values of μ_i 101

CHAPTER 6
IDENTIFICATION OF HAZARDOUS SITES USING A TRAFFIC CRASH PREDICTION MODEL (PREDICTION MODEL METHOD) 104
6.1 The limitation of the rate quality control method (or upper control limit method) 104
6.2 The concept of the prediction model method 107
6.3 Application and validation of the prediction model method 111
6.3.1 Illustration of the prediction model method 111
6.3.2 Validation of the prediction model method 114
6.4 Evaluation of Michigan freeway interchanges on the basis of the prediction model 119

CHAPTER 7
SUMMARY AND CONCLUSIONS 121
7.1 Summary 121
7.1.1 Traffic crash distribution 121
7.1.2 Traffic crash prediction model 122
7.1.3 The rate quality control method based on the Negative Binomial distribution 123
7.1.4 The Normal approximation method 124
7.1.5 The prediction model method 125

7.2 Conclusions......126

BIBLIOGRAPHY......127

APPENDICES......133
A. The results of accident prediction model calibration......134
B. The results of evaluating of freeway interchanges......145

LIST OF TABLES

Table 2.1 The correlation and residual values according to the distribution 16

Table 3.1 Interchange classification 24

Table 3.2 Summary of crashes per interchange (1996 ~ 1998) 25

Table 3.3 Injury percent by interchange type (1996 ~ 1998)...29

Table 3.4 Summary of mainline and ramp crashes (1996 ~ 1998) 31

Table 3.5 Summary of the crash types (1996 ~ 1998) 32

Table 3.6 Classification of independent variables 35

Table 3.7 Correlations between independent variables 37

Table 3.8 ANOVA for traffic effects 39

Table 3.9 ANOVA for geometric effects 41

Table 3.10 Cross tabulation of crashes by mainline volume and ramp volume 44

Table 3.11 The results of crash prediction model calibration (Interchange type 11) 60

Table 3.12 A comparison of model calibration results according to error structure
(Normal and Negative Binomial assumption) 64

Table 3.13 Sensitivity analysis (effect of main geometric variables)...............................67

Table 4.1 The estimation of parameters..87

Table 4.2 A comparison of hazardous sites according to the methods
(Diamond interchanges)..91

Table 4.3 A comparison of hazardous sites according to the methods
(Par Clo A or B 4 Q Interchanges)...92

Table 5.1 Estimation of parameters 96

Table 5.2 A comparison with the results using the Negative Binomial distribution 98

Table 6.1 A comparison of results 117

Table 6.2 The abnormal sites by the prediction model method 120

Table A.1 The results of accident prediction model calibration
(Interchange type 11) 135

Table A.2 The results of accident prediction model calibration
(Interchange type 12) 136

Table A.3 The results of accident prediction model calibration
(Interchange type 13) 137

Table A.4 The results of accident prediction model calibration
(Interchange type 14) 138

Table A.5 The results of accident prediction model calibration
(Interchange type 21) 139

Table A.6 The results of accident prediction model calibration
(Interchange type 31) 140

Table A.7 The results of accident prediction model calibration
(Interchange type 33) 141

Table A.8 The results of accident prediction model calibration
(Interchange type 35) 142

Table A.9 The results of accident prediction model calibration
(Interchange type 41) 143

Table A.10 The results of accident prediction model calibration
(Interchange type 51) 144

Table B.1 The results of evaluating of freeway interchanges
by the prediction model method 146

LIST OF FIGURES

Figure 2.1 Observed and theory variances by crash types (Poisson distribution) 11

Figure 2.2 Observed and theory variances by crash types (Negative Binomial distribution) 12

Figure 2.3 Observed and theory variances of annual crash frequency (Poisson distribution) 14

Figure 2.4 Observed and theory variances by annual crash frequency (Negative Binomial distribution) 15

Figure 3.1 Boundary of the interchange 23

Figure 3.2 Total crashes and injury crash percent 28

Figure 3.3 Mainline traffic volume and crashes 45

Figure 3.4 Ramp traffic volume and crashes 46

Figure 3.5 The process to calibrate coefficients & k parameter 57

Figure 3.6 Pearson Residuals and E(x) 62

Figure 4.1 The concept of rate quality control method under the Poisson distribution 73

Figure 4.2 Identification of hazardous sites by Bayes approach 76

Figure 4.3 A comparison of the Negative Binomial and Poisson assumption in applying rate quality control method 78

Figure 5.1 A comparison of the results of the Negative Binomial and Normal approximation approaches 99

Figure 5.2 The difference between the true and approximate upper control limits 103

Figure 6.1 Basic geometric classification tree for reference sites 106

Figure 6.2 Concept of prediction model method 108

Figure 6.3 An example of application of prediction model method (95 %) 113

Figure 6.4 A comparison of results of the rate quality control method and prediction model method....118

Chapter 1

INTRODUCTION

1.1 Background and problem identification

In response to limited budgets, it has become very important to ensure that funding available for road improvements is efficiently utilized. A typical safety program includes identification, diagnosis, and remediation of hazardous locations, and hence the success of the safety program can be enhanced by efficiently identifying hazardous locations.
A hazardous location is defined as a site where the observed number of crashes is larger than a specific norm (a record of crashes at locations with similar characteristics). That is, a site is deemed hazardous if its crash history over a given period exceeds a predetermined level which is based on the concept of confidence levels within the context of classical statistics (Witkowski 1988).

The observed number of crashes over a specific period at a specific site can usually be obtained from a database related to traffic crashes. However, several difficulties arise in determining a base for comparing this number to an expected number of crashes at reference sites that are defined as sites with similar geometric and traffic characteristics. Hauer (92) recognized that the identification of hazardous sites using reference sites causes conceptual and practical problem in nature. The main conceptual problem is that of choosing suitable reference sites, which is a matter of judgement. The practical problem is that if very similar sites are chosen to reduce the variations caused from the conceptual difficulties, the number of reference sites will usually be too small to allow for an accurate estimate of the hazard at a given site. These same questions were also raised by Mahalel (1982), Hauer and Persaud (1987), and Mountain and Fawaz (1989).

There are 397 interchanges along the four main Interstates (I-69, I-75, I-94 and I-96) in Michigan. In order to define reference sites for the evaluation of a given interchange in Michigan, the interchanges were first classified according to their geometry; such as interchange type, the number of ramps, shoulder width, the number of lanes, ramp length et al., and second according to traffic conditions. However, with this level of stratification, it was not possible to obtain enough reference sites to guarantee a significant level of accuracy for each type of interchange. To overcome these difficulties, a method using a crash prediction model to identify hazardous sites was examined in this study.

The basic concept of the prediction model method is that the expected value of crashes at the reference sites $E(\theta)$ can be obtained by developing a crash prediction model rather than on the basis of reference sites. A specific site is deemed to be hazardous if the probability of the number of observed crashes occurring at the site is smaller than some predetermined values (i.e.,0.05). That is, a hazardous location is one in which the deviation from the expected crash frequency $E(\theta)$ is large. The prediction model method is a technique to identify hazardous sites, based on an expected value which is calculated by accident prediction models. Thus, if this method is to be accurate, it is important to develop the traffic crash prediction models under appropriate rationale.

There are generally two kinds of crash prediction models which differ according to the assumption of the error structures. One is the conventional linear regression model with a constant normal error structure, the other is a regression model with a non normal and heterogeneous error structure (i.e., Poisson and Negative Binomial distribution). In this research, we have examined the error structures of crash occurrences in various respects on the basis of the observed data, and verified that crashes on freeway interchanges follow the Negative Binomial distribution rather than a Normal or Poisson distribution. Accordingly, the model parameters should be calibrated under the assumption of the Negative Binomial error structure.

The classical rate quality control method has been used by many transportation agencies to identify hazardous sites since it was first proposed in the transportation field in 1956 (Stokes and Mutabazi 1996). This method uses a statistical test to determine whether the crash rate of a site is abnormally high, compared with that of reference sites. Therefore, if the crashes follow the Negative Binomial distribution, the rate quality control method should be reexamined because it is based on the assumption that the probability of traffic crash occurrences can be approximated by the Poisson distribution (Norden et al. 1956, Morin 1967, and Stokes and Mutabazi 1996).

1.2 Proposed research objectives

The four major objectives of this research are:

1) to verify that the freeway traffic crashes follow the Negative Binomial distribution rather than the Poisson distribution,
2) to develop crash prediction models for freeway interchanges using the Negative Binomial distribution,
3) to provide a new framework for the rate quality control method for identifying hazardous sites on the basis of the Negative Binomial distribution, and
4) to propose a method for the identification of hazardous sites using a traffic crash model calibrated on the basis of the Negative Binomial distribution.

Even though there are several objectives for this research, each is based on the assumption that the error structure follows a Negative Binomial distribution. First, this research will describe how traffic engineers can apply the rate quality control method based on the Negative Binomial distribution. However, there are interchanges where this method can not be applied because an insufficient number of reference sites are available to allow for an accurate evaluation. To solve this kind of problem, a method for identifying hazardous sites using a crash prediction model is proposed. The prediction model method can be used to evaluate a freeway interchange without reference sites, and to determine the sites in need of remedial actions.

1.3 Structure of this dissertation

The background and objective of this research have been discussed in the first chapter. The issues related to the distribution of crash occurrences are analyzed in chapter 2. In chapter 3, the effort is focused on parameter calibration of the crash prediction models for freeway interchanges based on the Negative Binomial error structure. This chapter includes the description of independent variables, such as traffic and geometric features, the model structures, methods to converge nonlinear regression models, and measures of model accuracy. In addition, sensitivity analyses of the models is discussed in this chapter.

Chapter 4 presents the problem resulting from applying the Poisson error assumption in the existing rate quality control method, and develops a new framework for the rate quality control method on the basis of the Negative Binomial error assumption. Chapter 5 focuses on how this rate quality control method based on the Negative Binomial distribution can be simplified through a Normal approximation for the purpose of user convenience. This chapter also demonstrates that the Normal approximation method produces the same results as the proposed rate quality control method.

Chapter 6 describes how the prediction model method can be used as an alternative for the identification of hazardous sites when the number of reference sites is insufficient to allow for significant results. Based on the prediction model method, about 200 interchanges along Michigan freeways are evaluated. A summary and conclusions occupy the last chapter of this dissertation.

Chapter 2

THE PROBABILITY DISTRIBUTION OF TRAFFIC CRASHES AT FREEWAY INTERCHANGES

2.1 General

The most appropriate distribution of crash occurrences is a fundamental question that often arises in the traffic safety field. For example, the Poisson distribution frequently appears in articles identifying hazardous locations using control limit charts, because of its simplicity caused from the assumption that the variance is the same as the mean (Norden et al 1956, Hauer 1996). It has also been recognized that the Poisson distribution provides a better fit to traffic crash data than the Normal distribution (Miaou et al 1992, Jovanis and Chang 1993).

However, in studying the injury severity to the front seat occupants of vehicles in crashes, Hutchinson and Mayne (1977) realized that there appeared to be more variability of different severity levels occurring in different years than would be expected on the hypothesis of the Poisson distribution. When there is greater variability than expected by Poisson' law, we call this phenomenon over-dispersion. Issues related to this over-dispersion are also implicit in the works of earlier researchers (Benneson and McCoy 1997, Vogt and Bared 1999).

Consequently, two distributions (Poisson and Negative Binomial) have been assumed for traffic crash occurrences. However, no researcher has yet provided a full discussion of the issue, even though the assumption of the probability distribution for crash occurrence is very important for the identification of hazardous sites necessary for highway safety programs and for the calibration of crash prediction models.

For example, with the rate quality control method, a site is identified as hazardous if its observed crash rate exceeds the upper control limit, which is the mean crash rate of reference sites plus a multiple of the standard deviation of the site crash rates (Stokes and Mutabazi 1996). Herein, the standard deviation is equal to the square root of the mean for a Poisson distribution and the square root of the (mean + mean ^{2}/k) for the Negative Binomial distribution, respectively (Rice 1997).

Three distributions have generally been assumed for the calibration of traffic crash prediction models (i.e., constant normal, Poisson and Negative Binomial). However, recently there is an implicit agreement between traffic engineers that the Poisson or Negative Binomial distributions are more desirable assumptions than the constant normal distribution. Crash prediction models with a heterogeneous error structure such as the Poisson or Negative Binomial distribution, are generally calibrated using weighted least squares (Seber and Wild 1989). In weighted least square regression, data points are weighted by the reciprocal of their variances. Thus, in calibrating traffic crash models, the assumption of error structures is a very critical issue in determining the accuracy of coefficients. Because of the importance of the distribution of crash occurrences, the year to year variability in the number of crashes is examined and discussed in this chapter.

2.2 Concept of the Poisson distribution and Negative Binomial distribution

The Poisson distribution is often the first option considered for random counts; it has the property that the mean of the distribution is equal to the variance (Rice 1997) and the following frequency function:

$$p(X = x) = \frac{\exp(-m)(m)^x}{x!} \tag{2.1}$$

where,
m = mean

However, when the variance of the counts is substantially larger than the mean, consideration is given to the Negative Binomial distribution, which is a discrete distribution with the following frequency function (Rice 1997):

$$f(x/m,k) = \left(1+\frac{m}{k}\right)^{-k} \frac{\Gamma(k+x)}{x!\Gamma(k)} \left(\frac{m}{m+k}\right)^x \tag{2.2}$$

where,
m = mean
k = negative binomial parameter

2.3 Phenomena of over-dispersion over time

In examing the freeway interchange crash data over time, there appeared to be more variability than would be expected under the hypothesis of the Poisson distribution. The large variability could be expected because there are many factors to cause the annual crash frequency to vary, including maintenance activities, the weather and traffic changes.

The Negative Binomial distribution might be considered as a model for the situation in which the rate varies over time or space(Rice 1997). The Negative Binomial distribution has been assumed to explain various physical phenomena; the distribution of insect counts if the insect hatch from the depositions of larvae(Rice 1997). Thus, it is not unique to apply the Negative Binomial distribution in analyzing discrete random counts.

Two kinds of data sets are utilized to test the over-dispersion. One is the number of crashes classified by type, the other is the number of crashes per interchange per year across 84 interchanges. Analyses of the over-dispersion were performed for the crashes during the 5 year-period 1994-1998.

2.3.1 Analysis by crash types

To test over-dispersion of the crashes which occurred in freeway interchanges, crash frequencies of each of 24 types of crashes were obtained separately for each of 5 years from 1994 to 1998. The variance and the mean annual number of crashes were calculated on the basis of the crashes that occurred over the 5 years.

To test whether the crash occurrences follow the Poisson distribution, the observed variances of the annual number of crashes were plotted against the annual mean value. Therefore, there are 24 points corresponding to the 24 types of crashes. In **Figure 2.1**, the solid line is the variance that would be expected on the hypothesis of the Poisson distribution. If the Poisson distribution is a good fit, the observed variances should lie along the solid line. However, the figure shows that there is larger variability than would be expected under the Poisson distribution.

There is a much larger variability in the most common types of crashes (rear end, sideswipe) than for the less common types of crashes (backing, fixed object). This phenomenon was discussed in previous research (Hutchinson and Mayne 1977).

Noting that the Negative Binomial distribution is an alternative to reflect the phenomenon of over-dispersion, the maximum likelihood estimate of k was determined to be about 71 by fitting the data to the Negative Binomial distribution. In **Figure 2.2**, the solid line is the variance that would be expected on the hypothesis of the Negative Binomial distribution. This figure shows that the Negative Binomial distribution fits the data much better than the Poisson distribution shown in **Figure 2.1**.

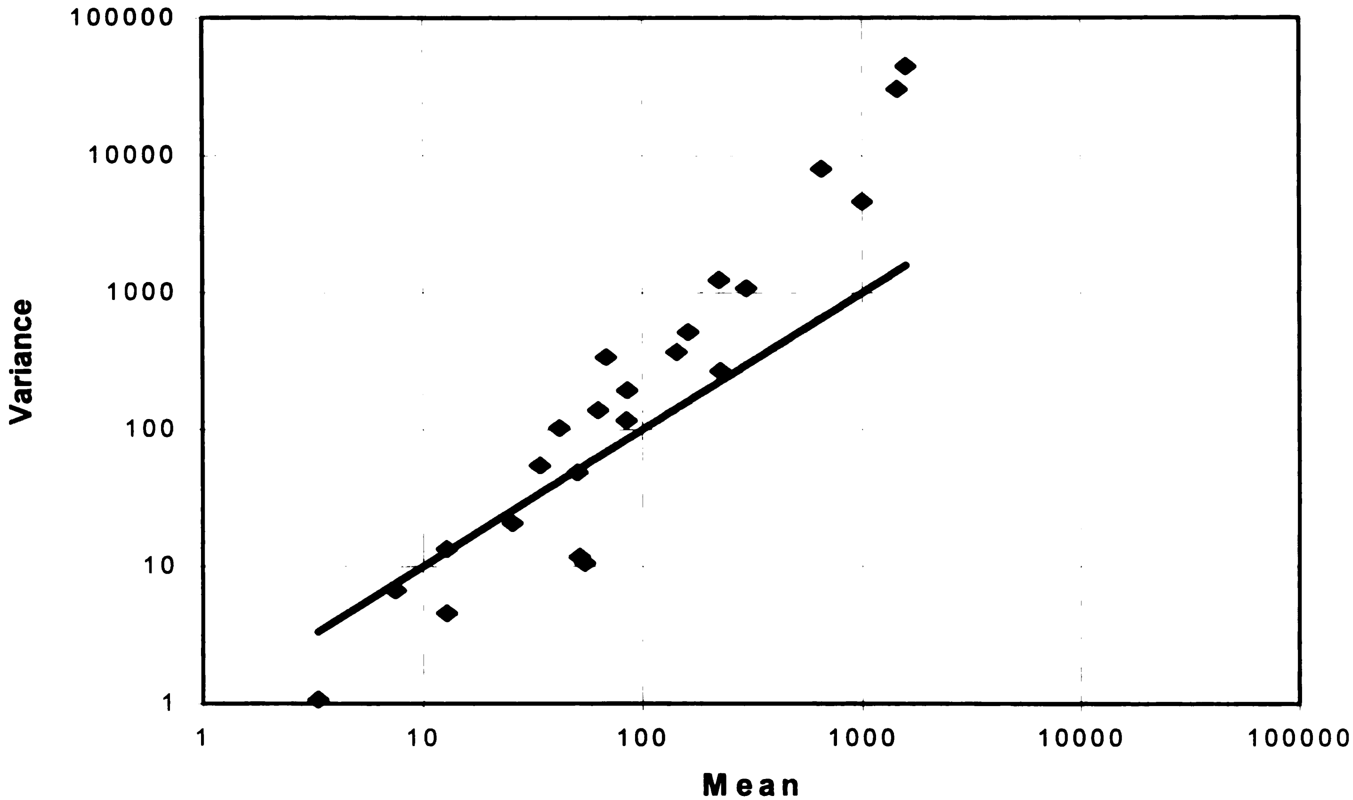

Figure 2.1 Observed and theory variances by crash types (Poisson distribution)

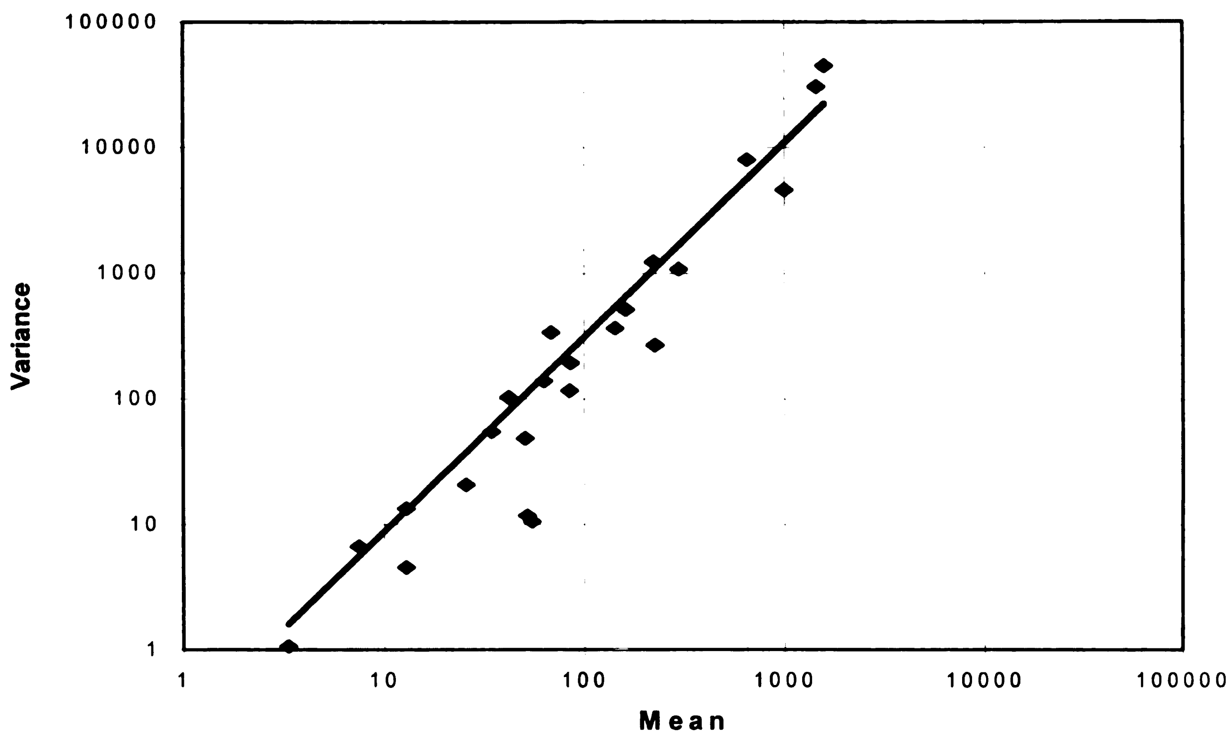

Figure 2.2 Observed and theory variances by crash types (Negative Binomial distribution)

2.3.2 Analysis by annual crash frequency per interchange (Diamond interchanges)

To see how widely this relationship applies, a similar approach was used for Diamond interchanges, which is the most common type of freeway interchange in Michigan.

The variance and the mean annual number of crashes were calculated from the total number of crashes that occurred on the same 84 interchanges from 1994 through 1998. The observed variances in the annual numbers of crashes were also plotted against the mean annual numbers, with a data point corresponding to each of the 84 interchanges. In **Figure 2.3**, the solid line is the variance that would be expected on the hypothesis of the Poisson distribution, and we see that there is also greater variability than expected by the Poisson distribution, as in the previous case. When the data were fit to the Negative Binomial distribution, it was found that the maximum likelihood estimate for k is about 21. **Figure 2.4** shows that the Negative Binomial distribution fits the data much better than the Poisson distribution.

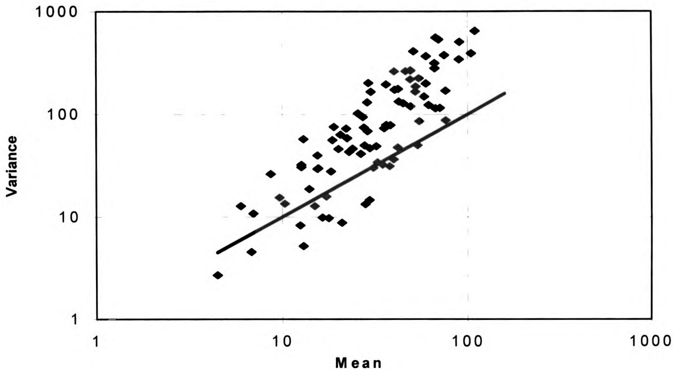

Figure 2.3 Observed and theory variances of annual crash frequency (Poisson distribution)

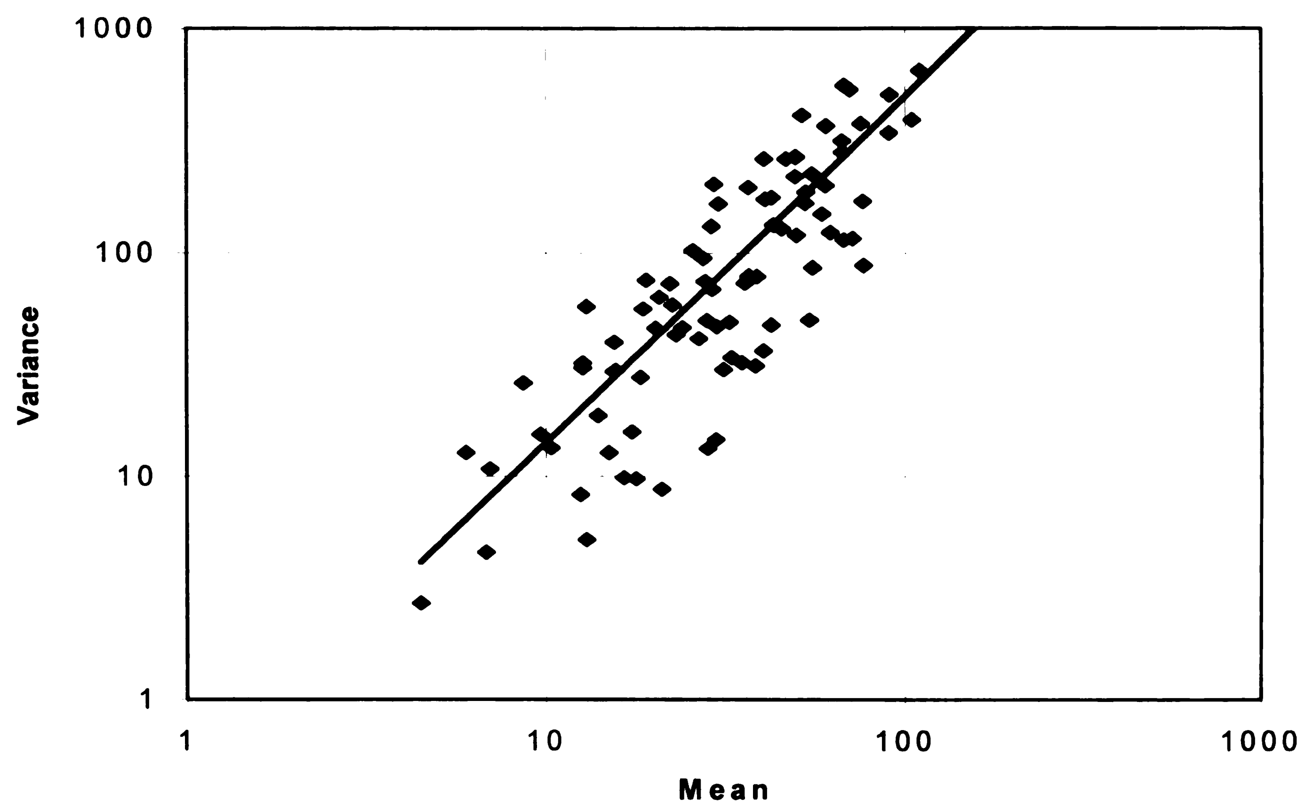

Figure 2.4 Observed and theory variances of annual crash frequency (Negative Binomial distribution)

2.3.3 The results

For theoretical support of these results, correlation coefficients and squared residuals were calculated for the data in **Figure 2.1** through **Figure 2.4**. As shown in **Table 2.1**, the correlation coefficients between the observed and the expected variances increased from 0.91 to 0.97 and from 0.84 to 0.90 in the analysis of 24 crash types and annual total crashes, respectively, when the Negative Binomial distribution was assumed. Squared residuals were calculated using the observed variances and expected variances. The residuals were reduced by more than 80 % when the Negative Binomial distribution was assumed as shown in **Table 2.1**.

Thus, we can conclude that the Negative Binomial distribution is a more reasonable assumption for the distribution of freeway interchange crashes than the Poisson distribution.

Table 2.1 The correlation and residual values according to the distribution

	Poisson	Negative Binomial	
	Correlation coefficient	Correlation coefficient	Squared Residual
Accident type	0.91	0.97	87%⇓
Annual crash frequency	0.84	0.90	84%⇓

Chapter 3

A TRAFFIC CRASH PREDICTION MODEL FOR FREEWAY INTERCHANGES

3.1 General

There have been several studies which purpose was to develop crash prediction models using the relationship between traffic crashes and various independent variables.
In all such studies, the first issue is selection of the independent variables. Using characteristics of a county, Maleck (1980) and Tarko et al (1996) developed models for predicting the expected annual crashes for a county. Independent variables in these models consist of a subset of the following factors: the number of licensed drivers, the number of registered vehicles, population, median family income, road mileage, and percentage of state roads over all ones.

Mcguigan (1981), Maher and Summersgill (1996), Persaud and Nguyen (1998), Rodriguez and Sayed (1999), Bonneson and McCoy (1997), Lau and May (1988), and Belanger (1994) developed crash prediction models for signalized or unsignalized intersections. These models include one or more of the following independent variables; major road traffic volume, minor road traffic volume, pedestrian volume and channelization on the main road. The main road traffic and minor road traffic have been found to be the most significant variables.

Hauer and Griffith (1994), Vogt and Bared (1999), Seder and Livneh (1981), and Moutain et al (1996) developed crash prediction models for road sections using only the traffic volume. In addition, Hauer and Persaud (1987) used traffic volume and train volume for crash models of rail-highway grade crossings, and Miaou et al (1992) modeled truck crashes using geometric characteristics and truck ADT. A few researchers modeled the effects of independent variables on traffic crashes on freeways. Kim (1989) used interchange types, traffic volume, population and the number of ramps to develop a crash prediction model for freeway interchanges. All of these models would be classified as macroscopic models because they use average daily traffic (ADT), rather than the traffic volume at the time of the crash.

Persaud and Dzbik (1993) developed a microscopic model to estimate crashes on freeway sections. Microscopic models relate crash occurrences to the specific flow at the time of the crash rather than to the average daily traffic (ADT). Hence a freeway with intense flow during rush hour periods would have a higher crash potential than a freeway with the same ADT, but with flow more evenly distributed during the day.

As noted above, traffic volume is considered the main contributing factor in predicting traffic crashes in most of the models, with additional geometric variables chosen based on the objective of modeling.

The second issue in the development of an accident prediction model is how to calibrate the model parameters, which usually depend on the error structure. There are two approaches that are often used when calibrating model parameters. One is a conventional linear regression approach, with its assumption of a normally distributed and homogeneous error structure. The linear regression approach has been recognized to be lacking the distribution properties to adequately describe the discrete, nonnegative, and sporadic traffic crash events with a low mean value (Mahalel 1986, Miaou and Lum1993). Before the Poisson approach was introduced, most models were developed on the basis of multi linear regression, with the assumption of a normal distribution. For example, McGuigan (1981), Kim (1989), and Lau and May (1988) used the normal error structure to calibrate their crash prediction models.

The other approach is the use of a regression model, with a non -normal and heterogeneous error structure. These include the Poisson, Negative Binomial and Gamma distributions. It has been generally recognized that crash frequencies better fit a model using the assumption of a Poisson distribution rather than a Normal distribution. For example, Miaou et al. (1992, 1994) proposed the Poisson model to develop the relationship between truck crashes and geometric design. Jovanis and Chang (1993) also used the Poisson model to relate crashes to mileage and environmental variables.

However, the Poisson model also has its weakness. For example, the Poisson model assumes that the variance is the same as the expected number, and hence it can not reflect the phenomenon of "over-dispersion" which often occurs in traffic crashes. In order to overcome this problem, Persaud and Nguyen (1998), and Rodriguez and Sayed

(1999) have proposed regression models with the Negative Binomial error structure to predict signalized intersection crashes.

The phenomenon of over-dispersion on freeway crashes has been verified and discussed in chapter 2. In this chapter, a crash prediction model for freeway interchanges will be developed under the assumption of a Negative Binomial error structure.

3.2 Dependent variable description

The focus on freeway interchange crashes requires a working definition of the boundary of an interchange. In this study, the interchange is composed of ramps and mainlines. The ramps include on- ramps and off-ramps, and the mainlines are defined as the section within 500 feet from the beginning of the off- ramp to 500 feet from the end of the on-ramp as shown in **Figure 3.1**. This definition is the same as that of the Michigan DOT interchange inventory file. The crashes on cross roads are not included in this study because of the practical barrier that traffic volume for the cross road is not available, and the engineering intuition that the crashes on the cross road may have very different characteristics (i.e., low severity, high percentage of angle crashes).

May (1964) found that there is little to be gained by using a study period longer than three years. Subsequently, many previous researchers have used three years of crash data in developing crash prediction models (Miaou and Lum 1993, Bonneson and Macoy 1993, Persaud and Nguyen 1998). Noting that data older than three years may not reflect the current situations, the number of crashes that occurred in the past 3 years (1996 through 1998) are used as the dependent variable for this study.

The accident rate will not be used as the dependent variable since accurate volume data for each element of the interchange is not available. The original source of the crash data is the "Official Michigan Traffic Accident Report' (form UD-10). The crash data are summarized in section 3.2.2

3.2.1 Classification by Interchange type

A lack of homogeneity refers to the understanding that different relationships may hold between variables on the basis of the values of various characteristics (i.e., geometry, control, traffic, and so on). In many cases, tree structures which are easily understood and interpreted, are built describing the main factors and interactions between factors (Lau and May 1988). However, the tree structures can be used only in the case of large samples, and hence this method may be inadequate in developing crash prediction models for freeway interchanges, even though it is a conceptually powerful and systematic tool.

In this study, a total of 199 interchanges are grouped into 10 categories as shown in **Table 3.1**. We can not classify the interchanges more specifically because of the limitation of sample sizes, even though the Michigan interchange inspection file includes 22 categories of interchanges. In the approach to grouping interchanges, the independent variables (i.e., traffic volume, ramp length, et al) were explicitly excluded from the features which were used in the classification of interchange types.

As shown in **Table 3.1**, the number of type 11 and type 31 interchanges is relatively large compared with those of other types.

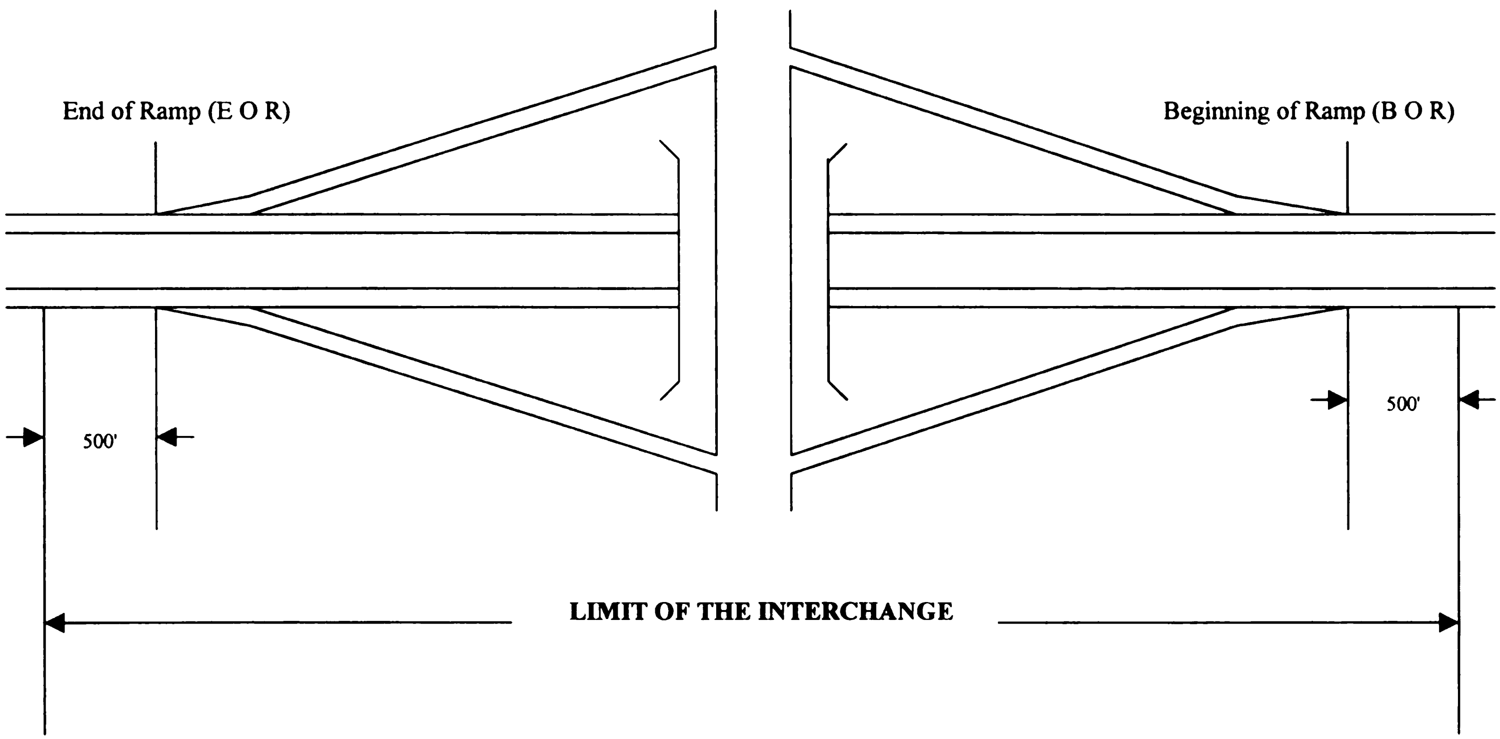

Figure 3.1 Boundary of the interchange

Table 3.1 Interchange classification

CLASSIFICATION		INTERCHANGE TYPE	SAMPLE SIZE
1. DIAMOND INTERCHANGE	Type 11	• Diamond	34
	Type 12	• Tight Diamond • Modified Tight Diamond	19
	Type 13	• Partial Diamond • Partial Tight Diamond	24
	Type 14	• Split Diamond • Modified Diamond	14
2. T-INTERCHANGES	Type 21	• Trumpet – A • Trumpet – B	9
3. CLOVER LEAFS	Type 31	• Partial Clover A • Partial Clover B • Partial Clover A 4 Quadrant • Partial Clover B 4 Quadrant	41
	Type 33	• Partial Clover AB • Partial Clover AB 4 Quadrant	21
	Type 35	• Clover • Clover with CD	8
4. DIRECTIONAL	Type 41	• Full Directional • Partial Directional • Directional Y • Partial Directional Y	21
5. OTHERS	Type 51	• Others	8
TOTAL			199

3.2.2 Crash data summary

3.2.2.1 Summary of crashes per interchange

The summary statistics describing the crashes that have occurred over 3 years in each interchange are provided in **Table 3.2**. As listed in the table, an average of 126 crashes occurred in each interchange, 28 % of which were injury crashes. The average number of crashes is highest in Directional interchanges, and lowest in T-interchanges.

Table 3.2 Summary of crashes per interchange (1996~1998)

Interchange type		Total crashes			Injury crashes (include fatal crashes)		
		Max	Min	Average	Max	Min	Average
Diamond	Type 11	321	24	132	93	6	39
	Type 12	492	42	123	156	6	33
	Type 13	252	18	120	84	3	33
	Type 14	393	24	99	135	3	27
T-interchange	Type 21	156	21	75	69	6	24
Clover-leaf	Type 31	402	33	135	99	6	33
	Type 33	237	24	84	54	3	21
	Type 35	405	51	168	138	12	48
Directional	Type 41	408	21	186	111	3	54
Others	Type 51	408	21	180	45	6	21
Total		492	18	126	156	3	36

3.2.2.2 Summary of injury data

Figure 3.2 shows the relationship between total crashes and injury crash percentage. As shown in the figure, the smaller the total number of crashes, the greater the scatter of injury crash percentage. Therefore, total crashes are a more reliable dependent variable than injury crashes, because there is always implicit variability in injury crashes. In the case of the interchanges with a small number of crashes, this variability may inappropriately model the effects of the independent variables on crashes.

Table 3.3 contains summary statistics of injury crashes that occurred in the past 3 years. It is not surprising that the percent of injury crashes is relatively high for T-interchanges and Directional interchanges (30.8 % and 29.2 % respectively), considering that the vehicle operating speeds on these types of interchanges are high compared with those on other types of interchanges.

The coefficient of variation V(x) is a stable measure of the variability of a random variable x, which is defined as (Harr 1996):

$$V(x) = \frac{\sigma(x)}{E(x)} \times 100 \qquad (\%)$$

The higher the coefficient of variation V(x), the greater will be the scatter. As a rule of thumb, coefficients of variation below 15 % are thought to be low, between 15 and 30 % moderate, and greater than 30 % high (Harr 1996).

As shown in the last row of the **Table 3.3**, the coefficient of variation of injury percent across the interchange types is 10. 8 %, which is low. This implies that interchange types are related to the number of crashes, but not the severity of the crashes. Thus, for this study, the total number of crashes is used as the dependent variable for the development of traffic crash prediction models.

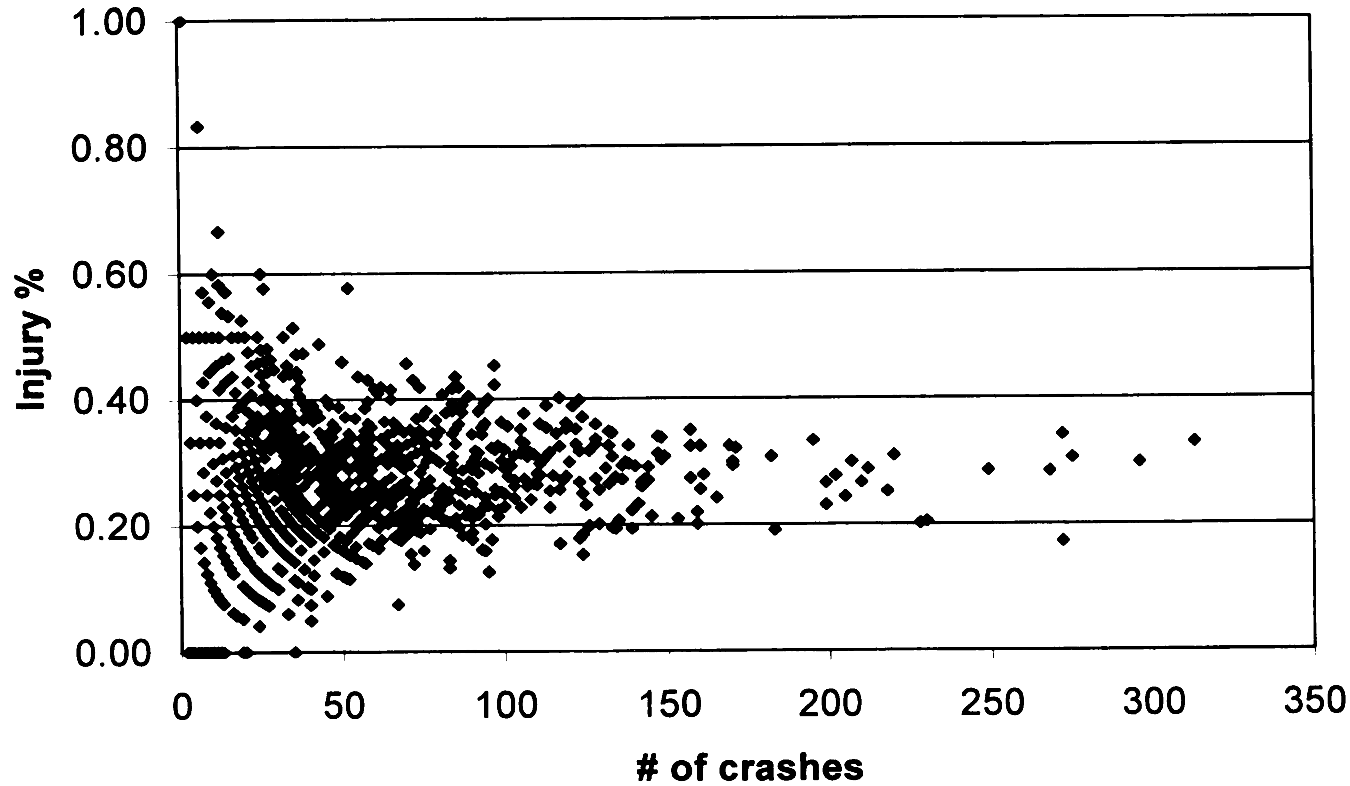

Figure 3.2 Total crashes and injury crash percent

Table 3.3 Injury percent by interchange type (1996~1998)

Interchange type		Total crashes	Injury crashes	Injury (%)
Diamond	Type 11	4479	1272	28.4
	Type 12	2211	600	27.1
	Type 13	2886	822	28.5
	Type 14	1380	393	28.5
T-interchange	Type 21	681	210	30.8
Clover-leaf	Type 31	5388	1380	25.6
	Type 33	1779	453	25.5
	Type 35	1347	381	28.3
Directional	Type 41	4074	1188	29.2
Others	Type 51	699	177	25.3
Total		24924	6876	27.6
V(x)		-	-	10.8

3.2.2.3 Summary of mainline and ramp crashes

Table 3.4 presents a statistical summary of mainline and ramp crashes that occurred from 1996 to 1998. Ramp accidents are about 4300 of the total 25000 crashes, or about 17 %. There is a large variability in the percent of ramp crashes across the interchange types, as shown in the table. That is, the coefficient of variation is 344 %, which is extremely high. This implies that different explanatory variables are needed when developing crash prediction models by interchange type.

Table 3.5 presents data on the crash type according to the interchange type. In our sample sites, rear end crashes account for 39.7 % of total crashes. Rear end crashes are especially high in Type 11(Diamond) and Type 35 (Clover leaf) interchanges, and low in Type 33(Partial Clover AB or Partial Clover AB 4 Q). Fixed object and sideswipe crashes are 20.9 % and 14.1 %, respectively, as shown in the table. The coefficients of variance of a special type of crash percent across interchange types range from 53 % to 172 %, which are high. Accordingly, one recognizes that the different types of interchanges are associated with different types of crashes.

It is very important to analyze crash type by interchange type because the crash type provides clues for treatment of a hazardous site. For example, if there were a high percent of sideswipe crashes in an interchange, traffic engineers would analyze in detail the merge section to find the solution. If there were many rear end crashes at an interchange, one possibility is that the ramp length is too short to accelerate to freeway speeds before vehicles enter the mainline. If there are many over turn crashes at an interchange, one possibility is that there may be an imbalance between the radius of the ramp curve and the exit speed limit onto the ramp. Thus it is valuable to classify crashes according to the crash type.

Table 3.4 Summary of mainline and ramp crashes (1996~1998)

Interchange type		Total crashes	Mainline		Ramp	
			Crashes	%	Crashes	%
Diamond	Type 11	4479	3780	84.4	699	15.6
	Type 12	2211	1872	84.7	339	15.3
	Type 13	2886	2634	91.3	252	8.7
	Type 14	1380	1329	96.3	51	3.7
T-interchange	Type 21	681	486	71.4	195	28.6
Clover-leaf	Type 31	5388	4323	80.2	1065	19.8
	Type 33	1779	1470	82.6	309	17.4
	Type 35	1347	960	71.3	387	28.7
Directional	Type 41	4074	3135	77.0	939	23.0
Others	Type 51	699	642	91.8	57	8.2
Total		24924	20631	82.8	4293	17.2
V(x)		-	-	-	-	344

Table 3.5 Summary of the crash types (1996~1998)

Interchange type		Total crashes	Rear end		Fixed obj (over turn)		Sideswipe		Others	
			#	%	#	%	#	%	#	%
Diamond	Type 11	4479	2247	50.2	908	20.3	463	10.3	861	19.2
	Type 12	2211	910	41.2	450	20.3	360	16.3	492	22.2
	Type 13	2886	963	33.4	615	21.3	467	16.2	842	29.2
	Type 14	1380	604	43.7	321	23.3	102	7.4	353	25.6
T-interchange	Type 21	681	205	30.0	184	27.1	88	12.9	205	30.0
Clover-leaf	Type 31	5388	2122	39.4	1275	23.7	816	15.1	1175	21.8
	Type 33	1779	400	22.5	434	24.4	374	21.0	571	32.1
	Type 35	1347	694)	51.5	234	17.4	130	9.7	289	21.5
Directional	Type 41	4074	1527	37.5	625	15.3	590	14.5	1332	32.7
Others	Type 51	699	233	33.3	153	21.9	117	16.8	196	28.0
Total		24924	9905	39.7	5199	20.9	3509	14.1	6314	25.3
V(x)				172		53		118		85

3.3 Independent variable description

Independent variables used for this study consist of traffic data and geometric data. The traffic data are:

1) Mainline traffic volume,
2) Ramp traffic volume, and
3) Truck traffic volume and truck percent.

Average daily traffic (ADT) on mainlines of freeways has been shown to be an important contributing factor in predicting interchange traffic crashes. The Michigan Department of Transportation (MDOT) maintains about 100 permanent traffic recorders located at various sites throughout the state. The traffic volume data at these counter locations are used to estimate the ADT on all highway segments each year.

Ramp ADT is also considered to be an important independent variable for model development. The ramp ADT are traffic volumes on every on and off ramp (including loop) within the Ramp Counting program jurisdiction (Detroit Metropolitan area, Flint, Lansing, Grand Rapids, Jackson, etc). Any missing ramp data is estimated by reviewing previous years' traffic volumes and adjacent ramps. This adjustment implies an assumption that if traffic exits a freeway, it will return through the same intersection, going the opposite way.

Truck percent was also included, based on engineering intuition that truck ADT and mainline ADT, or truck ADT and ramp ADT may have the same mechanistic origin, which causes multicollinearity in crash prediction models.

Geometric data were obtained from the sufficiency rating file(1994) and freeway interchange inventory file(1997), which are maintained by the Michigan DOT. **Table 3.6** presents all variables that are intuitively thought to effect crash frequency, and are possible to obtain. An analysis of variance (ANOVA) of all independent variables was performed to determine which variables have a significant effect on the dependent variable (i.e., crash frequency). The results of this preliminary analysis are discussed in detail in section 3.4.

Table 3.6 Classification of independent variables

	Independent variables	
	Variable type 1	Variable type 2
Traffic effects	• Mainline traffic(ADT) • On ramp traffic(ADT) • On and Off ramp traffic(ADT) • Truck traffic(Truck ADT) • Truck percent (%)	
Geometric effects	• Interchange length (miles) • Average spread - ramp length (miles) • Average loop- ramp length (miles)	• Number of lanes • Number of on ramps • Total number of ramps • Shoulder width(feet) • Lighting condition

3.4 Preliminary analyses of variables

3.4.1 Correlation analysis

There is an implicit assumption in statistical model development that the independent variables are mutually independent. It is generally accepted that multicollinearity exists when a linear combination of independent variables is highly correlated, and that it is difficult to identify independent variable effects on the dependent variable (Neter et al. 1992, Sever and Wild 1989). Therefore, explanatory variables with low collinearity should be selected in the process of modeling.

To evaluate the mutual independence between variables, a correlation table was produced. As shown in **Table 3.7**, some of the independent variables are identified as relatively highly correlated. For example, the correlation between the ramp traffic volume and the interchange size, and the correlation between the mainline traffic volume and shoulder width are 0.454 and - 0.411 respectively. Those are not high enough to be excluded in the first stage of model developments. However, these variables are carefully dealt with in the detailed process of modeling.

Table 3.7 Correlations between independent variables

	Mainline traffic (per lane)	Ramp traffic volume	Truck traffic volume	Truck percent	Interchange length	Average loop-ramp length	Average spread- ramp length	Number of lanes	Number of on and off ramps	Lights	Shoulder width
Mainline traffic volume (per lane)	1.00										
Ramp traffic volume	0.375	1.00									
Truck traffic volume	0.028	0.003	1.00								
Truck percent	-0.404	-0.384	0.438	1.00							
Interchange length	-0.011	0.454	0.052	-0.040	1.00						
Average loop - ramp length	-0.104	-0.122	0.076	-0.054	0.067	1.00					
Average spread - ramp length	-0.282	0.080	0.194	-0.105	0.226	0.160	1.00				
Number of lanes	0.200	0.131	-0.434	0.077	-0.121	-0.206	-0.197	1.000			
Number of on and off ramps	0.078	0.409	0.045	0.047	0.384	-0.093	0.127	-0.133	1.000		
Lights	0.422	0.102	-414	-0.060	-0.158	-0.212	-0.298	0.248	-0.172	1.00	
Shoulder width	-0.411	-0.245	0.079	0.052	-0.058	0.000	0.238	-0.094	-0.046	-0.320	1.00

3.4.2 Analysis of variance (ANOVA)

Analysis of variance (ANOVA) techniques are a useful tool for analyzing the statistical relationship between a dependent variable and independent variables. In fact, these may be considered as a special case of linear regression. However, ANOVA models allow analyses of statistical relations from a different perspective than with linear regression, and therefore are widely used. In this section the ANOVA is used for the preliminary analyses of the relationship between the independent variables and a dependent variable. The independent variables are categorized into several groups before the ANOVA models are applied (i.e., for mainline ADT, 1: under 10000, 2:10000~15000, 3: 15000~20000, 4: over 20000).

The next step is to carry out a test whether or not the category means μ_j are equal. The hypothesis for this test is the following (Neter et al. 1992)

$$H_0 : \mu_1 = \mu_2 = \mu_3 \ldots\ldots = \mu_r$$

$$H_1 : \text{Not all } \mu_j \text{ are equal}$$

Here, H_0 implies that all of the probability distributions have the same mean, and thus there are no factor effects. Alternative H_1 implies that the means are not equal, and hence that there are factor effects. The F- test statistic and p-value are used as a decision rule for this test, and statistical package SPSS (9.0 version) is used to investigate the ANOVA.

3.4.2.1 ANOVA for traffic effects

When $\alpha = 0.05$, $F(0.95; 3, 195)$ is equal to 2.65. For mainline ADT from **Table 3.8**, the F- test statisitic=17.578>2.65. Thus we conclude H_1- that the mean crash frequency is not the same for the different mainline ADT categories. Similarly, ANOVA of ramp ADT and truck percent result in the same interpretion as that of mainline ADT. However, for truck ADT, the F-test statistic 0.244 is less than the critical value of 3.04, and hence we conclude H_0 -that the mean crash frequencies are the same for different truck ADT. The large p-value of the test in this table provides strong evidence that the sample data are in accord with equal mean frequencies for the different truck ADT. Mainline ADT, ramp ADT, and truck percent are thus expected to be contributing factors in the crash prediction models

Table 3.8 ANOVA for traffic effects

Source of variance		d.o.f	Mean square	F-test		P-value
				Statistic	Critical Value (α=0.05)	
Mainline ADT	Hypothesis	3	12887	17.578	2.65	0.000
	Error	195	733			
Ramp ADT	Hypothesis	2	28635	45.134	3.04	0.000
	Error	196	634			
Truck ADT	Hypothesis	2	225	**0.244**	**3.04**	**0.784**
	Error	196	924			
Truck percent	Hypothesis	2	10434	12.722	3.04	0.000
	Error	196	820			

3.4.2.2 ANOVA for geometric effects

Table 3.9 presents the results of ANOVA for geometric effects. For the variables of interchange size and average spread ramp length, the F-test statistics are 6.760 and 3.901, respectively, which exceed the critical value of 3.04. This implies that the mean accidents are not the same for the different length of interchange, or the different length of spread ramps. However, for average loop ramp length, the F-test statistic 0.146 is very small, compared to the critical value of 3.11, and hence we conclude H_0 - that the mean crashes are the same for the different length of loop ramps. The small P-value of the test in this table provides strong evidence of this conclusion.

On the other hand, the number of lanes and shoulder width are expected to be important independent variables for the prediction models based on F-test statistics that exceed critical values at α 0.05. However, for lighting, the F-test statistic (1.953) is less than the critical value of 3.04, and hence we can not conclude that mean accident frequencies are not the same for the different lighting conditions. In addition, the F- test statistic for the number of on-off ramps is 1.818, which is close to the critical value of 1.93.

Thus, the number of on and off ramps, the number of lanes, shoulder width, interchange length and average spread ramp length are expected to be contributing factors. However, there are no factor effects caused by lighting condition and average loop ramp length, and thus no further analyses which include these variables is required.

Table 3.9 ANOVA for geometric effects

Source of variance		d.o.f	Mean square	F-test		P-value
				Statistic	Critical value (α=0.05)	
A. Variable type 1						
Interchange length	Hypothesis	2	5860	6.760	3.04	0.001
	Error	196	866			
Average loop ramp length	Hypothesis	2	115	**0.146**	**3.11**	**0.703**
	Error	83	782			
Average spread ramp length	Hypothesis	2	3565	3.901	3.04	0.021
	Error	193	902			
B. Variable type 2						
Number of on and off ramps	Hypothesis	9	1608	**1.818**	**1.93**	**0.067**
	Error	189	884			
Number of lanes	Hypothesis	4	2206	2.477	2.42	0.046
	Error	194	890			
Shoulder width	Hypothesis	1	17458	20.950	3.89	0.000
	Error	197	833			
Lighting	Hypothesis	2	1703	**1.953**	**3.04**	**0.144**
	Error	196	872			

3.5 Model structure

Model structure is another issue in building an accident prediction model. However it is very difficult to choose the form of model equations because modeling remains, partly at least, an art (McCullagh and Nelder 1989). There are, however, some principles related to model structures which are summarized as follows.

(McCullagh and Nelder 1989):

- A good model is one that fits the observed data very well.
- Simplicity is a desirable feature of any model; we should not include parameters that we do not need.
- Models should make sense intuitively.
- If main effects are found from several studies bearing on the same phenomenon, the main effects should usually be included whether significant or not.

The above principles were used in the process of choosing model structures for this study. There are a few research papers on freeway interchanges, as mentioned in section 3.1. But these may not be appropriate guides for this study, since the models are based on a normally distributed and homogeneous error structure. For this reason, the findings from these studies related to traffic crash estimation at intersections have been reviewed based on the engineering intuition that the crash patterns at interchanges would be similar to those at intersections.

Several studies (Maher and Summersgill 1996, Persaud and Nguyen 1998, Bonneson and McCoy 1997, Vogt and Bared 1999) found that nonlinear relation is mainly proposed, and traffic volume belongs in the main effect group among the various variables.

To confirm the model structure, the cross tabulation between crash frequency and traffic volume were produced as shown in **Table 3.10.** This approach was performed in a similar manner by Bonneson and McCoy (1993), and Hauer et al.(1988). In **Table 3.10**, the traffic ranges were selected such that the same traffic ranges are located in each row, or each column, in order to obtain equal weight in calculating the average number of crashes per interchange. Therefore, 52 interchanges with traffic volumes that exceed these ranges were excluded in building the table.

The cells give the average number of crashes that have occurred for 3 years at interchanges with mainline volume and ramp volume given in the left-most column and the upper row. The brief examination of the row and column summaries indicates a positive relation between crashes and both mainline volume and ramp volume as shown in **Figure 3.3** and **Figure 3.4**. However, the rate of increase may be different, depending on the traffic volume.

For example, while crashes are always increasing over all ranges of mainline ADT, the increase is very small between mainline ADT 10000~15000 and 15000 ~20000, compared with other ranges of mainline ADT. This implies that the increase of crashes with mainline ADT is nonlinear, and the increase can be captured by a function such as V^B, where V is mainline ADT and B is a coefficient larger than 0.0.

Table 3.10 Cross tabulation of crashes by mainline volume and ramp volume

Ramp volume / Mainline volume	5000 ~ 15000	5000 ~ 15000	5000 ~ 15000	5000 ~ 15000	Summary Row
5000 ~ 10000	50[1] 903[2]/18[3]	88 1233/14	66 132/2	62 186/3	66 2454/37
10000 ~ 15000	55 721/13	100 2091/21	108 1624/15	148 1038/7	98 5474/56
15000 ~ 20000	57 454/8	122 1095/9	90 270/3	133 1065/8	103 2884/28
20000 ~ 25000	116 815/7	170 851/5	178 1420/8	175 1049/6	159 4135/26
Summary Column	63 2893/46	108 5270/49	123 3446/28	139 3338/24	102 14947/147

1): Average number of crashes per interchange
2): Total crashes
3): The number of interchanges

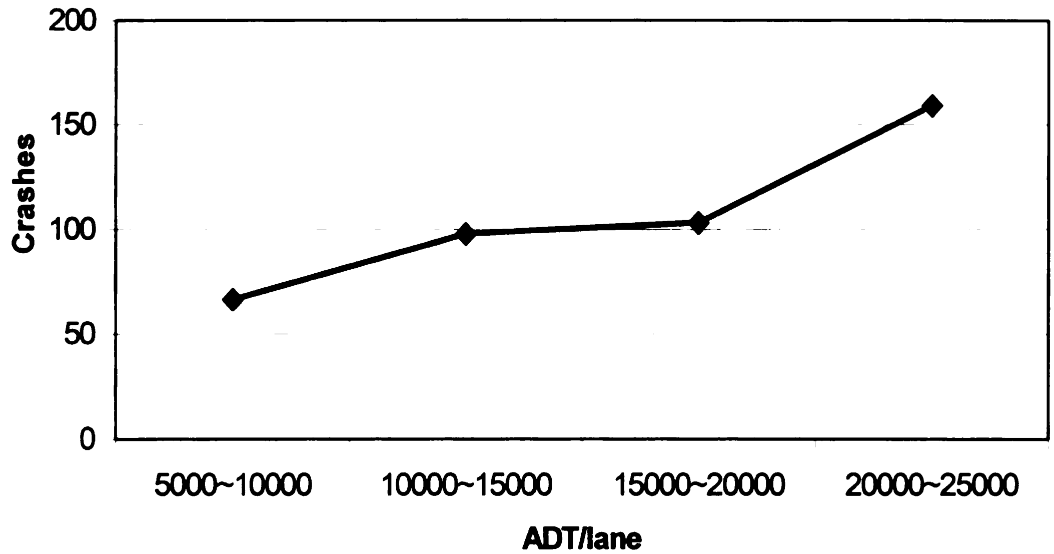

Figure 3.3 Mainline traffic volume and crashes

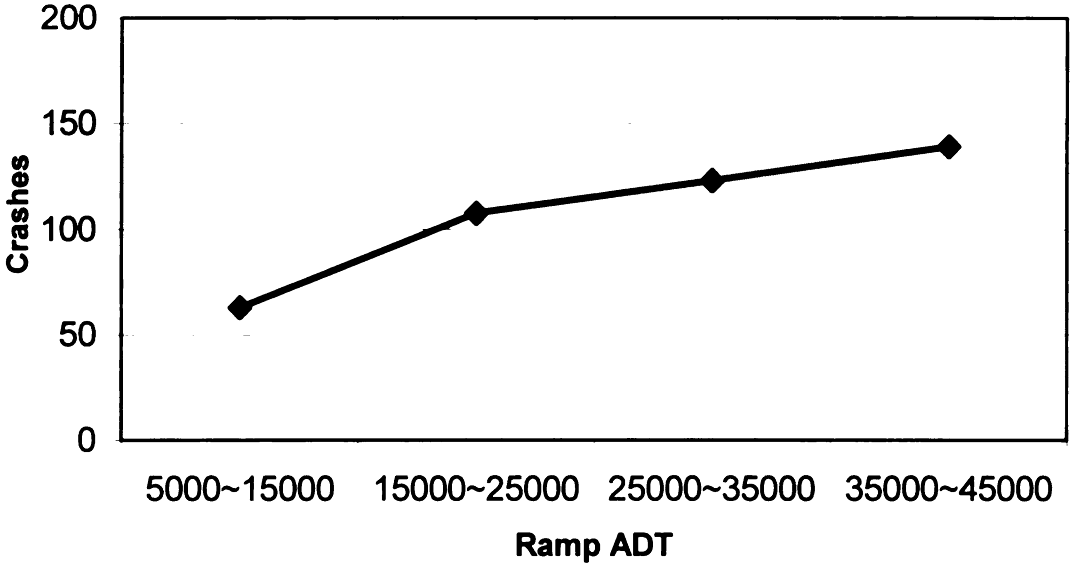

Figure 3.4 Ramp traffic volume and crashes

We can also determine from **Table 3.10** that there is a nonlinear relationship between crash frequency and traffic volume. For example, in the first column, crash frequencies increase sharply from 57 to 116 when the mainline volumes are changed from 15000~20000 to 20000 ~25000, whereas the crash frequencies increase only slightly (from 50 to 57) when the mainline volumes are changed from 5000~10000 to 15000~20000. These combinations can be found in other cells in **Table 3.10**, which is conceptually consistent with the nonlinear product of flows to power formulation as follows:

$$E(\theta) = A \times V_1^{B_1} \times V_2^{B_2} \tag{3.1}$$

where,

$E(\theta)$: Expected number of crashes
V_1 : Mainline volume
V_2 : Ramp volume
A, B_1, B_2 : Parameters

In principle, one should seek a model structure that best fits each interchange type. However, in this case, the model structure would be based on too small of a sample size to allow for finesse. Therefore, we regard equation (3.1) as the basic model structure describing the main effects of traffic variables on the interchange crash frequency.

The range of geometric variables is also an issue in choosing the appropriate model structure. The previous research found that the expected number of crashes can be represented by a product of geometric variables raised to various powers (Mountain et al. 1996), or by an exponential applied to a linear function of the geometric variables (Vogt and Bared 1999, Mahel and Summersgill 1996).

The effect of the range of possible geometric variables can not be evaluated efficiently, and hence, iterative tests of the model structures were performed. The results showed that a product of variables raised to various powers is appropriate for variables of type 1 (such as the size of interchanges), whereas an exponential applied to a linear function is appropriate for variables of type 2 (such as the number of on and off ramps).

On the basis of the literature review, the principles of model structures, and the results of the analyses, the general model structure for this study was finally determined to be of the following form:

$$E(\theta) = A \times V_i^{B_i} \times G_j^{C_j} \times \exp \sum (C_k \times G_k) \qquad (3.2)$$

where,

$E(\theta)$: Expected number of crashes

V_i : Traffic variables

G_j : Geometric variables(type 1)

G_k : Geometric variables(type 2)

A, B_i, C_j, C_k : Parameters

3.6 Model calibration and analysis

In section 3.4, the results of a preliminary analysis used to determine which variables have a significant effect on crash occurrences were discussed. The basic model structure that has been proposed in section 3.5 includes the independent variables that are significant as a result of the ANOVA. However, a variable can be insignificant when we put the variable into a nonlinear model structure stratified by interchange type, even though it has been evaluated to be significant in the preliminary analysis, because the preliminary ANOVA was performed independent of the interchange type. This issue is related to the simplicity of the model.

Simplicity is a desirable feature of any model as noted by McCullagh and Nelder (1989). This means that we should not include insignificant parameters in a model, noting that not only does a simple model enable the researchers to think about their data, but the model that involves only the correct variables gives better predictions than one that includes unnecessary variables. In this stage, the irrelevant terms from the general model structure are excluded, and the models are calibrated through checks on the fit of a model to the data, for example by residuals and other statistics.

A nonlinear regression model was proposed in the preceding section, and it has been verified that the crash occurrences follow a Negative Binomial distribution in chapter 2. Therefore, it is necessary to calibrate the coefficients of the crash prediction models and the Negative Binomial distribution parameter k simultaneously. There are two methods used to calibrate nonlinear regression models with a heterogeneous error structure (such as the Negative Binomial distribution): transformation of the model and generalized linear models (GLIM).

However, the transformation of models causes a change of scale in the data (Sever and Wild 1987, and McCullagh and Nelder 1989), which results in a violation of the Negative Binomial error assumption. Therefore, the analyses that follow are performed on the original scale of the data. This feature is a characteristic of generalized linear models (McCullagh and Nelder 1989). Previous researchers have suggested that the generalized linear models can be a technique to overcome the shortcomings of the conventional normally distributed error assumption in describing random, discrete and non-negative events which often occur in the traffic crash field (Rodriguez and Sayed 1999).

3.6.1 Link functions for the Generalized Linear Model (GLIM) approach

Recognizing that traffic crashes follow the Negative Binomial distribution as mentioned in chapter 2, the GLIM approach is utilized for model calibration. The GLIM approach used herein is based on the work of McCullagh and Nelder (1989), and Lawless (1987). The generalized linear modeling technique introduces a link function η that relates the linear equation to the expected value of an observation. This link function equates the non- linear relationship to a linear one.

On the other hand, there is a specific link function that is associated with the error structure of a distribution. This is defined as the natural link function. For example, natural link functions can be described for Normal, Poisson and Negative Binomial distributions as follows (McCullagh and Nelder 1989):

$$Normal : \eta = E(\theta)$$

$$Poisson : \eta = \ln[E(\theta)]$$

$$Negative\ Binomial : \eta = \left[\frac{E(\theta)}{K + E(\theta)}\right]$$

In order to describe the use of the Poisson link function, equation (3.2) in section 3.5 can be changed into a linear predictor as follows:

$$\begin{aligned}\eta &= \ln[E(\theta)] \\ &= \ln\left[A \times V_i^{B_i} \times G_j^{C_j} \times \exp\sum(C_k \times G_k)\right] \\ &= \ln A + B_i \ln V_i + C_j G_j + \sum(C_k \times G_k)\end{aligned}$$

Now, this is a linear predictive equation after applying the Poisson link function. However, our models for crash occurrence are based on the Negative Binomial distribution, and it is much harder to calculate a linear predictor from the natural link function for the Negative Binomial distribution. In fact, it is not algebraically possible to derive the linear predictor using the natural link function for the Negative Binomial distribution (Bonneson and Macoy 1997). Therefore, the Poisson link function is utilized instead, recognizing that the use of a natural link function is not a requirement for the GLIM approach (McCullagh and Nelder 1989).

In order to calibrate the prediction model, a dispersion parameter (D_p) will be utilized. That is, if D_p is greater than 1.0, then the data has a greater dispersion than is explained by the Poisson error assumption, and further analysis using the Negative Binomial error structure is required. In this case, the parameters are estimated in the iterative process using the maximum likelihood method. The model calibration procedures are explained in section 3.6.3.

3.6.2 Assessing the goodness of fit of the model

This section describes a basis of measuring the model significance. To make understanding easier, the following notations are used:

y_i : the observed number of crashes at a site i

$E(\theta)_i$: the expected number of crashes at a site i

$\bar{E}(\theta)$: the average expected number of crashes

$Var(y_i)$: estimated variance in crashes at a site i

n: sample size

p: the number of parameters

Several measures are used to assess the model fit and the significance of the model parameters, based on the studies of McCullagh and Nelder (1989), and Bonneson and McCoy (1997). One such measure is the generalized Pearson χ^2 statistic, which is calculated as:

$$Pearson\ \chi^2 = \sum_{i=1}^{n} \frac{(y_i - E(\theta)_i)^2}{\mathrm{var}(y_i)}$$

where var(y_i) is estimated from the variance equation of the Negative Binomial distribution which has been shown in equation (2.2). McCullagh and Nelder(1989) indicate that the generalized Pearson χ^2 statistic has the exact χ^2 distribution for a Normal linear model, while asymptotic results are available for other distributions. The asymptotic results may not be relevant to statistics calculated from a small sample size. Therefore it sometimes can not be used as an absolute measure for assessing the fit of a model.

A second measure of model fit is the Dispersion parameter (D_P), which can be calculated as:

$$Dispersion\ parameter(D_P) = \frac{Pearson\ \chi^2}{n - p}$$

As shown in the above formula, D_P can be obtained by dividing the Pearson χ^2 by n - p. McCullagh and Nelder(1989) indicated that it is a useful measure for assessing the fit of a model. A D_P value near 1.0 means that the error assumption of the model is equivalent to that found in observed data. If D_P is greater than 1.0, then the observed data has greater dispersion than is assumed in the model. This concept will be utilized in estimating the " k parameter " in the Negative Binomial distribution and the coefficients of the accident prediction models. This will be described in detail in the following section.

The third measure of model fit is the coefficient of determination R^2, which can be calculated as:

$$R^2 = 1 - \frac{SSE}{SST}$$

where

$$SSE = \sum_{i=1}^{n} \left[E(\theta)_i - y_i \right]^2$$

$$SST = \sum_{i=1}^{n} \left[y_i - \overline{E}(\theta) \right]^2$$

This measure is commonly used for the fit of a linear regression model based on the normally distributed error assumption. Nevertheless, this statistic can still be useful in assessing the model fit, recognizing the findings that the coefficient of determination R^2 is still efficient in assessing a model calibrated under a non normal error structure (Kvalseth 1985).

The fourth measure of model fit is the Pearson Residual, which can be calculated as:

$$\text{Pearson Residual}\,(\text{PR}_i) = \frac{E(\theta)_i - y_i}{\sqrt{\text{var}(y_i)}}$$

As shown in this formula, this is defined as the difference between the predicted and observed data divided by the standard deviation. The Pearson Residual will be discussed again in section 3.6.5.

In addition to these measures, the standard error and t-value are used for assessing the significance of variable coefficients. The t-value is the ratio between the variable coefficient and its standard error. The detailed descriptions of these statistics are not presented here since the concepts are commonly applied in measuring the fit of linear regression models.

3.6.3 The procedures used in parameter calibration

The calibration of model parameters was performed based on the works of Lawless (1987). The calibration for this research is a multi-step process as shown in **Figure 3.5**.

First, the model parameters are estimated based on the Poisson error structure that the variance equals to the expected value. Using the expected number being calculated in the first step, the second step is to estimate the "k" parameter. If 1/k is not greater than 0.0, then there is no over-dispersion in the observed data and the procedure stops. If 1/k is greater than 0.0, then a third step is to calculate new model coefficients under the Negative Binomial error structure using the k from the second step. In this step, the maximum likelihood estimates of the model coefficients are obtained by iterative weighted least squares. The final step is to calculate the Dispersion parameter (D_P). If D_P does not equal 1.0, the k parameter is increased (or decreased) and then a feedback loop is performed to the third step. The analysis is repeated in an iterative manner until the Dispersion parameter (D_P) converges to 1.0.

Models with Negative Binomial errors can not be calibrated using conventional statistical packages (i.e., SPSS, SYSTAT), and thus a statistical package for Generalized Linear Interactive Modeling (GLIM), which is specially designed to calibrate models with special types of errors (i.e., Negative Binomial, Poisson and Gamma), was used. Rodriguez and Sayed (1999) used a similar process in calibrating the traffic crash prediction models for urban unsignalized intersections.

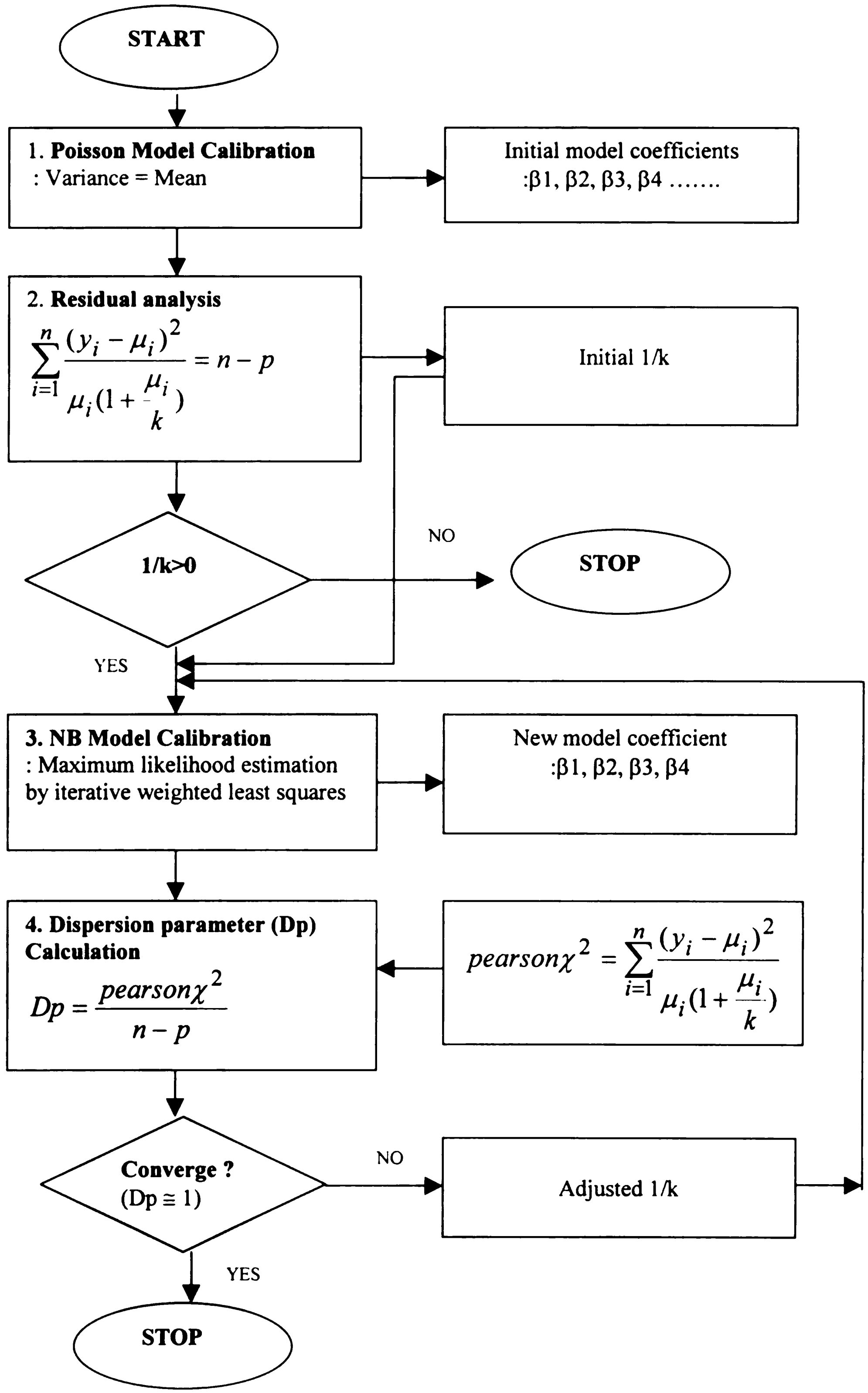

Figure 3.5 The process to calibrate coefficients & k parameter

3.6.4 Results of the model calibration

On the basis of the procedures for assessing the model fit explained in the previous sections, the crash prediction models have been calibrated. The logarithmic link function has the following basic form, as mentioned in section 3.6.1:

$$\ln[E(\theta)] = \ln A + B_i \ln V_i + C_j G_j + \sum (C_k \times G_k) \qquad (3.3)$$

This equation can be rewritten in a more useful form as:

$$E(\theta) = A \times V_i^{B_i} \times G_j^{C_j} \times \exp \sum (C_k \times G_k) \qquad (3.4)$$

where,

$E(\theta)$: Expected number of crashes
V_i : Traffic variables
G_j : Geometric variables
G_k : Geometric variables
A, B_i, C_j, C_k : Parameters

The model calibration process starts with individual models according to the interchange types that have been classified in section 3.2. **Table 3.11** presents several statistics relating to the calibrated crash prediction model for interchange type 11. In determining the significance of the variable coefficients, the 95 percent confidence level is used with a few exceptions. In the second row of the table, the statistic for the constant terms does not have any meaning since the logarithm results in a change of scale.

The table indicates that several variables have a significant effect on the frequency of interchange crashes. These variables are mainline traffic, ramp traffic, truck percent, interchange size, spread ramp length, and shoulder width. However, the number of lanes and the number of total ramps are not included in this model because the effect of these variables is not significant. The calibrated coefficients can be applied to the equation (3.4) that is the basic model structure, in order to predict the number of traffic crashes that would be expected for 3 years in interchange type 11. The resulting model can be written as follow:

$$E(\theta) = 3.448\, V_1^{1.401}\, V_2^{0.186}\, V_3^{0.620}\, G_1^{0.738}\, \exp(-1.267\, G_2 - 0.156\, G_5)$$

where,

V_1 : Mainline traffic volume per lane
V_2 : Ramp traffic volume
V_3 : Truck percent
G_1 : Interchange length
G_2 : Average spread - ramp length
G_5 : Shoulder width

A k parameter of 8.05 is found to yield a dispersion parameter of 1.0. The Pearson χ^2 is 28.84, and the degrees of freedom are 27(n-p-1=34-6-1). This statistic is less than $\chi^2_{0.05,\,27} = 40.11$, and hence the hypothesis that the model fits the data can not be rejected. It implies that the model is consistent with the observed data.

Several statistics associated with the calibrated crash prediction models for other interchange types are included as Appendix 1.

Table 3.11 The results of crash prediction model calibration

(Interchange type 11)

Coefficient	Variable definition	Unit	Estimate	Std error	t - statistic
A Log(A)	Constant	-	3.448 (1.238)	(0.67)	(1.85)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	1.401	0.30	4.66
B_2	V_2 : Ramp traffic volume	(ADT/1000)	0.186	0.12	1.55
B_3	V_3 :Truck percent	(%)	0.620	0.19	3.26
C_1	G_1 : Interchange length	(Mile)	0.738	0.15	4.92
C_2	G_2 : Average spread- ramp length	(Mile)	-1.267	0.97	-1.31
C_3	G_3 : The number of lanes	-			
C_4	G_4 : The number of total ramps	-			
C_5	G_5 : Shoulder width	(Feet)	-0.156	0.12	-1.30

Model statistic		
D_p	Dispersion parameter	1.0
X^2	Pearson chi -square	28.84 ($\chi^2_{0.95,\ 27} = 40.11$)
R^2	Coefficient of determination	0.60
K	Negative Binomial parameter	8.05

3.6.5 Pearson Residuals

A useful subjective measure of the model fit is the Pearson Residuals(PR),which are normalized residuals in the context that Pearson Residuals are the difference between the predicted and observed data divided by the standard deviation as described in section 3.6.2. One can visually assess the goodness of model fit by plotting the Pearson Residuals versus the estimates of the expected number of crashes. A good model will have the Pearson Residuals centered around 0.0.

Pearson Residuals are plotted against the expected frequency for the 199 interchanges in **Figure 3.6**. As shown in the figure, Pearson Residuals are centered around 0.0 for the entire range of expected frequency, which indicates that the calibrated models fit the observed data well. The advantage of the Negative Binomial error assumption in crash model development will be examined again in section 3.7.

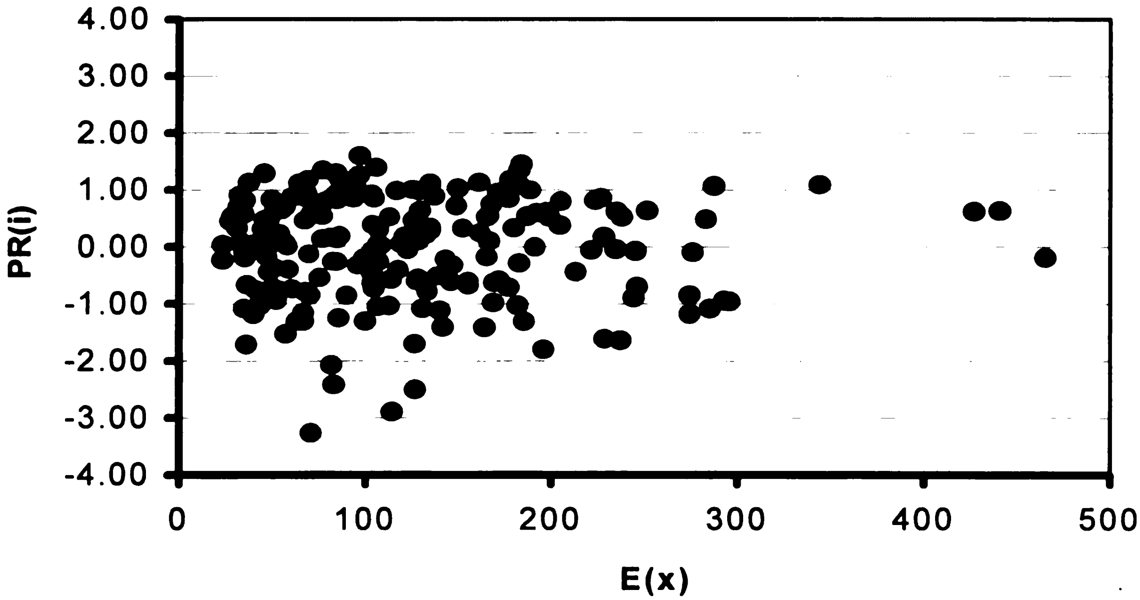

Figure 3.6 Pearson Residuals and E(x)

3.7 A comparison of model calibration results according to the error structure (Normal versus Negative Binomial assumption)

In section 3.1, it was noted that there are two error structures used in calibrating traffic crash prediction models. One is a normally distributed and homogeneous error structure, the other is a non- normal and heterogeneous error structure. Recently, Poisson or Negative Binomial error structures are most often assumed in modeling traffic crashes. Nevertheless, literatures reviewed did not contain a full description of the advantages relating to this approach. In order to examine the advantages of the Negative Binomial error assumption, the results of model calibrations for interchange type 11 and type 12 are compared in **Table 3.12**. The parameter estimates and their standard errors are very sensitive to the error structures assumed.

An extremely useful relative measure of the scatter of a random variable is its coefficient of variation V(x) (Harr 1996). This implies that the coefficient of variation is a measure of the reliability of the calibrated model coefficients. There were large reductions in the coefficient of variation, as shown in the table, when the models were calibrated using the Negative Binomial distribution instead of the Normal distribution. This reduction in the coefficient of variation occurs in all coefficients that have been calibrated as shown in the table, with a maximum reduction of 80 % and an average reduction of 30 %. These results support the hypothesis that the Negative Binomial distribution is a desirable assumption in calibrating crash prediction models relating to freeway interchanges.

Table 3.12 A comparison of model calibration results according to error structure (Normal and Negative Binomial assumption)

Type11	Normal			Negative Binomial			Reduction of V(x)
Parameter	Estimate (A)	Std.err (B)	V(x) (B/A×100) (%)	Estimate (C)	Std.err (D)	V(x) (D/C×100)	(%)
B1	1.118	0.37	**33**	1.401	0.30	**21**	**36**
B2	0.108	0.15	**142**	0.186	0.12	**64**	**55**
B3	0.425	0.24	**55**	0.620	0.19	**30**	**45**
G1	0.558	0.20	**36**	0.738	0.15	**20**	**44**
G2	-0.600	1.13	**189**	-1.267	0.97	**77**	**50**
G5	-0.297	0.26	**87**	-0.156	0.12	**77**	**11**
Type 12							
B1	1.003	0.26	**26**	0.946	0.24	**25**	**4**
G1	0.570	0.22	**39**	0.933	0.36	**39**	**0**
G2	-0.705	1.18	**167**	-3.842	1.31	**34**	**80**
							Max : 80 Min : 0 Avg : 36

3.8 Sensitivity analysis of the crash prediction model

There are two objectives associated with a sensitivity analysis: One is to examine the possibility that the crash prediction model violates conceptual rules. For example, if a model were designed such that its predicted crashes would decrease with an increase in ramp volume, the model should be rejected because it violates a conceptual rule. The other objective is to determine the effects of individual variables on the crash frequency at freeway interchanges.

The sensitivity analysis is performed for the major geometric variables, but not for the traffic variables because it is possible to change the geometry, but changing traffic is difficult. During the sensitivity analysis of a specific variable, other design parameters are assumed to be a constant. For this analysis, an experimental matrix was established, which includes 3 experiments (A: 0.1 mile shorter than mean, B: mean, C: 0.1 mile longer than mean) for interchange length, 3 experiments (A: 0.1 mile longer than mean, B: mean, C: 0.1 mile shorter than mean) for spread -ramp length, and 2 experiments (A: 12 feet and B: 10 feet) for shoulder width.

Table 3.13 illustrates the results of the sensitivity analyses. In the sensitivity analysis of interchange size, when the interchange size is increased by 0.1 mile, traffic accidents increase in all interchange types which use this variable as a model component. The average increase is 14 %.

In the sensitivity analysis of the spread- ramp length, traffic crashes increase by an average of 26 % when the spread- ramp length is decreased by 0.1 mile. The traffic crashes increase most rapidly for interchange type 12 (Tight diamond interchanges), which increases by 47 %. The crash frequency is very sensitive to shoulder width for both interchange types that include this variable, and especially for type 41(Directional interchanges). In the sensitivity analyses, no violation of conceptual rules of traffic crashes were found.

Table 3.13 Sensitivity analysis(effect of main geometric variables)

Parameter	Interchange type	Experiment (A)	Experiment (B)	Experiment (C)	Effects
Interchange Length	Length	0.534 mile	0.634 mile	0.734 mile	0.1 Mile (↑)
	Type 11	0.629	0.714	0.796	1.12
	Type 12	0.557	0.654	0.749	1.16
	Type 13	0.599	0.689	0.777	1.14
	Type 14	0.438	0.549	0.666	1.23
	Type 31	0.819	0.865	0.906	1.05
	Type 33	0.549	0.647	0.744	1.16
	Mean	0.599	0.686	0.773	1.14
Spread-ramp length	Length	0.33 mile	0.23 mile	0.13 mile	0.1 mile (↓)
	Type 11	0.658	0.747	0.848	1.14
	Type 12	0.281	0.413	0.607	1.47
	Type 14	0.472	0.592	0.744	1.26
	Type 33	0.438	0.563	0.723	1.28
	Mean	0.462	0.579	0.730	1.26
Shoulder width	Width	12 ft	10 ft		2.0 feet(↓)
	Type 11	0.154	0.211		1.37
	Type 41	0.057	0.093		1.63
	Mean	0.106	0.152		1.50

Chapter 4

THE DEVELOPMENT OF A METHOD TO IDENTIFY HAZARDOUS SITES BASED ON THE NEGATIVE BINOMIAL DISTRIBUTION

4.1 General

The traditional rate quality control method is based on the assumption that the probability of crash occurrences follow a Poisson distribution(Zegeer and Deen 1977) in which the mean and the variance are equal. The normal approximation to the Poisson provides a control chart without tedious interpolation from the table of the Poisson distribution(Orlansky and Jacobs 1956), and this chart has been commonly used for the identification of hazardous locations.

In chapter 2, the fact that the variance of crash occurrences at freeway interchanges is substantially larger than the mean was discussed, based on the observed data. This over-dispersion can be better explained by using the Negative Binomial distribution. A control chart constructed under the assumption of the Poisson distribution, can not reflect the phenomenon of over-dispersion in identifying hazardous locations.

The purpose of this chapter is to describe a technique to overcome the limitation of the rate quality control method based on the Poisson distribution. One statistician

(Rice 1997) suggests that the Negative Binomial distribution might be considered as a model for situations in which the rate varies over time and over space.

Thus, in this chapter, the rate quality control method will be developed under the assumption of the Negative Binomial distribution.

4.2 A review of the statistical methods identifying hazardous sites

4.2.1 The rate quality control method

The rate quality control method is one of the most common methods used to identify hazardous sites. This method was originally developed as a means to control the quality of industrial production(Norden et al.1956). This approach uses a statistical test to determine whether the traffic accident rate for a particular location is abnormally high compared with the rate of reference sites with similar properties. The statistical test is based on the assumption that traffic crashes are rare, hence the probability of their occurrences follows a Poisson distribution(Zegeer and Deen 1977).

There have been changes in the original equations based on a comparison of the errors between real values and estimated values obtained from the rate quality control method formula. The following is a brief description of changes of the rate quality control method.

The rate quality control method was proposed as a way to analyze crash data on highway sections in 1957 using the following formula:

$$UCL = \lambda + 2.576\sqrt{\frac{\lambda}{V_i} + \frac{0.829}{V_i} + \frac{1}{2V_i}} \tag{4.1}$$

$$LCL = \lambda - 2.576\sqrt{\frac{\lambda}{V_i} + \frac{0.829}{V_i} - \frac{1}{2V_i}} \tag{4.2}$$

where

UCL : upper control limit

LCL : lower control limit

λ : average accident rate of reference sites($\sum N_i / \sum V_i$)

N_i : the number of accidents at site i

V_i : the number of vehicles at site i

A decade later, it was recommended that the correction term($0.829/V_i$) be eliminated to improve the validity of the equations(Morin 1967). Thus, the following equations are currently in use to calculate the upper and lower limits for the rate quality control method.

$$UCL = \lambda + z\sqrt{\frac{\lambda}{V_i} + \frac{1}{2V_i}} \tag{4.3}$$

$$LCL = \lambda - z\sqrt{\frac{\lambda}{V_i} + -\frac{1}{2V_i}} \tag{4.4}$$

where

z : predetermined significance level

With the rate quality control method, a site is identified as hazardous if its observed accident rate exceeds the mean accident rate of similar sites plus a multiple of the standard deviation of the site accident rate, which is called the critical accident rate. The critical accident rates can be calculated for each site by applying the following equation:

$$RC_i = \lambda + z\sqrt{\frac{\lambda}{V_i}} + \frac{1}{2V_i} \tag{4.5}$$

where

RC_i : Critical rate for site i

In the above equation, the first two terms result from the Normal approximation to the Poisson distribution, the third term is a correction factor necessary because only integer values are possible for the observed number of accidents. The coefficient of the second term describes a probability factor determined by the level of statistical significance desired for RC_i. The FHWA, however, proposes the following equation for calculating the critical accident rate(Stokes and Mutabazi 1996).

$$RC_i = \lambda + z\sqrt{\frac{\lambda}{V_i}} - \frac{1}{2V_i} \tag{4.6}$$

where

RC_i : Critical rate for site i

Equation(4.6) is different from equation(4.5) in that the sign of last term is negative, and the difference results from whether a probability should be included or excluded if the rate is equal to the critical rate. The method of identifying hazardous sites by the rate quality control method can be visually explained by Figure 4.1, where filled stars correspond to the hazardous sites chosen under the rate quality control method.

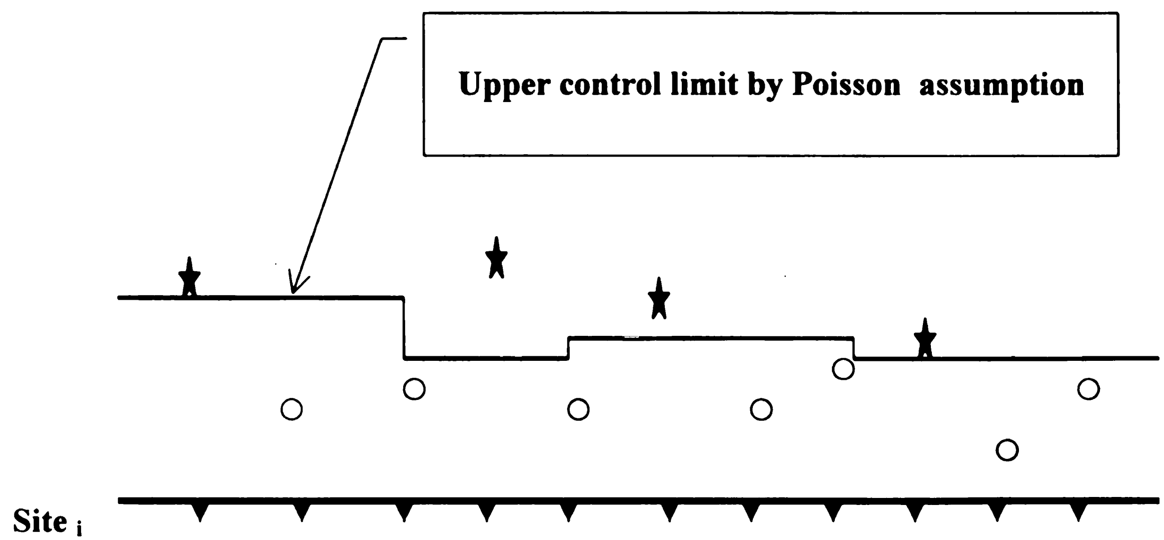

Figure 4. 1 The concept of rate quality control method under the Poisson distribution

4.2.2 The Bayes approach

Higle and Witkowski(1988) introduced and illustrated how Bayesian theory can be used to identify the hazardous intersections from among many signalized intersections. Hauer(1986) then proposed the application of Emperical Bayesian(EB) theory in traffic safety problems, based on Robbin's work(Robbin 1977,1979,1980) and this method has subsequently been used by many researchers (Maher and Summersgill 1996, Persaud 1993, Belanger 1994). Both the Bayesian approach and the EB method which are described above, have the following assumptions(Higle and Witkowski 1988).

Assumption 1: At a given site, when the average accident rate of reference sites(λ) is known, the count of accidents(N) obeys the Poisson probability law with expected value(λV_i).

$$p(N = N_i / \lambda V_i) = \frac{\exp(-\lambda V_i)(\lambda V_i)^{N_i}}{N_i} \qquad (4.7)$$

where,

λ : average accident rate of reference sites($\sum N_i / \sum V_i$)

λV_i : the expected value at site i

Assumption 2: The accident rate of reference sites(which the given site belongs to) can be described by a Gamma probability density function such as:

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \qquad (4.8)$$

where,

$f(\lambda)$: gamma probability density function of reference sites

α, β : parameters

Equation (4.8) is also denoted as the probability density function of the prior distribution in terms of Bayes theory. Here, parameters α and β can be estimated through the method of moment estimates(MME) or the maximum likelihood estimates(MLE).

Under the assumption that α and β were calibrated, if the observed data at a given site i are N_i and V_i, the probability density function of the posterior distribution can be described as:

$$f(\lambda / N_i, V_i) = \frac{(\beta + V_i)^{\alpha + N_i}}{\Gamma(\alpha + N_i)} \lambda^{\alpha + N_i - 1} e^{-(\beta + V_i)\lambda} \tag{4.9}$$

Equation (4.9) is the posterior distribution using Bayesian theory in its original meaning, and we can evaluate the hazard of a given site using this equation. That is, the probability that Bayes accident rate at site i, λ_i, exceeds an average accident rate of reference sites, λ_{avg}:

$$P(\lambda_i \geq \lambda_{avg}) = 1 - P(\lambda_i < \lambda_{avg})$$

$$= 1 - \int_0^{\lambda_{avg}} \frac{(\beta + V_i)^{\alpha + N_i}}{\Gamma(\alpha + N_i)} \lambda^{\alpha + N_i - 1} e^{-(\beta + V_i)\lambda} \, d\lambda \tag{4.10}$$

Figure 4.2 shows a graphical representation of the equation (4.10)

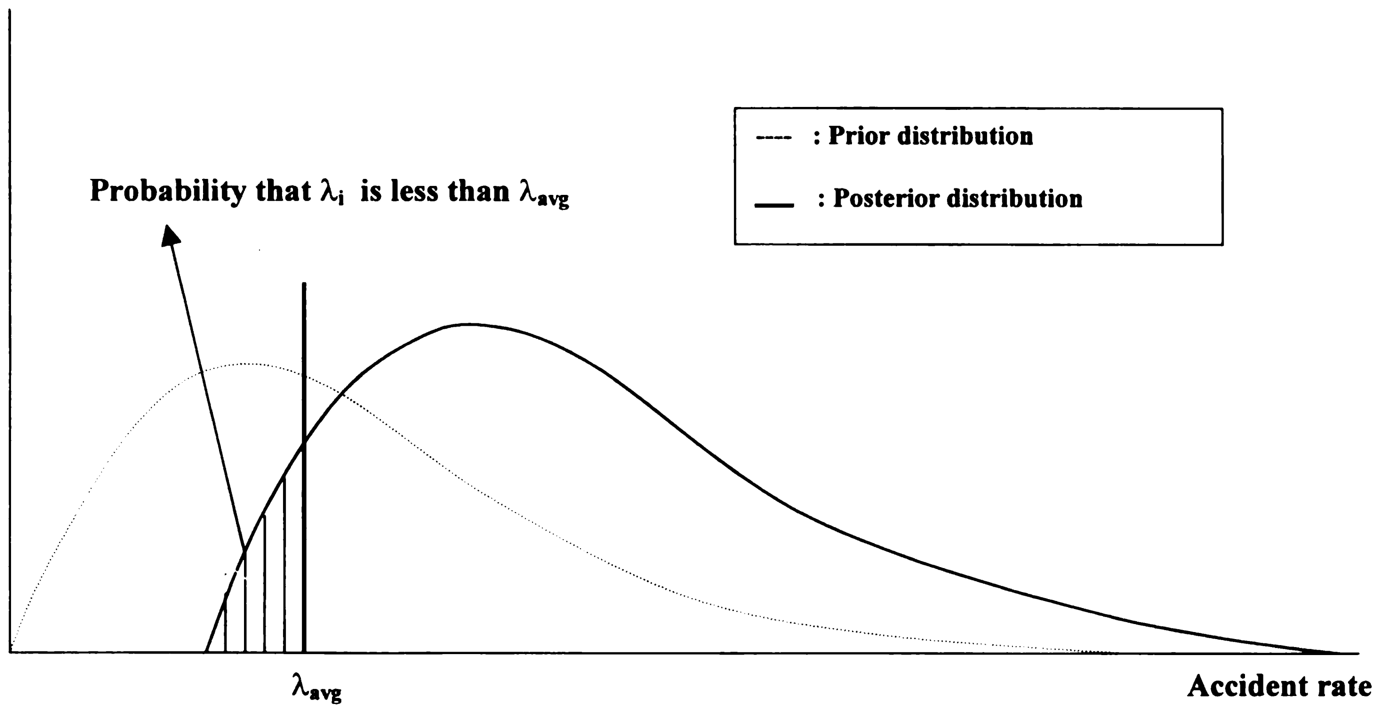

Figure 4. 2 Identification of hazardous sites by Bayes approach

4.2.3 The problems resulting from the use of the Poisson distribution

The rate quality control method as used in practice identifies a site as hazardous if its observed accident rate over a given period exceeds its critical accident rate, which is the average accident rate over reference sites plus a multiple of the standard deviation of the accident rate of the site over the same period. This rate quality control method is based on the Poisson distribution as mentioned is section 4.1. However, the Negative Binomial distribution fits the freeway interchange crash data much better than the Poisson distribution. Thus, this identification of hazardous sites may not be valid.

Recognizing that the variance of the Poisson distribution equals the mean, whereas the variance of the Negative Binomial distribution is (mean+mean ^{2} / k), the existing approach under the Poisson assumption will identify more sites as hazardous than would be expected under the Negative Binomial assumption. For example, in Figure 4.3, the solid line is the upper control limit chart based on the Poisson distribution while the dotted line is the upper control limit chart based on the Negative Binomial distribution. The stars correspond the hazardous sites which are chosen under the Poisson or Negative Binomial distribution.

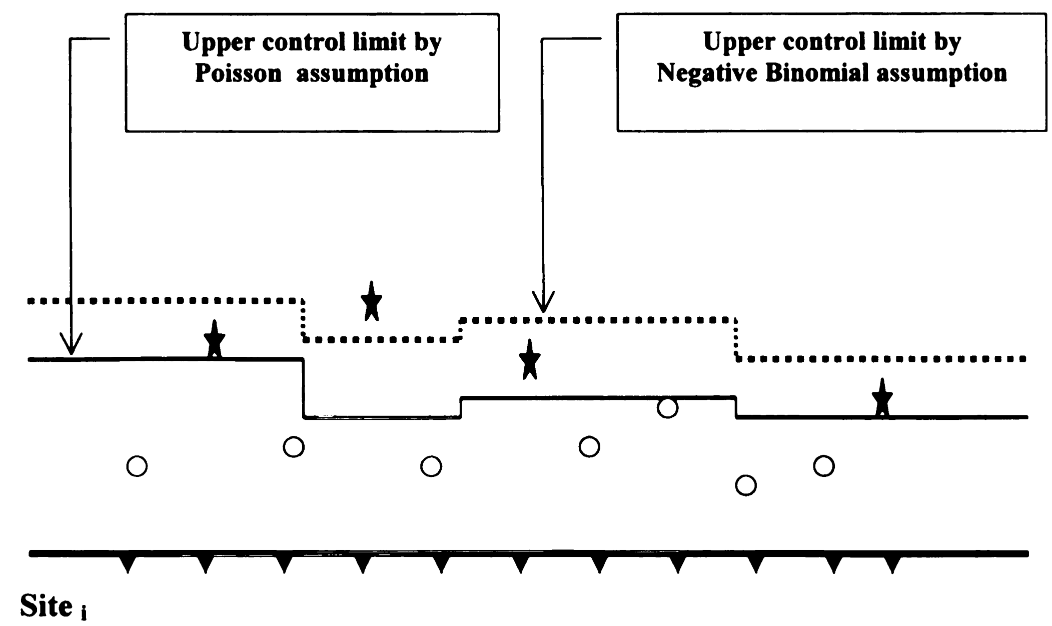

Figure 4. 3 A comparison of the Negative Binomial and Poisson assumption in applying rate quality control method.

4.3 The identification of hazardous sites based on the Negative Binomial distribution.

4.3.1 Concept

In equation (4.1) and (4.3), the true value of the upper control limit for the Normal approximation to the Poisson distribution on traffic crash frequency can be computed from the following equation:

$$P = \sum_{y=0}^{U-1} \frac{e^{-m} m^{y}}{y!} \tag{4.11}$$

where,

P : predetermined probability level
U : true upper control limit
m : expected value

In equation(4.11),

$$m = \lambda V_i ,$$

$$\lambda = \frac{\sum N_i}{\sum V_i}$$

where,

N_i = the accident frequency at site i
V_i = the number of vehicles at site i

Under the condition that the average accident rate of reference sites, λ, is known, the true control limit for a site i can be calculated by selecting the value that meets the predetermined probability levels (i.e., 0.9, 0.95) from equation (4.11).

The above concept can be utilized for the Negative Binomial distribution. That is, when crash occurrences follow the Negative Binomial distribution, the formula for the true upper control limit is as follows:

$$P = \sum_{y=0}^{U-1} \left(1 + \frac{m}{k}\right)^{-k} \frac{\Gamma(k+y)}{y!\Gamma(k)} \left(\frac{m}{m+k}\right)^{y} \qquad (4.12)$$

where,

U : the true upper control limit

m : mean

k : parameter

From equation (4.12), we can compute the true upper control limit under a desired probability level based on the Negative Binomial distribution for a given site. However, for the Negative Binomial distribution, the estimates of parameters (m ,k) are not as simple as those of the Poisson distribution. The method for these estimates will be described at length in section 4.3.2.

4.3.2 Estimation of the parameters

4.3.2.1 Derivation of maximum likelihood equation

There are two methods for estimating the parameters of the Negative Binomial distribution: one is a moment method, the other is a maximum likelihood method (Rice 1997). As mentioned in section (4.1), over-dispersion occurs over time, which implies that each site has its own distribution, and hence in fact, there are many distributions in a sample. Thus, for this research, the maximum likelihood method is used, recognizing that the parameter k can not be determined with an acceptable efficiency by a moment method for multiple distributions(Bliss 1953).

In order to estimate the parameters, two equations were derived from the maximum likelihood function, based on the Lawless' work(1987). For simplicity, equation (2.2) is transformed as equation (4.13), and thus the parameter k equals 1/a.

$$f(y/a,m) = \frac{\Gamma(y+a^{-1})}{y!\,\Gamma(a^{-1})}\left(\frac{am}{1+am}\right)^{y}\left(\frac{1}{1+am}\right)^{a^{-1}} \qquad (4.13)$$

The likelihood function is :

$$L(y_i,a,m) = \prod_{i=1}^{n} \frac{\Gamma(y_i+a^{-1})}{\Gamma(a^{-1})}\left(\frac{am}{1+am}\right)^{y_i}\left(\frac{1}{1+am}\right)^{a^{-1}} \qquad (4.14)$$

let $m_i = m_0 V_i$

$$L(y_i,a,m_0,V_i) = \prod_{i=1}^{n} \frac{\Gamma(y_i+a^{-1})}{\Gamma(a^{-1})}\left(\frac{am_0V_i}{1+am_0V_i}\right)^{y_i}\left(\frac{1}{1+am_0V_i}\right)^{a^{-1}} \qquad (4.15)$$

Noting that for any $c > 0, \Gamma(y+c)/\Gamma(c) = c(c+1)(c+2)\cdots\cdots(c+y-1)$,

$$\frac{\Gamma(y_i + a^{-1})}{\Gamma(a^{-1})} = \frac{1}{a}\left(\frac{1}{a}+1\right)\left(\frac{1}{a}+2\right)\cdots\cdots\cdots\left(\frac{1}{a}+y_i-1\right)$$

Now, we can write the log likelihood function, log $L(y_i, a, m_0, V_i)$ *as*

$$l(y_i, a, m_0, V_i) = \sum_{i=1}^{n} \log\left[\frac{\Gamma(y_i + 1/a)}{\Gamma(1/a)}\left(\frac{am_0V_i}{1+am_0V_i}\right)^{y_i}\left(\frac{1}{1+am_0V_i}\right)^{a^{-1}}\right]$$

$$= \sum_{i=1}^{n}\left[\log\frac{\Gamma(y_i+1/a)}{\Gamma(1/a)} + \log\left(\frac{am_0V_i}{1+am_0V_i}\right)^{y_i} + \log\left(\frac{1}{1+am_0V_i}\right)^{a^{-1}}\right] \tag{4.16}$$

In equation (4.16)

The first term,

$$\sum_{i=1}^{n}\log\frac{\Gamma(y_i+1/a)}{\Gamma(1/a)} = \log\left[\frac{1}{a}\left(\frac{1+a}{a}\right)\left(\frac{1+2a}{a}\right)\left(\frac{1+3a}{a}\right)\cdots\cdots\left(\frac{1+(y_i-1)a}{a}\right)\right]$$

$$= \log 1 + \log(1+a) + \log(1+2a)\cdots\log(1+(y_i-1)a) - y_i\log a$$

$$= \sum_{j=0}^{y_i-1}\log(1+aj) - y_i\log a \tag{4.17}$$

The second term,

$$\sum_{i=1}^{n} \log\left(\frac{am_0V_i}{1+am_0V_i}\right)^{y_i} = y_i \log(am_0V_i) - y_i \log(1+am_0V_i)$$

$$= y_i \log a + y_i \log m_0 + y_i \log V_i - y_i \log(1+am_0V_i) \quad (4.18)$$

The third term,

$$\sum_{i=1}^{n} \log\left(\frac{1}{1+am_0V_i}\right)^{\frac{1}{a}} = -\frac{1}{a}\log(1+am_0V_i) \quad (4.19)$$

Thus, from equations (4.17), (4.18) and (4.19) the log likelihood function can be summarized as follow :

$$l(y_i, a, m_0, V_i)$$

$$= \sum_{i=1}^{n}\left[\sum_{j=0}^{y_i-1} \log(1+aj) + y_i \log m_0 - y_i \log(1+am_0V_i) - \frac{1}{a}\log(1+am_0V_i)\right] \quad (4.20)$$

The simplest way to obtain (m_0, a) is to maximize $l(y_i, a, m_{0.} V_i)$ with respect to (m_0, a), and thus, we need to set the partial derivatives of the $l(y_i, a, m_{0.} V_i)$ equal to zero.

That is,

$$\frac{\partial l(y_i, a, m_{0.} V_i)}{\partial m_0} = 0 \tag{4.21}$$

$$\frac{\partial l(y_i, a, m_{0.} V_i)}{\partial a} = 0 \tag{4.22}$$

$$\frac{\partial l(y_i, a, m_{0.} V_i)}{\partial m_0} = \sum_{i=1}^{n} \left[\frac{y_i}{m_0} - \frac{y_i a V_i}{1 + a m_0 V_i} - \frac{V_i}{1 + a m_0 V_i} \right]$$

$$= \sum_{i=1}^{n} \left[\frac{y_i - m_0 V_i}{m_0 (1 + a m_0 V_i)} \right] \tag{4.23}$$

$$\frac{\partial l(y_i, a, m_{0.} V_i)}{\partial a} = \sum_{i=1}^{n} \left[\sum_{j=0}^{y_i - 1} \left(\frac{j}{1 + aj} \right) - \frac{y_i m_0 V_i}{1 + a m_0 V_i} + \frac{1}{a^2} \log(1 + a m_0 V_i) - \frac{1}{a} \frac{m_0 V_i}{1 + a m_0 V_i} \right]$$

$$= \sum_{i=1}^{n} \left[\sum_{j=0}^{y_i - 1} \left(\frac{j}{1 + aj} \right) + \frac{1}{a^2} \log(1 + a m_0 V_i) - \frac{m_0 V_i (y_i + a^{-1})}{1 + a m_0 V_i} \right] \tag{4.24}$$

From equation (4.23),

$$m_0 = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} V_i} \tag{4.25}$$

From equation (4.24),

$$\sum_{j=0}^{y_i-1}\left(\frac{j}{1+aj}\right) + \frac{1}{a^2}\log(1 + am_0V_i) - \frac{m_0V_i(y_i + a^{-1})}{1 + am_0V_i} = 0 \tag{4.26}$$

The maximum likelihood estimate of " m_0" can be easily calculated using equation (4.25), but that of " a " is not simply obtained because it is not a closed form. Therefore, a numerical approach is used to solve equation (4.26).

4.3.2.2 Parameter estimation

It is not easy to verify that one method is superior to another in the absence of perfect information, which can not be obtained naturally in the traffic safety field. Therefore, in this section the results of the Negative Binomial approach are compared with those of the existing two methods that are commonly used in the traffic safety field: the rate quality control method and the Bayes identification method. Because the two methods have already been reviewed at length in section 4.2 and 4.3, the parameters are estimated and the results obtained without expanding on them here.

There is a limitation in choosing the sample sites to examine the Negative Binomial approach because sites with similar geometry should be used for reference sites. Therefore, two data sets are selected for this study. The first data set includes 16 diamond interchanges with similar geometric properties (i.e., Diamond type, 6 lanes, 10 ft shoulder width, 4 ramps). The second data set includes 14 partial clover A or B 4 Quadrant interchanges which have similar geometric characteristics (i.e., 6 lanes, 10 ft shoulder width, 6 ramps). It is not possible to get a data set with exactly the same geometric conditions in practice, and the more classification variables used, the smaller the sample size.

With the above data sets, the new approach has been tested and compared with the results of the existing two methods. The parameters for Bayes approach have been estimated through the method of moment estimates (MME), and the parameters for the rate quality control method based on the Negative Binomial distribution have been estimated through the procedures described in the preceding sections. The estimates of the parameters are summarized in **Table 4.1**. Note that the k parameter for Diamond interchanges is larger than that of Partial Clover 4 Quadrant interchanges. It implies that the variance of Diamond interchanges is less than that of Partial Clover 4 Quadrant interchanges.

Table 4.1 The estimation of parameters

Method	Parameters	Estimates	
		Diamond	Par clo A or B 4 Q
Negative Binomial	m_0	0.0010	0.00117
	a	0.105	0.095
	k(=1/a)	9.52	10.52
Poisson	λ	0.0010	0.00117
Bayes approach	α	6.51	10.98
	β	6160.40	10194.61

4.3.3 Application and validation of the Negative Binomial approach

In order to choose hazardous sites based on the estimates of the parameters, the following scenarios are developed.

Scenario 1: a site i is hazardous if the observed accident rate (N_i/V_i) exceeds the upper control limit which is a function of the average accident rate of reference sites and a desired probability level. This scenario will be used to test the rate quality control method based on the Negative Binomial distribution.

Scenario 2: a site i is hazardous if the observed accident rate (N_i/V_i) exceeds the upper control limit which is a function of the average accident rate of reference sites and a desired probability level. This scenario will be used to test the rate quality control method based on the Poisson distribution.

Scenario 3: a site i is hazardous if the probability that its Bayes accident rate exceeds the average accident rate of reference sites is greater than a predetermined probability level.

For this study, a 95 % probability level is applied for all scenarios. **Table 4.2** presents the identification of hazardous sites for Diamond interchanges. An asterisk (*) corresponds to the sites that have been identified as hazardous on the basis of the above scenarios. For example, under the existing methods (rate quality control method based on the Poisson distribution, and Bayes approach), 7 sites (i.e., sites: 1, 2, 4, 7, 8, 9, 11) out of 16 are identified as hazardous, whereas under the new method (rate quality control method based on the Negative Binomial distribution), 2 sites(i.e., sites: 2 and 4) are identified as hazardous.

In each of the these scenarios (1, 2 and 3), a 95 % probability level was used, which implies that there is approximately 1 abnormal or hazardous site out of 20 random sites in the statistical sense. Therefore, 7 sites out of 16 is an unreasonably high percentage in the context of the 95 % probability level. Thus, the new method is clearly more conceptually persuasive in identifying hazardous locations than the existing methods.

For Par Clo A or B 4 Q interchanges, 4(i.e., sites: 7, 12, 13, 14) of 14 sites were chosen with the existing methods, whereas only 1 site was identified as hazardous when assuming the Negative Binomial error as shown in **Table 4.3**. Thus similar conclusions can also be reached.

The disagreement between the existing methods and the new method is probably best explained in the context of the underlying assumptions. The existing methods are both based on the widely accepted assumption that crashes occur according to the Poisson distribution, whereas the new method is based on the assumption that the occurrence of accidents follows a Negative Binomial distribution.

The upper control limits, which are functions of the average accident rate of the reference sites and the variance in the accident rate at the given site, were lower when a Poisson distribution instead of a Negative Binomial distribution is assumed. This causes the procedure to identify more hazardous sites than are expected, which was shown in **Figure 4.3**.

Chapter 2 verified and discussed at length that the variance of crashes is substantially larger than the mean, and hence the Negative Binomial distribution is an appropriate assumption for the occurrence of crashes at freeway interchanges. Thus, the rate quality control method based on the Negative Binomial distribution would be an effective measure for the identification of hazardous sites, in these cases where the variance of accidents exceeds the mean (over-dispersion).

Table 4.2 A comparison of hazardous sites according to the methods (Diamond interchanges)

Diamond interchanges						
Site(i)	Accidents (N_i)	Vehicles (V_i)	Observed rate (N_i/V_i)	Upper control limit		Bayesian Approach
				Negative Binomial	Poisson	
1	213	150529	0.00142	0.00153	0.00114 *	1.000 *
2	322	166103	0.00194	0.00154 *	0.00114 *	1.000 *
3	137	123131	0.00111	0.00154	0.00116	0.875
4	196	128354	0.00153	0.00153 *	0.00115 *	1.000 *
5	193	240902	0.00080	0.00153	0.00111	0.001
6	194	227992	0.00085	0.00153	0.00112	0.010
7	247	171559	0.00144	0.00153	0.00113 *	1.000 *
8	164	137889	0.00119	0.00154	0.00115 *	0.981 *
9	207	161545	0.00128	0.00154	0.00114 *	1.000 *
10	160	183034	0.00087	0.00154	0.00113	0.039
11	242	179213	0.00135	0.00153	0.00113 *	1.000 *
12	102	207928	0.00049	0.00153	0.00112	0.000
13	111	203816	0.00054	0.00153	0.00112	0.000
14	158	207045	0.00076	0.00153	0.00112	0.000
15	161	217719	0.00074	0.00153	0.00112	0.000
16	121	209255	0.00058	0.00153	0.00112	0.000

Table 4.3 A comparison of hazardous sites according to the methods (Par Clo A or B 4 Q)

Par Clo A or B 4Q						
Site(i)	Accidents (N_i)	Vehicles (V_i)	Observed rate (N_i/V_i)	Upper control limit		Bayes Approach
				Negative Binomial	Poisson	
1	39	39434	0.00099	0.00183	0.00147	0.127
2	45	65947	0.00068	0.00179	0.00140	0.000
3	53	86205	0.00061	0.00177	0.00137	0.000
4	62	94429	0.00066	0.00178	0.00136	0.000
5	127	100303	0.00127	0.00177	0.00135	0.764
6	120	103676	0.00116	0.00177	0.00135	0.407
7	157	108898	0.00144	0.00177	0.00135 *	0.990 *
8	111	110946	0.00100	0.00177	0.00134	0.041
9	103	112020	0.00092	0.00177	0.00134	0.005
10	117	127060	0.00092	0.00176	0.00133	0.003
11	131	133995	0.00098	0.00177	0.00133	0.016
12	226	169887	0.00133	0.00176	0.00131 *	0.959 *
13	286	202817	0.00141	0.00176	0.00130 *	0.998 *
14	403	235275	0.00171	0.00170 *	0.00129 *	1.000 *

Chapter 5

A SIMPLIFIED APPROACH FOR OVER- DISPERSION (NORMAL APPROXIMATION METHOD)

5.1 General

The rate quality control method based on the Negative Binomial distribution was discussed in the preceding chapter, and it was found that this method produces reasonable results in the statistical sense. However it is not easy for traffic engineers to apply this technique in the safety field because the parameters can not be estimated as simply as those of the Poisson distribution. This chapter provides a simple approach for the identification of hazardous sites when the Negative Binomial distribution should be assumed because of the phenomenon of over-dispersion.

5.2 Concept

The Negative Binomial approach can be simplified using the Normal approximation as:

$$N_i \sim N(\mu_i, d\mu_i) \tag{5.1}$$

where,

$\mu_i = \lambda V_i$

$$\lambda = \frac{\sum N_i}{\sum V_i}$$

N_i = the number of accidents at site i

V_i = the number of vehicles at site i

$d\mu_i$ = variance ($d \geq 1$)

In equation (5.1), the variance is larger than the mean, which is conceptually consistent with the error structure of the Negative Binomial distribution.

On the other hand, under ideal conditions, the Poisson distribution can be approximated by the Normal distribution for large values of μ_i, because the probability mass function of the Poisson distribution becomes more symmetric and bell-shaped as μ_i increases (Rice 1997). Let N_i be a sequence of Poisson random variables with the corresponding parameters. Then, $E(N_i) = Var(N_i) = \mu_i$. If we wish to approximate the Poisson distribution by a Normal distribution, the Normal distribution should have the same mean and variance as the Poisson, and hence the random variables can be standardized by letting,

$$X_i = \frac{N_i - \mu_i}{\sqrt{\mu_i}} \tag{5.2}$$

then, $E(X_i) = 0, Var(X_i) = 1$.

That is, $X_i \sim N(0, 1)$

However, the assumption is :

$E(X_i) = 0, Var(X_i) = d$. That is,

$$X_i \sim N(0, d), \tag{5.3}$$

Therefore, the random variables with over - dispersion can be standardized as follow :

$$Z_i = \frac{N_i - \mu_i}{\sqrt{d\mu_i}} \tag{5.4}$$

Then $E(Z_i) = 0,\ Var(Z_i) = 1$. That is,

$$Z_i \sim N(0,1) \tag{5.5}$$

Thus, Z_i can be applied to the identification of hazardous sites based on the traffic crash data with over-dispersion.

5. 3 Application and validation of the Normal approximation method

The Normal approximation method is an alternative to solving the difficulties associated with the estimation of parameters in applying the rate quality control method based on the Negative Binomial distribution. Therefore, the results should be analogous to those of this rate quality control method. The validity of the Normal approximation method was tested using the two data sets (16 Diamond interchanges and 14 partial clover A, or B 4 Quadrant interchanges) which were used in chapter 4.

5.3.1 Estimation of parameters

In order to apply the Normal approximation method for over-dispersion, two parameters need to be estimated. One is the average accident rate, λ. The other is a parameter relating to the over-dispersion, d. Here, the average accident rate is the same as that of the Poisson or Negative Binomial distributions described in chapter 4, whereas d is the variance of random variables that have been standardized by formula (5. 2). **Table 5.1** shows the results of the estimated parameters. The "d" estimates are 30.39 and 11.96 for Diamond, and Par Clo A or B 4 Q interchanges, respectively.

Table 5.1 Estimation of parameters

Method	Parameters	Estimates	
		Diamond	Par Clo A or B 4 Q
Normal approximation method	λ d	0.0010 30.39	0.00117 11.96

5.3.2 Validation of the Normal approximation approach

Based on the estimates of parameters in **Table 5.1**, the Normal approximation method was tested for the 2 data sets. **Table 5.2** shows that the use of the Normal approximation method produced similar results with those of the approach using the Negative Binomial distribution. In this table, asterisks (*) correspond to the sites that have been identified as hazardous on the basis of the two approaches.

For example, under a 95 % probability level, the Normal approximation method identifies 2 sites as hazardous, whereas the rate quality control method under the Negative Binomial distribution identifies 3 sites as hazardous out of 30 sites. That is, site 4 is not identified as a hazardous site by the Normal approximation method, whereas it is hazardous based on the Negative Binomial approach. However, this difference is not substantial with the probabilities being 0.94 and 0.95 respectively, when we use each approach.

Figure 5. 1 presents a comparison of the probabilities with which a site is identified as hazardous by the two methods. As shown in the figure, results of both methods are consistent, even though there are a few sites that disagree slightly. It is expected that the differences would be reduced even further with larger data sets.

Thus, for the identification of hazardous sites, the Normal approximation method can be used as an alternative to solving the difficulties associated with the estimation of the parameters for the rate quality control method. There is only a slight loss of accuracy as discussed here.

Table 5.2 A comparison with the results using the Negative Binomial distribution

Site (i)	Interchange type	μ_i	X_i	Z_i	Probability	
					Approximation	Negative Binomial
1	Diamond	151	5.03	0.91	0.82	0.91
2	Diamond	167	12.02	2.18	0.99*	0.99*
3	Diamond	124	1.20	0.22	0.58	0.72
4	Diamond	129	5.91	1.53	0.94	0.95*
5	Diamond	242	-3.14	-0.57	0.38	0.36
6	Diamond	229	-2.31	-0.42	0.33	0.42
7	Diamond	172	5.69	1.03	0.85	0.92
8	Diamond	138	2.17	0.39	0.65	0.79
9	Diamond	162	3.52	0.64	0.74	0.85
10	Diamond	184	-1.75	-0.32	0.38	0.45
11	Diamond	180	4.63	0.84	0.80	0.90
12	Diamond	209	-7.39	-1.34	0.09	0.05
13	Diamond	205	-6.55	-1.19	0.12	0.09
14	Diamond	208	-3.46	-0.63	0.26	0.31
15	Diamond	219	-3.90	-0.71	0.24	0.28
16	Diamond	210	-6.15	-1.12	0.13	0.11
1	Par Clo A 4 Q	46	-1.06	-0.31	0.39	0.43
2	Par Clo A 4 Q	77	-3.67	-1.06	0.14	0.12
3	Par Clo B 4 Q	101	-4.77	-1.38	0.08	0.07
4	Par Clo A 4 Q	111	-4.62	-1.34	0.09	0.09
5	Par Clo A 4 Q	117	0.88	0.25	0.61	0.77
6	Par Clo B 4 Q	121	-0.13	-0.04	0.48	0.59
7	Par Clo B 4 Q	128	2.61	0.76	0.78	0.83
8	Par Clo A 4 Q	130	-1.66	-0.48	0.32	0.55
9	Par Clo A 4 Q	131	-2.46	-0.71	0.34	0.33
10	Par Clo B 4 Q	149	-2.61	-0.75	0.33	0.33
11	Par Clo A 4 Q	157	-2.07	-0.60	0.38	0.39
12	Par Clo A 4 Q	199	1.92	0.55	0.71	0.75
13	Par Clo A 4 Q	237	3.15	0.91	0.82	0.81
14	Par Clo B 4 Q	276	7.68	2.22	0.99*	0.95*

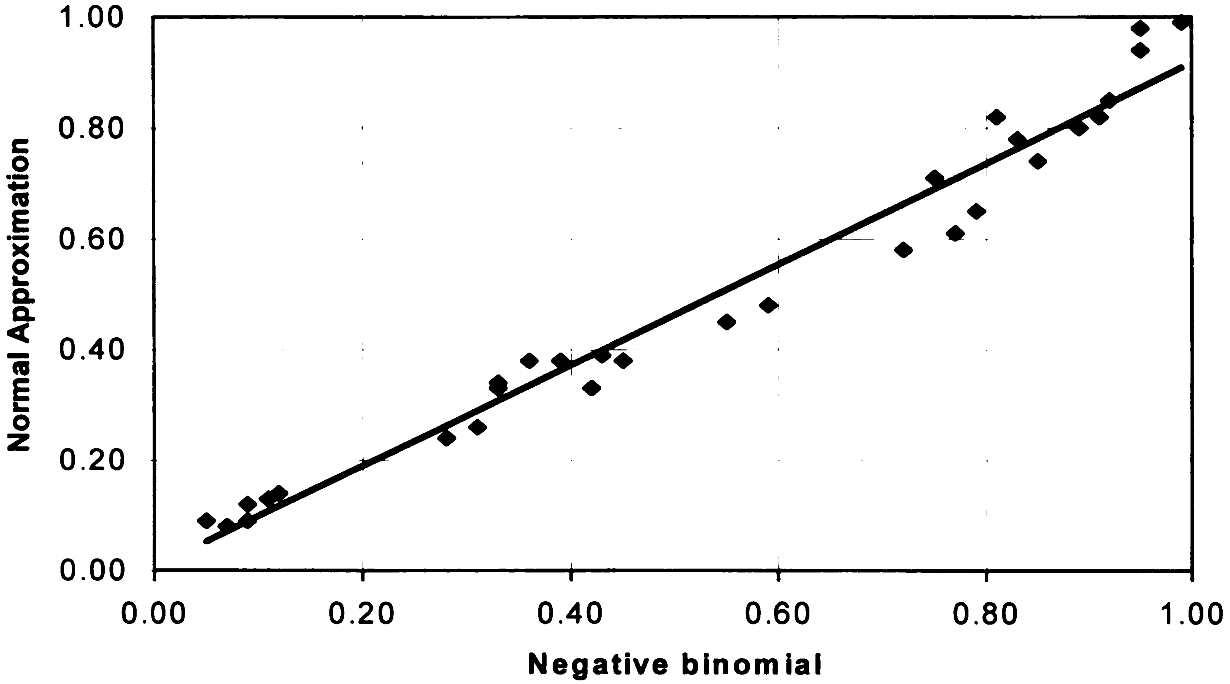

Figure 5.1 A comparison of the results of the Negative Binomial and Normal approximation approaches

f

t

(

a

The

the

respo

5.4 An examination of the assumptions

In the preceding sections, the following assumptions were made: 1) random variables X_i which are standardized follow a Normal distribution, and 2) the expected values μ_i should be large enough to be approximated by the Normal distribution. Therefore, these assumptions need to be examined.

5.4.1 Goodness of fit of the Normal distribution

In section 5.2, we assumed that X_i follows a Normal distribution without any verification. In order to enhance the credibility of this method, we need to test the goodness of fit of the random variable X_i to a Normal distribution. The Chi-square test was used to conduct this test after partitioning a Normal distribution into eight intervals of equal probability (Neter et al. 1992). Thus, if H_0 holds (that is, X_i is Normally distributed), then X^2 follows an approximate χ^2 distribution with n-p-1=8-2-1=5 degrees of freedom.

For $\alpha=0.05$, we require $\chi^2(0.95; 5)=11.07$.

Hence, the decision rule is as follows:

If $X^2 \leq 11.07$, conclude H_0

If $X^2 > 11.07$, conclude H_1

The analysis of Diamond interchanges and Par-Clo A, or B 4 Q interchanges, which are the same data sets as used in the previous section, found X^2 values of 6.00 and 5.43, respectively. Thus, a Normal distribution is a reasonable assumption for X_i.

5.4.2 Large values of μ_i

In section 5.2, we assumed that μ_i should be large enough to be approximated by the Normal distribution. This approximation method can not be used, if μ_i is a small number. Accordingly, the effects of various values of μ_i were tested to determine the limits of the approximation method. The analysis focuses on the calculation of the difference between the true and approximate upper control limits over the values of μ_i, as computed from the following equations.

$$P = \sum_{y=0}^{U-1} \frac{e^{-\mu} \mu^{y}}{y!} \tag{5.6}$$

$$Ua = \mu + k\sqrt{\mu} \tag{5.7}$$

where,

P : predetermined probability level
U : true upper control limit
Ua : approximated upper control limit
k : standard normal variate corresponding to the predetermined probability level

Figure 5.2 shows the difference between the true and approximate upper control limits for a range of expected frequencies (μ_i) from 0 to 60 crashes using the 95 probability level (k=1.645). Note that the difference is very large when the expected values are less than 5, then the curve flattens for expected values in the range from 5 to 15.

The difference is very small when the expected values are larger than 40. For this research, μ_i is large enough to be approximated, recognizing that the minimum value of μ_i is 151 and 46, for Diamond and Par Clo A or B 4 Q data sets, respectively as shown in **Table 5.2.**

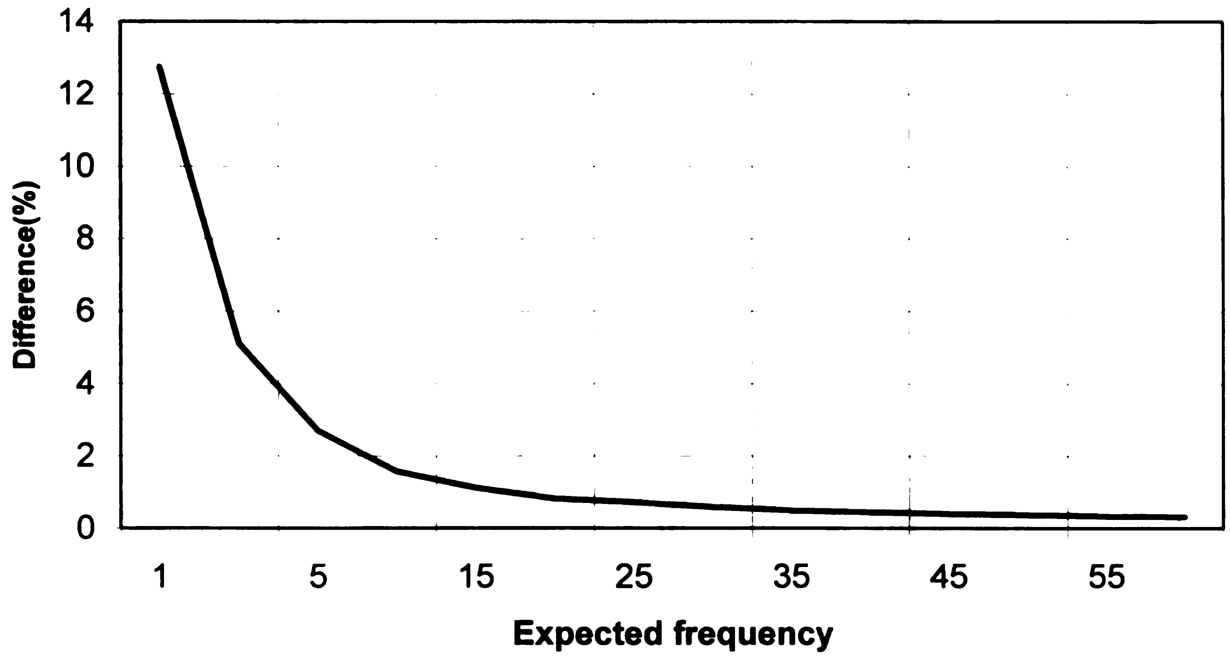

Figure 5.2 The difference between the true and approximate upper control limit (95 %)

Chapter 6

IDENTIFICATION OF HAZARDOUS SITES USING A TRAFFIC CRASH PREDICTION MODEL (PREDICTION MODEL METHOD)

6.1 The limitation of the rate quality control method (or upper control limit)

As mentioned before, a rate quality control method is commonly used for identification of hazardous sites. In order to overcome the problem caused by over-dispersion, the rate quality control method based on the Negative Binomial distribution rather than the Poisson distribution has been examined and proposed as an alternative. Nevertheless, in identifying the hazardous sites using reference sites we recognize that there are still limitations as defined by others (Elvik 1988, Mountain and Fawaz 1989, Hauer 1992):

First, the selection of reference sites is a matter of judgement, and hence the same site can be evaluated differently, depending on the researchers. Second, the number of reference sites will likely not be large enough to permit the accurate identification of hazardous sites in practice.

For example, suppose that the objective was to evaluate the safety of all interchanges in Michigan using the rate quality method. The first step is to classify all interchanges to find the reference sites with similar properties. **Figure 6.1** presents classification trees considering only the basic contributing factors to traffic crashes, and 1056 groups (= $22\times2\times2\times3\times4$) are produced, even though other contributing factors (i.e., ramp length, interchange size, et al) are not considered in grouping the interchanges. This implies that we need reference sites for 1056 groups to evaluate all the interchanges using the rate quality control method described in chapter 4. Thus, it is not possible to identify hazardous interchanges using reference sites, recognizing that there are a total of only 397 interchanges along the four main freeways (I-69, I-75, I-94, I-96) in Michigan. There are many interchange types for which a sizeable number of reference sites does not exist.

An alternative to the use of reference sites would be to use data from other states for the evaluation of freeway interchanges in Michigan. However, this approach causes several linked difficulties. For example, the definition of traffic crashes is different across states (i.e., total damage of $400 in Michigan, $500 in New Mexico and $1000 in Wisconsin: Michigan, New Mexico and Wisconsin traffic crash facts (1998)), and interchange crashes are sensitive to weather conditions (i.e., in Michigan, winter crashes are approximately 15 % higher than in other seasons). In addition, it is not easy to obtain well defined geometric and traffic data from other states. Thus it is obviously not a good approach to use data from other states for the identification of hazardous sites. In this chapter, a method to search for hazardous sites using an accident prediction model is examined.

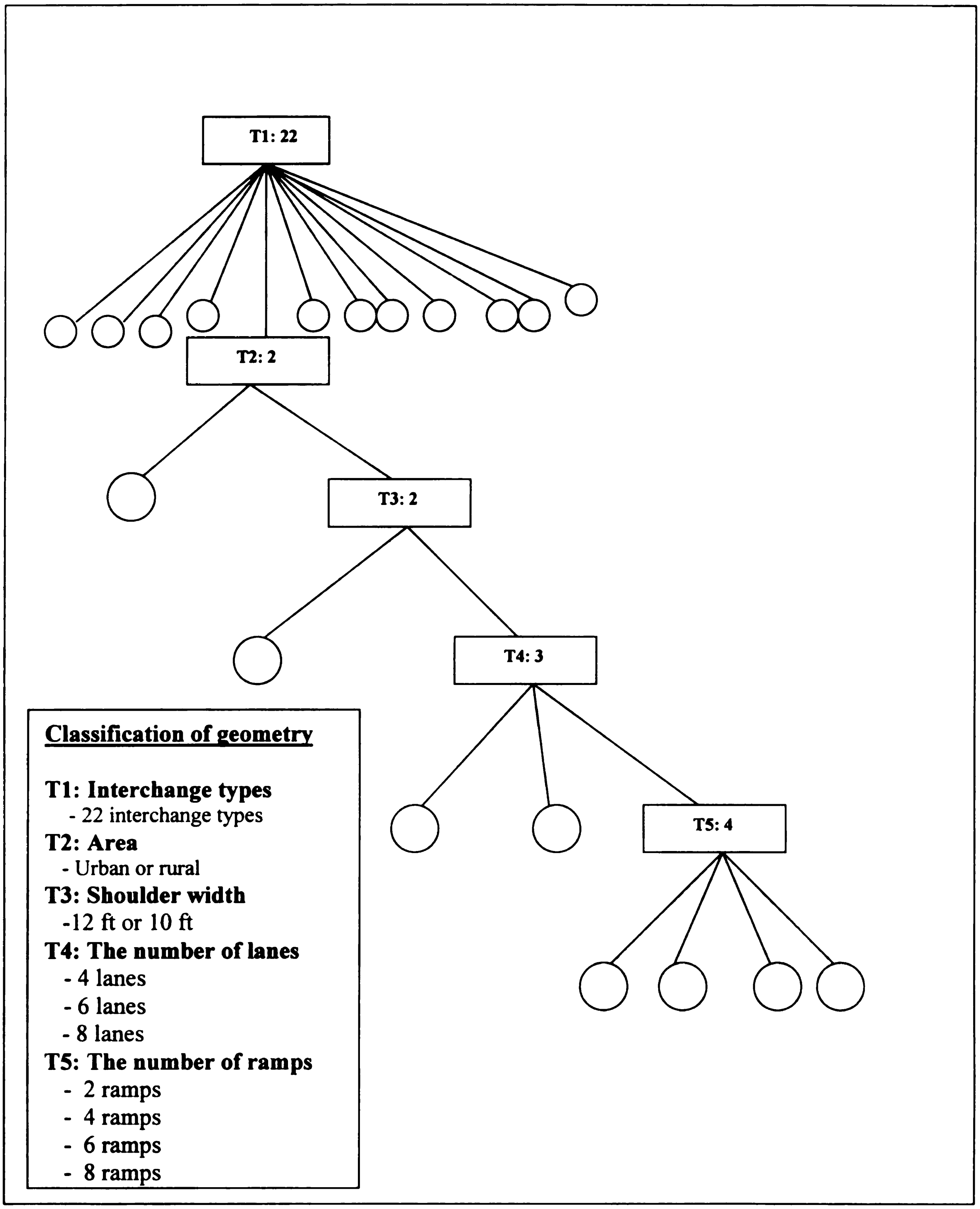

Figure 6.1 Basic geometric classification tree for reference sites

6.2 The concept of the prediction model method

In section 6.1, several limitations to the use of the rate quality control method for the identification of hazardous interchanges was discussed. The conceptual problem can be solved by diminishing the scope of individual judgement through logical procedures, whereas the practical problem can be treated by estimating the effects of special contributing factors to traffic crashes at a given site through an analysis of relevant traits at other sites. Previous researchers (Jorgensen 1972, Flak and Barbaresso 1982) have recommended that hazardous sites be estimated by the difference between the observed accident frequency (B) of a site and the expected frequency (A) as predicted by an accident prediction model as shown in **Figure 6.2**. McGuian (1981) noted that this difference represents the size of the potential crash reduction when a safety improvement project is implemented at the site. These ideas can be updated to solve both the conceptual problem and the practical problem which have been identified.

Suppose that the goal is to estimate the hazardness of site i using a statistical concept like the rate quality control method. In order to evaluate site i using the rate quality control method, reference sites with similar properties should be selected, and the accident rate of the site i compared with that of the reference sites. However, in the strict sense, there are no reference sites which exactly reflect the site i. Thus, the idea of the prediction model method is that the value of $E(\theta)$ obtained from the crash prediction model can be used instead of the average crashes of the reference sites to which the site i belongs. Using this approach, the reference sites match exactly the traits of the site i (these are imaginary reference sites as denoted by Hauer (1992).

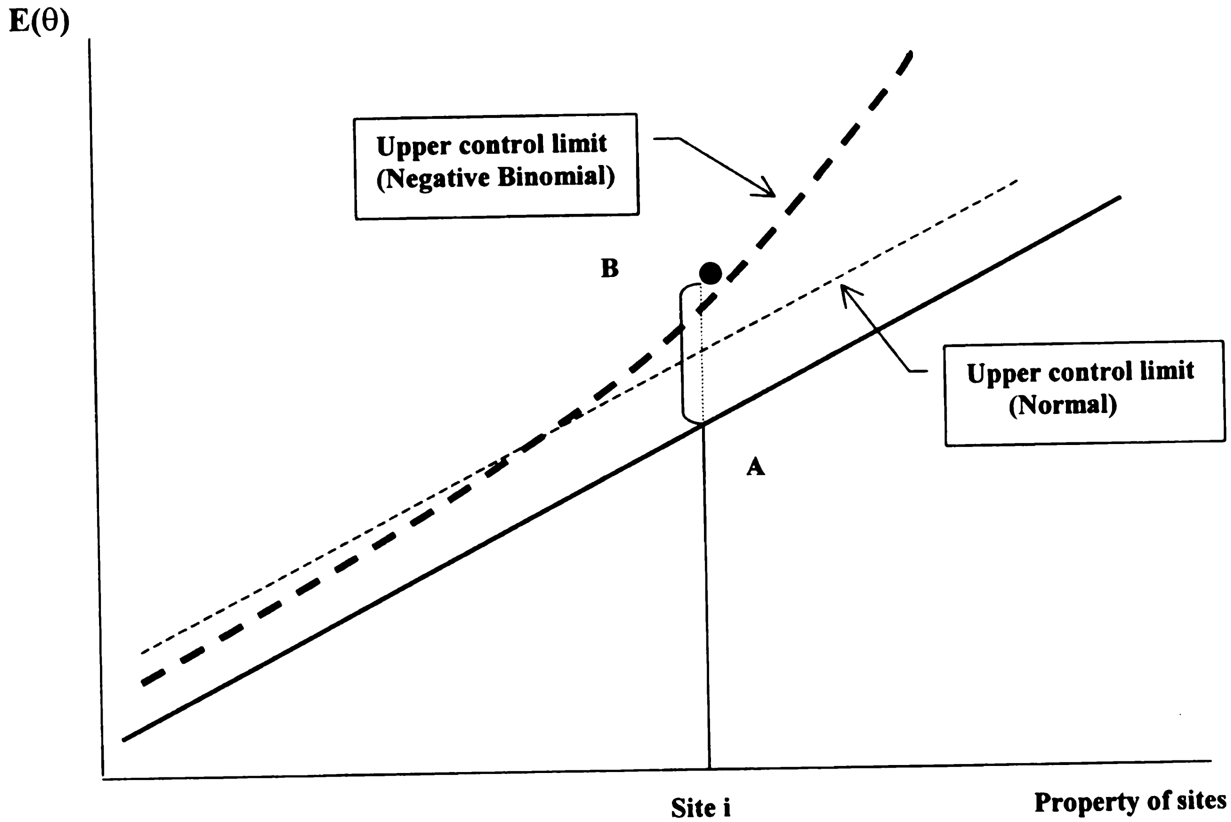

Figure 6.2 Concept of prediction model method

This approach is similar to the rate quality control method in the sense that both use the mean and standard deviation for identification of hazardous sites. However, the difference is that the mean is the expected value E(θ), based on a calibrated model for the prediction model method, whereas the mean is the average of the reference sites for the rate quality control method. This is why "E(θ)" instead of "m" is used in formula 6.1. Therefore, the calibration of the crash prediction model based on the correct error structure is extremely important to the identification of hazardous sites.

It has already been shown that the desirable assumption for freeway crash models is the Negative Binomial rather than the Normal or Poisson error structure. In order to illustrate the prediction model method for the identification of hazardous sites, the Negative Binomial distribution function is again mentioned as equation (6.1).

$$P = \sum_{x=0}^{U-1} \left(1 + \frac{E(\theta)}{k}\right)^{-k} \frac{\Gamma(k+x)}{x!\Gamma(k)} \left(\frac{E(\theta)}{E(\theta)+k}\right)^{x} \qquad (6.1)$$

where,

U : the true upper control limit
E(θ) : expected values
k : parameter

In equation (6.1), $E(\theta)$ would be obtained from the crash prediction model and the parameter k would be estimated in the process of calibrating coefficients of the crash prediction model, which were discussed in detail in chapter 3. From equation (6.1), therefore, the upper control limit for identification of hazardous sites at a desired probability level can be computed.

The variance of the Negative Binomial distribution is $E(\theta)+E(\theta)^2/k$, as discussed in chapter 2 and chapter 3, and hence the upper control limit will increase sharply with $E(\theta)$ as shown by the thick dotted line in Figure 6.2.

However, if an accident prediction model is developed under a constant normal error structure, the upper control limits would be a constant distance from the accident prediction line as shown by the thin dotted line in the Figure. This approach is similar to that of previous research (Jorgensen 1972, Flak and Barbaresso 1982).

6. 3 Application and validation of the prediction model method

6. 3.1 Illustration of the prediction model method

Suppose that the goal is to estimate the safety of a special site using a crash prediction model that has been calibrated under the Negative Binomial error structure.

Again, k can be estimated by the parameter calibration procedure described in chapter 3, and $E(\theta)$ can be computed from the crash prediction model using several independent variables of the site. Thus, the true upper control limit 'U' can be found from equation (6.1) for a given site under the desired probability level.

For example, consider site 1 in **Table 6.1**. Using the crash prediction model developed in chapter 3, the expected value at site 1,E (θ) is

$$= 3.448\, V_1^{1.401}\, V_2^{0.186}\, V_3^{0.620}\, G_1^{0.738} \exp(-1.267\, G_2 - 0.156\, G_5) \qquad (6.2)$$

$= 141.6$ accidents/3years

The standard deviation at site 1

$$= \sqrt{E(\theta) + E(\theta)^2 / k}$$

$= 51.3$ accidents/ 3years

The parameter k was determined to be equal to 8.05 in chapter 3. In equation (6.1), (6.2) and (6.3), the upper control limit ' U ' is 233 crashes for 3 years under the 95 percent probability level as follows:

$$P = \sum_{x=0}^{U-1} \left(1 + \frac{141.6}{8.05}\right)^{-8.05} \frac{\Gamma(8.05 + x)}{x!\Gamma(8.05)} \left(\frac{141.6}{141.6 + 8.05}\right)^{x} \qquad (6.3)$$

However, there were only 213 crashes over 3 years at the given site. Thus this site is not identified as hazardous under the 95 percent probability level as shown in **Figure 6.3.** Thus, we can test the hazardness of each site on the basis of various desired probability levels using these results.

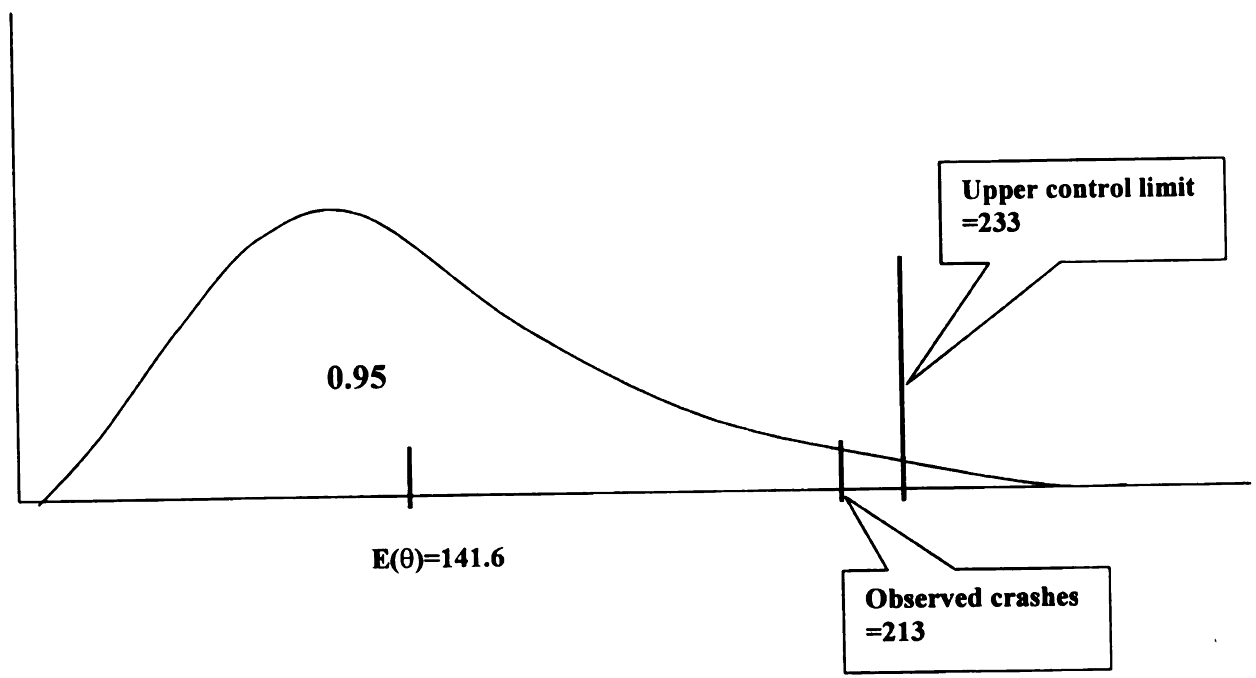

Figure 6. 3 An example of application of prediction model method(95 %)

6.3.2 Validation of the prediction model method

The conceptual foundation for identifying hazardous sites using an accident prediction model are straightforward as discussed in the previous section. There are two main advantages of this method over the rate quality control method. First, it diminishes the scope of individual judgement through a logical procedure. Second, a large number of reference sites for any particular site are not required.

Despite its advantages, the prediction model method can cause unreasonable results since there may be significant errors in choosing the model structure and calibrating the model parameters. For these reasons, it is important to illustrate empirically that the prediction model method and reference method produce similar results. However, we can not expect that the results of both approaches will be coincident, because in the strict sense, the imaginary reference sites for the prediction model method is a subset of the reference sites for the rate quality control method.

To demonstrate the results of both approaches, the two data sets that were analyzed in chapter 4 were used. In Table 6.1, the 5th column presents the probability that observed crashes exceed the expected crashes at a given site under the prediction model method. The 6th column represents the probability that the observed accident rate exceeds the reference accident rate under the rate quality control method. There is some disagreement between the methods as expected. When these sites are identified at a high probability level (i.e., 0.95), 3 sites out of 30 are identified by the rate quality control method (marked by a "*" in the table), whereas there are no sites identified when using the prediction model method. At a lower probability level (i.e., 0.90), 6 and 4 sites out of 30 are identified using the rate quality control method and prediction model method respectively (noted by a "♦" in the table).

In the prediction model method, the model parameters are calibrated through a minimization of the sum of squared residuals, and hence there may be underestimates of the variances for the special sites which have a larger values than the average sites as shown in **Table 6.1**. Moreover, not all geometric elements (i.e., interchange size, ramp length, et al) and traffic elements (mainline traffic, on and off ramp traffic, truck traffic, et al) were used in classifying the reference sites to design the upper control limit, whereas the imaginary reference sites for the prediction model method match exactly the characteristics of a special site.

Accordingly, it can be expected that the results of both approaches will be similar, but not coincide in every cell in **Table 6.1**. To test similarity of the results by the rate quality control method and prediction model method, the percentiles of sites were calculated and were plotted in **Figure 6.4**. The results of both approaches are highly correlated (correlation coefficient = 0.96).

All the sites were ranked by the probability and the top 10 sites were chosen from the two data sets (5 sites at Diamond interchanges, and 5 sites at Par-Clo A or B 4 Q interchanges). As shown in **Table 6.1**(noted by a "∨" in the table), the prediction model method identifies the same sites as the rate quality control method for the Diamond interchanges. It also identifies 4 sites out of the 5 identified by the rate quality control method for the Par-Clo A or B 4 Q interchanges. It is surprising that there is so little difference between the rate quality control method and the prediction model method in terms of determining the hazard ranking of several sites.

A practical application of the above results is that if the goal is to prioritize several sites for a highway safety program, the prediction model method can be used as a tool to produce very similar ranks as the rate quality control method. If the goal is to evaluate a specific site for a purpose, we can approximately evaluate the hazardness of the site under the desired probability level through the prediction model method. These advantages imply that we can overcome the conceptual and practical problem associated with the identification of sites where the crashes exceed the expected number of crashes as discussed in the previous sections, through the use of the prediction model method. The accuracy of this method depends on having the crash prediction model calibrated under the appropriate error structure.

Table 6.1 A comparison of results

Site(i)	Interchange type	The number of crashes (3 years)		Probability	
		Observed	Estimated	By upper control limit	By prediction model
1	Diamond	213	141.6	0.91 ◆ ∨	0.91 ◆ ∨
2	Diamond	322	204.2	0.99 * ◆ ∨	0.93 ◆ ∨
3	Diamond	137	113.8	0.72	0.75
4	Diamond	196	139.9	0.95 * ◆ ∨	0.87 ∨
5	Diamond	193	237.7	0.36	0.34
6	Diamond	194	251.5	0.42	0.29
7	Diamond	247	163.7	0.92 ◆ ∨	0.92 ◆ ∨
8	Diamond	164	138.7	0.79	0.73
9	Diamond	207	169.3	0.85	0.76
10	Diamond	160	166.5	0.45	0.51
11	Diamond	242	157.8	0.90 ◆ ∨	0.92 ◆ ∨
12	Diamond	102	177.7	0.05	0.10
13	Diamond	111	182.7	0.09	0.13
14	Diamond	158	179.5	0.31	0.42
15	Diamond	161	198.5	0.28	0.34
16	Diamond	121	188.4	0.11	0.16
1	Par Clo A 4 Q	39	44.8	0.43	0.43
2	Par Clo A 4 Q	45	69.1	0.12	0.20
3	Par Clo B 4 Q	53	87.1	0.07	0.15
4	Par Clo A 4 Q	62	93.3	0.09	0.21
5	Par Clo A 4 Q	127	85.4	0.77 ∨	0.89 ∨
6	Par Clo B 4 Q	120	135.1	0.59	0.44
7	Par Clo B 4 Q	157	127.8	0.83 ∨	0.76 ∨
8	Par Clo A 4 Q	111	102.5	0.55	
9	Par Clo A 4 Q	103	125.7	0.33	0.64 ∨
10	Par Clo B 4 Q	117	134.8	0.33	0.36
					0.42
11	Par Clo A 4 Q	131	166.1	0.39	0.33
12	Par Clo A 4 Q	226	221.5	0.75 ∨	0.58
13	Par Clo A 4 Q	286	275.9	0.81 ∨	0.59 ∨
14	Par Clo B 4 Q	403	285.3	0.95 * ◆ ∨	0.86 ∨

* : Identified sites under 95 percent probability level

◆: Identified sites under 90 percent probability level

∨ : Top 10 rankings (5 for Diamond, and 5 for Par Clo 4 Q)

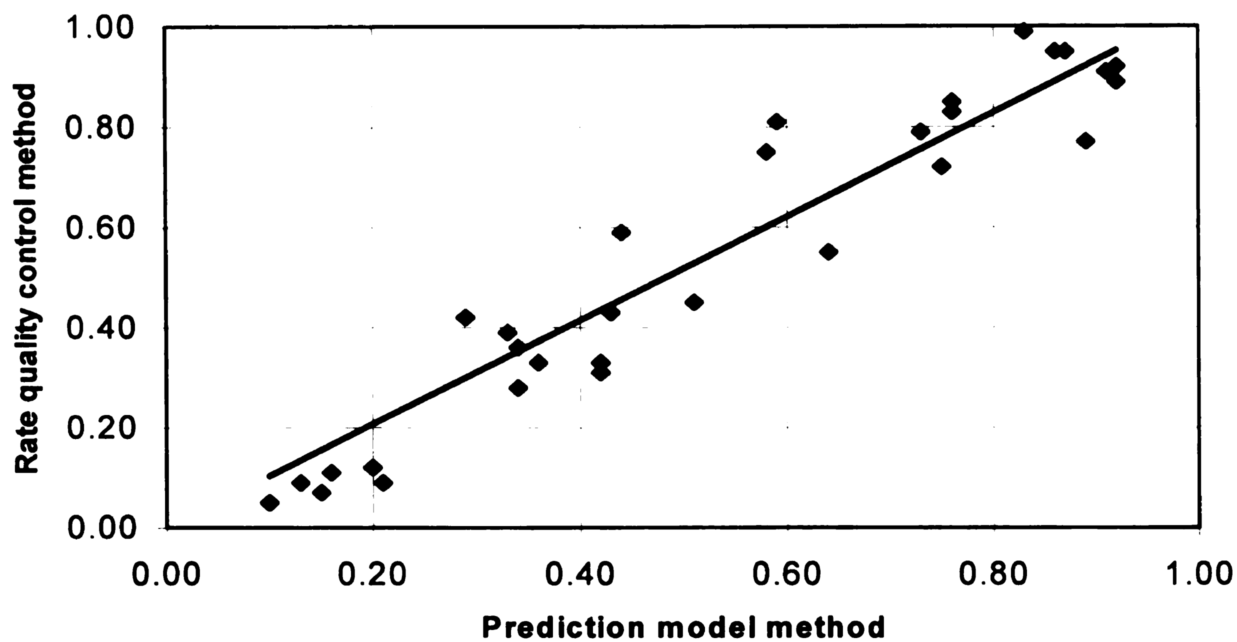

Figure 6.4 A comparison of results of the rate quality control method and prediction method

6. 4 Evaluation of Michigan freeway interchanges on the basis of the prediction model

As noted in the preceding section, the prediction model method can be used to identify hazardous sites without the use of reference sites. Using this approach, the 199 interchanges which were utilized in the crash prediction model development were assessed using the coefficients and k parameters estimated according to the interchange type in chapter 3.

The sites which exceed the thick dotted line in **Figure 6.2** are summarized in **Table 6.2**. Under the 95 % upper control limit, there is one such site out of the 10 interchanges on I-69, 4 sites out of 65 on I-75, 6 sites out of 90 on I-94, and 1 site out of 34 on I-96, respectively. Therefore, a total 12 sites are identified out of 199. These results are approximately consistent with the statistical concept that there may be 10 abnormal sites out of 200 random sites using the 95 % upper control limit. Under the 90 % upper control limit, 22 sites are chosen, which also supports the preceding conclusion. The results of evaluating all interchanges are presented in detail in the **Appendix**.

The identified sites are candidates for improvement under highway safety improvement program for freeway interchanges. These results could not be obtained through the existing rate quality method because there are not enough reference sites as discussed at length in section 6.1.

Table 6.2 The abnormal sites by the prediction model method

Route	Interchange ID	Interchange type	Cross road	Observed	Fitted	Probability	Identified sites	
							90 %	95%
I-69 (10)	69137	Full Direct	M-54 BR	165	83	0.98	*	*
I-75	75018	Diamond	Nadeau Rd	60	36	0.95	*	*
I-75	75026	Mod Tight Diamond	Huron River Rd	48	35	0.90	*	
I-75	75044	Part Diamond	Dearborn Rd	96	57	0.96	*	*
I-75	75074	Parclo AB	Adams Rd	208	114	0.99	*	*
I-75 (65)	75069	Clover w/C-D	Big Beaver Rd	406	229	0.95	*	*
I-94	94220 A	Diamond	French Rd	213	142	0.91	*	
I-94	94218	Diamond	Van Dyke Ave	247	164	0.91	*	
I-94	94230	Diamond	12 mile Rd	242	127	0.98	*	*
I-94		Tight Diamond	Pipestone Rd	142	106	0.90	*	
I-94	94217 B	Tight Diamond	Mt Elliott Ave	306	196	0.97	*	*
I-94	94214 A	Part Tight Daimond	Grand River Blvd	103	65	0.93	*	
I-94	94214 C	Part Tight Daimond	14th St	220	126	0.96	*	*
I-94	94214 B	Part Tight Daimond	Trumbull Ave	156	100	0.93	*	
I-94	94127	Mod Diamond	I-94 BL	98	66	0.91	*	
I-94	94215 C	Split Diamond	John R Rd	392	237	0.95	*	*
I-94	94034	Trumpet A	SB I-196	96	63	0.94	*	
I-94	94235	Trumpet B	Shook Rd	59	40	0.92	*	
I-94	94027	Parclo A	M-63	162	71	0.99	*	*
I-94	94028	Parclo B	Scottdale Rd	148	82	0.97	*	*
I-94 (90)	94219	Parclo B	Gratiot Ave	278	185	0.90	*	
I-96 (34)	96160	Other	Grand River Ave	107	66	0.95	*	*
Total (199)							22 sites	12 sites

( ): the number of sites which were assessed

Chapter 7

SUMMARY AND CONCLUSIONS

7.1 Summary

The Poisson distribution is a commonly accepted assumption in analyzing traffic crashes. When freeway interchange crash data were examined, it was found that there is substantially larger variability than would be expected if the distribution followed Poisson's law, and that the Negative Binomial distribution provides a better fit. This research focused on several linked issues which occur with the assumption that traffic crashes follow the Negative Binomial distribution rather than the Poisson distribution.

7.1.1 Traffic crash distribution

To test the distribution on freeway interchange crashes, the year to year variances were calculated for crashes that occurred during the 5 year-period 1994-1998. Throughout this study, it was found that there is greater variability than would be expected under the assumption of the Poisson distribution, and the Negative Binomial distribution fits the data much better than the Poisson distribution. That is,

- The correlation coefficients between observed and expected variances increased from 0.91 to 0.97 and from 0.84 to 0.90 in the analysis of 24 crash types and 84 Diamond interchanges, respectively, when the data were fitted to the Negative Binomial instead of the Poisson distribution.

- The squared residuals between observed and expected variances were reduced by more than 80 % when the Negative Binomial distribution is assumed.

7.1.2 Traffic crash prediction model

One objective of the research was to develop crash prediction models for freeway interchanges using the Negative Binomial error structure.

- Based on the results of ANOVA and correlation, mainline ADT, ramp ADT, and truck percent were selected as traffic variables that effect freeway interchange crashes. The number of on and off ramps, the number of lanes, shoulder width, interchange length, and average spread-ramp length were determined to be geometric variables that affect accidents.

- A non-linear regression model was selected as the model structure for the crash prediction model developed in this study, and the model is:

$$E(\theta) = A \times V_i^{B_i} \times G_j^{C_j} \times \exp\sum(C_k \times G_k)$$

where,

$E(\theta)$: Expected number of crashes

V_i : Traffic variables

G_j : Geometric variables(type 1)

G_k : Geometric variables(type 2)

A, B_i, C_j, C_k : Parameters

- To calibrate this model, the Generalized Linear Model (GLIM) approach that prevents the Negative Binomial error assumption from being violated was used.

- Using several measures of assessing the goodness of fit of models, such as Pearson Chi-square(χ^2), Dispersion parameter (D_p), Coefficients of determination (R^2), Pearson Residuals(PR) and so on, 10 crash prediction models were developed, one for each of the most common interchange types in Michigan.

- Large reductions in the coefficient of variation of parameter estimates were found when the traffic crash prediction models were calibrated based on the Negative Binomial error assumption. For example, the coefficient of variation of parameter estimates in the models for interchange type 11 and type 12 were reduced by an average of 36 percent when the models were calibrated under the Negative Binomial error assumption rather than the Normal one.

7.1.3 The rate quality control method based on the Negative Binomial distribution

Since the accidents follow the Negative Binomial distribution rather than the Poisson distribution, the rate quality control method needed to be reexamined because it is based on the Poisson distribution. The findings can be summarized as follows:

- The rate quality control method under the Poisson assumption identifies more sites as hazardous than should theoretically be expected because the variance of the Poisson distribution is equal to the mean, whereas the variance of the Negative Binomial distribution (and the observed data) equals the mean + mean ^{2}/k.

- The Negative Binomial distribution parameters that were necessary for the identification of hazardous sites were calculated using the maximum likelihood method of estimation.

- On the basis of the Negative Binomial error structure, the framework of a rate quality control method was proposed for the identification of hazardous sites. This framework produced more reasonable results than the existing approaches, such as the rate quality control method assuming the Poisson error structure, or Bayes approach.

7.1.4 The Normal approximation method

Even though the rate quality control method based on the Negative Binomial distribution produced conceptually more reasonable results than the existing approaches, the application of this method may not be efficient because the parameters can not be easily estimated.

- In order to overcome the difficulties associated with the estimation of parameters of the Negative Binomial distribution based rate quality control method, a Normal approximation method was proposed, and is shown to produce good results when identifying hazardous locations.

- The Normal approximation method identified hazardous sites with no loss of accuracy, even though it is a relatively simple method based on the Negative Binomial distribution.

- The validity of the Normal approximation method was shown to be contingent on two assumptions. These assumptions are: 1) the standardized random variables X_i follow the Normal distribution and 2) the expected mean μ_i is large enough.

- The testing of the two assumptions showed that X_i does follow the Normal distribution, and μ_i is large enough to allow for the accuracy of the Normal approximation method.

7.1.5 The prediction model method

In applying the rate quality control method to the identification of hazardous sites, two limitations were identified in this study. The conceptual problem is that the selection of reference sites is a matter of judgement. The practical problem is that a site can not be efficiently evaluated unless there is a sufficient number of reference sites to assure the accuracy of the results.

- To overcome the limitations of the rate quality control method, the prediction model method was tested, and it was found that there is little difference between the rate quality control method and the prediction model method in identifying hazardous sites. This implies that we can evaluate the safety of the sites in a statistical sense without reference sites.
- Recognizing the accuracy and the availability of the prediction model method, about 200 freeway interchanges in Michigan were evaluated, 12 sites were identified at the 95 % probability level.

7.2 Conclusions

This research focused on the issues occurring when an assumption is made that traffic crashes follow the Negative Binomial distribution rather than the Poisson distribution. The following is conclusions that were reached in this study.

- Crash prediction models for freeway interchanges can be efficiently calibrated under the assumption of the Negative Binomial error structure.

- The rate quality control method using the Negative Binomial distribution identified a more reasonable set of abnormal sites than the existing methods such as the Poisson based rate quality control method, or Bayes approach.

- The Normal approximation method proposed for user convenience identified hazardous sites without loss of accuracy, even though it is relatively simple compared to the Negative Binomial based rate quality control method.

- The prediction model method developed accurately identified the safety of sites in the statistical sense.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. A. H-S. Ang and W. H. Tang. "Probability concepts in engineering planning and design". John Willey & Sons, 1975.

2. A. Vogt and J. Bared. "Accident models for two-lane rural segments and intersections" Transportation Research Record, 1635, Transportation Research Board, National Research Council, Washing D.C., 1998, pp. 18-29.

3. A. P. Tarko, K. C. Sinha, and O. Farooq. "Methodology for identifying highway safety problem areas". Transportation Research Record, 1542, Transportation Research Board, National Research Council, Washing D.C., 1996, pp. 49-53.

4. A. Ceder and M. Livneh. "Relationships between road accidents and hourly traffic-I". Accident Analysis and Prevention, Vol. 14, No 1, 1982, pp. 19-34.

5. A. Ceder. "Relationships between road accidents and hourly traffic-II". Accident Analysis and Prevention, Vol. 14, No 1, 1982, pp. 35-44.

6. B. Persaud and L. Dzbik. "Accident prediction models for freeways" Transportation Research Record, 1401, Transportation Research Board, National Research Council, Washing D.C., 1993, pp. 55-60.

7. B. Persaud and T. Nguyen. "Disaggregate safety performance models for signalized intersections on Ontario provincial roads" Transportation Research Record, 1635, Transportation Research Board, National Research Council, Washing D.C., 1998, pp. 113-120.

8. B. Efron and C. Morris. "Data analysis using Stein's estimator and its generalizations". Journal of the American Statistical Association, Vol. 70, No 350, 1975, pp. 311-318.

9. B. Efron and C. Morris. "Stein's estimation rule and its competitors- An Empirical Bayes approach". Journal of the American Statistical Association, Vol. 68, No 341, 1973, pp. 117- 130.

10. C. Belanger. "Estimation of safety of four-leg unsignalized intersections". Transportation Research Record, Transportation Research Board, 1467, National Research Council, Washing D.C., 1994, pp. 23-29.

11. C. Morris. "Parametric Empirical Bayes inference: theory and applications" Journal of the American Statistical Association, Vol. 78, No 381, 1983, pp. 47-65.

12. D. Mahalel, A. S Hakkert and J. N. Prashker. "A system for the allocation of safety resources on a road network". Accident Analysis and Prevention, Vol. 14, No 1, 1982, pp. 45-56.

13. D. Mahalel, A. S. Hakkert and J. N. Prashker. "A system for the allocation of safety resources on a road network". Accident Analysis and Prevention, Vol. 14, No 1, 1982, pp. 45-56.

14. D. Mahalel. "A note on accident risk". Transportation Research Record, 1068, Transportation Research Board, National Research Council, Washing D.C., 1986, pp. 85-89.

15. D. R. D. Mcguigan. "Non-junction accident rates and their use in black-spot identification". Traffic Engineering and Control, Vol. 23, 1982, pp. 60-65.

16. D. R. D. Mcguigan. "The use of relationships between road accidents and traffic flow in 'black-spot' identification". Traffic Engineering and Control, Vol. 22, 1981, pp. 448-451.

17. E. Hauer and B. N. Persaud. "How to estimate the safety of rail-highway grade crossings and the safety effects of warning devices". Transportation Research Record, 1114, Transportation Research Board, National Research Council, Washing D.C., 1986, pp. 131-139.

18. E. Hauer and B. N. Persaud. "Problem of identifying hazardous locations using accident data". Transportation Research Record, 975, Transportation Research Board, National Research Council, Washing D.C., 1984, pp. 31-43.

19. E. Hauer and J. Lovell. "Estimation of safety at signalized intersection". Transportation Research Record, Transportation Research Board, 1185, National Research Council, Washing D.C., 1988, pp. 23-29.

20. E. Hauer, B. N. Persaud, A Smiley and D.Duncan. "Estimating the accident potential of an ontario driver". Accident Analysis and Prevention, Vol. 23, No 2/3, 1991, pp. 133-152.

21. E. Hauer. "Empirical Bayes approach to the estimation of 'unsafety': the multivariate regression method". Accident Analysis and Prevention, Vol. 24, No 5, 1992, pp. 457-477.

22. E. Hauer. "Identification sites with promise". Transportation Research Record, 1542, Transportation Research Board, National Research Council, Washing D.C., 1996, pp. 54-60.

23. E. Hauer. "On the estimation of the expected number of accidents". Accident Analysis and Prevention, Vol. 18, No 1, 1986, pp. 1-12.

24. E. Hauer, D. Terry, and M. S. Griffith. "Effects of resurfacing on safety of two - lane rural roads in New York state". Transportation Research Record, 1467, Transportation Research Board, National Research Council, Washing D.C., 1994, pp. 30-37.

25. E. Hauer, J. C. N. Ng, and J. Lovell. "Estimation of safety at signalized intersection". Transportation Research Record, 1185, Transportation Research Board, National Research Council, Washing D.C., 1988, pp. 48-60.

26. E. Willaimson and M. H. Bretherton. "Tables of the Negative Binomial probability distribution". John wiley & sons, 1963.

27. G. M. Carter and J. E. Rolph. "Empirical Bayes methods applied to estimating fire alarm probabilities". Journal of the American Statistical Association, Vol. 69, No 348, 1974, pp. 880-885.

28. H. R. Al-Masaeid, K. C. Sinha, and T. Kuczek. " Evaluation of safety impact of highway projects". Transportation Research Record, 1401, Transportation Research Board, National Research Council, Washing D.C., 1993, pp. 9-16.

29. H. Robbins. "Empirical Bayes estimation problem". Proc, Natl, Acad, Sci. USA, Vol.77, No.12, 1980, pp. 6988-6989.

30. H. Robbins. "Estimation and prediction for mixtures of Exponential distribution(Empirical Bayes/accident proneness". Proc, Natl, Acad, Sci. USA, Vol.74, No.7, 1980, pp. 2382-2383.

31. H. Robbins. "Prediction and estimation for the compound Poisson distribution(Empirical Bayes/accident proneness". Proc, Natl, Acad, Sci. USA, Vol.74, No.7, 1977, pp. 2670-2671.

32. J. A. Bonneson and P.T. McCoy. "Effect of median treatment on urban arterial safety : An accident prediction model". Transportation Research Record, 1581, Transportation Research Board, National Research Council, Washing D.C., 1997, pp. 27-36.

33. J. A. Bonneson and P. T. McCoy. "Estimation of safety at two-way stop controlled intersections on rural highways". Transportation Research Record, 1401, Transportation Research Board, National Research Council, Washing D.C., 1993, pp. 83-89.

34. J. L. Higle and J. M. Witkowski. "Bayesian identification of hazardous location" Transportation Research Record, Transportation Research Board, 1185, National Research Council, Washing D.C., 1988, pp. 24-29.

35. J. F. Lawless. "Negative Binomial and Mixed Poisson Regression". The Canadian Journal of Statistics, Vol. 15, No 3,1987, pp. 209-225.

36. J. A. Rice. "Mathematical statistics and data analysis: 2nd edition". Wadsworth, Inc, 1997.

37. J. Neter, W. Wasserman and G. A. Whitmore. "Applied statistics: 4th edition". A Division of Somon & Schuster, Inc, 1993

38. L. Moutain, B. Fawaz and D. Jarrett. "Accident prediction models for roads with minor junctions". Accident Analysis and Prevention, Vol. 28, No 6, 1996, pp. 695-707

39. L. F. Rodriguez and T. Sayed. "Accident prediction models for urban unsignalized intersections in British Columbia". University of British Columbia. 1999.

40. M. Norden, J. Orlansky and H. Jacobs. "Application of Statistical Quality-Control Technique to analysis of highway-accident data". Bulletin 117, HRB, National Research Council, Washing D.C., 1956, pp. 17-32.

41. M. A. Flak and J. C. Barbaresso. "Use of computerized roadway-information system in safety analyses". Transportation Research Record, 844, Transportation Research Board, National Research Council, Washing D.C., 1982, pp. 50-55.

42. M. J. Maher and I. Summersgill. "A comprehensive methodology for the fitting of predictive accident models". Accident Analysis and Prevention, Vol. 28, No 3, 1996, pp. 281-296.

43. M. Y. Lau and A. D. May, Jr. "Injury accident prediction models for signalized intersections". Transportation Research Record, 1171, Transportation Research Board, National Research Council, Washing D.C., 1988, pp. 58-67.

44. Michigan Department of State Police. " Michigan Traffic Crash Facts". 1997

45. M. E. Harr. "Reliability - based design in civil engineering". Dover Publication, Inc, 1996.

46. N. M. Laird and T. A Louis. "Empirical Bayes confidence intervals based on bootstrap samples". Journal of the American Statistical Association, Vol. 82, No 399, 1987, pp. 739-750.

47. R. E. Fay III and R. A Herriot. "Estimates of income for small places: An application of James-Stein procedures to census data". Journal of the American Statistical Association, Vol. 74, No 366, 1979, pp. 269-277.

48. R. W. Stokes and M. I. Mutabazi. "Rate-Quality Control Method of identifying hazardous road locations". Transportation Research Record, 1542, Transportation Research Board, National Research Council, Washing D.C., 1996, pp. 44-48.

49. Royal Statistical Society. "GLIM 4: "The Statistical system for Generalized Linear Interactive Modeling". Oxford University Press Inc, 1993.

50. S. Miaou, P. S. Hu, T. Wright, A. K. Rathi, and S. C. Davis. "Relationship between truck accidents and highway geometric design: A Poisson regression approach". Transportation Research Record, 1376, Transportation Research Board, National Research Council, Washing D.C., 1992, pp. 10-18.

51. S. Miaou and H. Lum. "Statistical evaluation of the effects of highway geometric design on truck accident involvement". Transportation Research Record, 1407, Transportation Research Board, National Research Council, Washing D.C., 1993, pp. 11-22.

52. S. Makridakis and S.Wheelwright. "Forecasting methods for management: 5th edition". John wiley & sons, 1989.

53. S. M. Ross. "Introduction to probability models: 6th edition". Academic Press, 1997.

54. S. L. Hui and J. O. Berger. "Empirical Bayes estimation of rates in longitudinal studies". Journal of the American Statistical Association, Vol. 78, No 384, 1983, pp. 753-759.

55. T. L. Maleck. " The development and evaluation of accident predictive models". Michigan State University: Thesis (Ph. D), 1980.

56. T. P. Hutchinson and A. J. Mayne. " The year to year variability in the numbers of road accidents". Traffic Engineering and Control, Vol. 18, 1977, pp. 432-433.

57. T. Kim. " Modeling freeway interchange accidents". Michigan State University: Thesis (Ph. D), 1989.

58. T. O. Kvalseth. " Cautionary note about R^2." American Statistician, Vol. 39, No 4, 1985, pp. 279-285.

APPENDICES

Appendix A - The results of accident prediction model calibration

Appendix B - The results of evaluating of freeway interchanges

APPENDIX A

THE RESULTS OF ACCIDENT PREDICTION MODEL CALIBRATION

THE RESULTS OF ACCIDENT PREDICTION MODEL CALIBRATION

Table A.1 The results of accident prediction model calibration (Interchange type 11)

Coefficient	Variable definition	Unit	Estimate	Std error	T - statistic
A Log(A)	Constant	-	3.448 (1.238)	(0.67)	(1.85)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	1.401	0.30	4.66
B_2	V_2 : Ramp traffic volume	(ADT/1000)	0.186	0.12	1.55
B_3	V_3 :Truck percent	(%)	0.620	0.19	3.26
C_1	G_1 : Interchange length	(Mile)	0.738	0.15	4.92
C_2	G_2 : Average spread ramp length	(Mile)	-1.267	0.97	-1.31
C_3	G_3 : The number of lanes	-			
C_4	G_4 : The number of total ramps	-			
C_5	G_5 : Shoulder width	(Feet)	-0.156	0.12	-1.30
Model statistic					
D_p	Dispersion parameter	1.0			
X^2	Pearson chi -square	28.84 ($\chi^2_{0.95,\ 27} = 40.11$)			
R^2	Coefficient of determination	0.60			
K	Negative Binomial parameter	8.05			

Table A.2 The results of accident prediction model calibration (Interchange type 12)

Coefficient	Variable definition	Unit	Estimate	Std error	t - statistic
A Log(A)	Constant	-	31.343 (3.445)	(0.73)	(4.72)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	0.946	0.24	3.94
B_2	V_2 : Ramp traffic volume	(ADT/1000)			
B_3	V_3 :Truck percent	(%)			
C_1	G_1 : Interchange length	(Mile)	0.933	0.36	2.59
C_2	G_2 : Average spread ramp length	(Mile)	-3.842	1.31	-2.93
C_3	G_3 : The number of lanes	-			
C_4	G_4 : The number of total ramps	-			
C_5	G_5 : Shoulder width	(Feet)			
Model statistic					
D_p	Dispersion parameter	1.0			
X^2	Pearson chi -square	14.66 ($\chi^2_{0.95,\ 14}$ = 23.68)			
R^2	Coefficient of determination	0.88			
K	Negative Binomial parameter	10.74			

Table A.3 The results of accident prediction model calibration (Interchange type 13)

Coefficient	Variable definition	Unit	Estimate	Std error	t - statistic
A Log(A)	Constant	-	3.614 (1.285)	(1.07)	(1.20)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	0.947	0.47	2.01
B_2	V_2 : Ramp traffic volume	(ADT/1000)	0.187	0.16	1.17
B_3	V_3 :Truck percent	(%)			
C_1	G_1 : Interchange length	(Mile)	0.816	0.22	3.71
C_2	G_2 : Average spread ramp length	(Mile)			
C_3	G_3 : The number of lanes	-	0.136	0.10	1.36
C_4	G_4 : The number of total ramps	-			
C_5	G_5 : Shoulder width	(Feet)			
Model statistic					
D_p	Dispersion parameter	1.0			
X^2	Pearson chi -square	19.82 ($\chi^2_{0.95,\ 19}$ = 30.14)			
R^2	Coefficient of determination	0.47			
K	Negative Binomial parameter	5.48			

Table A.4 The results of accident prediction model calibration (Interchange type 14)

Coefficient	Variable definition	Unit	Estimate	Std error	t - statistic
A Log(A)	Constant	-	17.531 (2.864)	(1.25)	(2.29)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	0.911	0.43	2.12
B_2	V_2 : Ramp traffic volume	(ADT/1000)	0.142	0.14	1.00
B_3	V_3 :Truck percent	(%)			
C_1	G_1 : Interchange length	(Mile)	1.315	0.33	3.98
C_2	G_2 : Average spread ramp length	(Mile)	-2.278	1.984	-1.15
C_3	G_3 : The number of lanes	-			
C_4	G_4 : The number of total ramps	-			
C_5	G_5 : Shoulder width	(Feet)			
Model statistic					
D_p	Dispersion parameter	1.0			
X^2	Pearson chi -square	9.07 ($\chi^2_{0.95,9}$ = 16.92)			
R^2	Coefficient of determination	0.65			
K	Negative Binomial parameter	6.38			

Table A.5 The results of accident prediction model calibration (Interchange type 21)

Coefficient	Variable definition	Unit	Estimate	Std error	t - statistic
A Log(A)	Constant	-	5.479 (1.701)	(1.02)	(1.67)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	0.467	0.43	1.09
B_2	V_2 : Ramp traffic volume	(ADT/1000)	0.470	0.18	2.61
B_3	V_3 :Truck percent	(%)			
C_1	G_1 : Interchange length	(Mile)			
C_2	G_2 : Average spread ramp length	(Mile)			
C_3	G_3 : The number of lanes	-			
C_4	G_4 : The number of total ramps	-			
C_5	G_5 : Shoulder width	(Feet)			
Model statistic					
D_p	Dispersion parameter	1.0			
X^2	Pearson chi -square	6.35 ($\chi^2_{0.95,6}$ = 12.19)			
R^2	Coefficient of determination	0.68			
K	Negative Binomial parameter	6.73			

Table A.6 The results of accident prediction model calibration (Interchange type 31)

Coefficient	Variable definition	Unit	Estimate	Std error	t - statistic
A Log(A)	Constant	-	3.494 (1.251)	(0.83)	(1.52)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	1.144	0.24	4.77
B_2	V_2 : Ramp traffic volume	(ADT/1000)	0.128	0.11	1.16
B_3	V_3 :Truck percent	(%)	0.138	0.12	1.15
C_1	G_1 : Interchange length	(Mile)	0.319	0.19	1.68
C_2	G_2 : Average spread ramp length	(Mile)			
C_3	G_3 : The number of lanes	-			
C_4	G_4 : The number of total ramps	-			
C_5	G_5 : Shoulder width	(Feet)			
Model statistic					
D_p	Dispersion parameter	1.0			
X^2	Pearson chi -square	37.68 ($\chi^2_{0.95,\ 35} = 51.00$)			
R^2	Coefficient of determination	0.72			
K	Negative Binomial parameter	7.02			

Table A.7 The results of accident prediction model calibration (Interchange type 33)

Coefficient	Variable definition	Unit	Estimate	Std error	t - statistic
A Log(A)	Constant	-	44.124 (3.787)	(0.87)	(1.20)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	0.515	0.24	2.15
B_2	V_2 : Ramp traffic volume	(ADT/1000)	0.244	0.12	2.03
B_3	V_3 :Truck percent	(%)			
C_1	G_1 : Interchange length	(Mile)	0.956	0.24	3.98
C_2	G_2 : Average spread ramp length	(Mile)	-2.500	0.98	-2.55
C_3	G_3 : The number of lanes	-			
C_4	G_4 : The number of total ramps	-			
C_5	G_5 : Shoulder width	(Feet)			
Model statistic					
D_p	Dispersion parameter	1.0			
X^2	Pearson chi -square	16.23 ($\chi^2_{0.95, 16}$ = 26.30)			
R^2	Coefficient of determination	0.82			
K	Negative Binomial parameter	13.85			

Table A.8 The results of accident prediction model calibration (Interchange type 35)

Coefficient	Variable definition	Unit	Estimate	Std error	t - statistic
A Log(A)	Constant	-	8.619 (2.154)	 (1.19)	 (1.81)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	0.736	0.82	0.90
B_2	V_2 : Ramp traffic volume	(ADT/1000)	0.270	0.41	0.66
B_3	V_3 :Truck percent	(%)			
C_1	G_1 : Interchange length	(Mile)			
C_2	G_2 : Average spread ramp length	(Mile)			
C_3	G_3 : The number of lanes	-			
C_4	G_4 : The number of total ramps	-			
C_5	G_5 : Shoulder width	(Feet)			
Model statistic					
D_p	Dispersion parameter	1.0			
X^2	Pearson chi -square	5.36 ($\chi^2_{0.95,5}$ = 11.07)			
R^2	Coefficient of determination	0.37			
K	Negative Binomial parameter	4.35			

Table A.9 The results of accident prediction model calibration (Interchange type 41)

Coefficient	Variable definition	Unit	Estimate	Std error	t - statistic
A Log(A)	Constant	-	28.247 (3.341)	 (2.344)	 (1.43)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	0.839	0.29	2.89
B_2	V_2 : Ramp traffic volume	(ADT/1000)	0.215	0.15	1.43
B_3	V_3 :Truck percent	(%)			
C_1	G_1 : Interchange length	(Mile)			
C_2	G_2 : Average spread ramp length	(Mile)			
C_3	G_3 : The number of lanes	-			
C_4	G_4 : The number of total ramps	-	0.182	0.06	3.03
C_5	G_5 : Shoulder width	(Feet)	-0.238	0.18	-1.32
Model statistic					
D_p	Dispersion parameter	1.0			
X^2	Pearson chi -square	17.99 ($\chi^2_{0.95,\ 17}$ = 27.59)			
R^2	Coefficient of determination	0.64			
K	Negative Binomial parameter	6.37			

Table A.10 The results of accident prediction model calibration (Interchange type 51)

Coefficient	Variable definition	Unit	Estimate	Std error	t - statistic
A Log(A)	Constant	-	3.658 (1.297)	(1.23)	(1.05)
B_1	V_1: Mainline traffic volume per lane	(ADT/1000)	0.478	0.65	0.73
B_2	V_2 : Ramp traffic volume	(ADT/1000)	0.506	0.33	1.53
B_3	V_3 :Truck percent	(%)			
C_1	G_1 : Interchange length	(Mile)			
C_2	G_2 : Average spread ramp length	(Mile)			
C_3	G_3 : The number of lanes	-			
C_4	G_4 : The number of total ramps	-			
C_5	G_5 : Shoulder width	(Feet)			
Model statistic					
D_p	Dispersion parameter	1.0			
X^2	Pearson chi -square	5.19 ($\chi^2_{0.95,\ 5}$ = 11.07)			
R^2	Coefficient of determination	0.47			
K	Negative Binomial parameter	4.86			

APPENDIX B

THE RESULTS OF EVALUATING OF FREEWAY INTERCHANGES

THE RESULTS OF EVALUATING OF FREEWAY INTERCHANGES

Table B.1 The results of evaluating of freeway interchanges by the prediction model method

Route	Interchange ID	Type	Interchange type	Cross road	Observed	Fitted	Probability	identified sites	
								90 %	95 %
I-69	69136	Type11	Diamond	Church St	76	81	0.49		
I-69	69128	Type31	Parclo B	Morrish Rd	41	57	0.27		
I-69	69129	Type31	Parclo B	Miller Rd	93	84	0.66		
I-69	69143	Type31	Parclo A 4 Q	Irish Rd	80	85	0.49		
I-69	69131	Type31	Parclo A 4 Q	Bristol Rd	48	106	0.05		
I-69	69141	Type31	Parclo A 4 Q	Belsay Rd	45	69	0.20		
I-69	69138	Type31	Parclo A 4 Q	M-54 dort Hwy	103	126	0.36		
I-69	69139	Type31	Parclo A 4 Q	Center Rd	62	93	0.21		
I-69	69135	Type31	Parclo B 4 Q	Hammerburg Rd	117	135	0.42		
I-69	69137	Type41	Full Direct	M-54 Br	165	83	0.98	*	*
I-75	75005	Type11	Diamond	Erie Rd	39	48	0.36		
I-75	75006	Type11	Diamond	Luna pier Rd	48	49	0.54		
I-75	75042	Type11	Diamond	Outer Dr	57	84	0.21		
I-75	75018	Type11	Diamond	Nadeau Rd	60	36	0.95	*	*
I-75	75045	Type11	Diamond	Springwells Ave	137	114	0.75		
I-75	75049A	Type11	Diamond	12th Ave	178	146	0.76		
I-75		Type11	Daimond	Clay Ave	36	35	0.59		
I-75	75052A	Type11	Diamond	Warren Ave	95	107	0.43		
I-75	75047	Type11	Diamond	M-3 Clark Ave	196	140	0.87		
I-75	75054	Type11	Diamond	Clay Ave	201	182	0.66		
I-75	75058	Type11	Diamond	7 mile Rd	194	251	0.29		
I-75	75057	Type11	Diamond	Mcnichols Rd	193	238	0.34		
I-75		Type11	Diamond	I-94	80	135	0.12		
I-75	75026	Type12	Mod Tight Diamond	Huron River Rd	48	35	0.90	*	
I-75	75036	Type12	Mod Tight Diamond	Eureka Rd	75	60	0.84		
I-75		Type12	Tight Diamond	M-50	47	97	0.05		

Table B.1 The results of evaluating of freeway interchanges by prediction model method (continued)

Route	Interchange ID	Type	Interchange type	Cross road	Observed	Fitted	Probability	Identified sites	
								90 %	95 %
I-75	75062	Type12	Tight Diamond	11 mile Rd	492	465	0.69		
I-75	75044	Type13	Part Diamond	Dearborn Rd	96	57	0.96	*	*
I-75	75051A	Type13	Part Tight Diamond	John St	240	168	0.89		
I-75	75047B	Type13	Part Tight Diamond	Lafayette Ave	109	96	0.74		
I-75	75052	Type13	Part Tight Diamond	Mack Ave	252	245	0.66		
I-75	75055	Type13	Part Tight Diamond	Holbrook Ave	51	85	0.25		
I-75	75055A	Type13	Part Tight Diamond	Caniff Ave	49	94	0.18		
I-75	75011	Type14	Diamond + loop	La plaisance Rd	23	23	0.62		
I-75	75040	Type14	Split Diamond	Northline Rd	148	226	0.25		
I-75	75060	Type14	Split Diamond	9 mile Rd	38	35	0.69		
I-75	75046	Type14	Split Diamond	Livernois Rd	77	129	0.20		
I-75	75020	Type21	Trumpet A	I-275	36	69	0.17		
I-75	75072	Type21	Trumpet A	Crooks Rd	69	105	0.30		
I-75	75081	Type21	Trumpet A	M-24	125	123	0.68		
I-75	75032	Type31	Parclo B	West Rd	60	77	0.33		
I-75	75067	Type31	Parclo A	Rochester Rd	311	246	0.78		
I-75	75084	Type31	Parclo B 4 Q	Baldwin Rd	120	135	0.44		
I-75	75089	Type31	Parclo A 4 Q	Sashabaw Rd	112	121	0.48		
I-75	75122	Type31	Parclo A 4 Q	Pierson Rd	127	85	0.89		
I-75	75041	Type31	Parclo A 4 Q	M-39 Soufield Rd	131	166	0.33		
I-75	75079	Type31	Parclo A 4 Q	University Dr	133	152	0.42		
I-75	75065	Type31	Parclo B 4 Q	14 mile Rd	403	285	0.86		
I-75	75063	Type31	Parclo A 4 Q	12 mile Rd	286	276	0.59		
I-75		Type33	Parclo AB	Bay city Rd	57	51	0.75		
I-75	75027	Type33	Parclo AB	Huron River Dr	25	23	0.72		
I-75	75021	Type33	Parclo AB	Newport Rd	39	40	0.60		
I-75	75029	Type33	Parclo AB 4 Q	Gibraltar Rd	47	48	0.60		
I-75	75009	Type33	Parclo AB	Otter Creek Rd	26	35	0.30		

Table B.1 The results of evaluating of freeway interchanges by the prediction model method (continued)

Route	Interchange ID	Type	Interchange type	Cross road	Observed	Fitted	Probability	Identified sites	
								90 %	95 %
I-75	75013	Type33	Palclo AB	M-50 Front Rd	28	34	0.40		
I-75	75014	Type33	Palclo AB	Elm Rd	44	36	0.84		
I-75	75074	Type33	Parclo AB	Adams Rd	208	114	0.99	*	*
I-75	75118	Type33	Parclo AB	M-56	105	107	0.60		
I-75	75083	Type33	Parclo AB 4 Q	Joslyn Rd	117	119	0.60		
I-75	75043	Type33	Parclo AB 4 Q	Schaefer Hwy	195	235	0.37		
I-75	75116	Type35	Clover -loop	M-121 Bristol Rd	220	172	0.81		
I-75	75069	Type35	Clover w/C-D	16 mile Rd	406	229	0.95	*	*
I-75	75077	Type35	Cloverleaf	M-59	135	223	0.28		
I-75	75048	Type41	Directional Y	Michigan Ave	106	104	0.63		
I-75	75051	Type41	Full Direct	Madison Ave	197	155	0.82		
I-75	75040A	Type41	Part Direct	Dix Hwy	123	129	0.57		
I-75	75125	Type41	Directional Y	I-475	64	103	0.22		
I-75	75034	Type41	Part Direct	Ext to Dix Hwy	57	48	0.76		
I-75	75117	Type41	Gen Directional	Miller Rd	193	155	0.80		
I-75	75075	Type41	Directional Y	Ext to WB I-75	118	107	0.70		
I-75	75059	Type41	Part Direct	Chrysler Rd	321	427	0.35		
I-75	75056	Type41	Full Direct	Davison to SB I-75	367	274	0.85		
I-75	75061	Type41	Full Direct	Ext to NB I-75	329	441	0.34		
I-75	75002	Type51	Other	Sumit Rd	21	29	0.49		
I-75		Type51	Other	Oakland Center Dr	35	52	0.43		
I-94	94157	Type11	Diamond	Jackson Rd	49	54	0.46		
I-94	94156	Type11	Diamond	Kalmback Rd	154	197	0.31		
I-94	94220A	Type11	Diamond	French Rd	213	142	0.91	*	
I-94	94218	Type11	Diamond	Van dyke Ave	247	164	0.91	*	
I-94	94230	Type11	Diamond	12 mile Rd	242	127	0.98	*	*
I-94	94228	Type11	Diamond	10 mile Rd	160	166	0.52		

Table B.1 The results of evaluating of freeway interchanges by the prediction model method (continued)

Route	Interchange ID	Type	Interchange type	Cross road	Observed	Fitted	Probability	Identified sites	
								90 %	95 %
I-94	94223	Type11	Diamond	Cadieux Ave	164	139	0.73		
I-94	94225	Type11	Diamond	Vernier Rd	322	244	0.83		
I-94	94224	Type11	Diamond	Moross Rd	207	169	0.77		
I-94	94227	Type11	Diamond	9 mile Rd	222	177	0.79		
I-94	94022	Type12	Tight Diamond	John Beers Rd	52	41	0.85		
I-94	94128	Type12	Tight Diamond	Michigan Ave	58	58	0.63		
I-94	94141	Type12	Tight Diamond	Elm Rd	59	84	0.25		
I-94	94085	Type12	Tight Diamond	Shafter 35th St	89	82	0.72		
I-94	94137	Type12	Tight Diamond	Airport Rd	43	67	0.19		
I-94	94030	Type12	Tight Diamond	Napier Ave	62	85	0.29		
I-94	94072	Type12	Tight Diamond	9th St	105	99	0.69		
I-94	94139	Type12	Mod Tight Diamond	M-106	58	78	0.31		
I-94		Type12	Tight Diamond	Pipestone Rd	142	106	0.90	*	
I-94	94075	Type12	Tight Diamond	Oakland Dr	115	149	0.33		
I-94	94217B	Type12	Tight Diamond	Mt Elliott Ave	306	196	0.97	*	*
I-94	94212B	Type13	Part Tight Daimond	30th St	132	103	0.83		
I-94	94211	Type13	Part Tight Daimond	Lonyo Ave	41	67	0.27		
I-94	94214A	Type13	Part Tight Daimond	Grand River Blvd	103	65	0.93	*	
I-94	94211C	Type13	Part Tight Daimond	Addison Ave	55	57	0.61		
I-94	94217	Type13	Part Tight Daimond	Chene Rd	112	177	0.28		
I-94	94213	Type13	Part Diamond	W Grand Blvd	163	112	0.90		
I-94	94222B	Type13	Part Diamond	Harper Ave	120	132	0.56		
I-94	94221	Type13	Part Tight Diamond	Outer Dr	39	55	0.37		
I-94	94214C	Type13	Part Tight Daimond	14th St	220	126	0.96	*	*
I-94	94211B	Type13	Part Tight Daimond	Cecil Ave	138	118	0.77		
I-94	94211A	Type13	Part Tight Daimond	Weir St	93	129	0.37		
I-94	94214B	Type13	Part Tight Daimond	Trumbull Ave	156	100	0.93	*	
I-94	94127	Type14	Mod Diamond	Concord Rd	68	59	0.74		
I-94	94023	Type14	Diamond+loop	Red Arrow Hwy	72	77	0.55		

Table B.1 The results of evaluating of freeway interchanges by the prediction model method (continued)

Route	Interchange ID	Type	Interchange type	Cross road	Observed	Fitted	Probability	Identified sites	
								90 %	95 %
I-94	94127	Type14	Mod Diamond	I-94 BL	98	66	0.91	*	
I-94	94175	Type14	Diamond+loop	Saline Rd	54	71	0.37		
I-94	94241	Type14	Mod Diamond	21 mile Rd	89	67	0.84		
I-94	94169	Type14	Mod Diamond	Zeeb Rd	87	104	0.45		
I-94	94199	Type14	Mod Diamond	Middlebelt Rd	122	103	0.76		
I-94	94222A	Type14	Split Diamond	Chalmers Ave	79	86	0.53		
I-94	94215C	Type14	Split Diamond	John Rd	392	237	0.95	*	*
I-94	94136	Type21	Trumpet A	M-60	46	51	0.58		
I-94	94034	Type21	Trumpet A	SB I-196	96	63	0.94	*	
I-94	94235	Type21	Trumpet B	Shook Rd	59	40	0.92	*	
I-94	94033	Type21	Trumpet B	I-94 BL	20	33	0.27		
I-94	94240	Type21	Trumpet B	Hall Rd	73	70	0.69		
I-94	94006	Type31	Parclo B	Union Pier Rd	32	33	0.54		
I-94	94012	Type31	Parclo B	Sawyer Rd	61	43	0.86		
I-94	94124	Type31	Parclo B	M-99	93	70	0.82		
I-94	94027	Type31	Parclo A	M-63	162	71	0.99	*	*
I-94	94028	Type31	Parclo B	Scottdale Rd	148	82	0.97	*	*
I-94	94232	Type31	Parclo B	Little Mack Ave	121	128	0.50		
I-94	94159	Type31	Parclo B	M-52	107	108	0.55		
I-94	94219	Type31	Parclo B	Gratiot Ave	278	185	0.90	*	
I-94	94001	Type31	Parclo A 4 Q	US-12	39	45	0.43		
I-94		Type31	Parclo A 4 Q	US-127	72	117	0.16		
I-94	94080	Type31	Parclo A 4 Q	Sprinkle Rd	116	120	0.52		
I-94	94234	Type31	Parclo B 4 Q	Harper Rd	157	128	0.75		
I-94	94078	Type31	Parclo A 4 Q	Kilgore Rd	90	137	0.20		
I-94	94076	Type31	Parclo B 4 Q	Westnedge Ave	90	150	0.14		
I-94	94236	Type31	Parclo A 4 Q	Metro Beach Rd	111	102	0.64		
I-94	94183	Type31	Parclo A 4 Q	Hamilton Rd	212	228	0.48		
I-94	94177	Type31	Parclo A 4 Q	State St	146	161	0.46		

Table B.1 The results of evaluating of freeway interchanges by the prediction model method (continued)

Route	Interchange ID	Type	Interchange type	Cross road	Observed	Fitted	Probability	Identified sites	
								90 %	95 %
I-94	94196	Type31	Parclo A 4 Q	Wayne Rd	226	222	0.57		
I-94	94243	Type31	Parclo A 4 Q	M-29	81	184	0.04		
I-94	94104B	Type33	Parclo AB	11 mile Rd	37	50	0.28		
I-94	94052	Type33	Parclo AB	Paw Paw Rd	67	52	0.88		
I-94	94237	Type33	Parclo AB	North River Rd	112	90	0.87		
I-94		Type33	Parclo AB 4 Q	Ford Plant Rd	39	46	0.43		
I-94	94181	Type33	Parclo AB 4 Q	US-12	43	64	0.19		
I-94	94208	Type33	Parclo AB 4 Q	Greenfield Rd	58	67	0.44		
I-94	94206	Type33	Parclo AB 4 Q	Oakwood Blvd	141	165	0.42		
I-94		Type33	Parclo AB 4 Q	Michigan Ave	94	134	0.20		
I-94	94004	Type35	Cloverleaf	US-12	51	47	0.69		
I-94	94180	Type35	Clover w/C-D	US-23&BL-94	169	147	0.73		
I-94	94074	Type35	Clover w/C-D	US-131	158	166	0.60		
I-94		Type35	Clover w/C-D	Pittsfield tw	72	161	0.14		
I-94	94198	Type35	Clover w/C-D	Merriman Rd	136	192	0.38		
I-94	94144	Type41	Part Direct Y	I-94	39	84	0.10		
I-94	94185	Type41	Part Direct Y	US-12	22	27	0.45		
I-94	94210	Type41	Part Direct	Michigan Ave	191	191	0.61		
I-94	94200	Type41	Part Direct	Ecorse Rd	74	125	0.19		
I-94	94220	Type41	Full Direct	Conner Ave	256	182	0.88		
I-94	94231	Type41	Part Direct Y	Gratiot Ave	44	86	0.13		
I-94	94202	Type41	Direct.w/loops	Telegraph Rd	228	283	0.41		
I-94	94214	Type41	Full Direct	Ext to I - 96	403	293	0.87		
I-94		Type41	Full Direct	M-10	409	296	0.87		
I-94		Type41	Full Direct	Dubois St	173	204	0.46		
I-94	94229	Type41	Full Direct	11 mile Rd	194	344	0.16		
I-94	94209	Type51	Other	Rotunda Dr	73	54	0.88		
I-94	94204	Type51	Other	Pelham Rd	157	143	0.78		
I-96	96187	Type11	Diamond	Grand River Ave	90	75	0.74		

Table B.1 The results of evaluating of freeway interchanges by the prediction model method (continued)

Route	Interchange ID	Type	Interchange type	Cross road	Observed	Fitted	Probability	Identified sites	
								90 %	95 %
I-96	96186	Type11	Diamond	Wyoming Ave	23	46	0.08		
I-96	96185	Type11	Diamond	Grand River Ave	27	31	0.44		
I-96		Type11	Diamond	Schaefer Rd	23	32	0.28		
I-96	96180	Type11	Diamond	Outer Dr	94	183	0.07		
I-96	96178	Type11	Diamond	Beach Daly	121	188	0.16		
I-96	96176	Type11	Diamond	Middlebelt Rd	158	180	0.42		
I-96	96175	Type11	Diamond	Merriman Rd	111	183	0.13		
I-96	96177	Type11	Diamond	Inkster Rd	161	199	0.34		
I-96	96174	Type11	Diamond	Farmington Rd	102	178	0.11		
I-96	96159	Type12	Mod Tight Diamond	Wixom Rd	120	171	0.24		
I-96	96188A	Type12	Tight Diamond	Livernois	174	165	0.69		
I-96	96184	Type12	Tight Diamond	Greenfield Rd	166	133	0.85		
I-96	96150	Type13	Part Diamond	Pleasant Valley Rd	18	37	0.17		
I-96	96191	Type13	Part Diamond	Myrtle Ave	113	168	0.32		
I-96	96190	Type13	Part Diamond	Warren Ave	253	213	0.78		
I-96	96188B	Type13	Part Tight Diamond	Joy Rd	37	61	0.26		
I-96	96189	Type13	Part Tight Diamond	W Grand Blvd	143	187	0.41		
I-96	96173B	Type13	Part Diamond	Levan Rd	154	287	0.18		
I-96	96182	Type14	Diamond+loop	Evergreen Rd	34	77	0.08		
I-96	96186B	Type21	Trumpet A	Davison Rd	156	128	0.82		
I-96	96151	Type31	Parclo B	Kensington Rd	90	113	0.34		
I-96	96155	Type31	Parclo A	Milford Rd	185	131	0.86		
I-96	96153	Type31	Parclo B 4 Q	Kent Lake Rd	53	87	0.16		
I-96	96162	Type31	Parclo A 4 Q	Novi Rd	49	96	0.09		
I-96	96170	Type31	Parclo A 4 Q	6 mile Rd	397	275	0.88		
I-96		Type33	Palclo AB	Holly Rd	61	50	0.84		
I-96	96169	Type33	Parclo AB 4 Q	7 mile Rd	237	234	0.64		
I-96	96160	Type51	Other	Grand River Ave	107	66	0.95	*	*

Table B.1 The results of evaluating of freeway interchanges by the prediction model method (continued)

Route	Interchange ID	Type	Interchange type	Cross road	Observed	Fitted	Probability	Identified sites	
								90 %	95 %
I-96	96179	Type51	Other	Telegraph Rd	36	67	0.29		
I-96		Type51	Other	To M-102	130	205	0.38		
I-96		Type51	Other	8 mile Rd	139	104	0.89		