Research Report No. UVACTS-14-5-36
May 2002

Determining the Safety Effects of Differential Speed Limits on Rural Interstate Highways Using Empirical Bayes Method

by

Xin Sun
Dr. Nicholas J.Garber

A Research Project Report
For the Mid-Atlantic Universities Transportation Center (MAUTC)
A U.S. DOT University Transportation Center

Xin Sun
Department of Systems and Information Engineering
Email: xs3g@Virginia.EDU

Dr. Nicholas J. Garber
Department of Civil Engineering
Email: njg@virginia.edu

Center for Transportation Studies at the University of Virginia produces outstanding transportation professionals, innovative research results and provides important public service. The Center for Transportation Studies is committed to academic excellence, multi-disciplinary research and to developing state-of-the-art facilities. Through a partnership with the Virginia Department of Transportation's (VDOT) Research Council (VTRC), CTS faculty hold joint appointments, VTRC research scientists teach specialized courses, and graduate student work is supported through a Graduate Research Assistantship Program. CTS receives substantial financial support from two federal University Transportation Center Grants: the Mid-Atlantic Universities Transportation Center (MAUTC), and through the National ITS Implementation Research Center (ITS Center). Other related research activities of the faculty include funding through FHWA, NSF, US Department of Transportation, VDOT, other governmental agencies and private companies.

CTS Website
http://cts.virginia.edu

Center for Transportation Studies
Department of Civil Engineering, University of Virginia
351 McCormick Road, P.O. Box 400742
Charlottesville, VA 22904-4742
434.924.6362

ABSTRACT

A differential speed limit is defined as being one limit for automobiles and a different limit for commercial motor vehicles ("trucks") whereas a uniform speed limit is defined as a single limit for cars and trucks. Because states enact differential speed limit (DSL) solely in order to improve safety, assessment of DSL's safety impacts is of significant importance to the transport community.

Previous Before-and-After studies could not fully investigate DSL's impact on crashes due to the limited periods of time used in these studies. A different genre of studies based on the comparison of safety effects at different physical sites, such as I-64 in the western portion of Virginia (UNI) and the adjacent section of I-64 in the eastern portion of West Virginia (DSL) were also inadequate because of the limited data available at the time.

Thirteen years have passed since the enactment of the Surface Transportation and Uniform Relocation Assistance (STURA), rendering a new set of data available for further study regarding the safety effects of DSL.

Using the Empirical Bayes method for before-after safety analysis, this study developed a multivariate crash estimation model (CEM) using the before treatment years data and predicted what the safety would have been if there was no DSL enactment for the after treatment years.

This study used data from seven states, which either kept the same speed limit strategy since 1990, or changed their strategy at least once. Six types of crashes (total number of crashes, total number of fatal crashes, total number of rear-end crashes, total number of crashes with truck involved, total number of fatal crashes with truck involved, total number of rear-end crashes with truck involved) were selected for analysis. The evaluations of DSL implementation was then carried out by comparing

the predicted “would have been” crash counts and the actual crash counts of the after treatment period.

A nonlinear relationship was found between crash counts and section length, and between crash counts and AADT. The results vary for different types of crashes through different states. The results, however, generally showed that as time passed, the actual total numbers of crashes for the after period were greater than the predicted “would have been” after total numbers of crashes. Whether this difference was caused only by the policy change of DSL or other factors that contribute to differing safety conditions is therefore not conclusive.

ACKNOWLEDGEMENTS

Many thanks are given to all those who have helped in giving both professional and moral support throughout this project. I especially thank Dr. Nicholas J. Garber, my advisor for his advice, guidance, and support to me, a newer in transportation field. I thank everyone at the Virginia Transportation Research Council and the University of Virginia for their support and assistance. Specially, I want to thank Mr. Lewis Woodson for his continuous help on both my project and my life. And I thank Dr. John Miller for his brilliant ideas and kindness on my research. Also I thank Mr. Dominique Lord of University of Texas for his experienced help and programs supplied for this project. A final thank is to my husband for always being there with me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

On April 2, 1987, the Surface Transportation and Uniform Relocation Assistance (STURA) Act was enacted which permitted individual state to raise speed limits from the previously mandated national speed limit of 88.5 km/h (55 mi/h) to 104.59 km/h (65 mi/h) on rural interstate highways. Based on this act, different states took various courses of actions for particular reasons. Some states enacted Differential Speed Limits in which different speed limits were set for passenger cars and trucks. Decision makers were, in most states, reluctant to permit trucks to be driven at such a high speed because they were worried that trucks traveling at such a high speed would increase the potential of crashes between trucks and other types of vehicles. The reason for their concern was that with the larger dimension, a truck needed more time and a longer distance to decelerate to a complete stop than a passenger car, given a particular speed. Based on this physical property, a lower speed limit was imposed on trucks. The differential speed limit (DSL) was then adopted by some states in order to lessen the impact of raising speed limit.

Also, some states adopted uniform speed limits for both passenger cars and trucks for the following reasons: First, these states argued that with the higher position in a truck, a truck driver has a greater sight distance than the passenger car driver. This allowed truck drivers to have more time and therefore a longer distance to react and decelerate to stop. This argument is also the reason why the American Association of State Highway and Transportation Officials (AASHTO) doesn’t have different minimum sight distance requirements for trucks and passenger cars. Secondly, the opponents of DSL also contended that differential speed limit might increase the variation of vehicle speeds, resulting in more conflicts between trucks and other types of vehicles, especially certain types of crashes, such as rear-end and side swipe collisions. Thirdly, the economic and time benefits of a higher speed limit for trucks are other factors emphasized by proponents of uniform speed limits. Therefore, some states increase the speed limit of trucks to 104.59 km/h (65 mi/h), the same as passenger cars.

During the years after the enactment of STURA, some states made changes in main speed limit policies for passenger cars and trucks, while other states maintained the same policies. The states can therefore be classified into four main groups based on their speed limit policies as shown in table 1.1. These are:

I. States with a uniform speed limit (UNI) since 1990

II. States with a Differential Speed Limits (DSL) since 1990

III. States that changed from UNI to DSL since 1990

IV. States that changed from DSL to UNI since 1990

Table 1.1. Four groups of states with their speed limit

Groups	State	Speed limit before treatment	Year of change	Speed limit after treatment	Notes
UNI	AZ	uniform		uniform	
	IA	65		65	
	MO	55		55 and 70	55 ? $\overset{1996}{?}$? 70
	NC	65		65 and 70	65 ? $\overset{1996}{?}$? 70
DSL	IL	65/55		65/55	
	CA	70-55/55		70-55/55	
	WA	65/60		65/60	

	IN	65/60		65/60	
UNI-DSL	AR	65	1996 Aug	70/65	
	ID	75	1998 July	75/65	65? ?[1996]? 75? ?[1998]? 75/65
DSL-UNI	VA	65/55	1994	65/65	

Note: UNI: always uniform speed limits for cars and trucks in the 90's

DSL: always differential speed limits for cars and trucks in the 90's

UNI-DSL: speed limit changed from uniform to differential speed limit in the 90's

DSL-UNI: speed limit changed from differential speed limit to uniform in the 90's

Previous Before-and-After studies could not fully investigate DSL's impact on crashes due to the limited periods of time used in these studies. A different genre of studies based on the comparison of safety effects at different physical sites, such as I-64 in the western portion of Virginia (UNI) and the adjacent section of I-64 in the eastern portion of West Virginia (DSL) were also inadequate because of the limited data available at the time.

Thirteen years have passed since the enactment of the Surface Transportation and Uniform Relocation Assistance (STURA), rendering a new set of data available for further study regarding the safety effects of DSL.

This study used data from several states, which has either kept the same speed limit policy since 1990, or has changed their policy at least once.

1.2 OBJECTIVES OF THIS RESEARCH

The objectives of this study are:

i. Determine the effect of the DSL for trucks and passenger cars on crashes.

ii. Determine the effect of the uniform speed limits for trucks and passenger cars on occurrence of crashes.

1.3 STRUCTURE OF THIS REPORT

This report is organized into 7 sections including this introduction as chapter one. Chapter two gives a brief review of the literature search, explaining different studies conducted and their conclusions. Chapter three introduces the criteria for data collection and data reduction for this study. Chapter four describes the Empirical Bayes method used in this study for crash data analysis. Chapter five illustrated how EB method was used step by step using one entity of Virginia as an example. In Chapter Six, EB method was applied to all states selected and results obtained were shown. Finally, chapter seven showed summaries of the results, conclusions and recommendation are also presented.

CHAPTER 2

LITERATURE REVIEW

This chapter gives a brief review of related studies that has been done recently. In the first section, study results of other researches on the safety effects of DSL are presented. In the second section, a summary of the literature of EB method is presented.

2.1 FORMER DSL STUDY RESULTS

Several studies have been conducted to evaluate the effect of Differential Speed Limit since the enactment of STURA in 1987. Following are the summaries of the literature reviewed related to safety analysis.

Council, Duncn and Khattack (1998) noted that for two vehicle rear-end collisions between cars and trucks, a high speed differential is one of the variables that increase the severity of the crash [1].

Harkey and Mera (1994) found that there are no statistically significant differences at the 95% confidence level for the number of crashes, the crash types, or crash severity [2].

Garber and Ravi Gadiraju (1988, 1989, 1991) have done a lot of statistical analysis on safety effect of DSL. In 1988, they found that accident rates increase with increasing speed variance for all classes of roads [3].

In 1989, a simulation study carried out in Virginia by Garber, also concluded that no safety benefits were observed as a result of imposing a DSL. That study did however report that there was a potential for increase in accident rates, especially on highways with high AADT and high percentage of trucks [4]. In 1991, the same authors found no evidence indicating that the increase of the speed limit to 104.59 km/h (65 mi/h) for trucks at the test sites resulted in a significant increase in fatal, injury and overall accident rates. Comparisons of crash rates in the adjacent states of Virginia and West Virginia showed relatively more rear-end crashes in Virginia and a lower number of two-vehicle crashes in West Virginia than in Virginia after the changes in speed limits. So no benefit was provided by the DSL over uniform speed limit [5].

In a Maryland study by J.W.Hall and L.V. Dickinson (1974), it was noted that speed differences contributed to the number of crashes, primarily to the number of rear-end and lane-changing crashes. Although the speed variation can be brought about by strictly enforced differential truck speed limits, the existence of a posted DSL was not found to be related to the occurrence of truck crashes. Also lower rates of truck crashes could be expected with higher speed limits and hence the study recommended an increase of truck speed limits from 88.5 km/h (55 mi/h) to 96.6 km/h (60 mi/h) or 104.59 km/h (65 mi/h) on highways carrying a higher truck percentage [6].

The Idaho DOT Evaluation of 2000 found that crash data didn't increase as a result of the speed limit change [7].

2.2 EMPIRICAL BAYES METHOD

The Empirical Bayes method was developed by Ezra Hauer as a new observational before-after studies method in road safety analysis. He indicated in his book [8] that there existed two "wrong" beliefs in safety effect analysis: Some believe "that unless one can conduct randomized experiments, the safety effect of a treatment can never be known". Others believe that they can evaluate "safety effect of a treatment by comparing accident rates of 'before' and 'after' implementation". By this method, he showed that “it is possible to learn from experience even if it does not confirm to the strictures of randomized experiments”. Statistically, estimation could be made from outcomes of repeated trials. But this estimation was based on one assumption that all related conditions remain fixed from trial to trial. In our safety analysis of interstate facility, the conditions change from case to case. The EB method can “use the evidence from a time series of annual crash counts to estimate the expected number of crashes in a certain year, given that the expected values are changing from year to year” [8].

In this method, crash frequency is used as a measurement that reflects road safety, which is different from the current engineering practice that uses crash rate to define safety. Crash rate is usually defined as following:

$$\text{Crash rate} = \text{Crash frequency}/(\text{AADT}*365*\text{length})$$

Thus in a safety study, which measure is a good indicator of road safety? Crash frequency or crash rate?

In his book, Ezra Hauer has explained why the conclusion may be paradoxical if made by comparing crash rates instead of using crash frequency. The main reason is that the crash rate can’t separate the effect of traffic flow on safety from the effect of treatment on safety. “This kind of error will be present always when the relationship between crash frequency and traffic flow is not proportional”. Interested readers are referred to Ezra Hauer 1997 [8] pp26-pp29 for a more detailed discussion. In this thesis, crash frequency is used as an indication of road safety. The crux of Hauer’s argument is that we cannot presume a linear relationship between crash counts and traffic flow. Rather, the relationship is estimated through regression analysis.

Some studies on the safety effects of a treatment has been conducted to compare the differences between the before and after crash counts. This is called the “naïve before-after comparison” by Hauer. It is right if nothing else changed on the studied entity except the treatment. Then the changes between the before and after crash counts is only caused by the treatment. But for road safety, the truth is that besides the treatment, there are many other factors that change with time. It can’t therefore be said that the changes between the before and after crash counts is only caused by the treatment. In short, even if the safety of an a roadway facility is defined as the number of crashes expected to occur on the facility, historical crash counts alone cannot be used to predict the number of crashes that will occur in the future.

Overall, the EB method offers the following advantages compared to traditional approaches such as hypothesis tests:

1. The linear regression analysis assumes that the error structure of the data follows a normal distribution with a mean of zero and a constant variance. However, several studies of crash data have concluded that the error structure of the Negative Binomial (NB) distribution would be a better description of the variation of crash frequency between sites [8,9,10,11]. There are two assumptions regarding the probability distribution. First, on a particular site, the distribution of crash counts ($K_{i,y}$) over multiple years obeys the Poisson Distribution. Second, for any particular year, the crash counts of $K_{i,y}$ between sites follow the Negative Binomial distribution.

   To ensure that the NB assumption was appropriate, a distribution analysis of Poisson assumption and NB assumption were conducted on VA, AZ, ID and NC data. This result showed conformity with the Negative Binomial distribution. Details of the distribution analysis are shown in APPENDIX A.

2. Traditional estimation analysis assumes that all the other relevant conditions except the treatment remain fixed between repeated measurements. This is not true because the estimated crash frequency (*m*) of a particular site doesn't remain constant from year to year. In fact, it changes from year to year. There are many other factors that contribute to the changes of the estimated crash frequency (*m*), such as traffic flow, precipitation, attitudes of drivers, etc. With multivariate model, the EB method captures the yearly change in the estimated crash frequency (*m*) in the following assumed rule [8]:

$$\frac{m_{i,y}}{m_{i,1}} \; ? \; \frac{E(m_{i,y})}{E(m_{i,1})} \; ? \; C_{i,y} \tag{4.1}$$

Where,

mi,y = the estimated crash frequency on site i in year y (y=1, 2, …Y+Z)

mi,1 = the estimated crash frequency on site i in the first year

E(mi,y) = the mean of the estimated crash frequency of site i in year y(y=1, 2, …Y+Z)

E(mi,1) = the mean of the estimated crash frequency of site i in the first year

Ci, y = yearly changing ratio with respect to the first year of the before period

Hauer [8] has stated clearly that it's not necessary to put the first year of before period in the denominator. The results are not affected by the choice of which year in the denominator. An analysis with the VA data set, choosing a different before year as a base in the denominator was also done, for evaluations, as well as a mathematical derivation (See APPENDIX B).

Several studies have applied the EB method developed by Hauer to different crash estimation and prediction studies.

Bhagwant N. Persaud, etc. (March, 2000) [12] gave a detailed example to show how the Empirical Bayes Method can be used to evaluate other safety effects from installing Roundabouts in America. However, the analysis was based on an averaged data for both the before and after periods instead of using yearly data. The disadvantage of using averaged data, e.g. averaged AADT for the 5 years before treatment and 3 years after treatment, is the assumption that the contributing factors, such as traffic flow, remains constant from year to year, which, in most cases is not the case. Therefore the yearly variation is not reflected accurately.

Bhagwant N. Persaud, Ezra Hauer, etc (1997) [9,11] described two basic issues of EB method. First, the error structure obeys Negative Binomial distribution which is considered to be more appropriate to describe the crash counts between sites than Poisson or normal distribution used in traditional regression modeling; Error is defined as residuals which is the difference between model estimated values and actual values. Second, how the estimated value is related with its variance in the reference group by a aggregation parameter and how to estimate this parameter using Maximum Likelihood procedure.

Model development is an important part in EB method. The model takes considerations of causal factors with corresponding parameters in the following format:

$$E(m) = f\,(?_i, X_i)$$

Where,

E(m)=expected number of crashes per time unit

Xi = contributing covariates

? i =parameters to be estimated

There exist many methods and software packages to estimate model parameters. Generalized Linear Modeling (GLM) has been recommended by Ezra Hauer, Persaud, Dominique, etc. as a good analysis tool to estimate model parameters [8,9,12,13,14]. The two main advantages of this method are the "link function" and "distribution" definition functions. The "link function" allows a linear transformation of the non-linear model format during regression. The "distribution" function makes it possible to describe the distribution of the responsible variable.

In 2000, Dominique discussed how to develop models considering temporal correlation (i.e. yearly trend) by comparing different procedures GEE and GLM. Later in chapter four, the validation of both methods will be done[15].

2.3 SUMMARY

From the related studies reviewed above, we can see that the safety effect of DSL is an important issue and has been researched using different methods. The EB method with many advantages over traditional methods shows a promising way for before-after observational safety studies. Details of EB method and how to apply it to our study will be described later in chapters four and five.

CHAPTER 3

DATA COLLECTION AND PREPARATION

In April, 2001, letters, e-mails and phone calls were used to different state DOTs requesting for crash, traffic flow, speed and geometric data. Upon received responses from the different state DOTs, the data were sorted and compiled and reduced into a series of spreadsheet for future analysis. Examples of initial data request and follow up data request letters are attached in appendix C.

3.1 CRASH DATA

To obtain crash data on VA rural interstate highways, we took all interstate systems including I-85, I-95, I-81, I-64, I-77 and divided them into sections with beginning and ending nodes for HTRIS. These nodes were approximately 10 mile long excluding urban locations, and also taking into account for different geometric and traffic operation characteristics.

Six types of annual crash data were obtained for this study. These are:

?? Total number of crashes involving passenger cars and trucks for all crash types

?? Total number of fatal crashes involving both passenger cars and trucks

?? Total number of rear-end crashes involving both passenger cars and trucks

?? Total number of crashes with trucks involved

?? Total number of fatal crashes with trucks involved

?? Total number of rear-end crashes with trucks involved

With the beginning and ending nodes available, the crash data of VA was then obtained from Highway Traffic Record Information System (HTRIS) on each interstate section. All the reported crashes on that section were displayed with basic characteristics, such as a microfilm number, year, date. Details of each crash could be traced further, including collision type, collision severity, vehicle types, weather and road conditions. Unfortunately, due to some discrepancies with HTRIS, a hard copy had to be produced with the microfilm numbers that can be traced, so that every type of accident was accounted for.

Since the data obtained from the other states was in a different format, they had to be put in a format that facilitated their use in this study. An example can be seen in appendix A.

3.2 TRAFFIC FLOW DATA

The traffic data used in this study was the Average Annual Daily Traffic (AADT). The AADTs for Virginia were obtained from the AADT books available for all the years from 1991 to 1999 except 1998. Some sections were long enough to have different AADT on different parts of it. A weighted average AADT was calculated based on the lengths that have different AADTs in one section. The 1998 AADT was estimated by averaging the AADTs of 1997 and 1999. Since the AADT data obtained from the other states was in a different format, they had to be put in a format that facilitated their use in this study. An example can be seen in appendix A.

3.3 SPEED DATA

The speed data required was the mean speed and the 85th percentile speed. Unfortunately, it was not possible to obtain annual speed data for a majority of the states for which the data were obtained. Some states didn't record speed data at all, while others only recorded the speed data for certain periods, and some only had records for sections of the highways. The speed data obtained were only adequate for the speed analysis, but couldn't be used in the EB method since the data was not available for each highway section in analysis. For example, an interstate section might have speed monitoring sites, which was enough to compare overall speeds but not enough to use in a crash estimation model.

3.4 GEOMETRY CHARACTERISTICS DATA

The first criterion for the data request was that they must be from rural interstate highways sections only. Rural interstate highways are defined by Federal Highway Administration (FHWA) as those segments of interstate highways located in census where the population is less than 50,000.

Several criteria were established for the sections of highways that traffic and crash data were requested. These were: the conditions on each section of these rural interstate highways must be homogeneous, i.e. speed and traffic flow not changed significantly along a given section; The length of each section should be long enough to ensure the occurrence of some crashes annually.

No interchanges were included within a section for which we obtained. A minimum length of 250 feet from the end of each section of an interchange was set.

3.5 DATA REDUCTION AND PREPARATION

Since the Crash Estimation Model (CEM) is estimated using yearly data, any missing values could cause a bias of the developed models, resulting in inaccurate predictions. Thus for those sections that had some AADTs missing, request was resent to see if it was possible to get the missing the data. If they were still not available, an estimation of the missing AADTs for a site was interpolated from the available AADTs. Table 3.1 summarized the data obtained from different stats.

Table 3.1 Summary of Available Data

Analysis Group	State	Year of Change	Years of Data	No. of sites	ADT[a]	All Crashes (cars and trucks)			Truck Crashes Only			Mean speed	85th Percentile Speed
						Total	Fatal	Rear End	Total	Fatal	Rear End		
Uniform Speed Limit Only	AZ		1991-2000	556	X	X	X	X	X	X	X	NA	NA
	IA		1993-1999	7	NA	X	X	X	X	X	X	X	X
	MO		1991-1999	3	X	X	X	X	X	NA	NA	X	X
	NC		1991-2000	26	X	X	X	X	X	X	X	some[b]	some[b]
Differential Speed Limit Only	IL		1993-1999	5	X	X	X	X	X	X	X	some[c]	some[c]
	CA		1991-2000	10	some[d]	X	X	X	X	X	X	some	some
	WA		1991-2000	9	X	X	X	X	some[e]	some[e]	some[e]	NA	NA
	IN		1995-1999	2	some[f]	X	X	X	X	NA	NA	some	some
Changed from Uniform to DSL	AR	Aug 1996	1991-1999	10	X	X	X	X	X	NA	NA	NA	NA
	ID	July 1998	1991-2000	32	X	X	X	X	X	X	X	some[f]	some[f]

	MT	Nov 1995	1993-2001	1[h]	X	X	X	X	X	X	X	NA	NA
Changed from DSL to Uniform	VA	1994	1991-1999	267	X	X	X	X	X	X	X	some[i]	some[i]

Note:

X: indicates we have the data

NA: indicates the data are not available

[a] ADT indicates Average Daily Traffic

[b] North Carolina speeds only available 1991-1994

[c] Illinois speeds not available for 1995 and 1996

[d] California ADT only available for 1999-2000

[e] Washington truck crashes only available for 1991-1996

[f] Indiana ADT only available for 1995 and 1998

[g] Idaho speeds available for 1997 – 2000

[h] Montana data aggregated for the entire system of interstates

[i] Speed data available for some Virginia sites from 1991-1992 only

However, not all of the data from all the states that had data were used in this study. Data from some states could not be used for different reasons. For example, Iowa sent us crash data, but no available AADT. Illinois sent a hard copy of crash and speed data in individual speed bins. AADT estimated from the binned speed data could not be validated by the IL DOT. To be safe in our study, we didn't use it. California was not used because AADT data was not adequate for analysis. Indiana is discarded because of too few data sites (only two sites) for regression analysis. Montana gave crash and speed data aggregated for the entire interstate systems instead of discrete sections. The states that provided enough data and were analyzed later are shown in Table 3.2.

Table 3.2 States that have been analyzed in this study

Groups	State
UNI	AZ
	MO
	NC
DSL	WA
UNI-DSL	AR
	ID
DSL-UNI	VA

Note: UNI: always uniform speed limits for cars and trucks in the 90’s

DSL: always differential speed limits for cars and trucks in the 90’s

UNI-DSL: speed limit changed from uniform to differential speed limit in the 90’s

DSL-UNI: speed limit changed from differential speed limit to uniform in the 90’s

CHAPTER 4

METHODOLOGY

The task of this chapter is to describe the basic theory of Empirical Bayes (EB) method used for our before-after study. It consists of four sections. The first section gives the basic idea of the EB method. The second section addresses the development of Crash Estimation Model (CEM) from the reference groups. Estimation of the expected crash frequency for the selected sites before speed limit change is described in the third section. The fourth section includes the prediction of the "would have been crash frequency" for the selected sites after the speed limit change.

4.1 DESCRIPTION OF EB METHOD

The basic premise of Hauer's EB approach **[8]** are shown in **Figure 4.1**

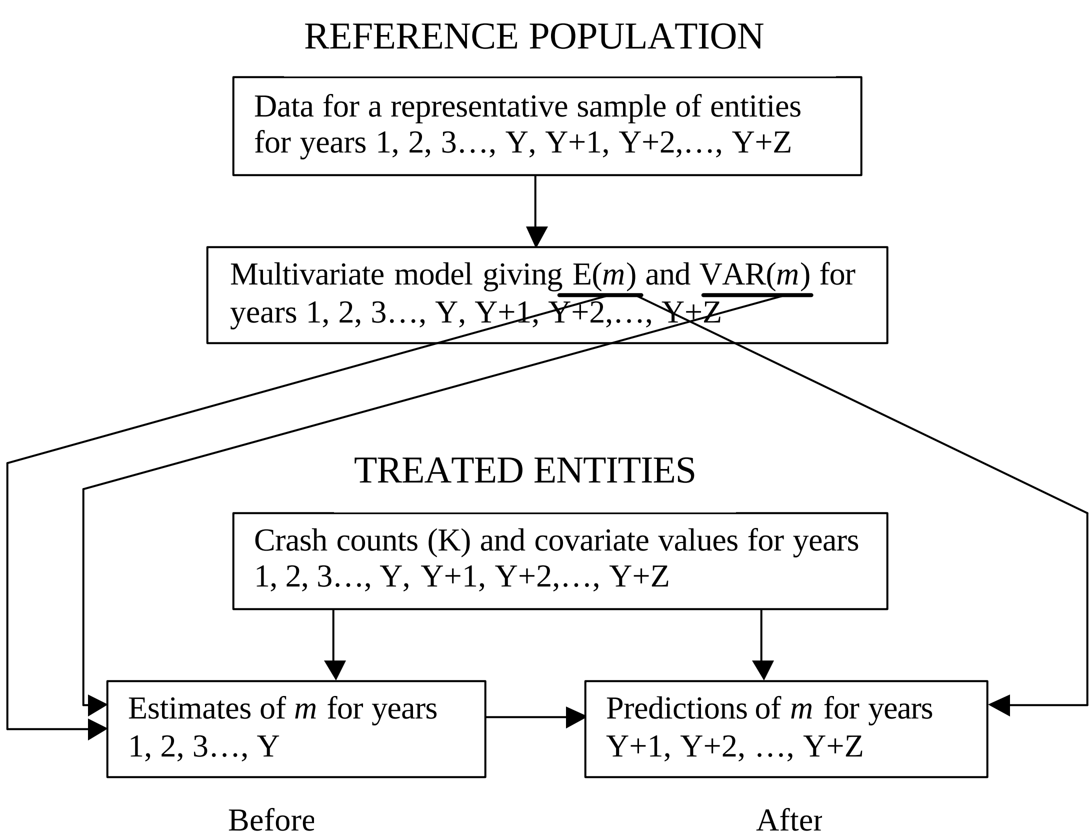


FIGURE 4.1 Basic thoughts of EB method

A treatment was enacted to "n" entities selected in this study. Year Y was the last year before treatment. Z was the number of years after treatment. For each treated entity, there was a sequence of crash counts K for a time series from year 1 to year Y. With the evidence from theses time series of annual crash counts, an estimate of crash count for each year noted as "*m*" in figure 4.1 was made.

To evaluate how this treatment affected safety, we need to predict what the expected crash frequency "*m*" would have been in the after period (year Y+1, Y+2, …Y+Z) if there had been no such treatment. To predict well, a fairly long time series of estimated crash counts *m* should be used.

Evaluation can be then made by comparing this prediction with the "what was" crash frequency of the after period with the treatment. How does EB method do this prediction and comparison?

i. A reference group of entities which are similar to the treated entity were chosen.

ii. Using the data of this reference group from year 1 to Y+Z, a multivariate model was developed to estimate the mean E(*m*) and the variance VAR(*m*) of the expected crash frequency *m* for the whole study period (i.e., before and after periods).

iii. For the treated entity with covariate values available, the multivariate model was applied to calculate the E(*m*) and VAR(*m*) for the before and after years.

iv. The expected crash frequency *m* was then calculated from the E(*m*) under the condition of a crash history of accounts (K) for the before years.

v. Finally, these *m's* of the before years then serve as a basis to get the predictions of the *m's* of the after years of treated entities, with the use of the multivariate model.

4.2 DEVELOPMENT OF CEM

4.2.1 REFERENCE GROUP

The function of multivariate Crash Estimation Models (CEM) in EB method is to estimate the mean of the expected frequency of crashes $E(m_{i,y})$.

To develop the multivariate models, we had to provide a reference group of entities with data from year 1 to year Y+Z, which are similar to the treated sites that we have selected. However, for many safety studies, like our study, this was not that easy. The treatment of our study are the enactments or removals of DSL policy on interstate highways. Hence, our treated entities are those interstate highway sections with a DSL treatment or otherwise. This before-after study is conducted within each state separately. For example, DSL 104.59/88.5 km/h (65/55 mi/h) was changed to UNI 104.59/104.59 km/h (65/65 mi/h) in 1994 on all VA rural interstate highways. It's almost impossible to find a reference group for the VA rural interstate highway sections because the treatment of DSL removal was for all the rural interstates in VA, resulting in no exact reference groups in VA that can be found to consist of rural interstates without the application of this treatment. The solution that was applied was to choose the before years data of these treated sites as a reference group to develop the CEM. An alternative solution would have been to use some states without such treatment as approximate reference groups, however, given the diversity among states, this alternative was not feasible. That is, the states had significant variation in terms of geometry, weather conditions, and terrain conditions.

Thus, we selected the data of the before period of the treated entities as a reference group. There are advantages and disadvantages of this choice. The advantage of using the treated entities of the before period data is that, this way, there is no approximation between the studied entities and the comparison group. The disadvantage is that we didn't have any after data of the treated entities if there had been no treatment. This way, the application of CEM which is developed from the before data of treated entities to the after data makes an assumption that the CEM can represent both the before and would have been after periods very well. This may not be the case because as time passes, many other changes of the after period years may not be reflected in the CEM. No test was run to see how much this assumption would affect the analysis results. This problem of not having an ideal reference group with both before and after data for this study, however, we would rather choose the compromise of using treated entities before data as a reference group rather than jumping to any other untreated states which seems to be in someway similar to the treated state. Once the reference group is chosen, we can start to develop the CEM. There are two things to be considered in the CEM development:

?? The choice of model form (CEM equation)

?? The estimation of model parameters

4.2.2 MODEL FORM

According to previous studies, the expected crash frequencies are related to some causal factors in a certain way. The CEM is an equation which defines the way in which the mean of the expected crash frequency is a function of the various covariates that represent the causal factors [8]. If all the causal factors could be included in the model, the model would reflect the most accurate relationship between the occurrence of crashes and contributing factors. But models should be simple and direct for a better understanding of the relationships. Also some of the factors that could be used can't be included because there was not enough data. Geometry is a causal factor; empirically, there tend to be more crashes on longer sections given the same other factors. It's usually assumed that crash counts are proportional to the section lengths. This, however, may not be true to all situations. Brown, H.C. and Mountain, L. have conducted studies to investigate this relationship [16,17]. For our study, we didn't make any assumption of this relationship. We wanted the data to show us whether there was any proportional relationship or not. Length was therefore considered as one contributing variable with a power exponent in the model equation. AADT was also included as an contributing variable in the equation because it's a major contributor to the crash occurrences. A power exponent for the AADT was also used in the model equation.

Since we are studying the issue of differential speed limit, it would have been beneficial to include the mean speed and the 85th speed. However, as we stated in chapter three, the speed data we obtained were enough to compare over speeds, but not adequate for crash estimation models which is used for each section in each year in this analysis. So speed data were not covered in the models. The following CEM form is taken for this study:

$$E(m)= ?_y (Length)^{?_1} (ADT)^{?_2} \tag{4.2}$$

Where,

m = the estimated annual crash numbers of one site

E{m} = the mean of the estimated annual crash numbers of one site

Length = the length of the section where the crash data are obtained

ADT = Average Daily Traffic of that segment analyzed

? y, ß1 , ß2 = parameters

The section length, ADT, Mean speed, 85%th speed are all covariates that contribute to the annual crash frequencies (K) happening on a particular section and have a certain kind of relationship with the crash numbers as expressed in the model equation by the parameters $?_y$, $ß_1$, $ß_2$, $ß_3$, $ß_4$.

4.2.3 PARAMETER ESTIMATION

The main task in this section was to estimate the parameters in the equation. Maximum likelihood method suggested by Ezra Hauer (1997) was used in the regression analysis of parameter estimation. Dominique Lord and Bhagwant Persaud (2000) [15] have studied to see why traditional ordinary least squares or weighted least squares regression methods can not be employed in the CEM parameter estimation. Two major reasons were found: First the discrete, non-negative nature of crash counts violated the assumptions of the least square regression methods; second, the variance in the number of crashes increases as the traffic flow increases also doesn't fit the assumption of the least square regression model. Details of this issue can be explored in Dominique Lord (2000).

The software Genstat 5, Release 4.21 was selected to estimate the parameters of CEM with the NB error structure. Two built-in procedures were considered. One is the Generalized Linear Model (GLM) (Dunlop) which uses a variant of the Newton-Raphson method to estimate parameters. The other is the Generalized Estimating Equation (GEE) which is classified as a multinomial analogue of a quasi-likelihood function (Liang and Zeger). Readers of interest are recommended to read the paper of Dominique Lord (2000) for details on these two methods. In this thesis, only the main similarities and differences are stated.

Similarities between these two procedures are:

?? Both procedures take a linear transformation of the CEM:

$$\mathrm{Ln}(E(m)) = \mathrm{Ln}(?_{y}) + ß_1 \mathrm{Ln}(Length) + ß_2 \mathrm{Ln}(ADT) \quad (4.3)$$

Where,

E{m} = the mean of the estimated annual crash numbers of one site

Length = the length of the section where the crash data are obtained

ADT = Average Daily Traffic of that segment analyzed

? y, $ß_1$, $ß_2$ = parameters

With the input of the crash frequency K as the response variable and the causal factors of Length, ADT as the contributing variables for each site each year, a link function of logarithm can be defined for the response variable to fit in the linear version by both procedures.

?? Also, the NB distribution can be used in both procedures. The Link function and flexibility in defining the distribution of raw data are the two major advantages of Genstat used in safety study.

There is also a big difference between the GEE and GLM packages as they are marked commercially. GEE procedure considers the temporal correlation that exists between the repeated crash counts year by year, while conventional GLM doesn't. The problem of temporal correlation exists when the model has different ? values for different years. To estimate these different ? values, each annual crash count is input as an observation. But these annual counts are correlated year by year. Each year has influence on the estimation of the ? values of other years. Thus this causes a problem of temporal correlation in the parameter estimation. It's not that the GLM mathematics can't consider the correlation, but the correlation structure is too complicated for the traditional GLM software. So with this correlation ignored by GLM, the variance of the parameters will be underestimated, resulting smaller standard errors of parameters. This may cause the improper selection of parameters because some parameters maybe wrongly accepted as significant with the underestimated variance. To overcome this difficulty, a GEE procedure proposed by Liang and Zeger was adopted in Genstat.

Another issue need to be addressed in the parameter estimation is that whether to develop different CEMs for different years or to develop only one general CEM that can be used for any year? This is called an issue of yearly trend of CEM by Dominique Lord (2000)[15]. Logically, the *ß's* are assumed to be constant from year to year. So the yearly trend of CEM here means do we estimate different *?'s* ($?_y$) for each year or do we keep a constant ? from year to year? For our study (analyzing each state separately), it seems no need to bother considering yearly trend because we didn't have a reference group for after-years data of treated entities. But to get a clear understanding of both procedures and for further study, parameter estimations with or without yearly trend were both considered.

So for both GLM and GEE procedures, two kinds of CEMs were developed. Totally, there are four types of CEM with model 1 and model 2 based on equation 4.4 and model 3 and model 4 based on equation 4.5:

?? Model 1--- GEE with yearly trend

?? Model 2--- GLM with yearly trend

$$E(m)= ?_y (Length)^{?}{}_1 (ADT)^{?}{}_2 \tag{4.4}$$

Where,

m = the estimated annual crash numbers of one site

E{m} = the mean of the estimated annual crash numbers of one site

Length = the length of the section where the crash data are obtained

ADT = Average Daily Traffic of that segment analyzed

$?_y$, $ß_1$, $ß_2$ = parameters

?? Model 3--- GEE without yearly trend

?? Model 4--- GLM without yearly trend

$$E(m)= ?\ (Length)^{?}{}_1 (ADT)^{?}{}_2 \tag{4.5}$$

Where,

m = the estimated annual crash numbers of one site

E{m} = the mean of the estimated annual crash numbers of one site

Length = the length of the section where the crash data are obtained

ADT = Average Daily Traffic of that segment analyzed

$?_y$, $ß_1$, $ß_2$ = parameters

All of these four models use the yearly data as input. In cases were the yearly data unavailable, the average data over the years of the before period were used in the GLM to estimate the parameters. This is model 5 and it's based on equation 4.6.

$$Ave\_E(m)= ?\ (Length)^{?}{}_{1}\ (ave\_ADT)^{?}{}_{2} \tag{4.6}$$

Where,

m = the estimated annual crash numbers of one site

ave_E{m} = the average of the means of the estimated annual crash numbers of one site over the period

Length = -the length of the section where the crash data are obtained

Ave_ADT = the average of the Average Daily Traffic of that segment analyzed over the period

? , $ß_1$, $ß_2$ = parameters

The drawback of model 5 is that it didn't reflect the variation of ADT and speed from year to year.

A total of five different models were developed and compared for some states in this study, and then one or two models were chosen for further analysis.

As we mentioned above in section 4.1, the CEM not only gives the mean of the expected crash frequency E(m), but also the variance VAR(m) of expected crash frequency for each section. And with the NB distribution of error structure, there exists the following relationship between the two:

$$\mathrm{VAR}(m)=[\ \mathrm{E}(m)]^2/k \tag{4.7}$$

Where,

VAR(m) = the variance of the expected crash frequency

E(m) = the mean of the expected crash frequency

k = aggregation parameter of NB distribution

There are different techniques to find the aggregation parameter k of a NB distribution. This study utilized a program provided by Persaud and Dominique, using maximum likelihood technique based on the following equations (1988, 1992) [10,11,18]:

$$VAR(K)=E(m)+[E(m)]^2/k \tag{4.8}$$

Where,

K = the actual crash frequency

VAR(K) = the variance of the actual crash frequency

E(m) = the mean of the expected crash frequency

k = aggregation parameter of NB distribution

VRA(K) is estimated by the squared residual, i.e. $(K-E(m))^2$. Details can be found in literature [10,11,15,18].

Equation 4.8 showed an interesting point of data distribution assumption. In our analysis, we assumed a Negative Binomial distributed structure of expected crash counts between sites. But as aggregation parameter k goes to infinity, VAR(K)=E(m), which indicates the Poisson distribution. This means that Poisson distribution is a special type of Negative Binomial Distribution.

The aggregation parameter k is calculated in an iteration process with a maximum likelihood program provided by Persaud and Dominique [15]. First, an initial value was given to the aggregation parameter k and the parameters were estimated by Genstat. The output file from the regression including the actual crash frequency K and the estimated frequency E(m) is then used as an input into the maximum likelihood program. A new value of aggregation parameter k is computed. This new aggregation parameter k is then fed back into Genstat for the second running of estimation. The iteration ends when the assumed aggregation parameter k converges to the same value.

The five types of models were developed for the total number of crashes using the Virginia and Arizona data. The results are shown in Tables 4.1 and 4.2 with the aggregation parameter k included in the last row.

Note that the comparison of GLM and GEE was only for the phase of model parameter estimation. Different models were compared for the purpose of choosing a best model that fits our data and our actual analysis situation. There is no comparison of GLM and GEE during the phase of prediction.

Table 4.1 Five Models for Total Number of Crashes on VA rural interstate highways

# of entities: 267										
Before years of crashes: 1991, 1992, 1993										
Parameters	Model 1		Model 2		Model 3		Model 4		Model 5	
	GEE with trend		GLM with trend		GEE without trend		GLM without trend		GLM without trend	
	yearly data		yearly data		yearly data		yearly data		3-year averaged data	
	E(m)= ? y (Length)? 1 (ADT)? 2		E(m)= ? y (Length)? 1 (ADT)? 2		E(m)= ? (Length)? 1 (ADT)? 2		E(m)= ? (Length)? 1 (ADT)? 2		ave_E(m)= ? (Length)? 1 (ave_ADT)? 2	
	estimate	Standard error	estimate	Standard error	estimate	Standard error	estimate	Standard error	estimate	Standard error
LN(? 1)	-3.774	0.5271	-3.774	0.41						
LN(? 2)	-3.848	0.5277	-3.848	0.415						
LN(? 3)	-3.732	0.528	-3.732	0.416						
LN(?)?	-3.78467	0.5276	-3.78467	0.413667	-3.828	0.5274	-3.829	0.41	-3.791	0.592
? ? ?	0.631	0.1167	0.6309	0.0526	0.632	0.1163	0.6323	0.0528	0.6117	0.0737
? ? ?	0.545	0.0588	0.5447	0.042	0.549	0.0585	0.5492	0.0418	0.5481	0.0607
k	5.62		5.62		5.61		5.61		5.56	

LN((? y): logarithm of ? parameter for year y
LN((?): average of all the LN((? y)

Table 4.2 Five Models for Total Number of Crashes on AZ rural interstate

# of entities: 556										
years of crashes: 1991-2000										
Parameters	Model 1		Model 2		Model 3		Model 4		Model 5	
	GEE with trend		GLM with trend		GEE without trend		GLM without trend		GLM without trend	
	yearly data(5560pts)		yearly data(5560pts)		yearly data(5560pts)		yearly data(5560pts)		10-year data(556pts)	
	$E(m)= ?_y (Length)^{?_1} (ADT)^{?_2}$		$E(m)= ?_y (Length)^{?_1} (ADT)^{?_2}$		$E(m)= ? (Length)^{?_1} (ADT)^{?_2}$		$E(m)= ? (Length)^{?_1} (ADT)^{?_2}$		$E(m)= ? (Length)^{?_1} (ADT)^{?_2}$	
	estimate	Standard error	estimate	Standard error	estimate	Standard error	estimate	Standard error	estimate	Standard error
$LN((?_1)$	-1.963	0.6397	-1.963	0.238						
$LN(?_2)$	-1.933	0.6373	-1.934	0.238						
$LN(?_3)$	-1.908	0.6381	-1.908	0.238						
$LN(?_4)$	-1.819	0.6381	-1.819	0.24						
$LN(?_5)$	-1.853	0.6434	-1.853	0.241						
$LN(?_6)$	-1.808	0.6511	-1.809	0.243						
$LN(?_7)$	-1.687	0.6506	-1.687	0.243						
$LN(?_8)$	-1.621	0.6493	-1.621	0.244						
$LN(?_9)$	-1.649	0.6523	-1.649	0.246						
$LN(?_{10})$	-1.657	0.6572	-1.657	0.249						
$LN(?)$	-1.7898	0.64571	-1.79	0.242	-2.328	0.6196	-2.33	0.233	-1.68	0.831
$?_?$	1.114	0.0609	1.1144	0.0242	1.133	0.0606	1.1325	0.0242	1.151	0.0826
$?_?$	0.24	0.0668	0.2403	0.024	0.297	0.0639	0.2974	0.0234	0.2242	0.0842
k	0.73		0.73		0.73		0.73		0.73	

$LN((?_y)$: logarithm of ? parameter for year y
$LN((?)$: average of all the $LN((?_y)$

In each table, the parameters for the GEE model 1 and GLM model 2 are the same and the parameters for model 3 and model 4 are also the same. This is because the same data were used as input for regression. But the parameter standard errors for Models 1 and 3 are much bigger than those of Models 2 and 4. This is consistent with the fact that GLM ignores the correlation between repeated measures which causes underestimate of the standard errors.

Test of the probability significance of estimated parameters was also given out in the process of Genstat regression for five models in the term of student-*t* values. All the *t* values for parameters in different models were much bigger than the critical *t* values according to corresponding degree of freedom, showing that the parameters are significant at a 5% significant level.

Since model 1 and model 2 are not suitable for our study and model 5 doesn't consider the yearly difference in the data, it seems that we may use Model 3 or Model 4. Based on the fact that Model 4 ignores the correlation between the repeated measurements, model 3 seems the ideal model for our analysis for states with multiple before years data. But if the before year data was only from one before year instead of multiple before years, there is no problem of yearly correlation, GLM model 4 is then used.

4.2.4 GOODNESS OF MODEL FIT

There are many ways to assess the quality of model fit. The cumulative residual method was used in this study [15]. Interested readers are referred to read the paper by Dominique Lord [15] for more discussions and details. The cumulative residual is the difference between actual crash counts and model estimated crash counts. The cumulative residuals for all the sites are plotted with the two plots of the two standard deviations (positive and minus). If the cumulative residuals oscillate around 0 within the range of the two plots of two standard deviations, a good quality of fit is reflected. The advantage of this method is that it doesn't depend on the number of observations as many traditional statistical procedures do. The validation of goodness of fit has been run for the VA total crashes for GLM model 3. Figure 4.2 shows the cumulative residual plots with respect to section length and AADT. Although there existed several sites that the cumulative residuals exceed the range of two standard deviations, the overall figures showed a good quality of fittness. The reason that there exist some outliers can be that we use one common ? parameter for different years while in fact there maybe a slight difference in ? values the from year to year. And also, estimation of those missing AADT may not reflect the fact very well. But overall, the CEM (GLM model 3 we choose as out model form) is a good reflection of reference population data.

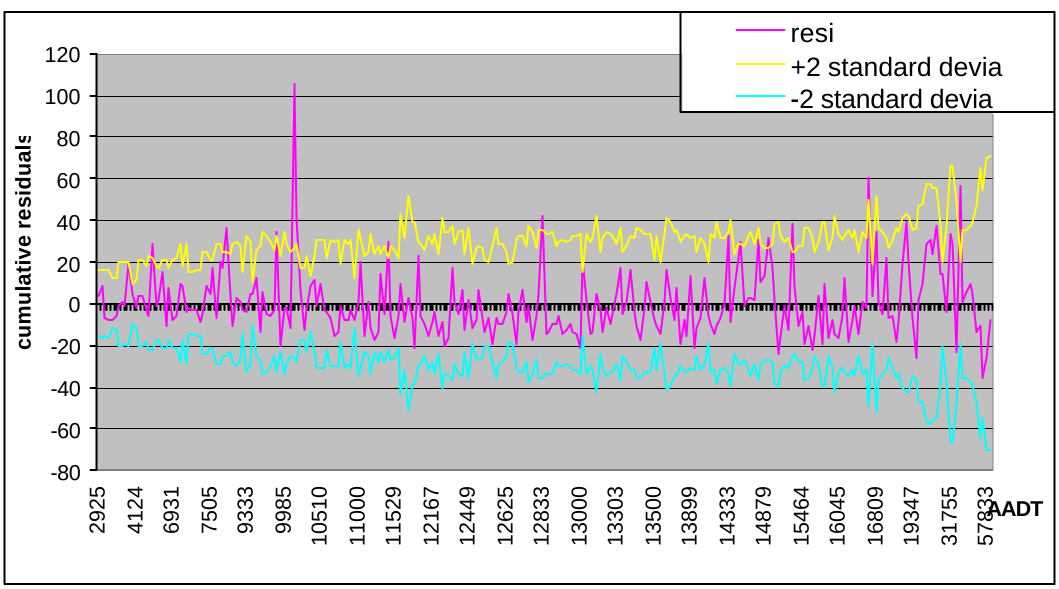


Figure 4.2 a) Cumulative residuals with respect to AADT

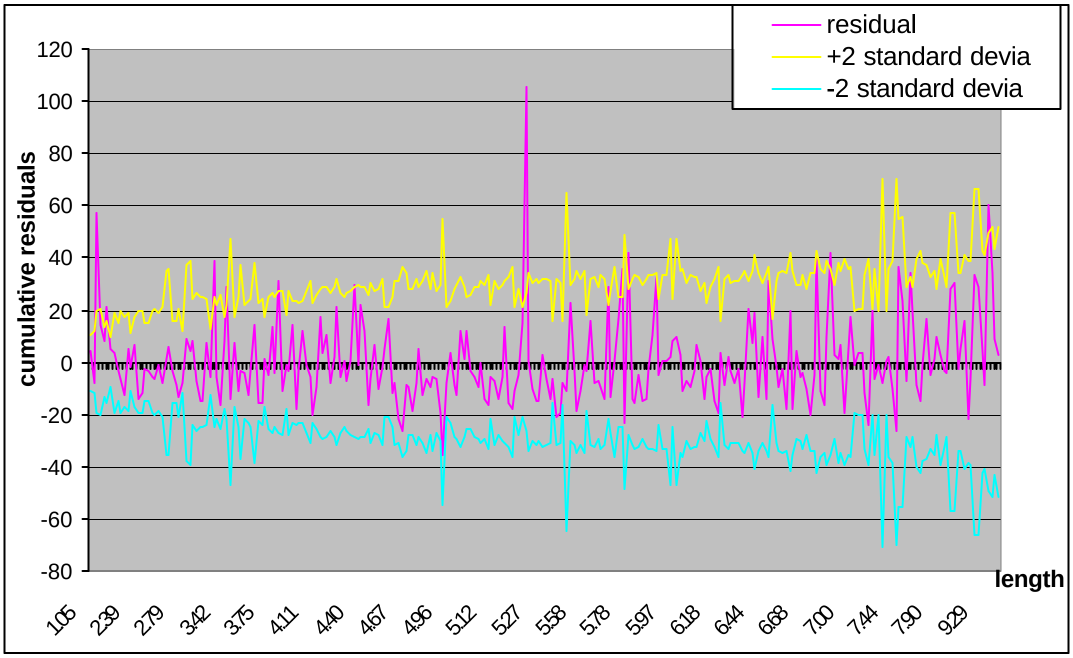


Figure 4.2 b) Cumulative residuals with respect to section length

4.3 ESTIMATION OF EXPECTED CRASH FREQUENCY m_1, m_2,… m_y

Before we go any further, we must understand why we are doing another estimation after we have developed the model.

Consider two sites that are exactly the same in all the contributing factors used in the CEM, their true crash counts can still be different, because the CEM couldn't account for all the factors that cause the difference in the crash potential [13]. The approach to account for this shortfall is to refine the E(m) by using actual crash counts as a condition to get a final estimate of crashes E(m|K). Following is the detailed steps of this refinement.

After the CEM was developed, we estimated the expected crash frequency m_1, m_2,.. m_y for the treated entities. First, the developed CEM was applied to all the before years data of each treated entity to get E($m_{i,y}$). Then, the ratio $C_{i,y}$ were computed using equation (4.1) for year y=1, 2, ….Y.

$$C_{i,y} = \frac{E(m_{i,y})}{E(m_{i,1})} \qquad y=1, 2, \ldots Y$$

The expected crash frequency $m_{i,y}$ and its variance VAR($m_{i,y}$) for years of y=1, 2, ….Y are then calculated by the following equations:

$$m_{i,1} = \frac{k + \sum_{y=1}^{Y} K_{i,y}}{\frac{k}{E(m_{i,1})} + \sum_{y=1}^{Y} C_{i,y}} \tag{4.9}$$

$$VAR(m_{i,1}) = \frac{k + \sum_{y=1}^{Y} K_{i,y}}{\left(\frac{k}{E(m_{i,1})} + \sum_{y=1}^{Y} C_{i,y}\right)^2} + \frac{m_{i,1}}{\frac{k}{E(m_{i,1})} + \sum_{y=1}^{Y} C_{i,y}} \tag{4.10}$$

$$m_{i,y} = C_{i,y}\, m_{i,1} \tag{4.11}$$

$$VAR(m_{i,y}) = C_{i,y}^2\, VAR(m_{i,1}) \tag{4.12}$$

Where,

$m_{i,1}$ = the estimated crash frequency on site i in the first year

VAR($m_{i,1}$) = the variance of the estimated crash frequency on site i in the first year

$m_{i,y}$ = the estimated crash frequency on site i in year y(y=1,…,Y)

VAR($m_{i,y}$) = the variance of the estimated crash frequency on site i in year y (y=1,…,Y)

k = aggregation parameter of NB distribution estimated with the CEM

$K_{i,y}$ = actual crash counts on site site i in year y (y=1,…,Y)

$E(m_{i,1})$ = the mean of the estimated crash frequency of site i in the first year

$C_{i,y}$ = yearly changing ratio of site i in year y(y=1,…,Y)

Y = the last year before treatment

4.4 PREDICTION OF m_{Y+1}, m_{Y+2},…m_{Y+Z}

With the expected crash frequency $m_1, m_2, \ldots m_y$ estimated for before period for each treated entity, we now are ready to predict the "would have been crash frequency" $m_{Y+1}, m_{Y+2}, \ldots m_{Y+Z}$ for the after period if there had been no such treatment.

First, the developed CEM was applied to all the after years data of each treated entity to get $E(m_{i,y})$ (y=Y+1, Y+2, …Y+Z). Then, the ratios $C_{i,y}$ are be computed using equation (4.1) for y=Y+1, Y+2, …Y+Z.

$$C_{i,y} ? \frac{E(m_{i,y})}{E(m_{i,1})} \qquad y=Y+1, Y+2, \ldots Y+Z.$$

The $m_{i,y}$ and it's variance $VAR(m_{i,y})$ for the years y= Y+1, Y+2, …Y+Z were then computed using the following equations:

$$m_{i,y} = C_{i,y}\, m_{i,1} \tag{4.13}$$

$$VAR(m_{i,y}) = C_{i,y}^{\ 2}\, VAR(m_{i,1}) \tag{4.14}$$

Where,

$m_{i,y}$ = the estimated crash frequency on site i in year y(y=Y+1,…,Y+Z)

$VAR(m_{i,y})$ = the variance of the estimated crash frequency on site i in year y (y=Y+1,…,Y+Z)

$C_{i,\,y}$ = yearly changing ratio of site *i* in year *y*(y=Y+1,…,Y+Z)

$m_{i,1}$ = the estimated crash frequency on site *i* in the first year

VAR($m_{i,1}$) = the variance of the estimated crash frequency on site *i* in the first year

4.5 EVALUATION OF SEFETY EFFECT

The safety effect of treatment was evaluated from the change between the "what would have been" crash counts and the "what was" crash counts over a period of time in the after period.

Clearly, the "what would have been" crash frequency is the predicted value $m_{i,y}$ we had calculated so far. However, there can be disagreements as to the "what was" crash frequency. To most people, especially engineers in the field, the "what was" crash frequency is apparently the actual crash number that is reported for the after period. But statistically, the "estimated" crash frequency calculated from the after data model would be more meaningful than just the random "actual" crash frequency. That is, with the actual crash data of the after period, develop a CEM to get the mean of the expected after crash frequency, then calculate the estimated after crash frequency. This estimated after crash frequency is then used as the "what was" after crash counts.

In our study, the actual crashes were used as the "what was" after crash counts. The results will be shown in chapter six.

Next, we'll talk about the index of change to evaluate the safety effect. Following Hauer's conclusion, let's notate the "would have been" crash number of a period of time during the after period on each site studied as ?$_{i}$, while notate the "what was" crash number on each site of a period of time during the after period as ?$_{i}$

The effect is evaluated on a group of composite sites. So we have to add the sites together. Sum all the ? *'s* and all the ? *'s* over all segments in the treated group (all the treated segments in one state in this report analysis). Notate them as ? and ?.

$$? \ ? \ \underset{i}{?} \ ?_{i} \tag{4.15}$$

$$? \ ? \ \underset{i}{?} \ ?_{i} \tag{4.16}$$

where,

? = the sum of all the "would have been" crash counts over all sites of the studied group for a period of time during the after period

? = the sum of all the "what was" crash counts over all sites of the studied group for a period of time during the after period

$?_{i}$ = the "would have been" crash counts on one site for a period of time during the after period

$?_{i}$ = the "what was" crash counts on one site for a period of time during the after period

Two methods are suggested to be used to evaluate the safety effect by Hauer [8]:

Method 1: Reduction in Expected Number of Crashes (?)

This method gives the difference between the "would have been" after crash data and the "what was" after crash data. Variance is also calculated to show the variation of this difference.

The difference between the sums of the before and after over all sites in a conversion group is notated as ?,

$$? = ? - ? \quad (4.17)$$

Where,

? = the difference between the "would have been" crash counts and the "what was" crash counts of the studied group for a period of time during the after period

? = the sum of all the "would have been" crash counts over all sites of the studied group for a period of time during the after period

? = the sum of all the "what was" crash counts over all sites of the studied group for a period of time during the after period

The variance of ? is given by:

$$Var(?)=Var(?)+Var(?) = ? Var(?_i) + ? Var(?_i) \quad (4.18)$$

Where,

Var(?)= the variance of the difference ?

Var(?)= the variance of ?

Var(?)= the variance of ?

Var(? i) = the variance of ? i

Var(? i) = the variance of ? I

So the empirical confidence bounds on ? are ? ? 2Var(?).

Method 2:Index of Effectiveness (?)

This method gives the ratio of the actual after data to the "would have been" after data. Variance is also calculated to show the variation of this ratio.

$$? = ?/? \qquad (4.19)$$

Where,

? = the ratio of the actual after data to the “would have been” after data over all sites of the studied group of the studied group for a period of time during the after period

? = the sum of all the “would have been” crash counts over all sites of the studied group for a period of time during the after period

? = the sum of all the “what was” crash counts over all sites of the studied group for a period of time during the after period

The percent change in crashes is 100*(1-?) which indicates the percentage of reduction in crashes. Hauer (1997) gives the unbiased estimate of ? as following:

$$?=(?/?)/\{1+Var(?)/?^{2}\} \qquad (4.20)$$

The variance of ? is given by:

$$Var(?)=?^{2}\{[var(?)/?^{2}]+[var(?)/?^{2}]\}/[1+var(?)/?^{2}]^{2} \qquad (4.21)$$

Where,

Var(?) = the variance of the ratio ?

Var(?) = the variance of ?

Var(?)= the variance of ?

? = the ratio of the actual after data to the would have been after data over all sites of the studied group of the studied group for a period of time during the after period

? = the sum of all the "would have been" crash counts over all sites of the studied group for a period of time during the after period

? = the sum of all the "what was" crash counts over all sites of the studied group for a period of time during the after period

So the empirical confidence bounds on ? are ? ? 2Var(?).

4.6 SUMMARY

This chapter gives a general description of the EB method. The assumption of NB distribution of the crash counts between different sites was evaluated and the results showed that NB distribution describes our crash counts data better than Poisson distribution. The two procedures GLM and GEE that were used in CEM parameter estimation were stated and compared using our data. In the next chapter, the EB method is applied to data of different states and the results will be shown. Also a detailed example of EB method will be illustrated.

CHAPTER 5

EXAMPLE APPLICATION OF THE EB METHOD

In Chapter four we talked about EB method. This chapter illustrates the use of the EB method with obtained data. A basic example was shown to go through the procedure step by step.

5.1 DEVELOPMENT OF CEM

As we said in chapter four, in theory, we would include as many causal factors as we can in the CEM. But in reality, this depends heavily on the data availability. We tried to collect speed data along with crash data. The speed data obtained were only adequate for the speed analysis, but couldn't be used in the EB method since the data was not available for each highway section in analysis. For example, an interstate section might have speed monitoring sites, which was enough to compare overall speeds but not enough to use in a crash estimation model. Thus we use length and ADT as the two contributing variables in the CEM:

$$E(m) = ?\ (Length)^{?}{}_{1}\ (ADT)^{?}{}_{2} \qquad (5.1)$$

Where,

m = the estimated annual crash numbers of one site

$E\{m\}$ = the mean of the estimated annual crash numbers of one site

Length = the length of the section where the crash data were obtained

ADT = Average Daily Traffic of that segment analyzed

$?$, $ß_1$, $ß_2$ = parameters

Let's take VA total crash data as an example. In 1994, the VDOT changed the speed limits on rural interstate highways from 104.59/88.5 km/h (65/55 mi/h) to 104.59/104.59 km/h (65/65 mi/h). The study group we choose consisted of 266 sites on VA rural interstate highways, which consisted of sections on I-85, I-95, I-81, I-64, I-77. The before period was from 1991, 1992, 1993. The after period was from 1995, 1996, 1997 and 1999. We didn't use any data for year 1998 because AADT was not available. We could have estimated the 1998 ADT using the linear assumption of yearly changing trend. But to avoid any misinterpretation of data used for CEM development, we chose to exclude this year.

The regression analysis to estimate the parameters of CEM was done by Genstat 5.1.

The parameters estimated using GLM model 3 with the before year data were as following in Table 5.1:

Table 5.1 CEM Parameters based on VA before data

K	5.9
?	0.02242775
$?_?$	0.62225762
$?_?$	0.54802324

k: the aggregation parameter

So our CEM takes the form of:

$$E(m)= 0.02242775* (Length)^{0.62225} (ADT)^{0.5480}$$

5.2 ESTIMATION OF EXPECTED CRASH FREQUENCY $m_1, m_2, \ldots m_y$ FOR THE BEFORE PERIOD

The CEM was developed using the three-year before data for VA, the next step is to estimate the expected crashes and predict the "would have been" crashes. To show the EB method procedure clearly, one site of VA data set was taken as an example to illustrate how the EB method was used step by step.

This section was located on VA I-64 eastbound from JB-WA/West Virginia state line (0 mile) to IS-64-E-7AR/To RT661 East and West (7.16 mile), with a length of 7.16 mile.

The before and after period data on this site were shown in Tables 5.2 and 5.3:

Table 5.2 Before Years Data

YEAR	ADT	total crash
1991	4000.000	8
1992	4250.000	11
1993	4300.000	7
sum		26

Table 5.3 After Years Data

YEAR	ADT	total crash
1995	4500	8
1996	4650	7
1997	4800	10
1999	4825.56	5
Sum		30

First, use the CEM developed above ($E(m) = 0.02242775 * (Length)^{0.62225}$ $(ADT)^{0.5480}$) to calculate the $E(m_{1,y})$ for each year. The results were shown in Table 5.4

$$E(m_{1,1991}) = 0.02242775 * (7.16)^{0.62225} (4000)^{0.5480} = 7.191$$

$$E(m_{1,1992}) = 0.02242775 * (7.16)^{0.62225} (4250)^{0.5480} = 7.434$$

$$E(m_{1,1993}) = 0.02242775 * (7.16)^{0.62225} (4300)^{0.5480} = 7.481$$

Table 5.4 Estimation Results for the Before Years

YEAR	ADT	total crash	$E(m_{1,y})$	$C1_{,y}$	$m_{1,y}$	$VAR(m_{1,y})$
1991	4000.000	8	7.191	1	8.190599	2.103007
1992	4250.000	11	7.434	1.033782	8.467292	2.247493
1993	4300.000	7	7.481	1.040429	8.521739	2.27649
sum		26		3.074211		

Then, the ratio $C_{i,\ y}$ was calculated using equation (4.1) as following for each before year and the results were shown in the right portion of Table 5.4:

$$C_{i,y} = \frac{E(m_{i,y})}{E(m_{i,1})}$$

$C_{1,1991} = E(m_{1,1991})/ E(m_{1,1991})=1$

$C_{1,1992} = E(m_{1,1992})/ E(m_{1,1991})= 1.033782$

$C_{1,1993} = E(m_{1,1993})/ E(m_{1,1991})= 1.040429$

The next task was to calculate the expected crash counts $m_{i,y}$ on this site for each before year with their variance VAR($m_{i,y}$) using equations (4.9) to (4.12)

$$m_{i,1} = \frac{k + \sum_{y=1}^{Y} K_{i,y}}{\frac{k}{E(m_{i,1})} + \sum_{y=1}^{Y} C_{i,y}} \tag{4.9}$$

$$VAR(m_{i,1}) = \frac{k + \sum_{y=1}^{Y} K_{i,y}}{\left(\frac{k}{E(m_{i,1})} + \sum_{y=1}^{Y} C_{i,y}\right)^2} = \frac{m_{i,1}}{\frac{k}{E(m_{i,1})} + \sum_{y=1}^{Y} C_{i,y}} \tag{4.10}$$

$$m_{i,y} = C_{i,y}\ m_{i,1} \tag{4.11}$$

$$\mathrm{VAR}(m_{i,y}) = C_{i,y}^{\ 2}\ \mathrm{VAR}(m_{i,1}) \tag{4.12}$$

$m_{1,1991} =(5.9+26)/(5.9/7.191+3.074211)= 8.190599$

$VAR(m_{1,1991})=8.190599/(5.9/7.191+3.074211)= 2.103007$

$m_{1,1992} = C_{1,1992}\ m_{1,1991} =1.033782*8.190599=8.467292$

$VAR(m_{1,1992})= C_{1,1992}^{\ 2} * VAR(m_{1,1991})= 1.033782\text{^}2*2.103007=2.247493$

$m_{1,1993} = C_{1,1993}\, m_{1,1991} = 1.040429*8.190599=8.521739$

$VAR(m_{1,1993}) = C_{1,1993}{}^{2} * VAR(m_{1,1991}) = 1.040429\text{^}2*2.103007=2.27649$

The results were shown in Table 5.4

5.3 PREDICTION OF $m_{Y+1}, m_{Y+2},\ldots m_{Y+Z}$ FOR THE AFTER PERIOD

First, the CEM was applied to the after period data to calculate the mean of the expected "would have been" crash frequency for after period years.

$$E(m_{1,1995}) = 0.02242775*(7.16)^{0.62225}\,(4500)^{0.5480} = 7.670$$

$$E(m_{1,1996}) = 0.02242775*(7.16)^{0.62225}\,(4650)^{0.5480} = 7.809$$

The E(m) for the other years was calculated in the same way and the results were shown in Table 5.5.

Table 5.5 Prediction Results for the After Years

YEAR	ADT	Ki,y	$E(m_{1,y})$	$C_{1,y}$	$m_{1,y}$	VAR($m_{1,y}$)
1995	4500	8	7.670	1.066677	8.73672	2.392799
1996	4650	7	7.809	1.086018	8.895134	2.480358
1997	4800	10	7.946	1.105079	9.051255	2.568189
1999	4825.56	5	7.970	1.1083	9.077637	2.583182
sum		30	31.39535	4.366072	35.76075	10.02453
average		7.5	7.849	1.091518	8.940187	2.506132

Note: Ki,y: the "actual" after crashes for year y

$E(m_{1,y})$: the mean of the expected "would have been" after crashes for year y

$C_{1,y}$: the changing ratio for the "would have been" after crashes

$m_{1,y}$: the expected "would have been" after crashes for year y

VAR($m_{1,y}$): the variance of the expected "would have been" after crashes for year y

Secondly, calculate the ratio Ci,y using equation (4.1) and the results were also shown in Table 5.5.

$$C_{1,1995} = E(m_{1,1995}) / E(m_{1,1991}) = 1.066677$$

$$C_{1,1996} = E(m_{1,1996}) / E(m_{1,1991}) = 1.086018$$

Finally, using equation 4.13 and 4.14, we can predict the "would have been"expected crash counts for the after years.

$$m_{i,y} = C_{i,y}\, m_{i,1} \tag{4.13}$$

$$\mathrm{VAR}(m_{i,y}) = C_{i,y}^{\ 2}\, \mathrm{VAR}(m_{i,1}) \tag{4.14}$$

$m_{1,1995} = C_{1,1995}\, m_{1,1991} =$1.066677*8.190599=8.73672

VAR($m_{1,1995}$)= $C_{1,1995}^{\ 2}$* VAR($m_{1,1991}$)= 1.066677^2*2.103007=2.392799

$m_{1,1996} = C_{1,1996}\, m_{1,1991} =$1.086018*8.190599=8.895134

VAR($m_{1,1996}$)= $C_{1,1996}^{\ 2}$* VAR($m_{1,1991}$)= 1.086018^2*2.103007=2.480358

The results were shown in Table 5.5

5.4 EVALUATION OF SAFETY EFFECT ON THIS PARTICULAR SITE OF I-64

The effect of treatment was evaluated by comparing the actual after crashes Ki,y with the predicted after crashes $m_{1,y}$ for each year of the after period and the cumulative difference was also calculated in Table 5.6.

Table 5.6 Evaluation for the example site

YEAR	Ki,y	Cumulative $K_{1,y}$	$m_{1,y}$	Cumulative $m_{1,y}$	VAR($m_{1,y}$)	*Excess* $m_{1,y}$? Ki,y	Cumulative excess	Ki,y /$m_{1,y}$

1995	8	8	8.73672	8.73672	2.392799	0.73672	0.73672	0.915675
1996	7	15	8.895134	17.63185	2.480358	1.895134	2.631854	0.786947
1997	10	25	9.051255	26.68311	2.568189	-0.94874	1.683114	1.104819
1999	5	30	9.077637	35.76075	2.583182	4.077637	5.760751	0.550804
average	7.5		8.940187		2.506132	1.440187		0.839561

Note: Ki,y: the "actual" after crashes for year y

$E(m_{1,y})$:the mean of the expected "would have been" after crashes for year y

$C_{1,y}$: the changing ratio for the "would have been" after crashes

$m_{1,y}$: the predicted "would have been" after crashes for year y

VAR($m_{1,y}$):the variance of the expected "would have been" after crashes for year y

For the whole four-year after period, we can see the effect by calculating the difference and the ratio.

$$?_i = ? \; m_{1,y} = 8.73672+8.895134+9.051255+9.077637=35.76075$$

$$?_i = ? \; K_{1,y} = 8+7+10+5=30$$

$$?_i - ?_i = 35.76075 - 30 = 5.76075$$

$$?_i / ?_i = 30/35.76075 = 0.838909$$

Apparently, over the whole four-year after period, the actual crash number at this site on I- 64 (?) was 5.76 less than the predicted crash number (?) if there had been no UNI speed limit change. This difference means that with the speed limit changed to UNI from DSL, there was a decrease of 16.11% (1 –0.838909 = 16.11%) of the actual crash counts than "would have been" crash counts if it had remained DSL on this site.

The yearly change was shown in Table 5.6 and plotted in Figure 5.1. The four-year after period cumulative change can be seen in Table 5.6 and plotted in Figure 5.2.

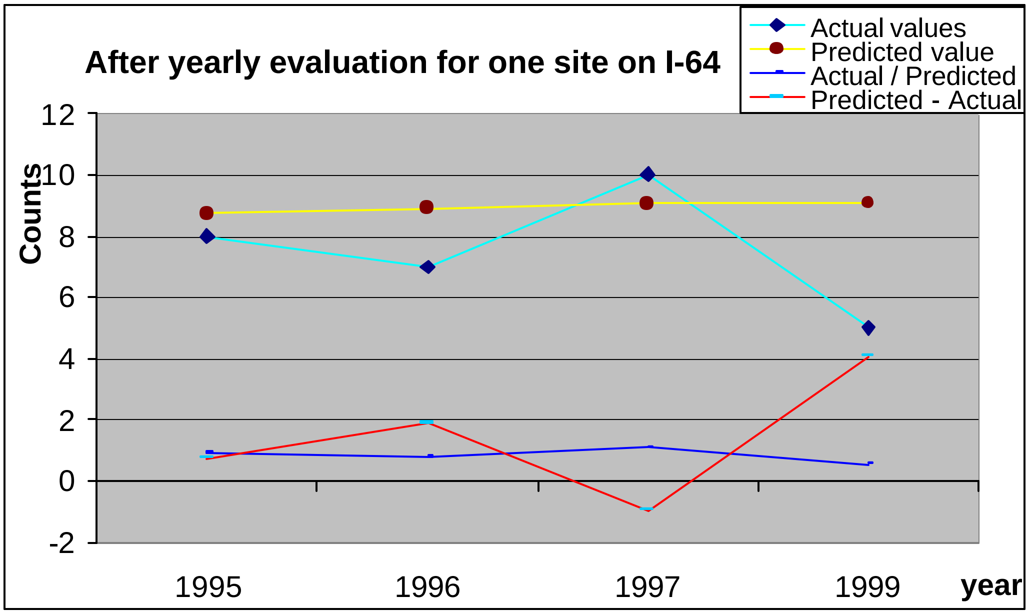


Figure 5.1 Yearly change for the example site

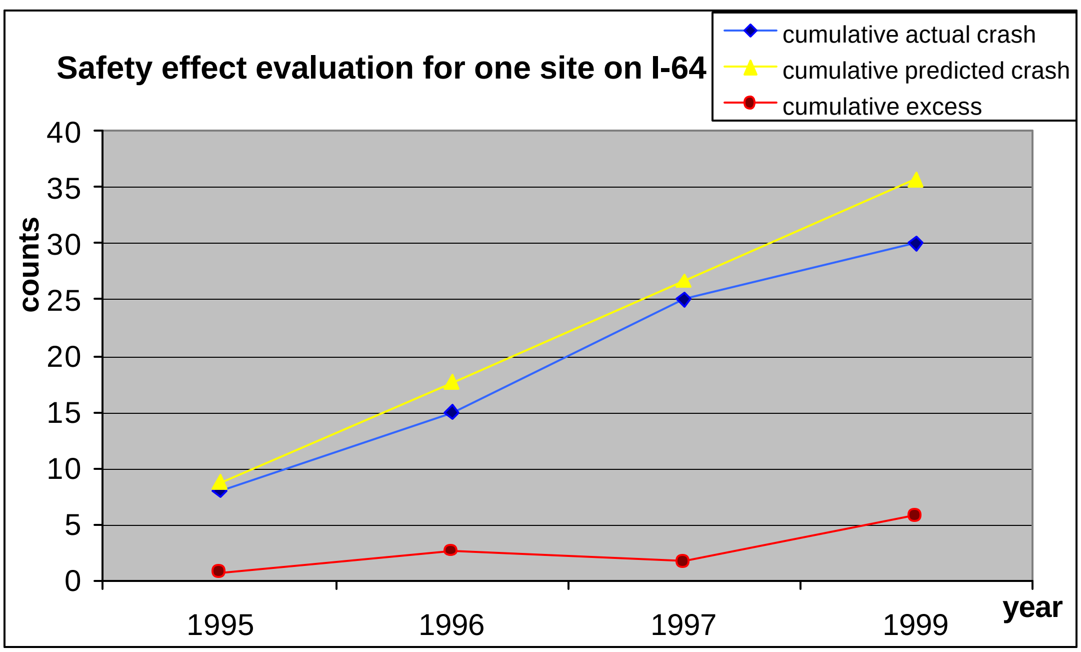


Figure 5.2 Four-year after period cumulative change for the example site

5.5 EVALUATION OF SAFETY EFFECT ON THE WHOLE GROUP

As we said in chapter four, the safety effect was evaluated by comparing the "what was" and "what would have been" crash counts of the after treatment period on the whole group studied. The example shown above was only one section of the whole group studied that consisted of 266 sections on 5 different interstates in VA. So to evaluate the safety effect of the removal of DSL in 1994 on VA interstates highways, all the 266 sections were used to compare the actual crashes and the "would have been crashes" of the after years. Following shows the two methods of evaluation. Detailed results will be shown in next chapter.

Reduction in Expected Number of Crashes (?):

$$?\ ?\ \underset{i}{?}\ ?_{\mathrm{i}}\ ?\ 13365.91$$

$$?\ ?\ \underset{i}{?}\ ?_{\mathrm{i}}\ ?\ 15377$$

$$? = ? - ? = 13365.91 – 15377 = -2011.09$$

The variance was calculated:

$$\mathrm{Var}(?) = \mathrm{Var}(?) + \mathrm{Var}(?) = ?\,\mathrm{Var}(?\ i) + ?\,\mathrm{Var}(?\ i) = 4246.913 + 15377 = 19623.91$$

The standard deviation was then:

$$?(?) = \{\mathrm{Var}(?)\}\text{^}0.5 = 140.0854$$

So we can say that for the four- year after period, the actual crash counts on the 266 sites (1424.32 miles totally) in VA was 2011.09 ? 2*140.0854 more than the predicted crash counts if there was no UNI enacted in 1994.

Index of Effectiveness (?):

$?=(?/?)/\{1+Var(?)/?^{2}\}$= (15377/13365.91)/(1+4246.913/13365.91^2)= 1.150437

The variance of ? was given by:

$Var(?) =?^{2}\{[var(?)/?^{2}]+[\ var(?)/?^{2}]\}/[1+ var(?)/?^{2}]^{2}$

=1.150437^2*{15377^2/15377+4246.913/13365.91^2}/[1+4246.913/13365.91^2]^2

=0.000118

So, the actual crash counts of the four after years was 15% ? 2*0.000118 more than the predicted crash counts. Assuming all the other factors remained constant except for the change of speed limit from DSL to Uniform, we refer from the ? being larger than 1.0 that this change from DSL to Uniform increased the total number of crashes on VA rural interstates.

5.6 SUMMARY

This chapter shows how to use EB method and to evaluate the effect through a particular site, as an example, step by step. In chapter six, the application of EB method to the different states, different types of crashes were shown with the results and summary.

Note that one drawback to this example - and to our method overall - is that the reference group (for the "would have been" after crashes) presumes the same temporal trend occurred during the before period.

CHAPTER 6

CALCULATION AND RESULTS OF DIFFERENT STATES

As noted in chapter one, there are four groups of states based on their speed limits changing policy in the 90's. The EB method was applied to the data for the following seven states: Virginia, Arkansas, Idaho, Arizona, Missouri, North Carolina and Washington. This chapter shows the EB method application and results of evaluation for all the seven states analyzed in this study.

6.1 VIRGINIA (DSL to UNI)

6.1.1 DATA CHARACTERISTICS

As we said in chapter five, in VA, the DSL104.59/88.5 km/h (65/55 mi/h) was changed to UNI 104.59/104.59 km/h (65/65 mi/h) on rural interstate highways in 1994. The study group consisted of 266 sites on rural VA interstate highways which consists of sections on I-85, I-95, I-81, I-64, I-77. Three years before 1994 was chosen as the before period. Four years after 1994 was chosen for the after period. The basic statistics of the data are shown in Table 6.1.1. For all the before and after period years, statistical characteristics of both the contributing factors (section length, ADT) and crash data are given. According to the available data, we had five types of crashes that were studied for VA.

Table 6.1.1 Data Statistics of VA for both before and after periods

Year	# of VA sites	Length (mile)	ADT	Total Crash	Fatal Crash	Rear-End Crash	Total Crash with	Fatal Crash with

							Truck	Truck
		Min-max-total	Min-max-average	Min-max-total	Min-max-total	Min-max-total	Min-max-total	Min-max-total
1991	266	1.050-14.25-1424.32	3066.88-55050-12719.34	0-55-2888	0-2-59	0-23-554	0-11-528	0-1-10
1992	266	1.050-14.25-1424.32	2633.762-145000-14967.756	0-43-2839	0-2-51	0-22-524	0-21-592	0-2-17
1993	266	1.050-14.25-1424.32	3075.322-62557.8-14637.69	1-53-3322	0-3-58	0-35-580	0-17-648	0-1-20
1995	266	1.050-14.25-1424.32	3421.704-63204.17-15499.571	0-82-3608	0-3-77	0-50-727	0-17-772	0-1-17
1996	266	1.050-14.25-1424.32	3669.293-65860.25-16134.708	1-73-3964	0-3-57	0-47-732	0-17-867	0-2-17
1997	266	1.050-14.25-1424.32	4216.282-74644.28-16940.482	0-58-3735	0-4-69	0-34-713	0-18-846	0-3-19
1999	266	1.050-14.25-1424.32	3600-61132.49-17693.307	1-76-4070	0-2-53	0-56-798	0-17-948	0-2-26

Note: Min: minimum value of all the sections of particular year

Max: maximum value of all the sections of particular year

Total: the summation of all the values of all the sections of particular year

Average: the average of all the values of all the sections of particular year

6.1.2 MODEL DEVELOPMENT

Model 3 (GEE without yearly trend: $E(m)= ?\ (Length)^{?}{}_{1}\ (ADT)^{?}{}_{2}$) was calibrated for each type of crash data in VA. The estimated parameters were listed in Table 6.1.2.

Table 6.1.2 Model parameters of VA Crash data

# VA of entities: 266					
years of crashes: 1991-1993; 1995, 1996, 1997,1999					
	Total	Fatal	Rear-End	Truck Total	Truck Fatal
k	5.9	5.58	1.71	3.99	15.0
?	0.02242775	0.00119143	0.00032469	0.00031610	0.00001710
$?_?$	0.62225762	0.84194822	0.53600821	0.81924963	1.22562971
$?_?$	0.54802324	0.39842408	0.82864036	0.78817569	0.63939524

Note: k: aggregation parameter of NB distribution

? ,$?_?$,$?_?$: model parameters

6.1.3 EVALUATION

Evaluation was done, as listed in Tables 6.1.3, for each crash type each year using the two methods stated in chapter four. The actual crash data for each year during the after period was listed in the second column noted as "?". The accumulated values of the ? over years were calculated in the third column noted as "Cumu ?". The predicted "would have been" crash data for each year during the after period were listed in the fourth column noted as "?". The accumulated values of the ? over years were calculated in the fifth column noted as "Cumu ?". The variance of the ? was calculated for each year during the after period and listed in the sixth column noted as "VAR(?)". The cumulative values of VAR(?) over the years were also listed in the column noted as "CumuVAR(?)". The evaluation was conducted using the difference between the "would have been" after crashes and the actual crashes were noted as ? in the column noted as "Excess ?" and also the accumulated ? over the years. The evaluation was also investigated by getting the ratio of the actual after crashes to the "would have been" after crashes for each year noted as "?" and the also the accumulated values of "Cum ?".

Variances for both the difference ? and the ratio ? are also calculated and listed in the table.

The evaluation showed that the ratio ? for each type of crashes in VA is bigger than 1.0. This means that the actual crash count was more than the predicted "would have been" crash counts. Assuming all the other factors remained constant except for the change of speed limit from DSL to Uniform and that the relationship between ADT and crash predicted by the CEM is accurate, we deduced from the ? being larger than 1.0 that this change from DSL to Uniform increased crash risk on VA rural interstates.

Table 6.1.3 a) Evaluation of VA Total Crash of after period

VA total crash evaluation															
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	3608	3608	3223.037	3223.037	985.575	985.575	-384.963	-384.963	1.119335	1.119335	4593.575	0.000466	4593.575	0.000466
1996	3964	7572	3296.327	6519.364	1031	2016.575	-667.673	-1052.64	1.202437	1.161408	4995	0.000502	9588.575	0.000242
1997	3735	11307	3383.393	9902.757	1088.097	3104.672	-351.607	-1404.24	1.103816	1.141767	4823.097	0.000442	14411.67	0.000156
1999	4070	15377	3463.153	13365.91	1142.241	4246.913	-606.847	-2011.09	1.175118	1.150437	5212.241	0.000471	19623.91	0.000117
Ave	3844.25		3341.478		4246.913		-502.773		1.150176					

Note: ? : he sum of all the "what was" crash counts over all sites of the studied group for a period of time during the after period

cumu ? : the sum of ? over years

? : the sum of all the "would have been" crash counts over all sites of the studied group for a period of time during the after period

cumu ? : the sum of ? over years

Var(?): the variance of ?

Cumu Var(?): the variance of Cumu ?

Excess ? : the difference between ? and ?

Cumu ? : the cumulative difference between ? and ?

?: the ratio of the actual after data to the “would have been” after data over all sites of the studied group of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

Var(?): the variance of the difference ?

Var(?): the variance of the ratio ?

Cumu Var(?): the variance of Cumu ?

Cumu Var((?): the variance of Cumu ?

Ave: the averaged value of the column

Table 6.1.3 b) Evaluation of VA Fatal Crash of after period

VA fatal crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	77	77	58.645	58.645	2.284	2.284	-18.355	-18.355	1.312114	1.312114	79.284	0.023471	79.284	0.023471
1996	57	134	59.57	118.215	2.355	4.639	2.57	-15.785	0.956223	1.133152	59.355	0.016626	138.639	0.010001
1997	69	203	60.82	179.035	2.459	7.098	-8.18	-23.965	1.133742	1.133605	71.459	0.019457	210.098	0.006611
1999	53	256	61.819	240.854	2.525	9.623	8.819	-15.146	0.856776	1.062708	55.525	0.014316	265.623	0.004597
Ave	64		60.2135		9.623		-3.7865		1.064713					

Table 6.1.3 c) Evaluation of VA Rear-end Crash of after period

VA rear - end crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	727	727	602.886	602.886	175.728	175.728	-124.114	-124.114	1.205284	1.205284	902.728	0.002698	902.728	0.002697
1996	732	1459	624.169	1227.055	188.312	364.04	-107.831	-231.945	1.172193	1.188738	920.312	0.002539	1823.04	0.001309
1997	713	2172	651.045	1878.1	205.569	569.609	-61.955	-293.9	1.094632	1.156301	918.569	0.002259	2741.61	0.000831
1999	798	2970	679.093	2557.19	224.676	794.285	-118.907	-412.807	1.174525	1.161289	1022.676	0.002398	3764.29	0.000617
Ave	743		639.298		794.285		-103.202		1.161658					

Table 6.1.3 d) Evaluation of VA Truck Total Crash of after period

VA truck total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	772	772	650.394	650.394	154.73	154.73	-121.606	-121.606	1.186539	1.186539	926.73	0.002337	926.73	0.002336
1996	867	1639	672.379	1322.773	165.298	320.028	-194.621	-316.227	1.28898	1.238837	1032.298	0.002522	1959.028	0.001216
1997	846	2485	699.914	2022.687	179.692	499.72	-146.086	-462.313	1.208277	1.228414	1025.692	0.00226	2984.72	0.000791
1999	948	3433	726.5	2749.187	194.031	693.751	-221.5	-683.813	1.304407	1.248618	1142.031	0.002419	4126.751	0.000597
Ave	858		687.297		693.751		-170.953		1.247051					

Table 6.1.3 e) Evaluation of VA Truck Fatal Crash of after period

VA truck fatal crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	17	17	16.746	16.746	0.051	0.051	-0.254	-0.254	1.014983	1.014983	17.051	0.060765	17.051	0.060764
1996	17	34	17.176	33.922	0.052	0.103	0.176	-0.078	0.989579	1.00221	17.052	0.057756	34.103	0.029626
1997	19	53	17.779	51.701	0.056	0.159	-1.221	-1.299	1.068487	1.025064	19.056	0.060269	53.159	0.019885
1999	26	79	18.265	69.966	0.067	0.226	-7.735	-9.034	1.423202	1.129068	26.067	0.078279	79.226	0.016193
Ave	20		17.492		0.226		-2.2585		1.124063					

6.2 ARKANSAS (UNI to DSL)

6.2.1 DATA CHARACTERISTICS

In the state of AR, before 1996, the speed limits were uniform 104.59 km/h (65 mi/h) for both passenger cars and trucks. From August of 1996, the speed limits were changed to DSL of 70/65mph.

The study group we choose was 10 sites on AR rural interstate highway I-40. Five before years 1991-1995 and three after years 1997-1999 data were available.

The basic statistics of the data are shown in Table 6.2.1. With the available data, we selected four types of crashes to study for AR.

Table 6.2.1 Data Statistics of AR for both before and after periods

Year	# of AR sites	Length (mile)	ADT	Total Crash	Fatal Crash	Rear-End Crash	Total Crash with Truck
			Min-max-average	Min-max-total	Min-max-total	Min-max-total	Min-max-total
1991	10	10	18000-26220-21406	14-36-275	0-2-5	3-12-74	5-14-107
1992	10	10	18968-22860-21956.3	15-65-326	0-2-6	1-25-85	5-23-118
1993	10	10	16250-28500-23313.2	25-54-380	0-5-11	4-23-88	9-21-153
1994	10	10	15780-31000-24600.1	25-53-374	0-1-4	5-24-113	8-24-143
1995	10	10	15000-28504-24133.1	15-57-387	0-2-5	4-23-99	5-32-150
1997	10	10	26288-33746-29711.9	26-61-414	0-2-6	7-22-113	8-30-194
1998	10	10	26000-39339-30182.6	31-59-459	0-5-14	6-21-119	14-28-197
1999	10	10	27000-35312-29675.9	26-74-476	0-3-10	3-44-173	11-32-195

Note: Min: minimum value of all the sections of particular year

Max: maximum value of all the sections of particular year

Total: the summation of all the values of all the sections of particular year

Average: the average of all the values of all the sections of particular year

6.2.2 MODEL DEVELOPMENT

Model 3 (GEE without yearly trend: $E(m)= ?\ (Length)^{?_1} (ADT)^{?_2}$) was calibrated for each type of crash data for AR. Since all the sections are approximately 10 miles long, the parameter $?_?$ for the length was assumed to be 1. The estimated parameters are listed in Table 6.2.2.

Table 6.2.2 Model parameters of AR Crash data

# of AR entities: 10				
years of crashes: 1991-1995; 1997-1999				
Parameter	Total	Fatal	Rear-End	Truck Total

k	19.03	25.0	14.86	18.2
?	0.002672306	0.059828087	0.000000016	0.023873781
??	1.0	1.0	1.0	1.0
??	0.714129486	0.003553155	1.773975952	0.401282965

Note: k: aggregation parameter of NB distribution

? ,? ?,? ?: model parameters

6.2.3 EVALUATION

Evaluation was done, as listed in Tables 6.2.3, for each crash type year by year using the two methods stated in chapter four.

Table 6.2.3 a) Evaluation of AR Total Crash of after period

AR total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	414	414	419.242	419.242	91.337	91.337	5.242	5.242	0.986984	0.986984	505.337	0.002856	505.337	0.002856
1998	459	873	425.143	844.385	94.155	185.492	-33.857	-28.615	1.079075	1.03362	553.155	0.00314	1058.492	0.001501
1999	476	1349	417.7	1262.085	90.641	276.133	-58.3	-86.915	1.138982	1.06868	566.641	0.003396	1625.133	0.001044
Ave	449.67		420.695		276.133		-28.9717		1.068347					

Note: ?: he sum of all the "what was" crash counts over all sites of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

?: the sum of all the "would have been" crash counts over all sites of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

Var(?): the variance of ?

Cumu Var(?): the variance of Cumu ?

Excess ?: the difference between ? and ?

Cumu ?: the cumulative difference between ? and ?

?: the ratio of the actual after data to the “would have been” after data over all sites of the studied group of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

Var(?): the variance of the difference ?

Var(?): the variance of the ratio ?

Cumu Var(?): the variance of Cumu ?

Cumu Var((?): the variance of Cumu ?

Ave: the averaged value of the column

Table 6.2.3 b) Evaluation of AR Fatal Crash of after period

AR fatal crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	6	6	6.204	6.204	0.137	0.137	0.204	0.204	0.963688	0.963688	6.137	0.156969	6.137	0.156969
1998	14	20	6.205	12.409	0.137	0.274	-7.795	-7.591	2.248245	1.608871	14.137	0.376346	20.274	0.133553
1999	10	30	6.204	18.613	0.137	0.411	-3.796	-11.387	1.606146	1.609867	10.137	0.265261	30.411	0.089252
Ave	10		6.204		0.411		-3.7957		1.606026					

Table 6.2.3 c) Evaluation of AR Rear-end Crash of after period

AR rear - end crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	113	113	143.314	143.314	34.342	34.342	30.314	30.314	0.787162	0.787162	147.342	0.006498	147.342	0.006498
1998	119	232	148.012	291.326	37.481	71.823	29.012	59.326	0.802616	0.795685	156.481	0.006493	303.823	0.003259
1999	173	405	141.658	432.984	33.517	105.34	-31.342	27.984	1.219215	0.934844	206.517	0.011038	510.34	0.002646
Ave	135		144.328		105.34		9.328		0.936331					

Table 6.2.3 d) Evaluation of AR Truck Total Crash of after period

AR truck total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	194	194	149.013	149.013	26.058	26.058	-44.987	-44.987	1.300374	1.300374	220.058	0.010676	220.058	0.010676
1998	197	391	150.208	299.221	26.503	52.561	-46.792	-91.779	1.309976	1.30596	223.503	0.010701	443.561	0.005357
1999	195	586	148.781	448.002	25.976	78.537	-46.219	-137.998	1.309115	1.307518	220.976	0.010774	664.537	0.003584
Ave	195.333		149.334		78.537		-45.9993		1.306488					

For Arkansas, the actual after crashes was more than the predicted "would have been" after crashes for each type of crash (except for the rear-end crashes for Arkansas). If this effect was only from the treatment of changing from UNI to DSL, we would have to say that the DSL caused less safety.

6.3 IDAHO (UNI to DSL)

6.3.1 DATA CHARACTERISTICS

In the 90's, the state of Idaho made two changes to its interstates speed limits. First change happened in May of 1996, the speed limits was raised from 104.59 km/h (65 mi/h) to 120.68 km/h (75 mi/h). Then in July of 1998, DSL was enacted to its interstate highways with a lower speed limit on trucks, which was 120.68/104.59 km/h (75/65 mi/h). Since we are only interested in the change of DSL or UNI enactment, we choose 1996 as the before period of DSL enactment, and 1999- 2000 as the after period. 32 sections on ID I-84, I-86, I-90, I-15 were chosen as a study group.

The basic statistics of the data are shown in Table 6.3.1. With the available data, we selected four types of crashes to study for ID.

Table 6.3.1 Data Statistics of ID for both before and after periods

Year	# of ID sites	Length (mile)	ADT	Total Crash	Rear-End Crash	Total Crash with Truck	Rear-End with Truck
		Min-max-total	Min-max-average	Min-max-total	Min-max-total	Min-max-total	Min-max-total
1997	32	0.1-10-86.058	2800-36782-11006.15625	0-15-152	0-10-33	0-2-8	0-1-2
1999	32	0.1-10-86.058	3050-40521.5-11881.0	0-44-237	0-42-70	0-2-21	0-2-6
2000	32	0.1-10-86.058	3100- 41685.5-12131.375	0-23-184	0-18-55	0-4-21	0-2-5

Note: Min: minimum value of all the sections of particular year

Max: maximum value of all the sections of particular year

Total: the summation of all the values of all the sections of particular year

Average: the average of all the values of all the sections of particular year

6.3.2 MODEL DEVELOPMENT

Model 4 (GLM without yearly trend: E(m)= ? (Length)^?1 (ADT)^?2) was calibrated for each type of crash data of ID. The estimated parameters are listed in Table 6.3.2.

Table 6.3.2 Model parameters of ID Crash data

# of ID entities: 32				
years of crashes: 1997; 1999-2000				
	Total	Rear-End	Truck Total	Truck Rear-End
k	5.21	25.0	25.0	25.0
?	0.00257736	0.00000005	0.00009057	1.27E-9
??	0.80459636	0.78225788	1.33445126	1.75629143
??	0.74482112	1.71657542	0.71664932	1.69845228

Note: k: aggregation parameter of NB distribution

? ,??,??: model parameters

6.3.3 EVALUATION

Evaluation was done, as listed in Table 6.3.3, for each crash type year by year using the two methods stated in chapter four.

Table 6.3.3 a) Evaluation of ID Total crashes of the after

ID total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1999	237	237	161.595	161.595	93.757	93.757	-75.405	-75.405	1.461383	1.461383	330.757	0.01656	330.757	0.01656
2000	184	421	162.94	324.535	95.552	189.309	-21.06	-96.465	1.1252	1.294913	279.552	0.011356	610.309	0.006972
Ave	210.5		162.2675		189.309		-48.2325		1.293291					

Note: ?: he sum of all the “what was” crash counts over all sites of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

?: the sum of all the “would have been” crash counts over all sites of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

Var(?): the variance of ?

Cumu Var(?): the variance of Cumu ?

Excess ?: the difference between ? and ?

Cumu ?: the cumulative difference between ? and ?

?: the ratio of the actual after data to the “would have been” after data over all sites of the studied group of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

Var(?): the variance of the difference ?

Var(?): the variance of the ratio ?

Cumu Var(?): the variance of Cumu ?

Cumu Var((?): the variance of Cumu ?

Ave: the averaged value of the column

Table 6.3.3 b) period Evaluation of ID rear-end crashes of the after period

ID rear - end crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1999	70	70	37.827	37.827	4.999	4.999	-32.173	-32.173	1.844087	1.844087	74.999	0.060041	74.999	0.060041
2000	55	125	39.153	76.98	5.465	10.464	-15.847	-48.02	1.399755	1.620936	60.465	0.042307	135.464	0.025569
Ave	62.5		38.49		10.464		-24.01		1.621921					

Table 6.3.3 c) Evaluation of ID Truck Total crashes of the after period

ID truck total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1999	21	21	8.503	8.503	0.19	0.19	-12.497	-12.497	2.463243	2.463243	21.19	0.303281	21.19	0.303281
2000	21	42	8.54	17.043	0.194	0.384	-12.46	-24.957	2.452493	2.461101	21.194	0.300812	42.384	0.151821
Ave	21		8.5215		0.384		-12.4785		2.457868					

Table 6.3.3 d) Evaluation of ID Truck rear-end crashes of the after period

ID truck rear-end crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1999	6	6	2.3	2.3	0.023	0.023	-3.7	-3.7	2.597403	2.597403	6.023	1.143782	6.023	1.143782
2000	5	11	2.354	4.654	0.024	0.047	-2.646	-6.346	2.114884	2.358441	5.024	0.906054	11.047	0.515488
Ave	5.5		2.327		0.047		-3.173		2.356143					

For Idaho, the actual after crashes was more than the predicted “would have been” after crashes for each type of crash. If this effect was only from the treatment of changing from UNI to DSL, we would have to say that the DSL caused less safety.

6.4 ARIZONA (always UNI)

6.4.1 DATA CHARACTERISTICS

The state of Arizona kept uniform speed limits for both passenger cars and trucks on its interstate system during the whole 90's. 556 sections on AZ I-8, I-10, I-40, I-17, I-15, I-19 was chosen as the study group. The basic statistics of the data are shown in Table 6.4.1. With the available data, we have selected six types of crashes to study for AZ.

Table 6.4.1 Data Statistics of AZ for the whole period

Year	# of AZ sites	Length (mile)	ADT	Total Crash	Fatal Crash	Rear-End Crash	Total Crash with Truck	Fatal crash with Truck	Rear-End with Truck
		Min-max-total	Min-max-average	Min-max-total		Min-max-total	Min-max-total		Min-max-total
1991	556	0.25-21.39-2078.52	1600-86000-12958.273	0-37-2873	0-3-78	0-12-367	0-11-587	0-2-12	0-5-141
1992	556	0.25-21.39-2078.52	1679.5-84000-13180.025	0-38-2926	0-2-66	0-10-400	0-9-633	0-1-13	0-4-131
1993	556	0.25-21.39-2078.52	1439.5-86743.5-13319.820	0-37-2962	0-2-72	0-11-395	0-10-601	0-1-10	0-5-130
1994	556	0.25-21.39-2078.52	1439.5-92263-13959.45	0-37-3273	0-3-78	0-20-414	0-8-650	0-1-15	0-4-131
1995	556	0.25-21.39-2078.52	1660.5-87533.5-14456.203	0-52-3273	0-3-87	0-16-452	0-15-638	0-1-19	0-5-139
1996	556	0.25-21.39-2078.52	1868-92765.5-15731.077	0-43-3465	0-3-89	0-11-419	0-10-636	0-1-16	0-3-125
1997	556	0.25-21.39-2078.52	1996.5-97648-16195.119	0-48-3901	0-3-92	0-11-585	0-11-765	0-1-21	0-4-182
1998	556	0.25-21.39-2078.52	1862-88131-16947.163	0-65-4269	0-3-118	0-19-637	0-11-831	0-2-16	0-5-172
1999	556	0.25-21.39-2078.52	1815.5-102775.5-18688.273	0-67-4134	0-3-105	0-21-699	0-16-845	0-2-27	0-8-196
2000	556	0.25-21.39-2078.52	1254-108236.5-20638.681	0-58-4234	0-4-124	0-22-640	0-13-834	0-2-26	0-4-181

Note: Min: minimum value of all the sections of particular year

Max: maximum value of all the sections of particular year

Total: the summation of all the values of all the sections of particular year

Average: the average of all the values of all the sections of particular year

Although our interest was on the safety effects of speed limit change from DSL to UNI or otherwise, we were still analyzing those states without such speed limit change for two reasons. First we wanted to see how crash counts changed over the years even if there was no DSL or UNI treatment; Secondly, we wanted to use the AZ data as a comparison to our VA data to see the effect caused by other factors. For states with UNI or DSL treatment, by comparing the “what was” and “what would have been” after crash counts, we could see that there existed a difference. But we can’t say that this difference was only caused by the treatment of DSL or UNI. There may have been other factors or changes that were not captured in the CEM that also contributed to the difference. To get an approximate idea how much these other factors contribute to the difference, we use Arizona state data to get an index. Assuming there was a certain treatment in AZ in 1994, a before period of 1991 to 1993 was used to develop a CEM for AZ. The “what would have been” after crash counts for 1995 to 1999 are then predicted. The changes caused not by treatment can be seen by comparing the “actual” after crash counts with the “what would have been” after crash counts. So for AZ, we had an assumed before and after periods similar to VA.

6.4.2 MODEL DEVELOPMENT

Model 3 (GEE without yearly trend: $E(m)= ?\ (Length)^{?}{}_{1}\ (ADT)^{?}{}_{2}$) was calibrated for each type of crash data for AZ. The estimated parameters are listed in Table 6.4.2.

Table 6.4.2 Model parameters of AZ Crash data

# of AZ entities: 556						
years of crashes: 1991-1993; 1995-1999						
	Total	Fatal	Rear-End	Truck Total	Truck Fatal	Truck Rear-End

k	1.13	4.51	0.98	1.61	1.02	1.39
?	0.41508313	0.01128276	0.00018516	0.04202020	0.00075255	0.00040185
??	1.08715749	1.04474868	1.09661697	1.07665328	1.26544756	1.20149493
??	0.12681739	0.11963191	0.75697386	0.20755823	0.17975495	0.53740628

Note: k: aggregation parameter of NB distribution

? ,??,??: model parameters

6.4.3 EVALUATION

Evaluation was done, as listed in Table 6.4.3, for each crash type year by year using the two methods stated in chapter four. Accumulated effects were calculated.

Table 6.4.3 a) Evaluation of AZ Total Crash of after period

AZ total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	3273	3273	2978.009	2978.009	947.683	947.683	-294.991	-294.991	1.098939	1.098939	4220.683	0.000498	4220.683	0.000498
1996	3465	6738	2999.62	5977.629	961.665	1909.348	-465.38	-760.371	1.155023	1.127143	4426.665	0.000527	8647.348	0.000256
1997	3901	10639	3001.939	8979.568	962.84	2872.188	-899.061	-1659.43	1.299355	1.184759	4863.84	0.000613	13511.19	0.000182
1998	4269	14908	3033.617	12013.19	983.233	3855.421	-1235.38	-2894.82	1.407081	1.240937	5252.233	0.000675	18763.42	0.000144
1999	4134	19042	3050.007	15063.19	994.127	4849.548	-1083.99	-3978.81	1.355262	1.264114	5128.127	0.00064	23891.55	0.000118
Ave	3808		3012.638		4849.548		-795.762		1.263132					

Note: ?: he sum of all the "what was" crash counts over all sites of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

?: the sum of all the "would have been" crash counts over all sites of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

Var(?): the variance of ?

Cumu Var(?): the variance of Cumu ?

Excess ?: the difference between ? and ?

Cumu ?: the cumulative difference between ? and ?

?: the ratio of the actual after data to the "would have been" after data over all sites of the studied group of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

Var(?): the variance of the difference ?

Var(?): the variance of the ratio ?

Cumu Var(?): the variance of Cumu ?

Cumu Var((?): the variance of Cumu ?

Ave: the averaged value of the column

Table 6.4.3 b) Evaluation of AZ Fatal Crash of after period

AZ fatal crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	87	87	72.991	72.991	2.854	2.854	-14.009	-14.009	1.19129	1.19129	89.854	0.017054	89.854	0.017054
1996	89	176	73.662	146.653	2.928	5.782	-15.338	-29.347	1.20757	1.199789	91.928	0.017153	181.782	0.008561
1997	92	268	73.688	220.341	2.9	8.682	-18.312	-47.659	1.247841	1.216079	94.9	0.017738	276.682	0.00578
1998	118	386	74.288	294.629	2.962	11.644	-43.712	-91.371	1.587561	1.309947	120.962	0.022687	397.644	0.004674
1999	105	491	74.757	369.386	3.003	14.647	-30.243	-121.614	1.403796	1.32909	108.003	0.019806	505.647	0.003787
Ave	98		73.8772		14.647		-24.3228		1.327611					

Table 6.4.3 c) Evaluation of AZ Rear-end Crash of after period

AZ rear - end crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	452	452	432.294	432.294	115.908	115.908	-19.706	-19.706	1.044937	1.044937	567.908	0.003089	567.908	0.003089
1996	419	871	449.577	881.871	125.406	241.314	30.577	10.871	0.931409	0.987366	544.406	0.002605	1112.314	0.001421
1997	585	1456	456.087	1337.958	128.39	369.704	-128.913	-118.042	1.281859	1.088001	713.39	0.003818	1825.704	0.001057
1998	637	2093	484.242	1822.2	144.908	514.612	-152.758	-270.8	1.314646	1.148434	781.908	0.003777	2607.612	0.000834
1999	699	2792	504.262	2326.462	157.838	672.45	-194.738	-465.538	1.385324	1.199956	856.838	0.003932	3464.45	0.000694
Ave	558		465.292		672.45		-93.1076		1.191635					

Table 6.4.3 d) Evaluation of AZ Truck Total Crash of after period

AZ truck total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	638	638	625.404	625.404	151.915	151.915	-12.596	-12.596	1.019745	1.019745	789.915	0.002032	789.915	0.002032
1996	636	1274	633.804	1259.208	156.232	308.147	-2.196	-14.792	1.003075	1.01155	792.232	0.001972	1582.147	0.001002
1997	765	2039	635.101	1894.309	156.61	464.757	-129.899	-144.691	1.204065	1.076243	921.61	0.002456	2503.757	0.000718
1998	831	2870	645.312	2539.621	161.687	626.444	-185.688	-330.379	1.287249	1.12998	992.687	0.002635	3496.444	0.000569
1999	845	3715	652.582	3192.203	165.501	791.945	-192.418	-522.797	1.294353	1.163683	1010.501	0.002632	4506.945	0.00047
Ave	743		638.441		791.945		-104.559		1.161697					

Table 6.4.3 e) Evaluation of AZ Truck fatal Crash of after

AZ truck fatal crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	19	19	11.885	11.885	0.36	0.36	-7.115	-7.115	1.59459	1.59459	19.36	0.139595	19.36	0.139595
1996	16	35	12.04	23.925	0.38	0.74	-3.96	-11.075	1.325429	1.461016	16.38	0.113805	35.74	0.063583
1997	21	56	12.064	35.989	0.365	1.105	-8.936	-20.011	1.736362	1.554705	21.365	0.150375	57.105	0.045148
1998	16	72	12.205	48.194	0.38	1.485	-3.795	-23.806	1.307602	1.493007	16.38	0.11066	73.485	0.032343
1999	27	99	12.328	60.522	0.388	1.873	-14.672	-38.478	2.184559	1.634933	27.388	0.187974	100.873	0.028338
Ave	19.8		12.104		1.873		-7.6956		1.629708					

Table 6.4.3 f) period Evaluation of AZ Truck Rear-end Crash of after period

AZ truck rear - end crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1995	139	139	144.638	144.638	21.804	21.804	5.638	5.638	0.960019	0.960019	160.804	0.007575	160.804	0.007575
1996	125	264	149.958	294.596	23.496	45.3	24.958	30.596	0.832697	0.895675	148.496	0.006258	309.3	0.003454
1997	182	446	150.855	445.451	23.602	68.902	-31.145	-0.549	1.205207	1.000885	205.602	0.009468	514.902	0.002592
1998	172	618	156.61	602.061	25.449	94.351	-15.39	-15.939	1.097131	1.026207	197.449	0.00823	712.351	0.001977
1999	196	814	161.598	763.659	27.179	121.53	-34.402	-50.341	1.211625	1.065699	223.179	0.008999	935.53	0.001631
Ave	162.8		152.7318		121.53		-10.0682		1.061336					

The evaluations showed that, for Arizona that always remained Uniform in the 90's, the actual after crashes were more than the predicted "would have been" after crashes for each type of crash (an average 28% more). This difference could have been caused by changes that were not included in the models.

As we said in chapter four, there was a drawback of our CEM development: in states with DSL treatment, we didn't have the reference group that can provide the after data if there had been no treatment because the treatment went to all interstate highways (We'll call the CEM developed only from the before data "before period CEM"). But for AZ, there was no treatment of DSL, so we can use both the "before" and "after" period data as a reference group to develop a CEM (We'll call this CEM a "whole period CEM"). The " what would have been" after crash prediction based on the "whole period CEM" can then be compared with the " what would have been" after crash prediction based on the "before period CEM" to see the difference that was caused by the drawback of the "before period CEM". Total Crashes in AZ was taken as an example to show this difference. The model parameters of both CEMs are shown in Table 6.4.4. The difference between the two "what would have been" after crash predictions are shown in Table 6.4.5.

Table 6.4.4 model parameters of both CEMs of Arizona

AZ	k	?	?$_?$	?$_?$
Before CEM	1.13	0.41508	1.08716	0.12682
Whole CEM	0.73	0.097481	1.132547	0.297259

Note: Before CEM: the CEM developed using the before years data of AZ (1991-1993)

Whole CEM: the CEM developed using the before and after years data of AZ (1991-2000) k: aggregation parameter of NB distribution

? ,? ?,? ?: model parameters

Table 6.4.5 Comparison of the two CEMs of Arizona

Year	?	?1	?2	?1-?2	?1/?2	Cumulative ?1	Cumulative ?2	Cumulative (?1-?2)	Cumulative (?1/?2)
1995	3273	2978.009	3500.547	-522.538	0.850727	2978.009	3500.547	-522.538	0.850727
1996	3465	2999.62	3555.101	-555.481	0.843751	5977.629	7055.648	-1078.02	0.847212
1997	3901	3001.939	3564.886	-562.947	0.842086	8979.568	10620.53	-1640.96	0.845492
1998	4269	3033.617	3654.451	-620.834	0.830116	12013.19	14274.98	-2261.79	0.841556
1999	4134	3050.007	3700.479	-650.472	0.82422	15063.19	17975.46	-2912.27	0.837986
ave	3808.4	3012.638	3595.093	-582.454	0.83818				0.84459

Note: ? : Actual crash for each year of AZ group

? 1: “what would have been” after crash predictions based on the “before period CEM”

? 2: “what would have been” after crash predictions based on the “whole period CEM”

? : difference between ? 1 and ? 2

Cumulative ? 1: cumulative sum of the “what would have been” after crash predictions based on the “before period CEM”

Cumulative ? 2: cumulative sum of the “what would have been” after crash predictions based on the “whole period CEM”

Cumulative ? : difference between Cumulative ? 1 and Cumulative ? 2

6.5 MISSOURI (always UNI)

6.5.1 DATA CHARACTERISTICS

The state of Missouri, during the 90's, kept its speed limits uniform for both passenger cars and trucks. But there was a speed limit increase from 88.5 km/h (55 mi/h) to 112.63 km/h (70 mi/h) in 1996. Although this was not a DSL or UNI treatment, we still did a before-after study to see what was the safety effect for this treatment of speed limit change. We had data on three sites on MO I-29, I-35, I-55 from 1991 to 1999. The before period was 1991 to 1995. The after period was 1997 to 1999.

The basic statistics of the data were shown in Table 6.5.1. According to available data, we have selected three types of crashes to study for MO.

Table 6.5.1 Data Statistics of MO for both before and after periods

Year	# of MO sites	Length (mile)	ADT	Total Crash	Rear-End Crash	Total Crash with Truck
		Min-max-total	Min-max-average	Min-max-total	Min-max-total	Min-max-total
1991	3	1.55-2.79-6.72	9761-13547-12166	2-12-26	0-1-1	0-2-3
1992	3	1.55-2.79-6.72	10265-13744-12517	1-12-17	0-2-3	0-2-3
1993	3	1.55-2.79-6.72	10635-14966-13411	2-18-30	0-4-5	0-2-4
1994	3	1.55-2.79-6.72	10917-15246-13676.67	2-11-23	0-2-3	0-2-3
1995	3	1.55-2.79-6.72	11365-15720-14187	4-9-21	0-2-2	0-1-1
1997	3	1.55-2.79-6.72	10864-17485-14147.67	3-24-48	0-5-7	2-11-18
1998	3	1.55-2.79-6.72	11136-17927-4502.67	5-17-38	0-3-4	1-5-11
1999	3	1.55-2.79-6.72	11424-18415-14887	2-20-32	0-6-7	1-5-10

Note: Min: minimum value of all the sections of particular year

Max: maximum value of all the sections of particular year

Total: the summation of all the values of all the sections of particular year

Average : the average of all the values of all the sections of particular year

6.5.2 MODEL DEVELOPMENT

Model 3 (GEE without yearly trend: E(m)= ? (Length)$^{?}$$_{1}$ (ADT)$^{?}$$_{2}$) was calibrated for each type of crash data for MO. The estimated parameters are listed in Table 6.5.2.

Table 6.5.2 Model parameters of MO Crash data

# of MO entities: 3			
years of crashes: 1991-1995; 1997-1999			
	Total	Rear-End	Truck Total
k	25.0	25.0	25.0
?	5.22E-15	6.95E-27	0.00013311
??	-0.52257092	-2.23979518	-2.0659391
??	3.71204715	6.45825792	1.08457264

Note: k: aggregation parameter of NB distribution

? ,? ?,? ?: model parameters

6.5.3 EVALUATION

Evaluation was done, as listed in Table 6.5.3, for each crash type year by year using the two methods stated in chapter four. Accumulated effects was also calculated.

Table 6.5.3 a) Evaluation of MO Total Crash of after period

MO total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	48	48	34.267	34.267	7.492	7.492	-13.733	-13.733	1.391884	1.391884	55.492	0.052056	55.492	0.052056
1998	38	86	37.58	71.847	9.014	16.506	-0.42	-14.153	1.004763	1.193173	47.014	0.032593	102.506	0.020972
1999	32	118	41.463	113.31	10.989	27.495	9.463	-4.69	0.766871	1.039166	42.989	0.021857	145.495	0.011415
Ave	39.33		37.77		27.495		-1.5633		1.054506					

Note: ? : he sum of all the “what was” crash counts over all sites of the studied group for a period of time during the after period

cumu ? : the sum of ? over years

? : the sum of all the “would have been” crash counts over all sites of the studied group for a period of time during the after period

cumu ? : the sum of ? over years

Var(?): the variance of ?

Cumu Var(?): the variance of Cumu ?

Excess ? : the difference between ? and ?

Cumu ? : the cumulative difference between ? and ?

? : the ratio of the actual after data to the “would have been” after data over all sites of the studied group of the studied group for a period of time during the after period

cumu ? : the sum of ? over years

Var(?): the variance of the difference ?

Var(?): the variance of the ratio ?

Cumu Var(?): the variance of Cumu ?

Cumu Var((?): the variance of Cumu ?

Ave: the averaged value of the column

Table 6.5.3 b) Evaluation of MO Rear-end Crash of after period

MO rear - end crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	7	7	7.123	7.123	1.206	1.206	0.123	0.123	0.959915	0.959915	8.206	0.146489	8.206	0.146489
1998	4	11	8.368	15.491	1.665	2.871	4.368	4.491	0.466909	0.701695	5.665	0.056945	13.871	0.049462
1999	7	18	9.947	25.438	2.354	5.225	2.947	7.438	0.687376	0.701935	9.354	0.075122	23.225	0.030851
Ave	6		8.479333		5.225		2.479333		0.704734					

Table 6.5.3 c) Evaluation of MO Truck Total Crash of after period

MO truck total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	Excess ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	18	18	3.191	3.191	0.156	0.156	-14.809	-14.809	5.555748	5.555748	18.156	2.122159	18.156	2.122159
1998	11	29	3.278	6.469	0.165	0.321	-7.722	-22.531	3.304955	4.448794	11.165	1.125858	29.321	0.821637
1999	10	39	3.374	9.843	0.175	0.496	-6.626	-29.157	2.918969	3.942025	10.175	0.953478	39.496	0.473148
Ave	13		3.281		0.496		-9.719		3.926558					

So, in Missouri, the actual after crashes was more than the predicted “would have been” after crashes for all types of crashes (except for the rear-end crashes). If this difference was only caused by the speed limit increase, we could say that increasing speed limit caused worse safety.

6.6 NORTH CAROLINA (always UNI)

6.6.1 DATA CHARACTERISTICS

The state of North Carolina, during the 90’s, kept its speed limits uniform for both passenger cars and trucks. But there was a speed limit increase from 104.59 km/h (65 mi/h) to 112.63 km/h (70 mi/h) in 1996. Although there was not a DSL or UNI treatment, we still did a before-after study to see the safety effect for this treatment of speed limit change. We have data on 25 sites on NC I-77, I-95, I-40 from 1991 to 2000. The before period was 1991 to 1995. The after period was 1997 to 2000.

The basic statistics of the data are shown in Table 6.6.1. With the available data, we have selected six types of crashes to study for NC.

Table 6.6.1 Data Statistics of NC for both before and after periods

Year	# of NC sites	Length (mile)	ADT	Total Crash	Fatal Crash	Rear-End Crash	Total Crash with Truck	Rear-End Crash with Truck
		Min-max-total	Min-max-average	Min-max-total	Min-max-total	Min-max-total	Min-max-total	Min-max-total
1991	25	4.02-9.18-143.8	8900-29900-16420	3-26-263	0-3-4	0-6-31	0-7-53	0-2-16
1992	25	4.02-9.18-143.8	9800-29000-17784	6-29-306	0-2-7	0-8-46	0-8-43	0-5-14
1993	25	4.02-9.18-143.8	8900-29300-16892	5-39-360	0-2-4	0-11-64	0-8-63	0-4-19
1994	25	4.02-9.18-143.8	10400-28400-18272	4-33-344	0-3-10	0-6-39	0-12-85	0-4-17
1995	25	4.02-9.18-143.8	10300-34900-20372	7-32-402	0-2-5	0-7-44	0-7-60	0-3-13
1997	25	4.02-9.18-143.8	14600-39500-24000	3-41-445	0-2-7	0-14-80	0-10-83	0-5-27
1998	25	4.02-9.18-143.8	15321-39863-24071.56	5-42-524	0-3-11	0-8-55	0-11-73	0-5-12
1999	25	4.02-9.18-143.8	18000-46000-25084	5-43-534	0-3-14	0-14-79	0-16-96	0-14-37
2000	25	4.02-9.18-143.8	16000-46000-25124	8-65-682	0-2-7	0-9-72	0-15-104	0-6-27

Note: Min: minimum value of all the sections of particular year

Max: maximum value of all the sections of particular year

Total: the summation of all the values of all the sections of particular year

Average: the average of all the values of all the sections of particular year

6.6.2 MODEL DEVELOPMENT

Model 3 (GEE without yearly trend: $E(m)= ?\ (Length)^{?_1} (ADT)^{?_2}$) was calibrated for each type of crash data for NC. The estimated parameters were listed in Table 6.6.2.

Table 6.6.2 Model parameters of NC Crash data

# of NC entities: 25					
years of crashes: 1991-1995; 1997-2000					
	Total	Fatal	Rear-End	Truck Total	Truck Rear-end
k	23.3	25.0	8.6	5.14	2.25
?	0.00105922	2.42E-14	0.00000005	0.00000010	0.00000001
??	0.77154758	0.77186377	0.85608372	0.15448450	0.13441037

??	0.82741168	2.87764717	1.61920550	1.70396521	1.83446495

Note: k: aggregation parameter of NB distribution

? ,? ?,? ?: model parameters

6.6.3 EVALUATION

Evaluation was done, as listed in Table 6.6.3, for each crash type by the year using the two methods stated in chapter four. Accumulated effects were calculated.

Table 6.6.3 a) Evaluation of NC Total Crash of after period

NC total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	445	445	424.647	424.647	80.459	80.459	-20.353	-20.353	1.047462	1.047462	525.459	0.002952	525.459	0.002952
1998	524	969	427.361	852.008	81.429	161.888	-96.639	-116.992	1.225583	1.13706	605.429	0.003533	1130.888	0.001622
1999	534	1503	442.105	1294.113	87.598	249.486	-91.895	-208.887	1.207317	1.16124	621.598	0.00338	1752.486	0.001098
2000	682	2185	441.422	1735.535	87.1	336.586	-240.578	-449.465	1.544317	1.258837	769.1	0.004559	2521.586	0.000902
Ave	546.25		433.8838		336.586		-112.366		1.25617					

Note: ?: he sum of all the "what was" crash counts over all sites of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

?: the sum of all the "would have been" crash counts over all sites of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

Var(?): the variance of ?

Cumu Var(?): the variance of Cumu ?

Excess ?: the difference between ? and ?

Cumu ?: the cumulative difference between ? and ?

?: the ratio of the actual after data to the “would have been” after data over all sites of the studied group of the studied group for a period of time during the after period

cumu ?: the sum of ? over years

Var(?): the variance of the difference ?

Var(?): the variance of the ratio ?

Cumu Var(?): the variance of Cumu ?

Cumu Var((?): the variance of Cumu ?

Ave: the averaged value of the column

Table 6.6.3 b) Evaluation of NC Fatal Crash of after period

NC fatal crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	7	7	12.788	12.788	0.464	0.464	5.788	5.788	0.545839	0.545839	7.464	0.043163	7.464	0.043163
1998	11	18	12.056	24.844	0.385	0.849	1.056	6.844	0.909998	0.723526	11.385	0.077066	18.849	0.029721
1999	14	32	13.669	38.513	0.545	1.394	-0.331	6.513	1.021237	0.830108	14.545	0.077086	33.394	0.02214
2000	7	39	14.273	52.786	0.61	2.004	7.273	13.786	0.488972	0.738301	7.61	0.034664	41.004	0.014348
Ave	9.75		13.197		2.004		3.4465		0.741512					

Table 6.6.3 c) Evaluation of NC Rear-end Crash of after period

NC rear - end crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	80	80	68.896	68.896	11.653	11.653	-11.104	-11.104	1.158327	1.158327	91.653	0.019967	91.653	0.019967
1998	55	135	68.752	137.648	11.308	22.961	13.752	2.648	0.798068	0.979575	66.308	0.013041	157.961	0.008251
1999	79	214	72.743	210.391	12.888	35.849	-6.257	-3.609	1.083376	1.016331	91.888	0.01763	249.849	0.005654
2000	72	286	74.956	285.347	13.772	49.621	2.956	-0.653	0.958215	1.001678	85.772	0.01493	335.621	0.004115
Ave	71.5		71.33675		49.621		-0.16325		0.999496					

Table 6.6.3 d) Evaluation of NC Truck Total Crash of after period

NC truck total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	83	83	95.923	95.923	22.021	22.021	12.923	12.923	0.863211	0.863211	105.021	0.01071	105.021	0.01071
1998	73	156	94.095	190.018	20.928	42.949	21.095	34.018	0.773982	0.819999	93.928	0.009577	198.949	0.005098
1999	96	252	98.884	288.902	23.59	66.539	2.884	36.902	0.968498	0.871573	119.59	0.011976	318.539	0.003614
2000	104	356	103.139	392.041	25.385	91.924	-0.861	36.041	1.005947	0.907526	129.385	0.012087	447.924	0.002803
Ave	89		98.01025		91.924		9.01025		0.90291					

Table 6.6.3 e) Evaluation of NC Truck Rear-end Crash of after period

NC truck rear - end crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1997	27	27	25.841	25.841	5.257	5.257	-1.159	-1.159	1.03669	1.03669	32.257	0.047514	32.257	0.047514
1998	12	39	25.381	51.222	4.93	10.187	13.381	12.222	0.469204	0.758447	16.93	0.019728	49.187	0.016852
1999	37	76	26.579	77.801	5.528	15.715	-10.421	1.801	1.381268	0.974322	42.528	0.065466	91.715	0.014878
2000	27	103	27.946	105.747	6.074	21.789	0.946	2.747	0.958693	0.972129	33.074	0.040555	124.789	0.010974
Ave	25.75		26.43675		21.789		0.68675		0.961464					

The evaluation showed that, in North Carolina, the actual after crashes was a little less than the predicted "would have been" after crashes for each type of crash (except for the total number of crashes). If this difference was only caused by the speed limit increase, we could say that increasing speed limit in NC didn't make safety worse.

6.7 WASHTINGTON (always DSL)

6.7.1 DATA CHARACTERISTICS

The state of Washington kept its speed limits DSL always 104.59/ 96.6 km/h (65/60 mi/h) during the 90's. There was data for 9 sections on I-90 of WA that was chosen as a study group. The basic statistics of the data are shown in Table 6.7.1. With the available data, total number of crashes was studied for WA.

Table 6.7.1 Data Statistics of WA for both before and after periods

Year	# of WA sites	Length (mile)	ADT	Total Crash
		Min-max-total	Min-max-average	Min-max-total
1991	9	0.97-6.15-29.07	6800-9800-8244.44	4-32-169
1992	9	0.97-6.15-29.07	6900-10000-8311.11	2-28-152
1993	9	0.97-6.15-29.07	7600-11400-8844.44	8-42-193
1994	9	0.97-6.15-29.07	7600-12700-9333.33	2-42-192
1995	9	0.97-6.15-29.07	9400-13100-11133.33	8-41-186
1996	9	0.97-6.15-29.07	9100-13300-11277.78	3-37-191
1997	9	0.97-6.15-29.07	9800-13900-11400	4-29-156
1998	9	0.97-6.15-29.07	10000-14100-11711.11	2-30-154
1999	9	0.97-6.15-29.07	10500-14700-12266.67	4-44-224
2000	9	0.97-6.15-29.07	11300-15300-12833.3	5-36-186

Note: Min: minimum value of all the sections of particular year

Max: maximum value of all the sections of particular year

Total: the summation of all the values of all the sections of particular year

Average: the average of all the values of all the sections of particular year

The reasons we studied WA although there was no DSL or UNI treatment enacted was very similar to the reason why AZ was analyzed. First we want to see how crash counts change over the years even there was no DSL or UNI treatment; Secondly, we want to use the WA data as a comparison to our VA data to see the effect caused by other factors except UNI treatment.

Assuming there was a UNI treatment in WA in 1994, we used the data from 1991 to 2000.

6.7.2 MODEL DEVELOPMENT

Model 3 (GEE without yearly trend: $E(m)= ?\ (Length)^{?_1} (ADT)^{?_2}$) was calibrated for each type of crash data for WA. The estimated parameters were listed in Table 6.7.2.

Table 6.7.2 Model parameters of WA Crash data

# of WA entities: 9	
years of crashes: 1991-1993; 1994-2000	
	Total
k	4.62
?	0.53094975
??	0.43973822
??	0.34043417

Note: k: aggregation parameter of NB distribution

$?$, $?_?$, $?_?$: model parameters

6.7.3 EVALUATION

Evaluation was done, as listed in Table 6.7.3, for total crashes year by year using the two methods stated in chapter four, for the after data 1994-2000. Accumulated effects was also calculated.

Table 6.7.3 Evaluation of WA Total Crash of after period

WA total crash evaluation														
YEAR	?	Cumu ?	?	Cumu ?	VAR(?)	Cumu VAR(?)	*Excess* ?	Cumu ?	Ratio ?	Cumu ?	Var(?)	Var(?)	Cumu Var(?)	Cumu Var(?)
1994	192	192	173.82	173.82	16.397	16.397	-18.18	-18.18	1.103992	1.103992	208.397	0.007002	208.397	0.007002
1995	186	378	184.034	357.854	18.364	34.761	-1.966	-20.146	1.010135	1.05601	204.364	0.006033	412.761	0.003251
1996	191	569	184.879	542.733	18.536	53.297	-6.121	-26.267	1.032548	1.048208	209.536	0.006153	622.297	0.002129
1997	156	725	185.705	728.438	18.697	71.994	29.705	3.438	0.839587	0.995145	174.697	0.004895	796.994	0.0015
1998	154	879	187.392	915.83	19.039	91.033	33.392	36.83	0.821361	0.959681	173.039	0.004741	970.033	0.001147
1999	224	1103	190.322	1106.152	19.639	110.672	-33.678	3.152	1.176315	0.99706	243.639	0.00692	1213.672	0.000991
2000	186	1289	193.202	1299.354	20.237	130.909	7.202	10.354	0.962201	0.991955	206.237	0.005474	1419.909	0.00084

Ave	184.143		185.622		130.909		1.479143		0.992306						

Note: ? : he sum of all the "what was" crash counts over all sites of the studied group for a period of time during the after period

cumu ? : the sum of ? over years

? : the sum of all the "would have been" crash counts over all sites of the studied group for a period of time during the after period

cumu ? : the sum of ? over years

Var(?): the variance of ?

Cumu Var(?): the variance of Cumu ?

Excess ? : the difference between ? and ?

Cumu ? : the cumulative difference between ? and ?

? : the ratio of the actual after data to the "would have been" after data over all sites of the studied group of the studied group for a period of time during the after period

cumu ? : the sum of ? over years

Var(?): the variance of the difference ?

Var(?): the variance of the ratio ?

Cumu Var(?): the variance of Cumu ?

Cumu Var((?): the variance of Cumu ?

Ave: the averaged value of the column

So for Washington state that had always DSL in the 90's, the actual total number of crashes were just a little less than the predicted "would have been" after crashes (almost the same). So we could say that under the situation of no DSL treatment, no effect happened on WA safety.

CHAPTER 7

SUMMARY OF RESULTS

The purpose of this study was to evaluate the safety effect of Differential Speed Limit policy change on rural interstate highways in the United States. We decided to use Empirical Bayes method for this study. Various models were discussed and compared to choose a best model for our data. Seven states of all the data we wound up were analyzed. These are: Virginia, Arkansas, Idaho, Arizona, Missouri, North Carolina and Washington. VA changed from DSL to Uniform in the 90's, AR and ID changed from Uniform to DSL in the 90's. MO and NC remained Uniform during the 90's. WA remained DSL during the 90's. All findings and conclusions for this study were based on the analysis and evaluation results that were done in chapter six.

7.1 FINDINGS AND OBSERVATIONS

For each of the seven states, evaluations were done. Table 7.1 shows the evaluations for each state based on the cumulative ratio ? for the whole after period, assuming that the DSL policy changing treatment was the only factor that has effect on safety.

Table 7.1 Summary of Evaluations for each state studied

Ratio ? \ State	DSL-UNI	UNI-DSL		UNI			DSL
Crash type	VA	AR	ID	AZ	MO	NC	WA
Total crash	1.15	1.07	1.29	1.26	1.04	1.26	0.99
Fatal crash	1.06	1.61	NA	1.33	NA	NA	NA

Rear-End crash	1.16	0.93	1.62	1.20	0.70	1.002	NA
Total crash with truck involved	1.25	1.31	2.46	1.16	3.94	0.91	NA
Fatal crash with truck involved	1.13	NA	NA	1.63	NA	NA	NA
Rear-End crash with truck involved	NA	NA	2.36	1.07	NA	0.97	NA

Note: NA: means that type of crash data were not analyzed in this study

Several Findings were made from the analysis:

1. The crash count was not proportional to the section length. The power exponents ($?_?$) for the length are almost always less than one (but some are greater than one). In case of $?_?$ being less than 1.0, it means that as length increases, the number of crashes forecasted by $E(m)$ will correspond to a lower crash rate. In case of $?_?$ being larger than 1.0, it means that as length increases, the number of crashes forecasted by $E(m)$ will correspond to a larger crash rate.
2. The crash count was not proportional to the AADT. The power exponents ($?_?$) for the length are almost always less than one (but some are greater than one). This means that as ADT goes up, the crashes forecasted by $E(m)$ will correspond to a lower crash rate.
3. In most cases, ? tended to be greater than 1.0 when $ß_2$ (the coefficient for ADT) was less than 1.0. This occurred because in most states, AADT increases over time, so with a $ß_2$ less than 1.0 the "would have been" crashes predicted by the $E(m)$ model will correspond to a lower crash rate. As $ß_2$ approaches zero, large increases in ADT tend to only increase crashes by a small amount, with the effect being that the "would have been" crash rate was so low that it was impossible for any policy change to show a ? less than 1.0 (an improvement in safety).

7.2 INTERPRETATION OF RESULTS

There are several reasons why the results are inconsistent and unexpected:

1. Due to the fact that there were no reference entities that could be found similar to the treated entities in the same state, the reference group used in our analysis was a group of treated entities with the data only from the years before the treatment. Because of this, it may have caused the models not to be reflective of the true underlying trend in the crashes that would have occurred without the treatment. For example, in Virginia, if the treatment of DSL removal was only applied to I-95 in 1994, with I-81 remaining uniform speed limit, then the reference group for the VA study could have consisted of I-81 from year 1991 to 1999 and I-95 from 1991-1993. Instead, the removal of DSL was applied to I-81 and I-95. So the reference group was simply an extrapolation of the temporal trend that occurred from 1991 to 1993. In other words, if some significant change occurred in, say, 1995 that increased the crashes but had nothing to do with speed limits policy change, then unfortunately this would not be captured in the reference group and therefore would not be reflected in the models.
2. Because sample size varied by state, one doesn't have equal confidence in each state's results. For example, one is less certain of the Missouri results (based on 3 sites) being representative of Missouri than one is of the Arizona results (based on 556 sites) being representative of Arizona.
3. A separate problem is the degree to which the selected sites are an unbiased sample. We didn't control how sites were selected and thus the randomness is a

function of how individual state set up their individual speed monitoring programs.

4. One big advantage in the EB method is that using a long series of years for the analysis, it increases the precision of model estimation and prediction. These durations of 1, 3 and 5 years in our study are relatively short in the sense that Hauer has pointed out that the power of the EB method may not be apparent with only 3 years of data.
5. Speed data were not available for each studied site of each year. If there had been speed data for each year and each site, this could have been covered as a contributing factor in the model. Then model would have reflected the truth better.

7.3 CONCLUSIONS

Since the index ? was the ratio of the actual after crashes to the "would have been" after crashes, a ratio ? being less than 1.0 means the actual after crashes was less than the "would have been" after crashes. If this change was only caused by the enacted treatment, we can say that this treatment was a good call toward improving safety.

In our study, the results vary for different types of crashes for different states. The study showed that usually as time passed, the actual number of crashes for the after period were larger than the predicted "would have been" total crashes. Usually this crash risk increase occurred whether the state had maintained uniform speed limits, maintained differential speed limits, changed from uniform to differential speed limits, or changed from differential speed limits to uniform speed limits. Thus these results are not conclusive. Two potential reasons for these inconclusive results are:

1. There may exit other factors not captured in the models that affect safety and
2. The appropriateness of the CEM which usually reflected a lower crash rate with a higher ADT.

7.4 LIMITATIONS AND RECOMMENDATIONS

1. Models were only developed and applied for the corresponding types of crash data for individual state within the certain data ranges (basic characteristics of AADT and section length for each state were all listed in chapter six). Uses of the models developed in our study to other states or other data ranges were not validated.
2. The purpose of this study was to evaluate the safety effects of DSL policy change on rural interstate highways. However, the models couldn't cover every factor that may have effects on road safety. So it's hard to make a definite conclusion that the safety effects were only caused by the DSL strategy.
3. Although we tried to include the mean speed and the 85th speed data in our models, the speed data we obtained were only adequate for the speed analysis. For example, an interstate section might have speed monitoring sites, which was enough to compare overall speeds but not enough to use in a crash estimation model. Since there was not enough speed data for each section in each year used in the crash analysis, our models only included section length and AADT as

variables. If there was more speed data available in the future, further studies could be conducted for a more accurate investigation.

REFERENCES

1. Council, F. M., Duncan, C.S., and Khattack, A. J. Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions, *Transportation Research Record 1635*, Transportation Research Board, National Research Council, Washington, D.C., 1998,pp. 63-71

2. Harkey, D.J. and Mera, R. *Safety Impacts of Different Speed Limits on Cars and Trucks*, The Scientex Corporation for the Federal highway Administration, McLean, Virginia, 1994, pp.30.

3. Garber N.J., and Ravi Gadiraju, Speed Variance and its Influence on Accidents, AAA Foundation for Traffic Safety, July, 1988.

4. Garber N.J., Ravi Gadiraju, Effect of Truck Strategies on Traffic Flow and Safety on Multilane Highways, AAA foundation for Traffic Safety, July 1989.

5. Garber N.J. and Gadiraju R., Impact of Differential Speed Limits on Highway Speeds and Accidents, Dept. of Civil Engineering, University of Virginia, Charlottesville, VA, 1991, pp13.

6. J.W. hall and L. V. Dickinson, An Operational Evaluation of Truck Speeds on Interstate Highways, Dpet. Of Civil Engineering, University of Maryland, February, 1974.

7. Idaho Transportation Department Planning Division, Evaluation of the Impacts of Reducing Truck Speeds on Interstate Highways in Idaho, -Phase III, Final Report Dec.,2000, National Institute fro Advanced Transportation Technology University of Idaho

8. Ezra Hauer, (1997) Observational Before-After Studies in Road Safety, estimating the effect of highway and traffic engineering measures on road safety. Department of Civil engineering, University of Toronto.

9. Bhagwant Persaud, Ezra Hauer, Richard Retting, Rama Vallurupalli and Kornel Muxsi, Carsh Reductions Related to Traffic Signal Removal in Philadelphia, Accident Analysis and Prevention, Vol. 29, No.6, pp803-810,1997

10. Ezra Hauer, Jerry C.N. Ng, (1988) And Jane Lovell, Estimation of Safety at Signalized Intersections, Transportation Research Record 1185. pp48-61

11. E. Hauer and B. N. Persaud,(1988) How to Estimate the Safety of Rail-Highway Grade Crossings and the Safety Effects of Warning Devices, Transportation Research Record 1114. pp131-140
12. Bhagwant N. Persaud, Richard A. Retting, Per E. Garber, Dominique Lord, Crash Reduction Following Installation of Roundabouts in the United States, Insurance Institute For Highway Safety, Ryerson Polytech University Toronto, Ontario, March 2000.

13. Bhagwant N. Persaud, Estimating Accident Potential of Ontario Road Sections, Transportation Research Record 1327. pp47-53

14. James A. Bonneson And Patric T. McCoy, Estimation of Safety at Two-way Stop-controlled Intersections on Rural Highways, Transportation Research Record 1401. pp83-89

15. Bhagwant N. Persaud and Dominique Lord (2000) Accident Prediction Models with and without trend: Application of the Generalized Estimating Equations (GEE) Procedure. Presented at the 79th Annual Meeting of the Transportation Research Board, TRB, Washington D.C.

16. Brown, H.C., and Tarko, A.P. (1999) The Effects of Access Control on Safety on Urban Arterial Streets. Presented at the 78th Annual Meeting of the Transportation Research Board, TRB, Washington D.C.

17. Mountain, L., Fawaz, B., and Jarrett, D. (1996) Accident Prediction Models for Roads with Minor Junctions. Accident Analysis & Prevention, Vol. 28, No. 6, pp. 695-707.

18. Hauer, E. (1992) Empirical Bayes Approach to the Estimation of "Unsafety": the Multivariate Regression Method. Accident Analysis & Prevention, Vol. 24, No. 5, pp. 457-477.

19. C. Abbess, D. Jarrett, and C.C. Wright, Accidents at blackspots: Estimating the effectiveness of Remedial Treatment, with special reference of the Regression-to mean effect, DTP Issues new Standard on Road Layout and geometry: Highway Link Design, pp535-543

20. Idaho Transportation Department Planning Division, Evaluation of the Impacts of Reducing Truck Speeds on Interstate Highways in Idaho, Finial Report, Nov. 1999.

21. Dickinson, L.V. and Hall, J.W. Truck Speeds and Accidents on Interstate Highways, *Transportation Research Record 486*, Transportation Research Board, National Research Council, Washington, D.C., 1974,pp. 19-32.

22. Baum, H.m., Esterlita, J.R., Penny, M., and Zador, P. Differential Speed Limits for Cars and Trucks: Do They Affect Vehicle Speeds, *Transportation Research Record 1318*, Transportation Research Board, National Research Council, Washington, D.C., 1991,pp. 3-7.

23. Jernigan J.D, Lynn, C.W., and Garber N. J., An Investigation of Issues Related to Raising the Rural Interstate Speed Limit in Virginia. Report No. VTRC 88-r17, March 1988.

APPENDIX A

DISTRIBUTION ANALYSIS OF EB METHOD

Ezra Hauer and Persaud [8,9,10,11] have clarified in details about the distribution assumptions of crashes for the before and after study. First, for a particular entity, the crash counts $K_{i,y}$ over years obey Poisson Distribution.

Secondly, for a particular year, the crash counts of a group $K_{i,y}$ between follow Negative Binomial distribution. And based on these two assumptions, the expected crashes of a group m_{i1} are Gamma distributed.

A direct view can be seen from the following chart. ($K_{i,y}$ is the actual crash count for site i and year y; $m_{i,y}$ is the expected crash counts for site i and year y)

Site	1991		1992		1993		1994		1995		1996	
1	K_{11}	m_{11}	K_{12}	m_{12}	K_{13}	m_{13}	K_{14}	m_{14}	K_{15}	m_{15}	K_{16}	m_{16}
2	K_{21}	m_{21}	K_{22}	m_{22}	K_{23}	m_{23}	K_{24}	m_{24}	K_{25}	m_{25}	K_{26}	m_{26}
3	K_{31}	m_{31}										
4	K_{41}	m_{41}										
5	K_{51}	m_{51}										
6	K_{61}	m_{61}										

K_{1y} : Poisson

K_{i1}: Negative Binomial

m_{i1} : Gamma

To test the Poisson and Negative binomial distribution of the data, different groups of data are extracted from some states.

1. Poisson Distribution Check

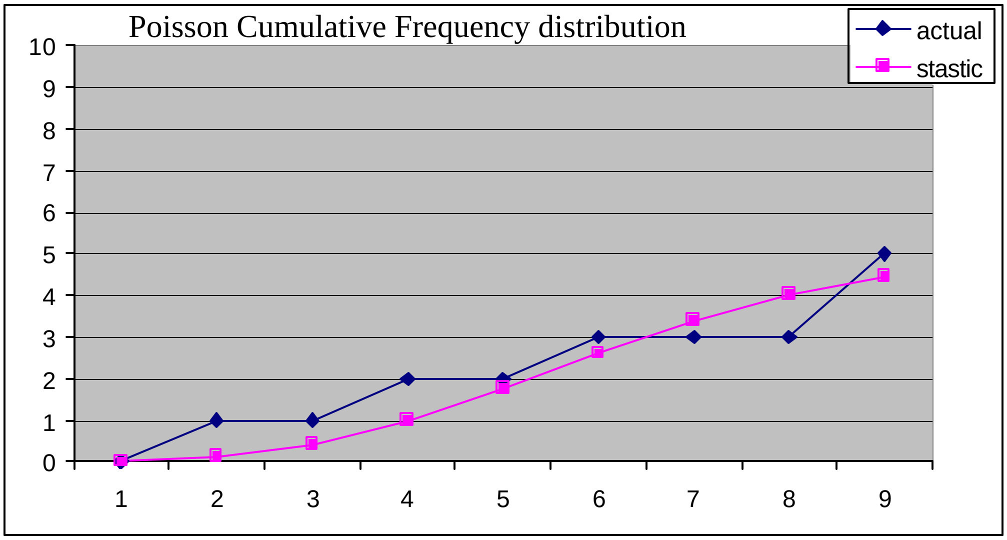
For Poisson distribution, tests were run for data from states VA and AZ. Two methods are used to check the distribution hypothesis. First frequency graphs were drawn to see the results directly. Then CHI Square was run to check the hypothesis by comparing the critical CHI Square value with the sample computed value. The results showed that this assumption fits well with our data.

Table 1 Poisson validation description and results

state	Test sites	sample size	data resource	crash type	?$^2_{cal}$	?$^2_{sta, 0.05}$	?$^2_{sta, 0.01}$	result at 5% level	result at 1% level
AZ	I-8 mp42.06 to 54.96	10	1991-2000	total	9.530	14.07	18.47	yes	yes
AZ	I-10 mp 19.79 to 26.65	10	1991-2001	total	6.738	15.51	20.08	yes	yes
VA	I-85 mp 19.52 to 24.73	5	1995-1999	total	8.764	11.07	15.09	yes	yes
VA	I-81 mp206.04 to	6	1995-2000	total	6.468	7.82	11.33	yes	yes

yes: the calculated ?^{2} value is less than the critical ?^{2} value, which means that the assumed distribution is accepted.

The following chart is shown as an example of how the Cumulative Frequency distribution looks like for both theoretical and the actual distribution. This site was chosen from one site on VA I - 85 from mp 19.52 to 24.73, about total crashes from 1995 to 1999.



2. Negative binomial distribution Check

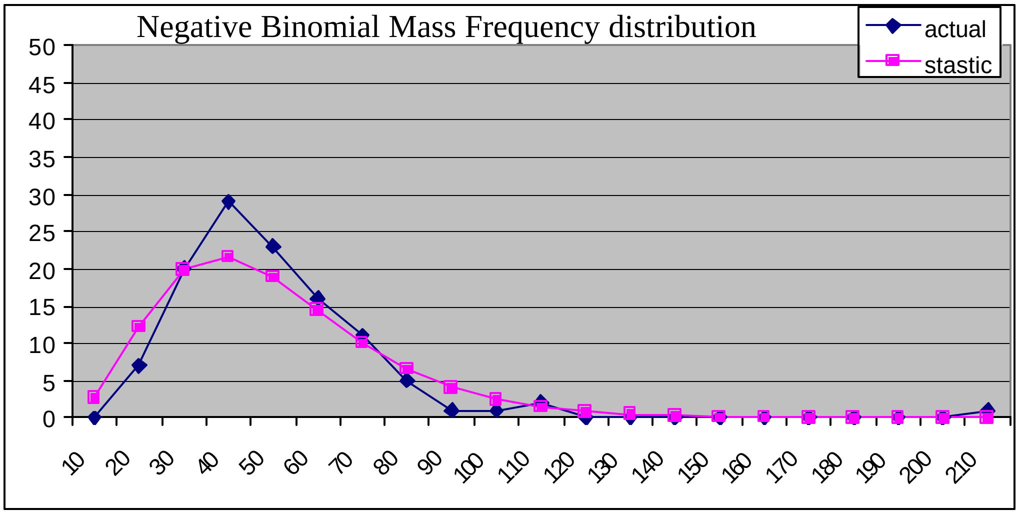
For Negative binomial distribution, tests were run for data from the states of VA, AZ, ID and NC. Instead of using the actual crash counts as variable, crash rate (10^8*crash/(365*ADT*Length)) was used because the crash segments have various lengths which would contribute to the crash numbers. Two methods are used to check the distribution hypothesis. First frequency graphs were drawn to see the results directly. Then CHI Square test was run to check the hypothesis by comparing the critical CHI Square value with the sample computed value. The results showed that this assumption can be used in our study.

Table 2 Negative Binomial validation description and results

state	year	sample size	Test site	crash type	$?^2_{cal}$	$?^2_{sta, 0.05}$	$?^2_{sta, 0.01}$	result at 5% level	result at 1% level
VA	1991	91	85n,95n,81n	total	42.326	44.8	60.1	yes	yes
VA	1992	90	85n,95n,81n	total	14.516	19.68	24.75	yes	yes

VA	1993	91	85n,95n,81n	total	10.730	20.08	14.07	yes	yes
VA	1995	117	85, 95, 81n	total	21.386	22.37	27.72	yes	yes
VA	1996	116	85, 95,81n	total	18.153	18.31	23.91	yes	yes
VA	1997	116	85, 95,81n	total	14.135	19.68	24.76	yes	yes
VA	1999	84	85n, 85s, 81n	total	29.852	27.59	33.44	yes	no(close)
NC	1993	26	40,95,77,85	total truck	4.005	6	9.22	yes	yes
NC	1999	25	40,95,77	Total truck	7.385	11.07	15.09	yes	yes
ID	1992	32	84,86,90,15	total	28.327	38.89	45.67	yes	yes
ID	1994	32	84,86,90,16	total	28.305	37.66	44.34	yes	yes
ID	1999	32	84,86,90,16	total	36.322	42.57	49.61	yes	yes
ID	2000	32	84,86,90,16	total	26.209	32.68	38.96	yes	yes
AZ	1991	277	8,10,15,17,19,40	total	24.151	21.03	26.25	yes	no(close)
AZ	1993	277	8,10,15,17,19,40	total	20.392	23.37	27.71	yes	yes
AZ	1994	278	8,10,15,17,19,40	total	17.697	26.3	32.03	yes	yes
AZ	1995	277	8,10,15,17,19,40	total	21.232	22.37	27.71	yes	yes
AZ	1996	278	8,10,15,17,19,40	total	19.977	22.37	27.71	yes	yes
AZ	1998	279	8,10,15,17,19,41	total	21.164	22.37	27.71	yes	yes
AZ	1999	280	8,10,15,17,19,42	total	20.149	23.69	29.17	yes	yes
AZ	2000	281	8,10,15,17,19,43	total	24.187	27.59	33.44	yes	yes

Mass Frequency distribution of total crashes on 116 sites from I 85, I 95, I81NB 99 of VA1997 for both theoretical and the actual distribution has been shown as an example in the following chart.



APPENDIX B

IS THERE ANY DIFFERENCE BY CHOOSING THE DIFFERENT YEAR AS THE DENOMINATOR IN THE CHANGING RATIO $C_{i,y}$?

In chapter four, we raised the question that whether changing the value in the denominator of $C_{i,y}$ will affect the analysis entirely. It was addressed here first using mathematics derivation. An example using VA data with different years in the denominator has also been run to illustrate there was no effect on the analysis.

As is the case with any data set, however, it is always possible that a single year could be an outlier. So this appendix only tested for the effect of changing the E(mi,1) shown in the denominator from year 1 to another year. It does not test for the effect of removing 1991 from the data set entirely.

Let's note that the yearly changing ratio with the first year as base in the denominator as $C_{i,y}^{1}$, while the yearly changing ratio with the third year as base in the denominator as $C_{i,y}^{3}$.

$$\frac{m_{i,y}}{m_{i,1}} \ ? \ \frac{E(m_{i,y})}{E(m_{i,1})} \ ? \ C_{i,y}^{1}$$

$$\frac{m_{i,y}}{m_{i,3}} \ ? \ \frac{E(m_{i,y})}{E(m_{i,3})} \ ? \ C_{i,y}^{3}$$

1. USE THE FIRST YEAR AS A BASE

If we use the first year as a base, the expected value of the first year crash count is estimated first as following:

$$m_{i,1} \; ? \; \frac{k ? \overset{Y}{\underset{y?1}{?}} K_{i,y}}{\frac{k}{E(m_{i,1})} ? \overset{Y}{\underset{y?1}{?}} C^{1}_{i,y}}$$

$$VAR(m_{i,1}) \; ? \; \frac{k ? \overset{Y}{\underset{y?1}{?}} K_{i,y}}{(\frac{k}{E(m_{i,1})} ? \overset{Y}{\underset{y?1}{?}} C^{1}_{i,y})^{2}} \; ? \; \frac{m_{i,1}}{\frac{k}{E(m_{i,1})} ? \overset{Y}{\underset{y?1}{?}} C^{1}_{i,y}}$$

The estimation of the expected values of crash counts of the other years was then calculated by multiplying the first year expected estimation of its changing ratio using the following equations:

$$m_{i,y} \; ? \; C^{1}_{i,y} m_{i,1} \; ? \; \frac{E(m_{i,y})}{E(m_{i,1})} m_{i,1}$$

$$VAR\;(m_{i,y}) \; ? \; (C^{1}_{i,y})^{2} m_{i,1} \; ? \; [\frac{E(m_{i,y})}{E(m_{i,1})}]^{2} m_{i,1}$$

Let’s calculate the third year expected value for an example,

$$\mathrm{m}_{\mathrm{i},3} \; ? \; \mathrm{C}_{\mathrm{i},3}\, \mathrm{m}_{\mathrm{i},1} \; ? \; \frac{E(m_{i,3})}{E(m_{i,1})} \mathrm{m}_{\mathrm{i},1}$$

$$? \frac{E(m_{i,3})}{E(m_{i,1})} \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{\frac{k}{E(m_{i,1})} ? \mathop{?}\limits_{y?1}^{Y} C_{i,y}^{1}} ? \frac{E(m_{i,3})}{E(m_{i,1})} \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{\frac{k}{E(m_{i,1})} ? \mathop{?}\limits_{y?1}^{Y} \frac{E(m_{i,y})}{E(m_{i,1})}}$$

$$? \; E(m_{i,3}) \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{E(m_{i,1})[\frac{k}{E(m_{i,1})} ? \mathop{?}\limits_{y?1}^{Y} \frac{E(m_{i,y})}{E(m_{i,1})}]}$$

$$? \; E(m_{i,3}) \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{k ? \mathop{?}\limits_{y?1}^{Y} E(m_{i,y})}$$

$$? \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{\frac{k}{E(m_{i,3})} ? \mathop{?}\limits_{y?1}^{Y} \frac{E(m_{i,y})}{E(m_{i,3})}} ? \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{\frac{k}{E(m_{i,3})} ? \mathop{?}\limits_{y?1}^{Y} C_{i,y}^{3}}$$

$$VAR(m_{i,3}) \; ? \; (C_{i,3}^{1})^2 VAR(m_{i,1}) \; ? \; [\frac{E(m_{i,3})}{E(m_{i,1})}]^2 VAR(m_{i,1})$$

$$? \; [\frac{E(m_{i,3})}{E(m_{i,1})}]^2 \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{[\frac{k}{E(m_{i,1})} ? \mathop{?}\limits_{y?1}^{Y} C_{i,y}^{1}]^2} \; ? \; [\frac{E(m_{i,3})}{E(m_{i,1})}]^2 \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{[\frac{k}{E(m_{i,1})} ? \mathop{?}\limits_{y?1}^{Y} \frac{E(m_{i,y})}{E(m_{i,1})}]^2}$$

$$? \; [E(m_{i,3})]^2 \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{[E(m_{i,1})]^2 [\frac{k}{E(m_{i,1})} ? \mathop{?}\limits_{y?1}^{Y} \frac{E(m_{i,y})}{E(m_{i,1})}]^2}$$

$$? \ [E(m_{i,3})]^2 \frac{k \ ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{[k \ ? \mathop{?}\limits_{y?1}^{Y} E(m_{i,y})]^2}$$

$$? \ \frac{k \ ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{[\frac{k}{E(m_{i,3})} ? \mathop{?}\limits_{y?1}^{Y} \frac{E(m_{i,y})}{E(m_{i,3})}]^2}$$

$$? \ \frac{k \ ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{[\frac{k}{E(m_{i,3})} ? \mathop{?}\limits_{y?1}^{Y} C_{i,y}^{3}]^2}$$

2. USE THE THIRD YEAR AS A BASE

If we use the third year as a base in the denominator, then the expected value of the third year crash count was estimated first as following:

$$m_{i,3} \ ? \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{\frac{k}{E(m_{i,3})} ? \mathop{?}\limits_{y?1}^{Y} C_{i,y}^{3}}$$

$$VAR(m_{i,3}) \ ? \frac{k ? \mathop{?}\limits_{y?1}^{Y} K_{i,y}}{(\frac{k}{E(m_{i,1})} ? \mathop{?}\limits_{y?1}^{Y} C_{i,y}^{3})^2} \ ? \frac{m_{i,1}}{\frac{k}{E(m_{i,1})} ? \mathop{?}\limits_{y?1}^{Y} C_{i,y}^{3}}$$

Compared the two results, we can see that the expected crash for the third using the third year as a base is just the same as calculated using the first year as a base and then multiply the changing ratio of the third year.

So from a mathematics point of view, it is very clear that choosing which before year as a base in the changing ratio denominator doesn't affect the results.

Also an example using the third year 1993 as a base in the changing ratio was run for the VA total crash data. The site was a section with a length of 6.13 mile from I-64 eastbound. The CEM was developed from GEE model 1. The results that from using the first year as the denominator of yearly changing ratio was shown in Table 1. The results from using the third year 1993 as the denominator of yearly changing ratio was shown in Table 2. The results for the estimation of the expected crashes are the same as using the first year 1991 as the base in the changing ratio denominator.

Table 1 Estimation of expected crashes using 1991 data as a base

year	$K_{i,y}$	$E_{i,y}$	$C^1_{i,y}$	$m_{i,y}$	$VAR(m_{i,y})$
1991	7	11.454	1	7.4	2.137
1992	4	10.372	0.906	6.701	1.753
1993	9	12.21	1.066	7.888	2.429
1995	13	11.35	0.991	7.332	2.099
1996	15	11.35	0.991	7.332	2.099
1997	6	12.428	1.085	8.029	2.516
1998	10	12.94	1.13	8.36	2.728
1999	10	13.435	1.173	8.68	2.941

Table 2 Estimation of expected crashes using 1993 data as a base

year	Ki,y	Ei,y	C3i,y	mi,y	VAR(mi,y)
1991	7	11.454	0.938	7.4	2.137
1992	4	10.372	0.849	6.701	1.753
1993	9	12.21	1	7.888	2.429
1995	13	11.35	0.93	7.332	2.099
1996	15	11.35	0.93	7.332	2.099
1997	6	12.428	1.018	8.029	2.516
1998	10	12.94	1.06	8.36	2.728
1999	10	13.435	1.1	8.68	2.941

APPENDIX C
DATA COLLECTION EXAMPLES

Example I: Data request letter

Dear Mr. Baldwin,

As I mentioned on the phone, the Virginia Transportation Research Council is working with the Federal Highway Administration to evaluate the potential safety impacts of differential speed limits for cars and trucks on Interstate facilities. I would like to request your assistance with obtaining crash and speed data (on Interstate highways) that can help us with this study. While we already have some limited data for the sites shown at the bottom of this letter, I'd like to obtain some additional data pertaining to these and other sites.

I realize that obtaining data can be time consuming so I am certainly willing to do whatever is possible to make it easier for you to fulfill our request. If at all possible, we would like to obtain the following crash and speed data in an electronic format. The crash data and the speed data may come from the same sites or they may come from different sites, whichever is easier (but all should be on Interstate Highways, with the speed limit shown. At sites with differential speed limits please list the speed limits for passenger cars and trucks separately).

For each year from 1991 to 2000, we would like to obtain the following <u>speed data elements</u> at each site:

1. average speed
2. individual vehicle speeds or speed bins (e.g. **x** vehicles between 51-55 mph, **y** vehicles 56-60 mph, etc.)
3. individual truck speeds or average truck speeds (if available)
4. individual car speeds and average car speeds (if available)
5. critical geometric data such as
   a. the number of lanes
   b. the number of interchanges (or the number of interchanges per mile)

The **speed data sites** that interest us are these plus any additional sites you recommend.

Route 29	Northbound	Milepost 29
Route 35	Northbound	Milepost 14

For each year from 1991 to 2000, I would also like to obtain the following <u>crash data elements</u>:

1. total number of crashes
   a. all crashes that do not involve trucks
   b. all crashes that do involve trucks
   c. all fatal crashes (regardless of truck involvement)
   d. all fatal crashes (of truck involvement)

e. total number of rear end crashes that do involve trucks
f. total number of rear end crashes that do not involve trucks

The crash sites should be the same from year to year, and can either be the sites shown above or be different from the speed sites. Each crash site should be a homogeneous section that can show some crashes (e.g. whether a crash site is 1 mile long or 10 miles long, it should be (a) big enough to obtain some crashes annually yet (b) small enough such that speeds and geometric characteristics for the site are homogeneous). We would like, if possible, up to **10 crash sites altogether.**

If possible, we would like the data to be in the following format. But we are also happy to have any format of data you sent to us.

route	year	Beginning mile post	Ending mile post	AADT	Total crash	Fatal crash	rear-end	truck total	truck fatal	truck RE	mean speed	85th speed

Finally, I would like to confirm that you have had since 1991 a uniform limit, that is, the same speed limit for cars and trucks.

Again, I sincerely appreciate your assistance. I would also be delighted to provide you with additional information about the purpose of this study, as well as any findings that result.

John Miller
Virginia Transportation Research Council
530 Edgemont Road
Charlottesville, VA 22903
(804) 293-1999 (voice)
(804) 293-1990 (fax)
millerjs@vdot.state.va.us (email)

Example II: Resent letter for more information

Dear Mr. Wyatt,

Thank you very much for the data you send to us! Once again we have another question: The accident data we got from you are from 1991 to 2000 as shown below

I77, From Future I-74 to Va. State Line
I77,From Yadkin Co. to Future I-74
I85,Randolph

I85,Vance
I40,Johnston
I40,Duplin
I40, Pender
I-95, Halifax
I-95,Nash
I-95,Wilson
I-85,Granville

I have four quick questions:

1. Can we get the ADT by year, from 1991 to 2000, for the accident sites shown above?

2. Do you have any crash data for
   (a) Rear-end crashes involving trucks, and
   (b) fatal crashes involving trucks?

3. When tabulating crashes, some states include only crashes on the mainline, and others include ramp crashes as well (regardless of causality). For North Carolina, which should we assume? If it is the case that ramp crashes are included, then do you know roughly how the number of ramp crashes compares to the number of mainline crashes?
4. Is the speed limit for cars and trucks 65 on North Carolina interstates, and has the limit changed since 1990?

John Miller
Virginia Transportation Research Council
530 Edgemont Road
Charlottesville, VA 22903
(804) 293-1999 (voice)
(804) 293-1990 (fax)
millerjs@vdot.state.va.us (email)

Example III: Different original data formats

Crash database data files from Arizona

Text one

00760818,1991-01-03 23:40:00,0799,03006,1,0,0,3,1,2,1,0,1,3,4,0300,7,2,1,0,0,1,00480,1991-03-27 00:00:00
00790265,1991-01-31 10:15:00,0799,04303,1,1,0,3,1,1,1,0,1,1,1,1500,6,3,3,0,0,1,04380,1991-02-27 00:00:00
00810582,1991-01-20 02:15:00,0799,04106,1,0,0,3,1,1,35,0,1,3,1,1500,3,2,1,0,0,0,02759,1991-03-14 00:00:00
00820188,1991-02-18 01:05:00,0799,04303,1,2,1,3,1,1,1,0,1,3,1,1500,7,3,5,0,0,1,06955,1991-03-21 00:00:00
00870874,1991-03-16 08:30:00,0799,04389,1,0,0,3,1,1,1,0,1,1,5,0300,7,3,1,0,0,1,10939,1991-04-23 00:00:00
00880287,1991-03-12 16:25:00,0799,04304,1,0,0,3,1,1,1,0,1,1,1,1500,2,3,1,0,0,0,10294,1991-04-29 00:00:00
00880326,1991-03-14 06:15:00,0799,04089,1,0,0,3,1,1,1,0,1,2,1,1500,1,3,1,0,0,1,10527,1991-04-29 00:00:00
00900205,1991-03-15 18:40:00,0799,03037,3,7,2,3,1,1,16,6,1,1,1,1500,1,3,5,0,0,0,10860,1991-04-22 00:00:00

Text two
00760818,1991-03-27 00:00:00,13,MIGRATION FROM IMS 960515 ,1997-10-13 11:17:00,1900-01-01 00:00:00
00790265,1991-02-27 00:00:00,16,MIGRATION FROM IMS 960515,1997-10-13 11:17:00,1900-01-01 00:00:00
00810582,1991-03-14 00:00:00,13,MIGRATION FROM IMS 960515 ,1997-10-13 11:17:00,1900-01-01 00:00:00
00820188,1991-03-21 00:00:00,16,MIGRATION FROM IMS 960515 ,1997-10-13 11:17:00,1900-01-01 00:00:00
00870874,1991-04-23 00:00:00,13,MIGRATION FROM IMS 960515 ,1997-10-13 11:17:00,1900-01-01 00:00:00
00880287,1991-04-29 00:00:00,2,MIGRATION FROM IMS 960515 ,1997-10-13 11:17:00,1900-01-01 00:00:00
00880326,1991-04-29 00:00:00,2,MIGRATION FROM IMS 960515 ,1997-10-13 11:17:00,1900-01-01 00:00:00

Text three
00820188,A,117,140
00820188,P,116,127
00900205,A,1851,1930
00900205,P,1851,1914
01060316,A,1850,1913
01060316,P,1850,1906

Text four
00810582,4,1
00840191,4,1
00880751,4,1
00940387,4,1
01010841,4,1
01110195,4,1

Crash database definition files from Arizona

File one
,Not Reported
0,Single Vehicle
1,Sideswipe (same)
2,Sideswipe (opposite)
3,Angle
4,Left Turn
5,Rear-End
6,Head-On
7,Backing
8,Other
A,Driveway/Alley Related
B,Non-Contact (mc)
C,Non-Contact (not mc)
D,U-Turn

File two
0,Not Reported
1,Overturning
2,Exhaust Fume Poisoning
3,Breakage of Vehicle
4,Explosion of Vehicle
5,Fire in Vehicle
6,Occupant Fall from Vehicle
7,Occupant Hit by Object
8,Injured from Moving Part of Vehicle

9,Object Falling from, or in Vehicle
10,Object Thrown towards, in, or on Vehicle
11,Object Fall on Vehicle
12,Toxic Chemical Leak
13,All Other Non-Collision
14,Collision with Pedestrian
15,Collision with Pedestrian Conveyance
16,Collision with other Motor Vehicle
17,Collision with Motor Vehicle Other Roadway
18,Collision with Motor Vehicle Parked Properly
19,Collision with Motor Vehicle Parked Improperly
20,Collision with Train, Forward
21,Collision with Train, Stopped
22,Collision with Train, Backward

Crash database table explanation files from Arizona

File one

TABLE NAME	COLUMN NAME	SEQ #	DATA TYPE	BYTE(S)	STATUS
airbag_defn	airbag	1	tinyint	1	0
airbag_defn	airbag_defn	2	varchar	21	0
alignment_defn	alignment	1	tinyint	1	0
alignment_defn	alignment_defn	2	varchar	12	0
body_style_defn	body_style	1	tinyint	1	0
body_style_defn	body_style_defn	2	varchar	53	0
change_log	microfilm	1	char	8	0
change_log	entry_date	2	smalldatetime	4	0
change_log	coder	3	tinyint	1	0
change_log	comment	4	char	40	0
change_log	upload_date	5	smalldatetime	4	0
collision_manner_defn	collision_manner	1	char	1	0

File two

---
* SMS Database sp970 Table Names *
---

sms01000 incident
01 collision_manner_defn
02 damage_severity_defn
03 daylight_defn
04 first_harmful_defn
05 injury_severity_defn
06 intersection_related_defn
07 junction_defn
08 nsc_reportable_defn
09 scene_defn
10 traffic_way_defn

11 weather_defn

sms02000 change_log

Example IV: Final database format prepared for use in this study

State	route	year	Beginning mile post	Ending mile post	Length	AADT	Total crash	Fatal crash	rear-end	truck total	truck fatal	truck RE	Mean speed	85th speed
AZ	I-8	1991	14.24	21.03	6.79	4950.0	8	0	2	2	0	1		
AZ	I-8	1992	14.24	21.03	6.79	4353.0	7	0	1	3	0	1		
AZ	I-8	1993	14.24	21.03	6.79	3461.5	6	0	2	2	0	1		
AZ	I-8	1994	14.24	21.03	6.79	4670.5	9	0	0	0	0	0		