U.S. Department
of Transportation

Federal Highway
Administration

Evaluation of Accident Analysis Methodology

Research and Development
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, Virginia 22101-2296

Publication No. FHWA-RD-96-039
November 1996

FOREWORD

This study improved a statistical methodology developed in a previous Federal Highway Administration (FHWA) effort. The methodology is an Empirical Bayes (EB) technique with modeling capabilities used to identify and correct regression-to-the-mean bias in two highway safety applications: (1) identifying and ranking problem locations, and (2) evaluating safety treatments. Several data sets from FHWA's Highway Safety Information System were analyzed with different variations of the EB methodology and other statistical techniques. The data sets were analyzed to assess the practical issues of the new methodology.

One major issue still not resolved is how to properly define the reference group that is a critical data requirement of the EB method. This issue was discussed at great length at a workshop with experts having a background in advanced statistics and highway safety research. The defining of the reference group appears to be less challenging for the application of identifying and ranking problem locations. An FHWA research effort planned for 1997 will further advance the application of identifying and ranking problem locations. The intended audience for this report is highway safety researchers with a strong background in advanced statistics.

A. George Ostensen
Director
Office of Safety and Traffic Operations
Research and Development

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-RD-96-039		
4. Title and Subtitle EVALUATION OF ACCIDENT ANALYSIS METHODOLOGY		5. Report Date November 1996
		6. Performing Organization Code
7. Author(s) Olga J. Pendleton		8. Performing Organization Report No.
9. Performing Organization Name and Address Texas Transportation Institute College Station, TX 77843-3135		10. Work Unit No. (TRAIS) 3A3b0052
		11. Contract or Grant No. DTFH61-92-R-00029
12. Sponsoring Agency Name and Address Office of Safety & Traffic Operations, R&D Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296		13. Type of Report and Period Covered Final Report September, 1992 - November, 1995
		14. Sponsoring Agency Code
15. Supplementary Notes Contracting Officer's Technical Representative (COTR): Michael S. Griffith, HSR-20		

16. Abstract

This study compared both traditional and empirical Bayes (EB) methodologies for ranking high hazard locations and evaluating highway safety treatments for five case studies. The case studies were either based on data collected for previous studies or data retrieved from the Highway Safety Information System (HSIS). These included two high hazard location ranking studies for New York and Illinois and three treatment evaluation studies: Michigan raised pavement markers, Michigan change in posted speeds, and Minnesota traffic signal installation. In addition, an extended empirical Bayes methodology was developed that allows for the use of explanatory variables (covariates) in a regression model along with a computer program for its implementation. The explanatory variables used in these case studies were roadway characteristics such as curvature, pavement width, shoulders, divided highway, and initial posted speed. The regression methodology also allowed for the examination of a non-linear relationship between vehicle-miles traveled (VMT) and accidents. There was strong evidence that such a non-linear relationship exists, casting doubt on the validity of accident rate-based methods. Although there was no significant difference among the methodologies, there was a significant increase in accidents for roads with initial posted speeds of <30 mi/h (48.3 km/h) in Michigan (23 to 27 percent) where speed limits were raised and a significant reduction in accidents (25 to 30 percent) in Minnesota after the installation of new traffic signals. Guidelines are presented for determining when the various methodologies are warranted and appropriate.

17. Key Words	18. Distribution Statement
Safety evaluation, high hazard location, regression-to-the-mean, before/after studies, raised pavement markers, signalization, speed change, empirical Bayes.	No restrictions. This document is available to the public through the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	102	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: Volumes greater than 1000 l shall be shown in m^3.				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5(F-32)/9 or (F-32)/1.8	Celcius temperature	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.71	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact)		
°C	Celcius temperature	1.8C + 32	Fahrenheit temperature	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

(Revised September 1993)

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

LIST OF TABLES (CONTINUED)

LIST OF TABLES (CONTINUED)

EXECUTIVE SUMMARY

The objective of this study was to compare accident analysis methodologies in highway safety analyses. Numerous methodologies have been applied to highway safety analyses. This study attempts to address the issue of which methodology is best and under what circumstances. This study examines three types of experimental designs commonly used in the field of highway safety. These analyses are applied to accident data from either previously researched studies or newly researched studies, supplemented by roadway inventory and accident history data available from the Highway Safety Information System (HSIS). Although the results of the data analyses are interesting, the main purpose of this research is to compare methodologies. Deficiencies in the data bases, violation of key assumptions, etc., make it difficult to support any strong conclusions based on the results. In each case, these deficiencies are exposed in the hopes that future research efforts will benefit from the understanding of these shortcomings.

Five data sets were analyzed using three accident analysis methodologies. Two of these methodologies exist in previous sources, but the third was developed in this research study. The methodologies are:

1. Classical (Class).
2. Empirical Bayes Without Covariates (EB).
3. Extended Empirical Bayes With Covariates (EBC).[1,2,3]

Two common practices in highway safety are the identification of potential high hazard locations and the evaluation of highway safety improvements based on site accident histories. Five data sets were identified for the application of these methods — two high hazard location analyses (Illinois and New York) and three highway safety evaluation studies (Michigan raised pavement marker study, Michigan speed study, and Minnesota signalization study).

Three of these data sets originated from previous studies and were analyzed for the Federal Highway Administration (FHWA) by other researchers. They are the New York data set, the Michigan speed study, and the Minnesota signalization study.[4,5,6] The remaining data sets were obtained from the HSIS by the Highway Safety Research Center (HSRC). Details of these case studies will be presented in this report, along with a summary of conclusions based on the comparison of the methodologies.

In addition to the application of these methodologies to case study examples, this report contains a historical review of the development of the methodologies, a description of the computational procedures of the methodologies, and an explanation of the required assumptions and shortcomings of the methodologies. The motivation behind such concepts as a comparison group and/or condition, a test for comparability, and regression-to-the-mean (RTM) bias is also presented, as well as recommendations on how to determine which method is most applicable to each individual case study.

SUMMARY OF FINDINGS

Illinois High Hazard Analysis

The data used in this study were produced from Illinois data maintained within the HSIS. Data were obtained from both the Illinois Roadlog File and the 1990, 1991, and 1992 Illinois Police Reported Accident File. The roadway functional class included major rural two-lane roads with average annual daily traffic (AADT) between 1,000 and 5,000. The new methodology developed in this study allowed for the use of explanatory variables or covariates. The covariates selected for this study were the roadway characteristics: (1) straight versus curved; (2) existence of shoulders; and (3) pavement width categorized as narrow (total pavement width < 32 ft [9.8 m]), moderate (total pavement width of 32 ft [9.8 m]), or wide (total pavement width of > 32 ft [9.8 m]).

Two variables were identified as causing a great deal of variability in exposure (vehicle-miles traveled [VMT] for this application), hence, a vital assumption required of the EB methods was potentially violated (exchangeability). One of these variables was posted speed and the other was section length. The data were, therefore, partitioned into two posted speed categories — < 55 mi/h (88.5 km/h) and ≥ 55 mi/h (88.5 km/h). Due to the existence of many short sections (almost 50 percent of the sections were < 0.1 mi [0.16 km]), only sections greater than this length were analyzed. Obviously, many of these extremely short segments represented intersections, but HSIS was not able to identify intersections.

Pavement width and curvature were significant covariates in only 1 year for the higher posted speed sections; otherwise, none of the roadway covariates were significant. The extended empirical Bayes (EB) methodology developed for this study reduced to the original EB methodology for this application. Because there was not a great deal of regression-to-the-mean bias in these data, there was little, if any, difference in the rankings of either accident frequencies or rates among these methodologies.

New York High Hazard Analysis

This case study used data that were originally for the evaluation of the effect of resurfacing on the safety of two-lane rural highways in New York.[4] Since there were no covariates available, the extended EB methodology simply reduced to the original EB procedure, so only the classical and the EBEST (empirical Bayes estimation of safety in transportation) methods were compared.

Thirteen years of non-intersection accident data (1975 through 1987) were available for 430 sites, along with AADT estimates. The original data set contained accident data reported in

0.1-mi (0.16-km) segments for each section. This data set contained a highly diverse set of sections with regard to traffic, and this fact may impinge on the strength of any conclusions drawn from the aggregation of such a diverse population.

The unit of 1 mi (1.61 km) was selected based on the accident frequencies in this data base. The classical and EB methods tended to agree as to which sites are ranked within the five highest accident frequency locations for all years. It is only within the lower five rankings that the methods tended to disagree. With regard to ranking by accident rate, there was very little agreement between the classical and EB methodologies. However, this was probably more of an indication of the variability of exposure (VMT) in these data than any true regression-to-the-mean adjustment. The EB method may be inappropriate for these data as the variability in VMT may be an indication of the violation of the exchangeability assumption.

Michigan Raised Pavement Marker Study

The data in this analysis was obtained from the Michigan data in the HSIS, maintained by the University of North Carolina Highway Safety Research Center. A total of 17 locations, comprising 56.11 mi (90.3 km) where raised pavement markers were installed on the center line in 1989, represent the treatment sites for this analysis. A total of 42 sites, comprising 146.28 mi (235 km) where raised pavement markers were not installed and no other significant changes were made in roadway geometrics or operations data after August 1, 1987, were selected as reference sites. Video discs maintained by the FHWA were reviewed to ensure that the reference sites were truly comparable with respect to roadway geometrics.

Accident data were collected for 2 years before and 2 years after the treatment. Daytime accidents at the same sites were selected as the comparison group for the treatment condition of nighttime accidents. Daytime and nighttime "exposures" were estimated by assuming that 25 percent of the AADT occurred at night and 75 percent occurred during the day.

Candidate covariates that were modeled using the extended EB methodology were: (1) major or minor arterial, (2) urban versus non-urban, (3) divided or undivided, and (4) narrow (total [left and right] shoulder width $\leq$ 16 ft [4.9 m] versus wide).

Models were also developed to explore the potential non-linear relationship between accident frequency and VMT by modeling non-linear functions of VMT as an explanatory variable in the extended EB methodology. There was substantial evidence that there is indeed a non-linear relationship between these variables. The only significant covariate appeared to be divided versus undivided highway — accident rates tended to be lower on divided highways.

All methods agreed that the treatment and comparison sites were comparable in the before and after periods, regardless of whether a comparison group (nighttime accidents at the reference site locations) or comparison condition (daytime accidents at the treated sights) was used. No significant change in accidents was found by any of the methods for either divided or undivided highways.

Michigan Speed Limit Study

The data in this analysis were obtained from a previous study entitled *A Comparison of Speed Zoning Procedures and Their Effectiveness*, sponsored by the Michigan Department of Transportation (MDOT) from September 1989 through September 1992.[5] Two analyses were conducted, one on 16 sites where the speed limits were raised, with a corresponding reference group of 22 sites where no speed change occurred, and a second analysis on 38 sites where speed limits were lowered, with a corresponding reference group of 47 sites.

The covariates used in these analyses were: (1) major or minor highway, (2) average lane width, and (3) posted speed before change. For sites where the speed limits were raised, initial posted speed categories were: (1) ≤ 30 mi/h (48.3 km/h), (2) from 35 to 40 mi/h (56.3 to 64.4 km/h), and (3) ≥ 45 mi/h (72.5 km/h). The categories for the sites where the speed limits were lowered were: (1) < 45 mi/h (72.5 km/h) and (2) ≥ 45 mi/h (72.5 km/h). The most significant variable was initial posted speed limit. There was also evidence of significant non-linearity between accident frequency and VMT.

None of the sites showed a significant treatment effect using any of the methods, except the sites with ≤ 30 mi/h (48.3 km/h) where the speed limits were raised. A 23 to 27 percent increase in accidents was estimated. However, the low frequencies and few sites in this category cast doubt on the degree of credibility of this conclusion.

Minnesota Signalization Study

The methodologies of this study were applied to Minnesota HSIS data by Jun Wang and were reported in an HSIS Summary Report (see appendix A). The treatment group consisted of 13 intersections where new traffic signal controls were installed. The reference group consisted of 79 sites that were defined to include intersections that were comparable to the treatment group with respect to daily entering traffic, number of approach legs, intersection configuration, etc. All study sites were examined using Minnesota's video disc system to verify comparability of sites with respect to roadway geometrics. The sites were examined using *Manual on Uniform Traffic Control Devices* (MUTCD) warrant criteria and it was found that 38 percent of the reference sites met at least one of the volume warrants and 14 percent met the accident warrants for signalization, though no signals were installed.

The study included accidents from the years 1985 through 1990, which occurred within 250 ft (76.2 m) of the test intersections on both major and minor roads. Accidents occurring during the construction period were excluded. The before and after treatment periods varied from 21 to 31 months.

The treatment sites had a higher accident rate than the reference sites before treatment, which is indicative of regression-to-the-mean potential. There was a 30 percent reduction in accidents after new traffic signals were installed, according to the classical methodology, and a 25 percent reduction with the EB method, both of which were statistically significant at the 0.05 level. For injury accidents only, the classical estimate was a 38 percent reduction and the EB estimate was a 23 percent reduction — again, both statistically significant. Although both methods found significant reductions, the EB estimate was considerably lower (more conservative) than that for the classical method.

A significant issue raised by this case study is the selection of reference sites. From a practical standpoint, an intersection in the reference group that does not satisfy at least one of the MUTCD warrants is not a site that could have been selected for treatment. However, does this have any statistical impact in violation of the exchangeability assumption? This issue warrants further investigation.

CONCLUSIONS AND FUTURE RESEARCH

One of the strongest conclusions to emerge from this study revolves around the collection of reference group data. Considerable effort was made to define and identify reference groups that were comparable using video discs and large automated accident and roadway inventory data bases. Yet, in most cases, there was little, if any, statistical difference in the conclusions drawn by either method. The question emerges then, is the cost and effort of collecting reference group data justified and, if so, under what conditions? That is, when is there sufficient RTM bias to warrant the additional effort and can a set of criteria be established to guide the researcher in this judgment before embarking upon this hefty data collection effort? It would be interesting to see a sensitivity analysis of just how sensitive (or robust) the classical methodology is to RTM biases.

Another area of concern that emerged in this study is the issue of the assumption of exchangeability. In many of the case studies, this assumption came into question with regard to traffic volumes, section length, or reference sites that were candidates for treatment, such as in the Minnesota signalization study. Again, a sensitivity analysis of the robustness of the EB methodology to departures from exchangeability might produce guidelines for the researcher prior to data collection to ascertain whether or not it is reasonable to assume that the reference group to be collected has any practical chance of being within acceptable ranges of exchangeability.

Several results indicated that there is some validity to the assumption of a non-linear relationship between accident frequency and VMT from the models examined in these case studies. Given that traditional accident analysis methods are based on accident rate, and that inherent in the calculation of rate is an assumed linear relationship between accident frequency and VMT, what is the impact of ignoring the non-linearity? Is there some range over which linearity can be assumed? If so, what is it? If there is a non-linear relationship, how does it impact traditional rate-based methodologies? This is a very important issue that has continuously surfaced in the literature, but which has never adequately been addressed. Given the results of these case studies and the strong support for non-linearity, it may be timely to conduct a thorough investigation of this relationship and its impact on the way things have always been done.

The issue of sample size once again emerged as a key consideration in accident analyses. There were some case studies where even significant treatment effects were questionable due to small sample sizes. But how small is too small? Can the sample size be too large so as to obscure practical significance? It would appear that a very useful tool would be a sample-size spreadsheet in which various entries could be changed (expected treatment effect, potential RTM bias, traffic volume ranges, etc.) and the table could automatically calculate recommended sample sizes. This effort would probably yield a low-cost, yet highly useful and visible, product.

The existing version of the BEATS (Bayesian estimation of accidents in transportation studies) program is not user-friendly and requires additional effort to even be useable by the statistically sophisticated researcher. The issue of who should use this complex methodology requires careful consideration. Because of the assumptions of exchangeability, etc., the methodology in the hands of someone who does not thoroughly understand its requirements has a potential for misuse and misapplication. The workshop participants seemed in agreement that two versions of the computer program should be developed. One version would be very user-friendly, similar to the format of the original BEATS program, and would only preform the ranking analyses. This version could be distributed to practitioners for use in high hazard location identification, etc. The second version could be less user-friendly, but easily executed by the researcher who has a good understanding of the statistical methodology and the impacts of the assumptions. This version should also include some graphics, such as likelihood functions and confidence intervals. This, too, could be a very low-cost effort at this point, but is really needed if this methodology is to be used at all. Given the effort and resources devoted to the development of this methodology, it would seem that this additional effort is highly warranted.

In conclusion, this study has attempted to apply existing accident analysis methodologies to five carefully selected data sets. The effort expended in data collection was substantial, yet from a practical standpoint, it did not make a difference in the conclusions reached by the methodologies. Before such effort is expended, guidelines should be developed for the researcher that indicate whether or not the collection of a reference group is warranted. The most critical need for future research appears to be in the area of developing such guidelines through sensitivity analyses, sample-size spreadsheets, etc. Also, the computer program should be enhanced for use by practitioners who do have the ability and desire to engage in this data collection effort and to use the extended EB methodology.

CHAPTER 1: INTRODUCTION

The objective of this study was to compare accident analysis methodologies in highway safety analyses. Numerous methodologies have been applied to highway safety analyses — some adopted from other disciplines such as medicine or psychology, and others developed specifically for highway safety research. The question of which methodology is best and under what circumstances has never been fully addressed. This study attempts to do this for three types of experimental designs that are commonly used in the field of highway safety. These analyses are applied to accident data from either previously researched studies or newly researched studies, supplemented by roadway inventory and accident history data available from the Highway Safety Information System (HSIS). Although the results of the data analyses are interesting, the main purpose of this research is to compare methodologies. Deficiencies in the data bases, violation of key assumptions, etc., make it difficult to support any strong conclusions based on the results. In each case, these deficiencies are exposed in the hope that future research efforts will benefit from the understanding of these shortcomings.

Five data sets were analyzed using three accident analysis methodologies. Two of these methodologies exist in previous sources, but the third was developed in this research study. The methodologies are:

1. Classical (Class).
2. Empirical Bayes Without Covariates (EB).
3. Extended Empirical Bayes With Covariates (EBC).[1,2,9]

Two common practices in highway safety are the identification of potential high hazard locations and the evaluation of highway safety improvements based on site accident histories. Five data sets were identified for application of these methods:

1. High hazard location studies.
   a. Illinois data.
   b. New York data.

2. Highway safety studies.
   a. Michigan raised pavement marker study.
   b. Michigan speed study.
   c. Minnesota signalization study.

Three of these data sets originated from previous studies and were analyzed for the Federal Highway Administration (FHWA) by other researchers. They are the New York data set, the Michigan speed study, and the Minnesota signalization study.[4,5,6] The remaining data sets were obtained from the HSIS by the Highway Safety Research Center (HSRC). Details of these case studies will be presented in this report, along with a summary of conclusions based on the comparison of the methodologies.

CHAPTER 2: HISTORICAL SUMMARY OF ACCIDENT ANALYSIS METHODOLOGIES

Traditional methodologies for either ranking data for high hazard location identification or highway safety evaluations are based on observed accident rates and frequencies. Rates are defined as accident frequencies divided by million vehicle-miles traveled (VMT), which is a measure of the annual average daily traffic (AADT) per mile of roadway. Ranking analyses are generally very simplistic — the candidate set of roadway sections or sites are ranked according to descending annual accident count and/or accident rate. Some analyses combine the rankings of the two in a weighted fashion.

Highway safety studies are typically based on accident frequencies before and after a highway treatment. One of the earliest methods used to evaluate a treatment effect is based on only two measurements — one before treatment and one for some period after treatment.[7] These measurements are frequently annual accident counts. Some time period is usually left between the before and after periods to allow for driver adjustment to the change, construction period, etc. This is by far the simplest, but weakest, methodology and there are serious deficiencies. For one thing, there is no "control" for any changes that might have occurred in the two time periods that were independent of the treatment, but which could affect the accident counts, like increase in traffic, severe differences in weather, etc. If a severe snowstorm occurred in the after period, accidents could have increased for this reason alone and a treatment that really improved safety would be masked by this unrelated event. There is also the issue of the regression-to-the-mean (RTM) phenomenon. Since it is highly likely that these sites were selected for treatment because they had an unusually high number of accidents before the treatment, accidents might decrease in the after period even without treatment, simply because these sites are "regressing" or returning to the true number of accidents. In these cases, a treatment that is really ineffective might be classified as effective because the accidents decreased — the decrease being merely an artifact of RTM and having nothing to do with the treatment.

Historically, although the RTM phenomenon was recognized for some time, it was the lack of control issue that first received attention.[8] The before/after evaluation with a comparison group or condition evolved. A comparison **group** is actually a group of sites selected as being similar enough to the treated sites to adequately reflect what would have happened to the treated sites had they not been treated. A comparison **condition** is an identified factor on which the accidents at the same sites can be partitioned so that one level of the partitioning reflects the treatment effect and the other represents conditions where the treatment would not be expected to have any effect. For example, in the evaluation of raised pavement markers, since the markers are thought to be more effective in reducing nighttime accidents than those occurring during the day, accidents at the same sites are summed over day and night periods. The daytime accident frequencies then serve as "controls" for the nighttime condition. Another often-used comparison condition is wet/dry, although it is more difficult to accurately define a "wet" accident from a "dry" accident.

The above methods assume that there is one point in time for both the before and after periods. When more than one point is available for the before period, the methodology can be extended to provide a check on the validity of the comparison group or condition.[1] Generally, this is done on yearly units. If two or more years of accident data can be retrieved, the before data can be subjected to a "test for comparability." If this test is failed, the safety treatment evaluation is questionable because the comparison group or condition is not behaving like the treatment group and may not be a valid control for before/after differences.

The methodology of empirical Bayes was applied to accident analyses to adjust for regression-to-the-mean. Although there are several variations on the application of this methodology, the basic philosophy is the same.[9,10,11] The above methodologies all function under the premise that the number of accidents observed in the before period are good "estimates" of what would be expected in the after period without the treatment. If regression-to-the-mean is a problem, i.e., if the most hazardous sites are treated and these sites were experiencing an unusually high number of accidents, then the before accidents do not represent good estimates of what to expect apart from treatment. The empirical Bayes methods all attempt to adjust for this by providing better estimates of what to realistically expect. These estimates adjust for regression-to-the-mean by providing some number between that observed and an estimate of the "true" mean for the treated sites. If the accidents are unrealistically high, the empirical Bayes estimate will "shrink" the observed mean closer to the "true" mean and hence the estimate of what to expect will be smaller than the observed before-accident count. This methodology is computationally quite complex and requires sophisticated software to compute.

CHAPTER 3: DESCRIPTION AND NUMERICAL COMPUTATIONS

Highway safety studies are typically based on accident frequencies before and after a highway treatment. The most elementary method is the simple before/after test. A test statistic is defined as:

$$T = \frac{D - C}{\sqrt{D + C}} \qquad \textbf{(1)}$$

where D = number of accidents at treatment sites after treatment and C = number of accidents at the treatment sites before treatment. This test statistic is asymptotically normally distributed. Any value exceeding 1.96 in absolute value is considered statistically significant at the 0.05 level. This method is extremely weak and not recommended; however, it was the general practice for some time. One of the many problems with this method is the lack of a control or comparison group to adjust for any confounding factors such as site differences over time, etc.

The before/after design with control or comparison condition has been the recommended practice. This analysis is based on accident frequencies in a two-by-two table ordered as below:

Table 1. Generic before/after with comparison group.

	Comparison	Treatment
Before	A	C
After	B	D

The "treatment effect" is measured using the odds ratio:

$$O.R. = \frac{A/B}{C/D} \qquad \textbf{(2)}$$

where A/B is the odds of the before-to-after accidents in the comparison group and C/D is the same odds for the treatment group. Any differences in the before-to-after period will cancel out, assuming that the same differences occurred in both groups. If the comparison group is a valid comparison group, this will be the case. The treatment effect is estimated as this ratio minus one (times 100 as a percent change); negative values imply a decrease in accidents due to the treatment. There are several candidate test statistics for testing statistical significance, but the one that will be chosen here is the likelihood ratio chi-square. The formula for the likelihood ratio chi-square is:

$$L.R.\chi^2 = 2\left(\sum_{i=1}^{2}\sum_{j=1}^{2} o_{ij} \ln\left(\frac{o_{ij}}{e_{ij}}\right)\right) \qquad \textbf{(3)}$$

where o_{ij} is the observed frequency and e_{ij} is the expected frequency. For the classical method, the expected frequencies are the same ones used in the Pearson chi-square, i.e., the product of the cells' corresponding marginal frequencies divided by the total frequency. For example, the expected number of before accidents in the comparison group for table 1 would be:

$$e_{11} = \frac{(A + C) * (A + B)}{(A + B + C + D)} \qquad \textbf{(4)}$$

Sometimes the comparison is not a group of sites, but the same sites under different conditions assumed not to be affected by the treatment. For example, in the evaluation of raised pavement markers, daytime accidents at the same site are sometimes used as a comparison group for nighttime accidents, assuming the markings affect nighttime safety more than day.

The assumption that the comparison group is valid can be tested if more than 1 year of data is available before treatment. This design is called a before/after design with test for comparability. Typically, the test for comparability consists of a chi-square analysis performed on the table of counts for the treatment and comparison group for the before period only. For example, suppose 3 years of accident counts are available before treatment for both the treated and comparison groups (conditions). The before data are represented by table 2, where B1, B2, and B3 represent the three before years.

Table 2. Generic before table for 3 years of before data.

	Comparison	Treatment
B1	A	D
B2	B	E
B3	C	F

Non-significance in this case would imply that there is no difference (comparability) between the treatment and comparison groups before treatment. This is obviously a desirable feature as the comparison group is intended to represent a group similar to the treatment group, but which was not treated. In the case of using daytime accidents at the same sites as a comparison (group) condition, this group represents accidents at the same sites, but for a condition (day) that is assumed to not be affected by the treatment.

The classical analysis would actually use the likelihood ratio chi-square rather than the Pearson chi-square because of the additive nature of that statistic. An Analysis of Variance (ANOVA)-like table can then be constructed that partitions the total chi-square for all data (before and after periods) into a portion representing comparability of the groups in the before period and a test for the treatment effect.

The test for the treatment effect is the likelihood ratio chi-square for the table collapsed over the years before and after (table 1). However, this test is only justified if the treatment and comparison groups pass the comparability test. In cases where only 1 year of data is available, this check for comparability cannot be done and one has to assume that the sites are comparable.

Empirical Bayes Methodologies and Regression-to-the-Mean

Although the comparison group concept was a good one for adjusting for differences over time, there is another very problematic factor in highway safety studies that the comparison group concept does not address. This very serious confounding factor is the regression-to-the-mean (RTM) phenomenon. This phenomenon occurs because of the non-random process of treatment-site selection. Because of this biased selection process, there exists a very high probability that a reduction in accidents might be observed even if these sites were left untreated.

Classical statistical methods assume that the before accident history of a site is a suitable estimate of what to expect at that site in the after period. This is clearly not the case. The fact that these sites were the most hazardous could be an indication that they were experiencing an unusually high number of accidents at this time, hence the before accident experience is inflated and not a good estimate of what one might expect had the treatment been withheld.

Significant progress has been made in developing statistical methodologies to adjust for this RTM phenomenon, which cannot be controlled for through experimental design. Hauer's pioneering efforts to address this problem have led to an entirely new application of the empirical Bayes methodology. An earlier FHWA study developed a methodology and computer program — BEATS (Bayesian estimation of accidents in transportation studies) — to adjust for regression-to-the-mean bias.[2] A brief description of other empirical Bayes methods will be discussed here.

The empirical Bayes estimation of safety in transportation (EBEST) methodology is an empirical Bayes method for estimating the expected accident rate during the after period that is more reliable than the classical estimate in that it adjusts for regression-to-the-mean bias.[2] The methodology was developed under a contract with the FHWA and serves as an initial effort in improving upon the previously proposed empirical Bayes procedure of Hauer.[9] It differs from the Hauer procedure in two ways:

1. The estimate is derived by the method of maximum likelihood estimation as opposed to the method of moments.

2. It incorporates a measure of exposure in the estimation process.

The simpler EBEST procedure has been extended and improved upon by the hierarchical Poisson regression model that enables the use of covariates such as roadway geometrics, environmental factors, etc., as well as a researcher-designated prior shrinkage value.[3] These extensions enable a more accurate estimate of the posterior moments that is especially important when the number of sites is small. This method also allows us to check the assumption of linearity between the exposure and accident counts and to incorporate non-linear relationships between these variables if warranted. This extended method was applied to the data sets in this study.

Arnold and Antle developed a methodology that uses the method of moments in an empirical Bayes approach analyzing driver accident liability.[10] These estimators are, however, modified according to the proposed method and these modifications cause the method to fail as a general approach in analyzing Poisson data.

Davis proposed a methodology that uses a hierarchical Bayes beta-binomial model to tackle the very difficult problem of accident analysis with unknown exposure values.[11] These types of problems typically arise when the safety entity of interest is not a road section where exposure estimates are feasible, but rather entities such as drivers or vehicle types, etc. In this specific application, driver age and culpability were the variables of interest. The "exposure" of the driver by age group is chiefly unattainable. To resolve this problem, an induced exposure model is used that makes many questionable assumptions. The key assumption is that the total number of two-vehicle accidents for which an older driver was at fault and the total number of two-vehicle accidents for which the driver was an innocent victim are independent. If defensive driving is effective, for example, this assumption seems unlikely. If, in fact, these variables are correlated, the model would be using the wrong benchmark for the relative exposure of the older group. Also, the variances would not be additive.

Hauer proposed an empirical Bayes method based on the method of moments and extended it for use with covariates.[4,12] This method differs from the EBEST regression method primarily in that it is not a maximum likelihood procedure and the covariate extension is not modeled in the same way.

Comparison of Methodologies

The before/after methodology without a comparison group is the weakest of all before/after analyses and should be avoided. It is only valid if there is absolute certainty that there are no before/after differences that could confound a treatment effect, such as changes in traffic volume, weather, etc. Otherwise, every attempt should, in some way, be made to control for the effect of any changes over time. If there is no reason to suspect a regression-to-the-mean effect, the comparison group/condition methodologies are fairly sound and established. There may be cases when the treatment group truly represents an unbiased and nearly random sample. For example, if the intended treatment is changing posted speed limits and the sites selected for the change were not selected based on their accident histories, but rather for policy or other reasons, regression-to-the-mean may be a non-issue.

The effect of not having exposure data has not really been tested. That is, a comparison between rate-based and non-rate-based analyses on the same data has not been explored. The above-mentioned study did find some evidence, as have others, that there is a non-linear relationship between accident frequency and vehicle-miles traveled.[12] If this is indeed the case, rate-based analyses could be erroneous since the very calculation of rate assumes a linear relationship between these variables. In this case, it is possible that methodologies that successfully adjust for exposure are actually preferable.

The comparison group methodology's weakest point is in defining and collecting data on a truly comparable group. The test for comparability can give some reassurance if enough before data are available and reliable. Sometimes other factors such as site modifications, construction, etc. make it difficult to obtain these data. The comparison condition methodology's weakest point is that the condition may be difficult to define — like wet versus dry or day versus night. What about dusk and dawn? Does the fact that some precipitation occurred during that day really mean the pavement was wet all day? In some cases, the comparison condition may be straightforward. An example of this would be the evaluation of the impact of the presence of a parked police car on accidents. If the data are partitioned into the hours the vehicle is present and the hours it is not, the comparison condition is clearly defined. However, other factors may confound the evaluation that have nothing to do with the statistical design — such as drivers avoiding the area, being forewarned by other drivers, etc.

CHAPTER 4: ASSUMPTIONS REQUIRED OF THE EBEST METHODOLOGY

Although regression-to-the-mean occurs to some extent in all cases, it only causes problems in studies where the treatment assignment is subject to selection bias, as when the sites selected for treatment generally have higher accident rates than those not selected. In the presence of this bias, the net accident rate at the treatment sites would be expected to drop, even in the absence of the treatment's effect. Hence, this amount must be estimated and removed. There are several tools for doing this, the key one being the availability of more information in the form of a legitimate "reference" group. The reference group represents the population of sites not selected for treatment, but which also might have been selected. Together, the treatment and reference groups must represent a good sample of the entire population of sites that the treated group came from.

Suppose a candidate set of sites has been proposed for some highway improvement. However, when examining the accident rates, it is clear that not all sites need the improvement; they were just similar in some characteristic to the ones that really needed the improvement. Typically, the sites would be ranked by accident rate or frequency and a certain number of sites with the highest rates would be selected. In this scenario, the pool of sites not selected may well represent a legitimate reference group. This is not always the case and the unselected sites could possibly be biased as well, so care must be taken in examining these sites.

A method of doing this would be to make a simple estimate of an upper bound for the magnitude of the selection bias by calculating the average accident rates in the treatment and in the reference groups. For roadway sections of various lengths, the rate is often defined as:

$$Rate = \frac{number\ of\ accidents}{VMT} \qquad (5)$$

where

$$VMT = \frac{AADT * 365 * sl}{1{,}000{,}000} \qquad (6)$$

AADT is the average annual daily traffic and sl is the section length.

Regression downward of the treatment data would occur only if the treatment group rate exceeds the reference group rate. Suppose the treatment group rate was 286/287.3 = 0.99 and the reference group rate was 656/2906.4 = 0.23. Since the treatment group rate is

Preceding Page Blank

higher than the reference group rate, some regression-to-the-mean can be assumed. The question is how much RTM and will it make a practical difference. In this example, if the reference rate is used to estimate the number of accidents that would have occurred at the treatment site apart from any regression-to-the-mean, then 287.3 times 0.23 or 66 accidents would be a lower bound for the RTM adjustment. That is, using the adjustment, the most we could expect is that the 286 accidents would be reduced to 66 accidents. If this reduction would appear to be substantial enough to change any conclusions that might be drawn about the treatment effect, then there is no point in proceeding with the regression-to-the-mean adjustment.

To determine whether or not this would change the statistical significance, this estimated maximum reduction can be inserted in table 1 in place of C, the odds ratio computed (equation 2) and the test of statistical significance run (equation 3). This RTM "pretest" would be a good indicator of whether or not to use the more complicated EB methodology. It would also be an indicator that the simpler, classical method will not be affected by RTM bias.

The simple method described above provides an upper bound for the amount of adjustment, mainly because full shrinkage (adjustment) to the mean does not occur in RTM situations due to true differences among the traffic sites. The EBEST procedure estimates a shrinkage coefficient, b_i, which is between zero and one and weights the amount of adjustment of the observed accident rate to the true accident rate for that site (true rate being based on an estimation using the treatment and reference group sites). The closer b_i is to one, the more the observed rate will be "shrunk to" the true rate. The closer this value is to zero, the less adjustment that will occur. Therefore, another indicator of the degree of RTM in the data is an examination of the average b_i's for the treated sites. However, this computation is only available after the method has been applied and, therefore, does not provide an advanced indicator of whether or not to proceed with the EB analysis.

An important underlying assumption in the EBEST methodology is exchangeability. This assumption basically requires that the reference and treatment groups represent samples from the same population. In practice, however, it is not always obvious when violations of this assumption exist. There are some exploratory techniques that can be used to check for the validity of this assumption. One method is to examine residual plots against the logarithm of exposures. The residuals are the deviations of the observed and expected accident frequencies based on the model. Exposure is generally VMT. If there are any "patterns" in these residuals (i.e., anything other than a shotgun scatter), some doubt is cast on the validity of the assumption. For example, if the residuals for the reference group are all clustered in one area and the residuals for the treatment group are in another area, the groups may not be comparable.

A second way is to put a dummy variable in the model on the right-hand side as zero for the reference group and one for the treatment group. This dummy variable should be in addition to the candidate covariates. If this coefficient is statistically significant, it means that the treatment and reference groups are "different" in some way. Finally, one can examine the mean and variance of VMT for each level of the covariates for the treatment and reference groups. These should be "comparable" and can actually be tested for comparability using an independent t-test between the treatment and reference group means.

These methods have not been fully explored and at the moment are only indicators, or red flags, to alert us as to the possibility of violation of the exchangeability assumption. The problem is that there may be some covariates that are not in the model. This is a problem in ordinary least-squares regression as well. There is always the chance that some key factor has been ignored that could account for any non-exchangeability.

Another concept that is critical to the successful application of EBEST is exposure. Exposure, as defined for variable-length sections (e.g., VMT), is just a specific type of exposure measure. Exposure is some measure that provides constants of proportionality for accidents when sites differ on some basic factors, such as traffic density, section length, etc. Sites with higher AADT's, for example, would experience more accidents for this reason alone, but not necessarily have higher accident rates. There has been some question as to whether or not exposure, specifically VMT, is linearly related to accident counts. If not, perhaps it should be modeled as a covariate, which will be discussed next, and allowed to assume non-linear functional forms. This was explored in the current study.

The original EBEST methodology just described has been extended by Morris and Christiansen to be used with covariates — roadway characteristics or variables that might result in differing measures of treatment effectiveness.[3] Covariates allow sites to differ depending on the covariate values. The logarithms of the accident rate are modeled in the new computer program as linearly dependent on these covariates, analogous to ordinary least-squares regression. Examples of potential covariates in highway safety studies are highway type (urban/non-urban), pavement width, divided or undivided highways, etc. In this method, the treatment and reference sites' true mean accident rate before treatment is assumed to be a function of the covariates in a regression-type model in the prior distribution. Tests of significance of these covariates can be made. In this way, an adjusted expected rate can be computed as a function of the covariates. This model allows for the incorporation of subsets that satisfy the exchangeability assumption when the total group may not. For example, let's say the pool of mixed urban and rural site rates are not exchangeable, but within each classification they are exchangeable. Thus, urban sites alone are exchangeable and rural sites alone are exchangeable. By incorporating the urban/rural covariate, the exchangeability assumption is met.

Both the treatment and reference group before rates are adjusted using this model and they are then inserted into the tables used in the classical methods. The classical methods are then performed as usual using the revised (or shrunken) estimates.

CHAPTER 5: ILLINOIS ROADWAY HIGH HAZARD ANALYSIS

The data used in this study were produced from Illinois data maintained within the HSIS. Data from both the Illinois Roadlog File and the 1990, 1991, and 1992 Illinois Police Reported Accident File were used. The roadway segments in this analysis were all classified as functional class 30, "non-urban and urban areas major highway." They were all rural, two-lane roads with AADT between 1,000 and 5,000. If shoulders were present, they were either paved or sod as defined by the variable shoulder type. The covariates selected for this analysis were:

1. CURV - 0 = straight and 1 = curved as defined by the variable horizontal curve radius. Any non-zero value was defined as curved.

2. SHLD - 0 = no shoulder, 1 = shoulder.

3. WIDE - 1 = total pavement width < 32 ft (9.8 m).
   2 = total pavement width of 32 ft (9.8 m).
   3 = total pavement width > 32 ft (9.8 m).

In preliminary analyses, two variables were identified as causing a great deal of variability in exposure (VMT for this application). Such variability violates the assumption of exchangeability required of the EB methodology and indicates a blending of non-homogeneous populations. One of these variables was posted speed and the other was section length. The data were partitioned into two posted speed categories — < 55 mi/h (88.5 km/h) and ≥ 55 mi/h (88.5 km/h).

The accident data were summed over road segments that were homogenous with respect to the above covariates. For example, accidents were summed along a road section until one of the covariates changed, e.g., a straight section became curved, lost the shoulder, etc. As a result, section lengths varied dramatically from many short sections (almost 50 percent) < 0.1 mi (0.16 km) in length to sections as long as 7 mi (11.3 km). Clearly, some of the short sections represented intersections. Unfortunately, the accident data did not contain information as to whether they occurred at intersections or not. The best that could be done was to try to isolate the intersections from the rest of the data by partitioning the data into two segments — "short" roadway sections ≤ 0.1 mi (0.16 km) and "long" sections > 0.1 mi (0.16 km). The distribution of section lengths and VMT is listed in table 3.

Preceding Page Blank

Table 3. Descriptive statistics on section length and VMT (in miles).

Variable	< 55-mi/h short		< 55-mi/h long		≥ 55-mi/h short		≥ 55-mi/h long	
	n	mean	n	mean	n	mean	n	mean
sect lngth	114	0.047	165	0.447	371	0.050	833	1.002
90 VMT	114	0.550	165	5.410	371	0.500	833	10.86
91 VMT	114	0.600	165	5.950	371	0.550	833	11.95
92 VMT	114	0.660	165	6.500	371	0.610	833	13.02

1 mi = 1.61 km

Regression Results

The extended empirical Bayes estimation of safety in transportation (EBEST) methodology provides a means of estimating the accident rate for a given site adjusted for regression-to-the-mean bias and incorporating roadway features (covariates) into the analysis. The site means are assumed to be a function of those covariates that serve the role of independent or x-variables in a regression model. These covariates can be tested for significance in the analysis. For the high hazard identification problem, the population of sites in the sample are assumed to represent both the potential treated sites and the reference sites in the "before" period of a traditional before/after study. These sites are then ranked according to their EB estimates as opposed to the observed accident rates or frequencies. The analysis presented here will compare the EB ranking results with the classical ranking methods for the 3 years, 1990, 1991, and 1992, for four subsets of the data:

1. Less than 55-mi/h (88.5-km/h) posted speed short segments.

2. Less than 55-mi/h (88.5-km/h) posted speed long segments.

3. Greater than or equal to 55-mi/h (88.5-km/h) posted speed short segments.

4. Greater than or equal to 55-mi/h (88.5-km/h) posted speed long segments.

The revised Bayesian estimation of accidents in transportation studies (BEATS) program was first run using all three covariates. Based on which covariates were significant, the program was rerun on those covariates only and the EB estimates from that model were used in the rankings. Since traditionally high hazard location rankings are done on both accident frequencies and accident rates, both EB estimated accident frequencies and rates were used in the ranking analyses. Table 4 presents the results from the EB analyses.

Table 4. Regression results of the EB analyses.

Year	Model	< 55-mi/h short		< 55-mi/h long		≥ 55-mi/h short		≥ 55-mi/h long	
1990		Est	T	Est	T	Est	T	Est	T
	Int.	-2.59	-2.18	-0.67	-1.94	-0.35	-0.67	-1.42	-6.54
	Curv	-0.85	-0.76	-0.84	-0.86	0.09	0.24	0.53	2.31
	Shld	-0.17	-0.22	-0.28	-1.85	0.79	1.61	0.08	0.38
	Wide	0.96	2.29	-0.55	-2.47	-0.65	-3.07	-0.16	-3.04
1991									
	Int.	-2.09	-2.25	-1.02	-3.09	-0.71	-1.16	-1.19	-6.26
	Curv	-0.08	-0.17	-0.80	-0.90	0.32	0.79	0.26	1.03
	Shld	0.20	-0.33	0.20	0.67	1.36	2.35	-0.32	-1.81
	Wide	0.48	1.47	-0.16	-1.11	-0.48	-2.12	-0.06	-1.19
1992									
	Int.	-2.15	-1.34	-1.13	-3.14	-	-	-1.46	-15.3
	Curv	-0.71	-0.51	-0.49	-0.53	-	-	0.27	1.09
	Shld	-0.39	-0.36	-0.02	-0.07	-	-	-0.17	-0.91
	Wide	0.87	1.50	-0.14	-0.91	-	-	-0.11	-2.10

The dash in the table indicates that the data would not run on BEATS.
1 mi/h = 1.61 km/h

The column labeled "T" is the coefficient estimate divided by its standard error and it can serve as a statistical test of significance. Any values greater than 1.96 were defined to be "statistically significant" and were interpreted as that coefficient being important in the prediction of the accident rates. The conclusions from these statistics are:

1. For short segments with posted speeds < 55 mi/h (88.5 km/h):

- In 1990, only pavement width was significant (and positive) for short segments with < 55-mi/h (88.5-km/h) posted speeds. Neither curvature nor the presence of shoulders was significant. The model was rerun with width only, and this model was used to generate the estimates used in the rankings. When the model was rerun with pavement width only (see table 5), the coefficient was still positive. This means that collinearity was not the cause for the counter-intuitive sign — i.e., with only one variable it is appropriate to interpret a positive sign as accidents increasing when pavement width increases. The reason for this is not known. It has been postulated (see appendix B) that with some sections 0.1 mi (0.16 km) long or less, these models may be meaningless on this data set. For this reason, results for the short sections will only be provided as examples of interpreting model results and no conclusions will be drawn for these sections.

- No variables were significant for 1991 and 1992. The original EB method without covariates was used to produce the estimates for the ranking.

2. For long segments with posted speeds < 55 mi/h (88.5 km/h):

- In 1990, only pavement width was significant. The signs for both shoulders and pavement width coefficients were negative, which would imply that the presence of shoulders and wider pavements reduce accidents. When the model was rerun using only pavement width, the coefficient for pavement width was negative (see final models, table 5).

- No variables were significant for long segments with posted speeds < 55 mi/h (88.5 km/h) for any of the years 1991 and 1992. The original EB method with no covariates was used to produce EB estimates for ranking these years.

3. For short segments ≥ 55 mi/h (88.5 km/h):

- In 1990, pavement width was significant for short segments with posted speeds ≥ 55 mi/h (88.5 km/h), but the sign was negative — opposite of the direction for the short segments with < 55-mi/h posted speeds. When the final model was rerun (table 5), the sign was still negative. It is possible that because these segments had posted speeds ≥ 55 mi/h (88.5 km/h), there were fewer extremely short sections to cause interpretation problems. The only way to really know this would be to be able to identify and isolate intersections from non-intersections — a variable that was not available on the HSIS.

- In 1991, both the presence of shoulders and pavement width were significant. Note that the coefficient for the presence of shoulders is positive, implying that accidents increase when shoulders are present, another counter-intuitive conclusion. However, this may be an artifact of the collinearity between the presence of shoulders and pavement width. Recall that the pavement width is **total** pavement width, which includes shoulder width. This is a good example of why it is wrong to directly interpret signs when more than one variable is in the model. Such indicators, however, do suggest that model diagnostics for the extent of the collinearity are warranted and should be developed in an enhanced version of the computer program. Note that for 1991, the presence of shoulders was still positive in the final model, but given the problem noted in comment 1 with the short-section data, it is difficult to tell if this is caused by the extremely short segments or the potential collinearity.

- We were not unable to get the 1992 data set to run on the computer program and we do not know why.

4. For long segments ≥ 55 mi/h (88.5 km/h):

- In 1990, both curvature and pavement width were significant in predicting accident rates on long segments with posted speeds ≥ 55 mi/h (88.5 km/h). Coefficient signs did not change in the final model (table 5).

- No variables were significant for this group in 1991 and only pavement width was significant (and negative) in 1992.

Table 5. Final models for cases where model variables were significant.

Model	1990 < 55 -mi/h short		1990 < 55 -mi/h long		1990 ≥ 55 -mi/h short		1990 ≥ 55 -mi/h long		1991 ≥ 55 -mi/h short	
	Est	T	Est	T	Est	T	Est	T	Est	T
Int.	-2.93	-2.94	-0.72	-2.51	0.39	-0.67	-1.35	-12.6	-0.62	-1.02
Curv	-	-	-	-	-	-	0.53	2.33	-	-
Shld	-	-	-	-	-	-	-	-	1.53	2.84
Wide	1.01	2.52	-0.27	1.84	-0.67	-1.52	-0.16	-3.03	-0.53	-2.39

1 mi/h = 1.61 km/h

Comparison of Methodologies

Due to the problems with the short-section data, the remainder of this report will focus on the long sections only. Tables 6 and 7 contain analysis results for long sections ≥ 55 mi/h (88.5 km/h). Table 6 lists the site numbers of the top 10 sites ranked according to accident frequency for the 3 methods, and table 7 lists the descriptive statistics for the top 10 EBC ranked sites. The greatest differences occurred in 1992, hence, we will focus on this year.

Four sites ranked by the observed frequencies were not included in the EBC or EB rankings (550, 351, 180, and 462 with observed frequencies of 15, 14, 12, and 12, respectively). Most notable were sites 351 and 462, which went from being ranked fifth and ninth by observed frequencies to numbers 41 and 44 by the EBC estimated frequencies, respectively.

Conversely, 4 sites that had high EBC ranks were not ranked in the top 10 by their observed ranks. They were 819, 594, 546, and 143, with observed accident frequencies of 12, 11, 11, and 11, respectively. Their rankings based on observed frequencies were 11, 24, 23, and 15, respectively. Examining the descriptive statistics for the top 10 ranked EBC sites in table 7, we note that although there was a difference in the rankings, there was not a tremendous difference in expected and observed frequencies, i.e., there was not much shrinkage at these sites. The average shrinkage for the entire data set was 0.574 however, and the rankings by rates will show a greater difference than the rankings by frequencies.

Table 6. Top 10 accident frequency sites for ≥ 55-mi/h (88.5-km/h) long sections.

	1990			1991			1992		
Rank	Class	EB	EBC	Class	EB	EBC	Class	EB	EBC
1	387	387	387	143	143	-	744	744	744
2	143	143	143	358	358	-	735	735	735
3	204	204	204	351	546	-	211	387	387
4	363	363	363	546	395	-	550	211	211
5	358	358	358	395	647	-	351	819	819
6	367	367	367	647	396	-	387	594	594
7	132	546	546	396	744	-	180	396	546
8	169	169	169	365	541	-	396	546	531
9	546	395	132	381	484	-	462	143	396
10	395	132	395	484	365	-	531	169	143

The dash in the table indicates that since there were no significant covariates, EBC estimates are the same as in the previous column, namely EB.

Table 7. Comparison of sites that differed in ranks for ≥ 55-mi/h (88.5-km/h) long sections (1992 frequencies).

	1992 Frequencies			Roadway Variables			
EBC rank	EBC	Class	EB	VMT	Curv	Shld	Wide
1	21.76	22	21.6	107.00	no	yes	<32 ft (9.8 m)
2	16.88	17	17.1	111.68	no	yes	<32 ft (9.8 m)
3	13.82	14	13.9	78.50	no	yes	32 ft (9.8 m)
4	13.36	15	13.4	41.43	no	yes	32 ft (9.8 m)
5	11.76	12	11.5	64.54	no	yes	32 ft (9.8 m)
6	11.40	11	11.4	109.17	no	yes	>32 ft (9.8 m)
7	11.31	11	11.3	89.62	no	yes	32 ft (9.8 m)
8	11.19	12	11.0	41.44	no	yes	<32 ft (9.8 m)
9	11.13	12	11.4	51.03	no	no	>32 ft (9.8 m)
10	11.12	11	11.3	85.53	no	yes	>32 ft (9.8 m)

1 mi = 1.61 km

Tables 8 and 9 provide analysis results for rankings on accident rates for long sections ≥ 55 mi/h (88.5 km/h). In 1990, 3 sites that were ranked in the top 10 according to the observed accident rate were not ranked in the top 10 by the EB methods, namely, sites 772, 573, and 716. Moreover, site 716 was ranked 41 by the EBC procedure. Four sites that were ranked in the top 10 by the EBC method were not ranked according to the observed rate rankings, namely sites 296, 334, 363, and 255. Of these, the most dramatic difference was the ranking of site 255, which went from 8 by EBC to 61 by the observed rates. It is of interest to note that the EB method without covariates differed more from the EBC method than the classical method since 6 of the sites ranked in the top 10 by EB were not in the top 10 using EBC. All sites were ranked within the top 25 by EBC however, so the difference in ranks was not as dramatic. In 1992, 5 sites that were ranked in the top 10 according to observed ranks were not in the EBC top 10 — the most dramatic difference being site 315, which was ranked 78 by EBC. Five of EBC's top 10 sites were not in the classical top 10. Since the most dramatic difference in the rate rankings occurred in 1990, table 8 focuses on 1990 rates. Note that the shrinkage of the estimated rates is substantial for most sites using EBC, but again, not as dramatic as EB. The average shrinkage for this EBC model was 0.554. Only curvature and width were significant covariates in this model. There was an even mixture of curved and straight roads in the top 10 EBC sites, but all sites had shoulders and all but 1 were < 32 ft (9.8 m) wide.

Table 8. Top 10 accident rate sites for ≥ 55-mi/h (88.5-km/h) long sections.

	1990			1991			1992		
Rank	Class	EB	EBC	Class	EB	EBC	Class	EB	EBC
1	187	262	187	78	351	-	493	351	351
2	262	17	262	313	468	-	224	111	462
3	52	776	496	315	17	-	366	462	111
4	496	52	52	17	78	-	351	366	550
5	17	153	296	351	341	-	277	234	234
6	772	363	334	341	288	-	303	550	366
7	153	25	363	307	61	-	174	277	277
8	573	2	255	468	348	-	111	201	201
9	716	222	17	288	439	-	700	493	493
10	131	76	131	61	557	-	315	326	326

The dash in the table indicates that since there were no significant covariates, the EBC estimates are equal to those in the previous column, namely EB.

Table 9. Comparison of sites that differed in ranks for ≥ 55-mi/h (88.5-km/h) long sections (1990 frequencies).

	1990 Frequencies			Roadway Variables			
EBC rank	EBC	Class	EB	VMT	Curv	Shld	Wide
1	1.01	4.86	0.56	0.62	yes	yes	< 32 ft (9.8 m)
2	0.91	3.88	0.84	1.55	no	yes	< 32 ft (9.8 m)
3	0.75	2.49	0.42	0.80	yes	yes	< 32 ft (9.8 m)
4	0.75	2.50	0.69	2.00	no	yes	< 32 ft (9.8 m)
5	0.70	1.78	0.41	1.12	yes	yes	< 32 ft (9.8 m)
6	0.67	1.44	0.40	1.39	yes	yes	< 32 ft (9.8 m)
7	0.67	0.78	0.65	25.58	no	yes	< 32 ft (9.8 m)
8	0.66	1.37	0.28	1.46	yes	yes	< 32 ft (9.8 m)
9	0.64	2.44	0.76	2.46	no	yes	≥ 32 ft (9.8 m)
10	0.63	1.92	0.58	2.08	no	yes	< 32 ft (9.8 m)

1 mi = 1.61 km

Summary and Conclusions

In summation, this data analysis has illustrated several important issues that arise in doing a high hazard analysis on roadway sections. These issues are:

1. Roadway section lengths need to be defined in such a way as to render a meaningful comparison. Although the extended BEATS methodology can adjust for site differences that may account for violation of the exchangeability assumption, the absence of important covariates that were missed or simply not in the data base can still cause problems with the exchangeability assumption. This was addressed in some of the general comments from the expert panel workshop reports (appendix B).

2. Collinearity can be a problem and it is important that some type of collinearity diagnostics be made available to the researcher to identify such problems. The research in this area, however, is still incomplete and it was beyond the scope of this project to address this issue. At present, the researcher can only go to some external ordinary least-squares package, such as the Statistical Analysis System (SAS), and try to identify collinearity in the independent variables by using a linear model on accident rates. The collinearity diagnostics for the independent variables should not be greatly affected by this procedure, although this is not a known certainty.

3. Ranking of sites by accident frequency is a questionable application based on EB estimates because the expected frequencies are generally non-integers. Ranking by accident rate is more plausible. Yet, in practice, high hazard location studies are often based on rankings of both frequency and rate.

4. The BEATS program is sensitive to the data and sometimes will not run. Further research is needed to explain why this occurs and what, if anything, can be done about it. The errors listed by the program when this occurs do not give any indication as to whether there is a problem in convergence, degeneracy, etc.

5. The HSIS data base is a useful source for producing large, multi-state data bases for studies such as this. However, the inability to identify intersections is a serious constraint for ranking purposes.

6. Basically, very little practical difference was found between the EB and classical methods in rankings within the top five positions. Although there were some dramatic differences in the ranking positions, there was almost no difference in expected frequency or expected rates. Perhaps this is due to not enough regression-to-the-mean to cause a serious problem. Because EB methods are difficult to apply to frequency rankings, the classical methods may be best when there is little RTM. On the other hand, if there is a lot of RTM or if there is a non-linear relationship between accident frequencies and VMT, the EB methods may be best. A validation method that was not applied to these data, but which has potential, is to develop models based on two consecutive years and to use them to predict estimates for the third. Then, a comparison can be made of the estimated ranks from these procedures to test the robustness of the rankings.

CHAPTER 6: NEW YORK ROADWAY HIGH HAZARD ANALYSIS

The data in this case study were originally used to evaluate the effect of resurfacing on the safety of two-lane rural highways in New York.[4] A group of 430 reference sites was selected for this evaluation. Hauer revisited these 430 sites and developed a method for identifying high hazard locations, which he terms PIL's (priority investigation locations). The purpose of this analysis is to apply the empirical Bayes methodology developed in this study to this same data set. Since there were no covariates available, the extended EB methodology simply reduces to the original EB procedure so that only the classical and EBEST methods will be compared.

The extent to which these results can be compared to Hauer's is questionable due to a difference in the definition of section length — as will be noted in the following report of the results. Thirteen years of non-intersection accident data (1975 through 1987) were available for these 430 sites, along with average annual daily traffic (AADT) estimates. The original data set contained accident data reported in 0.1-mi (0.16-km) segments for each section. The sections varied in length from 0.2 mi to 17.1 mi (0.32 km to 27.52 km). The AADT's ranged from 73 to 14,145, and million vehicle-miles traveled (MVMT) ranged from 0.8 to 516 (1.3 to 830.4 million vehicle-kilometers traveled [MVKT]) among these 430 sections. It is obvious, therefore, that this data set contained a highly diverse set of sections with regard to traffic and this fact may impinge on the strength of any conclusions drawn from the aggregation of such a diverse population when using VMT as a measure of exposure in the EBEST methodology.

In spite of these concerns, the task at hand was to apply the EBEST and classical methodologies to this data set and to compare the results. Had additional information been available regarding site characteristics, the covariate portion of the extended EBEST methodology may have been able to adjust for the traffic diversity among these sites. For example, if highway functional class had been available, this could have been treated as a categorical covariate. Sections within a given functional class would be expected to be more homogeneous in terms of traffic.

However, such data were not available, so all sites had to be treated as samples from the same population — a highly questionable assumption. Recall that violation of this assumption (hence violation of the assumption of exchangeability) can have serious consequences on the application of the EBEST method. Also, the huge variability in traffic (exposures) translates into highly variable and unstable empirical Bayes estimates. Indicators of this variability will be presented in this report.

The first issue to be addressed was section length. One-tenth of a mile (0.16-km) sections are far too short to produce any useful accident analyses. Furthermore, one has to question the accuracy of such accident reports. Those familiar with accident reporting know the limitations of being able to identify the accident location to within 0.1 mi (0.16 km). Hence, it is possible that an accident reported at mile point 2.1 (kilometer point 3.4) could have actually occurred at 1.9 (3.1) or 2.2 (3.5) (or even 2.9 [4.7]!).

For these reasons, analyses that are based on accidents along highway sections often deal with summations over these incremental section lengths and use a moving window to scan the accident frequencies. For example, a unit of say 1 mi (1.61 km) might be selected and accident sums generated for 1-mi (1.61-km) sections beginning at mile 0 (accidents are summed from mile 0.0 to 0.9, 1.0 to 1.9, etc. [0.0 to 1.45, 1.61 to 3.22, etc.]). The window is then moved up some amount, say 0.1 mi (0.16 km), and the summations are recalculated, i.e., accidents are summed from mile 0.1 to 1.0 (0.16 to 1.61), 1.1 to 2.0 (1.77 to 3.36), etc. Such a thorough analysis was not feasible or within the scope of this project, so only one pass of a 1-mi (1.61-km) window was used here.

The unit of 1 mi (1.61 km) was selected based on the accident frequencies in this data base. These sections, typical of two-lane rural non-intersection segments, had very low accident frequencies. In fact, for the 0.1-mi (0.16-km) partitionings, the data consisted mostly of zeroes and ones (i.e., rarely would more than one accident occur within the same 0.1-mi [0.16-km] section). Hauer used 0.3-mi (0.48-km) sections, which still resulted in very low accident frequencies. After examining various possible section lengths, it was decided that a minimum of 1-mi (1.61-km) sections would be needed to produce any meaningful results. Even with 1-mi (1.61-km) section lengths, for the year with the most accidents (1987 with a total of 2,169 accidents), the highest accident frequency was 18 with 99 percent of the sites having 7 accidents or less. Given the difference in the way Hauer defined section length, it is doubtful that the results of the two analyses can be directly compared. However, for the sake of academic curiosity, it may be of interest to see which sections proved to be the most hazardous using the various methods.

The results and conclusions of this study are not to be interpreted as the "answer" to the question of which sites are most hazardous for this data set given the problems just discussed with the variability of these sites. This analysis is provided solely for the purposes of demonstrating the use of this methodology for identifying high hazard locations.

Table 10 presents the descriptive statistics for the 13 years of data among the 2,137 1-mi (1.61-km) segments of the 430 sections. (Some segments were < 1 mi (1.61 km) — e.g., if a section contained 2.5 mi (4.02 km), the three segments for that section would contain accident frequencies for two 1-mi (1.61-km) lengths and a 0.5-mi (0.80-km) length. The vehicle-miles traveled (VMT) and vehicle-kilometers traveled (VKT) would consider the differences in section length in computing accident rates. Accident frequencies were adjusted for section length and accidents per mile were used in the accident count analyses.

Table 10. Descriptive statistics on accident frequencies and VMT (VKT).

Year	Accidents		Volume Measure (x 10^9)		
	Total	Max	Total VMT	Min VMT	Max VMT
1975	2054	13	1.657	0.799	516.3
1976	2106	12	1.687	0.832	524.2
1977	2010	11	1.678	0.843	520.0
1978	1871	10	1.688	0.854	522.1
1979	1871	11	1.623	0.832	500.5
1980	1796	12	1.567	0.821	482.1
1981	1702	14	1.588	0.843	487.3
1982	1885	12	1.616	0.865	500.8
1983	1872	11	1.663	0.909	525.6
1984	2079	13	1.748	0.964	563.4
1985	1940	20	1.815	1.018	596.0
1986	2078	13	1.910	1.084	638.8
1987	2169	18	2.096	1.204	713.6

1 mi = 1.61 km

Comparison of Methodologies

The EBEST methodology was used to rank the sites both by expected accident frequency and expected accident rate. Similarly, a straightforward ranking of the observed accident frequencies and rates was used and will be referred to as the classical analysis. Tables 11 through 13 list the site numbers of the top 10 sites ranked according to accident frequency for the 2 methods for each of the 13 years. The section numbers are interpreted as follows. The portion of the section number preceding the decimal point is the subsection number referred to by Hauer. The portion following the decimal is the 1-mi (1.61-km) segment of that subsection. For example, 1.2 means the second segment of subsection 1, i.e., miles 1.0 to 1.9 (kilometers 1.61 to 3.06). Section 7.4 is the fourth segment of subsection 7, i.e., miles 4.0 to 4.9 (kilometers 6.4 to 7.9), etc.

Table 11. Top 10 accident frequency sites — 1975-1978.

Rank	1975		1976		1977		1978	
	Class	EB	Class	EB	Class	EB	Class	EB
1	338.1	338.1	338.1	338.1	145.2	145.2	129.1	338.1
2	129.1	145.3	372.1	364.2	218.1	218.1	340.3	338.2
3	349.2	129.1	8.2	219.1	352.3	145.1	338.1	129.1
4	38.2	338.2	65.2	338.2	7.3	219.1	338.2	340.3
5	145.3	357.2	78.2	215.1	376.6	352.3	371.1	219.1
6	329.3	364.2	76.4	78.2	145.1	218.2	65.2	145.2
7	338.2	419.1	133.1	65.2	218.2	145.3	145.2	215.1
8	65.2	352.1	215.1	357.2	417.2	219.3	215.1	218.3
9	246.2	246.2	219.1	216.2	1.2	419.1	218.3	145.3
10	352.4	218.3	338.2	218.3	38.1	338.2	219.1	242.1

Table 12. Top 10 accident frequency sites — 1979-1982.

Rank	1979		1980		1981		1982	
	Class	EB	Class	EB	Class	EB	Class	EB
1	218.3	364.2	129.1	129.1	338.1	338.1	338.1	338.1
2	240.6	218.3	145.1	364.2	338.2	338.2	215.1	215.1
3	364.2	219.2	219.1	219.1	371.1	419.1	351.4	219.1
4	69.7	357.2	364.2	145.1	419.1	357.2	65.3	76.2
5	219.2	219.1	419.1	419.1	8.4	215.1	76.2	357.1
6	290.1	145.2	8.3	145.2	174.4	339.1	111.1	247.1
7	357.2	357.1	8.4	357.2	215.1	219.2	219.1	364.1
8	7.2	240.6	68.1	215.2	225.6	357.4	327.1	218.3
9	7.3	216.2	224.1	338.2	334.3	364.2	371.1	357.2
10	156.2	338.1	239.4	218.3	339.1	145.2	372.1	357.3

Table 13. Top 10 accident frequency sites — 1983-1987.

Rank	1983		1984		1985		1986		1987	
	Class	EB	Class	EB	Class	EB	Class	EB	Class	EB
1	218.3	247.2	339.1	339.1	339.1	339.1	218.3	218.3	338.2	338.2
2	247.2	218.3	352.1	247.2	352.1	338.2	371.3	419.1	372.1	338.1
3	215.2	215.2	247.2	352.1	338.2	352.1	376.6	219.1	338.1	339.1
4	339.1	339.1	76.4	145.1	377.4	377.4	419.1	247.1	339.1	219.1
5	419.1	145.2	129.3	377.4	76.2	338.1	219.1	357.2	216.2	357.1
6	145.2	338.1	133.1	338.2	145.1	145.1	247.1	145.2	129.3	419.1
7	376.6	208.4	145.1	364.2	145.2	145.2	138.2	377.4	351.3	241.1
8	76.3	219.1	371.2	129.3	338.1	364.2	145.2	364.2	419.1	129.3
9	215.1	215.1	211.1	357.1	76.3	218.1	352.1	376.6	111.1	357.2
10	338.1	340.3	65.3	145.2	107.5	218.1	357.2	352.1	129.1	377.4

Examining these tables, we see that the classical and EB methods tend to agree as to which sites are ranked within the five highest accident frequency locations for all years. It is only within the lower five rankings that the methods tend to disagree. In fact, the lower the ranking, the more often the two methods tend to select different locations. To visualize which sites would be classified as PIL sites (i.e., sites that are consecutively ranked in the top 10), the ranking results are presented in table 14. The icons in the body of the table represent the method that selected the sites in the top 10 at least twice. The bolded x denotes sites selected by EBEST, the black square with the dot in the center denotes sites selected by their observed accident frequencies, and the smiley face denotes sites selected by both methods as being in the top 10. From this table, we observe that site 145.2 was selected from 1977 through 1986 (with the exception of 1982) by the EBEST method, and the classical method agrees in five of these years. Sites 215.1, 338.1, 339.1, and 352.1 were selected 3 years in a row by both methods. Sites 219.1, 338.2, and 357.4 were selected more than 3 years in a row by one or both methods. The EBEST method also twice targeted site 364.2 as hazardous for 3 years in a row. This table also allows identification of hazardous "areas" or spots. When sections are consecutive, such as 338.1 and 338.2, this might be an indication that a larger stretch of highway is hazardous. Furthermore, if section 339.1 were adjacent to section 338.2, this might also be an indication of a problem "area." It would be interesting to generate such tables for moving windows over consecutive highway locations.

Table 14. Top-ranked accident count sites.

	Years												
Section	75	76	77	78	79	80	81	82	83	84	85	86	87
7.4			◘		◘								
8.4						◘	◘						
65.2	●	●		◘									
65.3								◘		◘			
76.3									◘		◘		
76.4		◘								◘			
111.1								◘					◘
129.1	●			●		●							◘
129.3										●			●
133.1		◘								◘			
145.1			✖			●				●	●		
145.2			●	●	✖	✖	✖		●	✖	●	●	

Table 14. Top-rank accident count sites (continued).

	Years												
	75	76	77	78	79	80	81	82	83	84	85	86	87
145.3	☻		✖	✖									
215.1		☻		☻			☻	☻	☻				
215.2						✖			☻				
216.1		✖			✖								
218.1			☻								☻		
218.2			☻								✖		
218.3	✖	✖		☻	☻	✖		✖	☻			☻	
219.1		☻	✖	☻	✖	☻		☻	✖			☻	☻
219.2					☻		✖						
247.1								✖				☻	✖
247.2									☻	☻			
338.1	☻	☻		☻	✖		☻	☻	☻		☻		☻
338.2	☻	☻	✖	☻		✖	☻			✖	☻		☻
339.1							☻		☻	☻	☻		☻
352.1	☻									☻	☻	☻	
352.3	◘		☻										
357.1					✖			✖		✖			✖
357.2	✖	✖			☻	✖	✖					☻	✖
357.3								✖	✖				
357.4							✖	✖		✖	☻	✖	✖
364.2	✖	✖			☻	☻	✖			✖	✖	✖	
371.1	◘			◘			◘	◘					
372.1		◘						◘		◘			◘
376.6			◘						◘			☻	
419.1	✖		✖			☻	☻		☻			☻	☻

The EBEST methodology was also used to rank the sites by expected accident rate. Again, the ranking of the observed accident rate will be referred to as the classical analysis. Tables 15 through 17 list the site numbers of the top 10 sites ranked according to accident rate for the 2 methods for each of the 13 years.

Table 15. Top 10 accident rate sites — 1975-1978.

Rank	1975		1976		1977		1978	
	Class	EB	Class	EB	Class	EB	Class	EB
1	147.3	38.2	73.4	372.1	11.7	417.2	23.15	23.15
2	399.3	371.1	157.1	417.3	37.2	37.2	74.3	371.1
3	157.1	411.2	12.10	8.2	178.4	38.1	406.7	347.3
4	154.6	39.1	362.4	23.15	23.16	205.8	261.6	347.7
5	74.3	373.1	51.12	368.1	421.2	430.5	379.1	334.3
6	276.5	352.1	55.5	8.4	21.3	334.3	111.3	66.5
7	337.3	129.1	23.15	273.5	406.4	116.2	11.7	351.4
8	411.2	173.3	248.6	194.1	196.5	7.3	182.3	8.3
9	12.1	347.7	261.6	180.2	12.1	368.1	427.1	8.6
10	121.1	312.1	180.2	186.2	12.9	311.1	100.9	366.1

Table 16. Top 10 accident rate sites — 1979-1982.

Rank	1979		1980		1981		1982	
	Class	EB	Class	EB	Class	EB	Class	EB
1	251.3	69.7	158.2	8.3	250.7	334.3	379.9	111.1
2	334.6	290.1	11.7	8.4	127.4	388.2	111.1	334.3
3	103.3	240.6	386.9	129.1	111.1	371.1	280.8	349.2
4	119.1	156.2	320.4	38.1	251.3	111.1	337.3	351.4
5	23.15	170.1	119.1	38.2	427.5	8.4	153.3	372.1
6	274.7	7.2	276.4	377.5	154.6	174.4	186.3	371.1
7	12.7	7.3	274.7	68.1	334.3	75.5	425.4	268.3
8	148.4	170.2	111.1	420.3	388.2	37.1	127.4	351.3
9	185.3	8.3	12.9	39.2	176.7	89.1	334.3	381.10
10	276.5	23.15	73.3	124.1	187.3	295.2	103.3	356.3

Table 17. Top 10 accident rate sites — 1983-1987.

Rank	1983		1984		1985		1986		1987	
	Class	EB	Class	EB	Class	EB	Class	EB	Class	EB
1	73.4	371.3	83.9	372.1	158.3	339.1	11.7	371.3	111.1	111.1
2	147.3	31.2	111.1	371.2	274.7	352.1	153.3	376.6	6.5	372.1
3	153.4	376.6	349.4	371.4	127.2	395.1	406.1	347.1	337.3	351.5
4	334.6	44.2	51.12	111.1	424.3	381.5	149.3	138.2	187.3	426.7
5	152.3	273.2	106.4	113.3	280.8	426.5	274.7	48.2	12.1	251.2
6	178.1	371.1	199.6	352.1	407.5	165.4	181.3	419.1	251.2	273.2
7	119.3	167.2	72.3	76.4	111.1	111.1	407.5	371.1	152.5	371.3
8	260.1	426.4	185.4	8.3	22.4	347.6	337.3	218.3	415.4	347.1
9	56.4	218.3	11.7	371.1	381.5	40.4	421.2	181.3	427.5	39.1
10	199.6	430.5	387.5	106.4	389.5	205.8	138.3	342.2	294.4	377.1

Examining these tables, we see that there is very little agreement between the EBEST and observed rate rankings. This is probably largely due to the tremendous variability in the rate denominator (VMT). For the years 1975, 1977, 1978, 1979, 1986, and 1987, only 1 site was selected by both methods as ranking in the top 10. These sites were 411.2, 37.2, 23.15, 23.15, 181.3, and 111.1, respectively. Two sites were mutually selected in 1976 (23.15 and 180.2), 1981 (111.1 and 388.2), 1982 (111.1 and 334.3), and 1984 (111.1 and 106.4). No sites were mutually selected in 1980 or 1983. The ratios of the maximum to minimum exposures ranged from a high of 645.93 (1975) to a low of 578.00 (1980), indicating that the exchangeability assumption is truly questionable for these data. The average shrinkage coefficient (b_i) ranged from a high of 0.802 (1983) to a low of 0.749 (1980). Recall that if this number is close to 1, less weight is placed on the observed rate and more weight is placed on the global (all sites) expected rate in computing the EBEST expected rate for a given site. Thus, when the shrinkage coefficient is large, as is the case for these data, there will be a greater discrepancy between the classical and EBEST methods. Another way to look at it is that when the shrinkage coefficient is large, there is more "regression-to-the-mean."

The dilemma is which method (if either) is truly identifying high hazard locations. Tables 18 through 20 contain similar information as table 14 but for the accident rate rankings. The first obvious difference in rate rankings, as opposed to count rankings, is that many more sites were flagged as ranking in the top 10 in at least 2 years by 1 or the other method. This occurs for two reasons:

1. Difference in the two methods as to which sites are selected.

2. Lack of consistency within either method in identifying similar sites over time.

If a PIL is defined as any section that is classified as ranking in the top 10 for 3 or more consecutive years in a row, only sites 111.1, 334.3, 371.1, and 371.3 would qualify. Of these, only 111.1 had any agreement between the two methods. The EBEST procedure seemed to target section 371, in general. Another example of a possible hazardous "spot" locator can be found in sections 38.1 and 38.2 for the years 1979 through 1980.

Table 18. Top-ranked accident rate sites.

	Year												
Section	75	76	77	78	79	80	81	82	83	84	85	86	87
7.3			✖		✖								
8.3				✖		✖				✖			
8.4		✖				✖	✖						
11.7			◘	◘		◘			◘	◘		◘	
12.1	◘		◘					◘	◘				
12.9			◘			◘		◘					
12.1		◘											◘
23.15		●		●	●								
23.16			◘	◘		◘							
38.1			✖		✖	✖					✖		
38.2	●				✖	✖							
39.1	✖												✖
39.2						✖						✖	
51.12		◘								◘			
55.5		◘							◘	◘			
73.4		◘							◘				
74.3	◘			◘									
76.4										✖	◘		
103.3					◘			◘					
111.1						◘	●	●		●	●		●
119.1					◘	◘							
119.3	◘								◘				

Table 19. Top-ranked accident rate sites (first continuation).

	Year												
Section	75	76	77	78	79	80	81	82	83	84	85	86	87
127.4							◘	◘					
129.1	✖			✖		✖							
147.3	◘								◘				
154.6	◘						◘						
157.1	◘	◘											
170.1					✖				✖				
185.4										◘			◘
187.3							◘						◘
199.6									◘	◘			
205.8			✖								✖		
218.3									✖		◘	✖	
251.2									◘				●
251.3					◘		◘						
273.2									✖				✖
274.7					◘	◘			◘			◘	
276.5	◘				◘								◘
290.1					✖				✖				
334.3				◘		✖	●	●					
334.6					◘				◘				
337.3	◘							◘		◘		◘	◘
339.1										✖	✖		
347.7	✖			✖									
347.1												✖	✖

Table 20. Top-ranked accident rate sites (second continuation).

	Year												
Section	75	76	77	78	79	80	81	82	83	84	85	86	87
349.4						◘	◘			◘		◘	
351.3	✖			✖				✖					✖
351.4	✖			✖				✖					
352.1	✖									✖	✖		
366.1				✖		✖							
368.1		✖	✖			✖							
371.1				✖			✖	✖	✖	✖		✖	
371.3		✖			✖				✖		✖	✖	✖
371.4									✖	✖			
372.1		✖						✖		✖			✖
376.6									✖		◘	✖	
381.1								✖		●			
407.5											◘	◘	
417.2			✖	✖									
419.1									✖			✖	
421.2			◘				◘					◘	
424.3											◘	◘	
425.4								◘					
426.4									✖			✖	
426.5											✖	✖	
427.1			✖	◘									
427.5							◘						◘
430.5			✖						✖				

With regard to whether or not either method has identified truly hazardous sites based on accident rates, it is difficult to support either one. Clearly, they differ. However, given the "hodgepodge" of sites that were thrown into this pool with very little forethought as to site differences, neither procedure can be recommended with any great degree of confidence.

High hazard location analysis should be approached with much more consideration of the experimental design. In today's powerful computing environment, it is all too easy to simply throw together huge data bases and blindly employ advanced statistical procedures to perform a task. The appropriate approach with the New York (or any) data set would have been to carefully classify homogeneous sections, at least with regard to traffic, if not other roadway features, and then to derive expected rates for these groupings. Then the expected rates can be ranked using all sites with variable moving windows and tables such as tables 18 through 20 generated to identify potential problem sites or areas.

Although the section-length differences between Hauer's analysis and this study differed, the results available in the body of Hauer's March 1995 report identified some of the top sites recognized as PIL's in Hauer's analysis of non-intersection accidents. According to tables 3 through 9 of this report, all sections that Hauer targeted were also ranked in the top 10 for either accident count or accident rank ratings at least once in the 13 years by at least 1 method (EBEST or classical). That is, these sites are represented in either table 14 or tables 18 through 20. However, the sections selected as most hazardous in this study are not represented in Hauer's table. It does appear that Hauer's method seems to select more sites more often and to produce a large quantity of information for the highway safety analyst to process. Of course, had a moving window been used with the study in this report, much more information would have been produced here as well. However, the author's recommendation would be to discontinue analyzing this data set as it stands, without some very careful considerations of identifying sites that were more homogeneous with regard to key highway safety factors.

Summary and Conclusions

In summation, the following issues emerged as a result of this case study analysis:

1. In the identification of high hazard locations on roadway segments of very short unit length, such as the 0.1-mi (0.16-km) segments of the New York data set, the issue of how long a segment should be to provide meaningful insight into the distribution of accidents along the roadway sections is not clear and has not adequately been addressed in research studies. A 0.1-mi (0.16-km) segment is clearly too short for most applications and results in a large number of zeroes and ones. Hauer used 0.3-mi (0.48-km) segments. In this study, 1-mi (1.61-km) lengths seemed to provide a fairly wide distribution of accident counts and resulted in a data base of manageable size. However, beyond this intuitive reasoning, there is no quantitative evidence that this is the most appropriate unit. It might be interesting to examine a data set such as this with varying unit lengths and then to look at the resulting distribution of accident rates and frequencies for the entire data set. This is just one of many aspects for future research that have emerged from these case study analyses. Such research would be very cost-effective since the data have been collected and this effort is generally the most costly portion of research.

2. The issue of variability in exposure in such large data sets emerged. This data set was highly variable and probably violated the assumption of exchangeability in the EB methods. Future research might include partitioning these data into more homogeneous subgroups for deriving the EB estimates and then recombining the estimated data to do a global ranking. Also, diagnostics for identifying lack of exchangeability could be developed.

3. At issue in ranking safety studies is the question of whether to use accident counts or rates. In practice, most researchers use both. The question of whether or not it makes sense to rank rates with EB methods since the expected rates are generally non-integers is also an issue.

4. The classical and EB methods tended to agree in ranking the top five sites by accident frequency and the two methods became increasingly disparate as the ranks decreased.

5. There was little agreement between the two methods when ranking by accident rates. There was also little agreement within a method in selecting the same sections from year to year. This is probably a reflection of the highly variable denominator — VMT.

6. Table 14 and tables 18 through 20 show promise of providing a graphical method for identifying potential trouble "areas," as well as PIL's — sites that are consistently hazardous over time. This study was only able to use a fixed window, but it would be interesting to revisit these data with a moving window and to regenerate these distribution tables. If this methodology is to be distributed to States for high hazard identification, the experts who participated in the workshop (see appendix B) have recommended that distribution tables such as these should be automated.

CHAPTER 7: MICHIGAN RAISED PAVEMENT MARKER STUDY

The data in this analysis were obtained from the Michigan data in the HSIS maintained by the University of North Carolina Highway Safety Research Center. A total of 17 locations, comprising 56.11 mi (90.3 km) where raised pavement markers were installed on the center line on undivided arterials and on lane lines on divided arterials between August 1 and September 30, 1989, represent the treatment sites for this analysis. A total of 42 sites, comprising 146.28 mi (235 km) where raised pavement markers were not installed and no other significant changes were made in roadway geometrics or operations data after August 1, 1987, were selected as reference sites. Video discs maintained by the FHWA were reviewed to ensure that the reference sites were truly comparable with respect to roadway geometrics.

Accident data were collected for 2 years before and 2 years after the treatment. Daytime accidents at the same sites were selected as the comparison group for the treatment condition of nighttime accidents. "Day" was defined as 1 hour after sunrise to 1 hour before sunset and "night" was defined as 1 hour after sunset to 1 hour before sunrise. Sunrise and sunset hours were selected based on a chart provided by the FHWA that contained daily readings for sunrise and sunset for Lansing, Michigan. Locations not close to Lansing were adjusted according to the estimated time differences. Since there was only an annual daily estimate for traffic (AADT) available, daytime and nighttime "exposures" were estimated by assuming that 25 percent of the AADT occurred at night and 75 percent occurred during the day.

These data provide us with an example for applying and comparing four of the treatment evaluation study designs described in the beginning of this report, namely:

1. Before/after without comparison group.

2. Before/after using daytime accidents as comparison group.

3. Before/after using nighttime accidents at the reference sites as comparison group.

4. Before/after with comparison group and test for comparability.

The three methodologies to be compared are:

- Classical (Class).
- EB without covariates (EB).
- EB with covariates (EBC).

Preceding Page Blank

These designs and methods will be referred to according to the numerical listing above. For simplicity, the notations Class, EB, and EBC will be used for classical, EB without covariates, and EB with covariates, respectively.

Table 21 gives descriptive statistics on the distribution of VMT and section lengths for these data. For the assumption of exchangeability required of the EB methods, it is important that the treatment and reference groups be similar with respect to the exposure measures (namely, VMT and section length). Overall, 5 percent of the sites were 1 mi (1.61 km) or less, with the shortest section being 0.83 mi (1.34 km).

Table 21. Descriptive statistics on VMT and section length.

n				mean	max	min	med
VMT	TRT Night	17	Before	16.89	65.48	3.07	10.99
			After	18.57	72.02	3.38	12.08
	REF Night	42	Before	17.29	58.38	3.03	15.05
			After	19.02	64.21	3.33	16.56
	TRT Day	17	Before	50.65	196.43	9.21	32.96
			After	55.71	216.07	10.14	36.26
	REF Day	42	Before	51.87	175.13	9.09	45.14
			After	57.05	192.64	10.00	49.66
LENGTH	TRT	17		3.30	7.88	1.02	1.85
	REF	42		3.48	8.80	0.83	2.82

TRT - Treatment condition.
REF - Reference condition.
1 mi = 1.61 km

Extended EB Regression Results

Candidate covariates that were modeled using the extended EB methodology were:

MIN - 0 for major and 1 for minor arterials.

URB - 0 for non-urban and 1 for urban sections.

DIV - 0 for undivided and 1 for divided roadways.

SWIDE - 0 for total (left and right) shoulder width $\leq$ 16 ft (4.9 m), 1 otherwise.

In addition, all interactions of the above four variables were included in the candidate models. Two interactions — min*swide and div*swide — were collinear and were dropped. This is logical since minor roads that are narrow are also undivided and undivided roads that are narrow are also minor arterials. The single interaction of min*div represented both of the interactions that were redundant. None of the other interactions were significant in any models. Total pavement width was also examined, but was identically correlated to divided and undivided roads, so it was replaced by the DIV variable.

Models were also entertained that attempted to identify any non-linearity between accidents and VMT. This was done by adding the quadratic term or the logarithm of VMT on the right-hand side of the models. EBC1 corresponds to estimates from the EB model where VMT and VMT^2 were modeled as covariates rather than as exposure (the exposure values were set to 1 in the data). EBC2 is the model with the natural logarithm of VMT that was modeled as a covariate and all exposures were set to 1. EBC3 is the model with VMT modeled as exposure (i.e., linearity is assumed).

Table 22 gives the model statistics for the significant variables in the models run on the treatment and reference group before data for nighttime accidents. In each case, all variables were put in the model and the significant variables were selected. Then the models were rerun with only the significant variables. For EBC1, both VMT and VMT^2 were significant, along with DIV (whether or not the lanes were divided), in predicting the before nighttime accidents. Because of the magnitude of the variables VMT and VMT^2, VMT was divided by 1000 in the data used for this model. Hence, to get the coefficient for VMT, the coefficients in the table must be divided by 1000.

Table 22. EB regression results.

Model		Estimate	Standard Error	T
EBC1	Int.	1.4266	0.1747	8.1637
	VMT	101.86	15.31	6.6541
	VMT^2	-964.61	266.95	-3.6135
	DIV	-0.8246	0.2918	-2.8262
EBC2	Int.	-1.1068	0.4004	-2.7644
	LVMT	0.9614	0.0999	9.6206
	DIV	-0.8281	0.2084	-3.9731
EBC3	Int.	0.1263	0.0636	1.9859
	DIV	-0.08711	0.1750	-4.9789

1 mi = 1.61 km

For EBC2, both the log of VMT (LVMT) and DIV were significant, and for EBC3, only DIV was significant. Note that the coefficients for DIV were all negative and of about the same magnitude regardless of how exposure was used in the model. The negative coefficient means that accident rates are lower on divided highways than on undivided highways. The fact that the magnitudes are the same mean that even though there is an indication of a non-linear relationship between VMT and accidents (due to the significance of VMT in the EBC1 and EBC2 models), the non-linearity was not significant enough to alter the relative importance of DIV. The average shrinkages for these models were 0.334, 0.382, and 0.380, respectively. Therefore, we would not expect significant RTM adjustment in these data. Table 22 represents the models for the combined 2 years of before data. However, for the test of comparability, these models had to be run on each year separately. Tables 23 and 24 represent these models.

Table 23. EB regression results for first before year.

Model		Estimate	Standard Error	T
EBC1	Int.	0.8943	0.1991	4.4901
	VMT	90.131	16.96	5.3148
	VMT^2	-716.94	292.71	-2.4494
	DIV	-0.9185	0.33	-2.7819
EBC2	Int.	-0.4344	0.3103	-1.4
	LVMT	0.9685	0.1149	8.4271
	DIV	-0.7574	0.2287	-3.311
EBC3	Int.	-0.5168	-0.0718	-7.1956
	DIV	-0.7935	0.1909	-4.1555

1 mi = 1.61 km

Table 24. EB regression results for second before year.

Model		Estimate	Standard Error	T
EBC1	Int.	0.5907	0.2139	2.7618
	VMT	112.95	18.28	6.1775
	VMT^2	-1223.64	321.26	-3.8089
	DIV	-0.7984	0.3394	-2.352
EBC2	Int.	-0.4975	0.3379	-1.4723
	LVMT	0.9535	0.1259	7.575
	DIV	-0.9454	0.25947	-3.6434
EBC3	Int.	-0.6197	-0.0788	-7.865
	DIV	-0.9949	0.2181	-4.5616

1 mi = 1.61 km

The average shrinkage coefficients were 0.477, 0.542, and 0.530 for the first before year for models EBC1, EBC2, and EBC3, respectively. For the second before year, they were 0.494, 0.509, and 0.515, respectively.

Before/After Without Comparison Group

Table 25 shows the accident frequencies to be used in the before/after comparison group analysis (equation 1) where the estimates for each methodology are inserted in place of C (for the cases where the expected frequencies are non-integers, all frequencies were multiplied by 10 in computing the test statistic):

Table 25. Before/after analysis with no comparison group.

	Class	EB	EBC1	EBC2	EBC3
Expected Before Accidents (C)	286	276	269.1	269.2	268.8
Estimated Treatment Effect	-4%	-0.4%	2.1%	2.1%	2.2%
T-Statistic	-0.464	-0.043	0.253	0.253	0.266

The classical method estimated a 4 percent reduction in nighttime accidents after the raised pavement markers were laid, the EB method without covariates estimated a 0.4 percent reduction, and all EB methods with covariates estimated around a 2 percent reduction. None of these reductions were statistically significant at the 5 percent level since none of the T-statistics were less than -1.96.

Before/After With Test for Comparability

Since 2 years of data were available before the installation of raised pavement markings, a test for comparability can be performed. Typically, the test for comparability consists of a chi-square analysis performed on the table of counts for the treatment and comparison group for the before period only. Non-significance in this case would imply that there is no difference (comparability) between the treatment and comparison groups before treatment. This is obviously a desirable feature as the comparison group is intended to represent a group similar to the treatment group, but which was not treated. In the case of using daytime accidents at the same sites as a comparison (group) condition, this group represents accidents at the same sites, but for a condition (daytime) that is assumed to not be affected by the treatment.

The classical analysis would actually use the methodology described in the beginning of this report. The EB estimates are inserted as D, E, and F in table 2. For non-integers, all frequencies are augmented by 10 in computing the test statistics. Because these frequencies were generally non-integers and the software designed for these tests (SAS Proc Freq) requires that they be integers, all counts in the tables were multiplied by 10. This has no effect on the estimate of the treatment effect (odds ratio) and inflates the likelihood ratio chi-square (L.R. χ^2) statistic by a factor of 10. Hence, to compute the actual statistic, the likelihood ratio chi-square for the table is simply divided by 10. The test for the treatment effect is conducted on the collapsed table with the EB methods replacing C in table 2 with the same adjustment for non-integers. However, this test is only justified if the groups pass the comparability test.

Table 26 lists the observed accident counts for each before and after year for nighttime and daytime accidents for the treatment and reference sites. The test for comparability using the before periods for nighttime accidents at the treated sites with daytime accidents at those same sites as the comparison condition would perform the chi-square test on the frequencies shown in table 27.

Table 26. Accident frequencies by separate before and after periods for daytime and nighttime accidents at the treated and reference sites.

	TRT DAY	TRT NIGHT	REF DAY	REF NIGHT
Before year 1	149	167	291	335
Before year 2	132	119	298	321
After year 1	143	124	274	293
After year 2	146	151	344	352

Table 27. Accident counts for test for comparability using daytime accidents.

	TRT DAY	TRT NIGHT
Before year 1	149	167
Before year 2	132	119

Tables 28 and 29 list the estimates for all of the methodologies and models and the resulting likelihood ratio statistics. These are to be compared with the value 3.84. If the statistics exceed this value, there is a significant difference in the groups, i.e., they are not comparable and the collapsed analyses are invalid. Table 28 uses the daytime accidents as the comparison condition and table 29 uses the nighttime accidents at the reference sites.

Table 28. Test for comparability analysis with daytime accidents as comparison group.

		Year	Class	EB	EBC1	EBC2	EBC3
All	Expected Before Accidents	Before 1	167	155	150.3	148.5	148.6
		Before 2	119	116.2	112.4	114.1	113.6
	L.R. χ^2		1.66	0.95	0.078	0.19	0.165
Undivided	Expected Before Accidents	Before 1	123	107.6	110.1	108.4	108.4
		Before 2	96	90.4	91.3	91.3	91
	L.R. χ^2		0.836	0.276	0.952	0.15	0.137
Divided	Expected Before Accidents	Before 1	44	37.6	40.2	40.1	40.2
		Before 2	23	22.4	21.1	22.7	22.6
	L.R. χ^2		1.3	0.124	0.001	0.006	0.039

Table 29. Test for comparability analysis with nighttime accidents as comparison group.

		Year	Class	EB	EBC1	EBC2	EBC3
All	Expected Before Accidents	Before 1	335	357.1	353.7	355.0	354.4
		Before 2	321	330.2	329.6	327.4	327.4
	L.R. χ^2		4.29	1.06	0.065	0.122	0.122
Undivided	Expected Before Accidents	Before 1	297	313.4	310.9	312.0	312.6
		Before 2	287	295.6	292.7	292.7	292.5
	L.R. χ^2		1.8	0.044	0.504	0.065	0.078
Divided	Expected Before Accidents	Before 1	38	43.9	42.8	42.4	42.3
		Before 2	34	35.1	36.9	34.8	34.8
	L.R. χ^2		2.39	0.118	0.013	0.07	0.065

Only one test for comparability failed — namely, when using all the data and the nighttime accidents at the reference sites as the comparison group for the classical methodology (table 29). The chi-square statistic for this was 4.29, which exceeds the 3.84 value at the 5 percent level of significance. However, all of the EB methods agreed that these sites were comparable. Again, the EB method without covariates showed the greatest difference from the observed frequencies. It can be concluded from this analysis that the sites were comparable and, in at least one case, the EB methods differed in conclusion from the classical.

Before/After With Daytime Accidents as Comparison Group

The classical analysis of before/after data with comparison group that was described in the beginning of the report on the data combined for the two before and after years is provided in table 30.

Table 30. Raised pavement marker accident frequencies.

		Day	Night
All	Before	281	286
	After	289	275
Undivided	Before	203	219
	After	207	220
Divided	Before	78	67
	After	82	55

Table 31 lists the expected accident frequencies and treatment effects for the three methodologies using daytime accidents as the comparison condition. None of the methods found a significant treatment effect (all L.R. χ^2's are less than 3.84), although the estimates of the treatment effect ranged from a 7 percent reduction using the classical method to less than a 1 percent reduction using the extended EB methodologies.

Table 31. Before/after analysis with daytime accidents as comparison group.

		Class	EB	EBC1	EBC2	EBC3
All	Expected Before Accidents	286	276	269.6	269.2	268.8
	Estimated Treatment Effect	-7%	-3.1%	-0.6%	-0.7%	-0.5%
	L.R. χ^2	0.32	0.07	0.003	0.003	0.002
Undivided	Expected Before Accidents	219	203.9	205.4	206.6	205.1
	Estimated Treatment Effect	-1.5%	5.8%	5.0%	4.4%	5.2%
	L.R. χ^2	0.12	0.166	0.126	0.098	0.134
Divided	Expected Before Accidents	67	63.7	63.7	62.6	63.7
	Estimated Treatment Effect	-22%	-18%	-18%	-16%	-18%
	L.R. χ^2	1.06	0.66	0.66	0.54	0.66

Before/After With Nighttime Reference Sites as the Comparison Group

This data set provides an alternative comparison group to the standard, i.e., daytime accidents at the same site. Since reference group sites were selected that were similar to treated sites in all aspects that are thought to affect safety (roadway geometrics, traffic, etc.), it is reasonable to assume that they could represent the comparison group in a before/after study. The EB methods are applied slightly differently, however. Since the reference sites are possibly also biased with respect to regression-to-the-mean (i.e., they were **not** selected for treatment, hence, they are possibly "low hazard" sites), the before accident frequencies for the reference sites are also adjusted by the EB estimates. That is, in the computation of the treatment effect and test statistic, C in equation 2 adjusted as well as A. Table 32 lists the observed frequencies for this analysis.

Table 32. Raised pavement marker night accident frequencies.

		Reference Night	Treatment Night
All	Before	656	286
	After	645	275
Undivided	Before	584	219
	After	576	220
Divided	Before	72	67
	After	69	55

Table 33 presents the results of the treatment evaluation analysis. The L.R. χ^2 in the table is the actual test statistic, not the p-value. Hence, in most cases, the very small values are not significant at the 0.0001 level of significance.

Table 33. Before/after analysis with nighttime accidents at the treated sites as comparison group.

			Class	EB	EBC1	EBC2	EBC3
All	Expected Before Accidents	Before 1	286	276	269.1	269.2	268.8
		Before 2	656	686	670.8	674.4	674.2
	Estimated Treatment Effect		-2	1.3%	4.8%	7.0%	7.0%
	L.R. χ^2		-0.05	0.017	0.499	0.710	0.717
Undivided	Expected Before Accidents	Before 1	219	203.9	205.4	206.6	205.1
		Before 2	584	602.1	598.4	597.4	598.6
	Estimated Treatment Effect		0.019	12.8%	11.3%	10.4%	11.5%
	L.R. χ^2		0.027	1.13	0.89	0.77	0.92
Divided	Expected Before Accidents	Before 1	67	63.7	63.7	62.6	63.7
		Before 2	72	79.1	78.8	77.5	75.8
	Estimated Treatment Effect		-0.14	-1.0%	-1.4%	-1.3%	-5.1%
	L.R. χ^2		0.39	0.002	0.032	0.003	0.045

Although none of the methods found a significant treatment effect, it is of interest to note the differences in the estimates. For all data combined, the classical methodology estimated a small reduction in accidents (2 percent), whereas all four EB procedures estimated an increase (1.3, 4.8, 7.0, and 7.0, respectively). For undivided highways, all methods estimated an increase in accidents after treatment, but the EB estimates were much larger (around 11 percent) than the classical (1.9 percent). For the divided highways, all methods estimated a reduction, but the classical method estimated a much larger reduction (14.3 percent) than the EB methods (between 1 and 5 percent). Caution must be used in attributing much meaning to these signs as the statistical tests are not significant and thus these may be just random fluctuations about zero.

Summary and Conclusions

In summation, the following issues emerged as a result of this data analysis:

1. Exposure measures are generally difficult to obtain for control-condition studies such as this one. In this study, the control condition was daytime accidents. Although many sites presently have continuous-count stations located proximally to study sites (see specific comments on raised pavement marker study in the Workshop Report of appendix B), this information was not available from the HSIS data base. Hence, the assumption had to be made that daytime AADT is approximately 75 percent of total AADT — an assumption that has been used in the past, but has not, to my knowledge, been rigorously tested. For comparison conditions such as wet/dry, the exposure measure becomes even more difficult to attain since it is not even clear what the exposure should be.

2. The extended EB methodology seemed to uniformly produce treatment effectiveness estimates that were lower than the classical estimates, yet higher than the EB method without covariates. In other words, the classical method would appear to overstate the case and the EB method without covariate adjustment understates the case, whereas the extended EB method is somewhere in-between.

3. There appears to be some justification for questioning the linear relationship between accidents and VMT. Since this assumption is inherent in the computation of accident rates, one wonders what effect this has on accident rate-based analyses that have been and still are the most commonly used analyses in highway safety research. Further research is definitely warranted in this area. A sensitivity analysis on the robustness of the linearity assumption in rate-based analyses would be extremely interesting. It is possible that within certain ranges of AADT, the assumption is more robust and valid than in others.

4. The models in this study seemed to indicate that the most significant covariate in raised pavement marker treatments is whether or not the highways are divided or undivided. The negative coefficient in the regression models reflects the fact that there are fewer accidents on divided highways.

5. The issue of which is a more reliable comparison group — daytime accidents at the same sites or nighttime accidents at the reference sites — is debatable. None of the analyses found a significant treatment effect as evidenced by the extremely low likelihood ratio chi-squares. However, the treatment effect estimates differed considerably depending upon which comparison group was used. When daytime accidents were used, all methods were fairly consistent in the treatment effect estimates. However, when nighttime accidents were used, the EB methods tended to agree, but the classical method was extremely different from the EB methods. For example, for undivided highways, the EB methods estimated a treatment effect of

around an 11 percent increase in accidents, while the classical method estimated a 2 percent increase. For divided highways, the effects were reversed with the EB methods estimating around a 1 percent decrease and the classical method estimating a 14 percent decrease. This is probably due more to the fact that with the EB methods, the reference group before estimate gets adjusted for RTM as well as the treatment group, so there are two estimated quantities in the computation of the odds ratio as opposed to when daytime accidents are used. At any rate, this does raise the issue of which is the more appropriate comparison group when more than one is available.

6. The check for comparability can be used with the EB methods. All methods agreed that daytime accidents were comparable to the treatment sites. However, the classical method found nighttime accidents to lack comparability to the daytime accidents, whereas the EB methods did not find significant lack of comparability. This could mean that the classical method is more sensitive to lack of comparability or, said another way, the EB methods may be more robust against lack of comparability.

CHAPTER 8: MICHIGAN SPEED LIMIT STUDY

The data in this analysis were obtained from a previous study entitled *A Comparison of Speed Zoning Procedures and Their Effectiveness*, sponsored by the Michigan Department of Transportation (MDOT) from September 1989 through September 1992.[5] This case study analysis represents only a portion of these data. Two analyses were conducted, 1 on 16 sites where the speed limits were raised, with a corresponding reference group of 22 sites where no speed change occurred, and a second analysis on 38 sites where speed limits were lowered, with a corresponding reference group of 47 sites. Some of the original sites were omitted from the present analysis (as will be explained); thus, the accident frequencies in these analyses will be lower than those of the MDOT study.

There were two main reasons that all sites could not be included in this analysis. The intent of this study is to demonstrate the application of particular methodologies. Sites had to have sufficient information as required by the methodologies in order to be included. The first reason was due to the test for comparability. For example, the test for comparability required that sites have more than 1 year of before data. The majority of sites had 3 years of before data. Thus, any sites with less than 3 years of before data were eliminated from this study. Secondly, the extended regression EB methodology was based on covariates and not all sites had complete information on all covariates from the HSIS data base.

This elimination of sites based on data quantity did not bias the treatment evaluation, but it did result in a reduction of total accidents as compared with the MDOT study. The fact that there were low numbers of sites to begin with, however, could impact conclusions from this analysis. For the lowered speed limit sites, the accident counts were not substantially lower than those reported by the MDOT study and conclusions for this analysis may not be significantly affected. However, for the raised speed limit sites, a substantial portion of the accident frequency in the reference group sites is omitted and thus conclusions based on this portion of the study may be weaker.

The covariates used in these analyses were:

MIN - 0 = major highway.
1 = minor highway.

WIDE - Average lane width (only three widths were represented 10, 11, and 12 ft [3.05, 3.35, and 3.66 m]).

SL - Posted before speed limit.

The posted before speed limit could be treated in two ways in the model — as a continuous variable or as a categorical variable. However, the distribution of accident frequencies was very low for some posted speed values, so the continuous variable approach was abandoned and categories were defined to capture the best distribution of accident frequencies. In general, the continuous variable approach is superior to categorizing because categorization results in a loss of information.

The categories for the second method for sites where the speed limits were raised were:

1. Less than or equal to 30 mi/h (48.3 km/h).

2. From 35 to 40 mi/h (56.3 to 64.4 km/h).

3. Greater than or equal to 45 mi/h (72.5 km/h).

The categories for the sites where the speed limits were lowered were:

1. Less than 45 mi/h (72.5 km/h).

2. Greater than or equal to 45 mi/h (72.5 km/h).

The number of sections in these categories were 10, 14, and 14, respectively, for the raised speed sites and 2, 53, and 30, respectively, for the lowered speed limit sites. Since there were so few ≤ 30-mi/h (48.3-km/h) sites among the lowered speed study, this regression variable representing speed was essentially only bi-leveled — < 45 mi/h (72.5 km/h) and ≥ 45 mi/h — in the models.

These data provide us with an example for applying and comparing two methodologies for three commonly designed safety studies. The study designs are:

1. Before/after without a comparison group.

2. Before/after with comparison group and test for comparability.

3. Before/after using reference sites as a comparison group.

The methodologies to be compared are:

1. Classical (Class).

2. EB with covariates (EBC).

(The results for the EB method without covariates were so similar to the method with covariates that it was not included in this report.) For simplicity, the notations Class and EBC will be used for classical and EB method with covariates, respectively.

Table 34 gives descriptive statistics on the distribution of VMT for the raised speed limit sites, table 35 gives the descriptive statistics for the lowered speed limit sites, and table 36 gives descriptive statistics on the distribution of section lengths. For the assumption of exchangeability required of the EB method, it is important that the treatment and reference groups be similar with respect to these variables, which both represent exposure.

Table 34. Descriptive statistics on VMT at raised speed limit sites.

Group	Time	Year	Mean	Max	Min	Med
TRT n=16	Before	BY1	48.64	431.30	0.197	21.94
		BY2	51.20	545.03	0.208	23.10
		BY3	51.71	458.57	0.210	23.33
	After	AY1	52.18	462.69	0.212	23.54
		AY 2	52.75	467.78	0.214	23.80
		AY3	53.22	471.99	0.216	24.01
REF n=22	Before	BY1	55.46	186.99	3.87	36.40
		BY2	58.38	196.84	4.08	38.32
		BY3	58.96	198.81	4.12	38.70
	After	AY1	59.50	200.59	4.16	39.05
		AY2	60.15	202.80	4.21	39.48
		AY3	60.69	204.63	4.24	39.83

1 mi = 1.61 km

Table 35. Descriptive statistics for VMT at lowered speed limit sites.

Group	Time	Year	Mean	Max	Min	Med
TRT n=38	Before	BY1	45.35	390.80	2.14	16.48
		BY2	47.93	411.37	2.25	17.36
		BY3	48.41	415.48	2.28	17.53
	After	AY1	48.84	419.23	2.30	17.69
		AY2	49.38	423.83	2.32	17.88
		AY3	49.83	427.65	2.34	18.04
REF n=47	Before	BY1	68.10	509.33	4.61	44.08
		BY2	71.67	536.13	4.85	46.40
		BY3	72.40	541.49	4.90	46.86
	After	AY1	72.45	550.41	4.99	46.95
		AY2	73.86	552.38	5.00	47.80
		AY3	74.52	557.35	5.05	48.23

1 mi = 1.61 km

Table 36. Descriptive statistics on section length.

Site	Group	n	Mean	Max	Min	Med
Raised	TRT	16	0.733	2.35	0.011	0.686
	REF	22	1.0000	2.31	0.390	0.822
Lowered	TRT	38	0.887	2.62	0.310	0.651
	REF	47	0.887	2.63	0.310	0.651

1 mi = 1.61 km

Examining the mean VMT values of tables 34 and 35, we note that the treatment and reference groups are very similar for the raised speed limit sites, but different for the lowered speed limit sites. The average VMT was significantly higher for the reference group then for the treatment group for the lowered speed limit sites (at the 5 percent level of significance using a two-sample independent t-test). Section lengths were comparable for both groups.

Regression Results

There is one candidate EB estimate with covariates that correspond to the EBC model. EBC corresponds to estimates from the EB model where the only significant covariate was initial posted speed. This variable was used both as a continuous variable (SL) and as the categorized variables (BS1 and BS2) defined as follows:

BS1 - 1 if initial speed was ≤ 30 mi/h (48.3 km/h), 0 otherwise.

BS2 - 1 if initial posted speed was ≥ 45 mi/h (72.5 km/h), 0 otherwise.

for the raised speed limit model; and BS defined as:

BS - 1 if initial posted speed was < 45 mi/h (72.5 km/h), 0 otherwise.

for the lowered speed limit model.

By defining the dummy covariates in this way, a test for the equivalence of the accident rates for these groups compared to the "omitted" group — initial speeds of 35 and 40 mi/h (56.3 and 64.4 km/h) for the raised model and ≥ 45 mi/h (72.5 km/h) for the lowered model — is available. The test statistics for the coefficient of BS1 compares the ≤ 30-mi/h (48.3-km/h) group with the 35- and 40-mi/h (56.3- and 64.4-km/h) group, and the test for the coefficient for BS2 compares the ≥ 45-mi/h (72.5-km/h) group with the 35- and 40-mi/h (56.3- and 64.4-km/h) group. The test statistic for the coefficient BS compares the < 45-mi/h (72.5-km/h) group and the ≥ 45-mi/h group.

For example, the interpretation of their corresponding T-statistics is as follows:

BS1 - If T exceeds 1.96, the average accident rate for roads with low speeds (speeds < 35 mi/h [56.3 km/h]) was greater than the average accident rate for roads with medium posted speeds (35 and 40 mi/h [56.3 and 64.4 km/h]) in the before period before speeds were raised.

BS2 - If T is less than -1.96, the average accident rate for roads with high speeds (> 40 mi/h [64.4 km/h]) was less than the average rate for roads with medium posted speeds (35 and 40 mi/h [56.3 and 64.4 km/h]) in the before period before speeds were raised.

BS - If T is less than -1.96, the average accident rate for roads with high speeds (≥ 45 mi/h [72.5 km/h]) was less than the average accident rate for roads with low speeds (< 45 mi/h [72.5 km/h]) in the before period before speeds were lowered.

The regression results for this model for all years combined for the raised speed limit sites are listed in table 37. The BEATS program would not run on the combined years for the lowered speed limit sites. Other models, including the models with VMT, VMT^2, and log of VMT modeled on the right-hand side, were attempted, but would not run on the BEATS program (reason not known).

Table 37. EB regression results for raised speed limit sites.

Model		Estimate	Standard Error	T
EBC	Int.	-0.3567	0.074	-4.83
	BS1	0.699	0.113	6.191
	BS2	-0.55	0.106	-5.204

The coefficients for each group separately can be derived from the model coefficients as follows:

Low initial speeds: -0.356 + 0.699 = 0.343

Medium initial speeds: -0.356

High initial speeds: -0.356 - 0.550 = -0.906

Note that according to the magnitude of these coefficients, as the initial posted speeds increased, the accident rates decreased before the speed change was enacted.

The EBC model with covariates using the initial speed categorical variables is presented in tables 38 through 43. The models were run with all covariates and a variety of significant models resulted, depending upon the years. An attempt will be made to summarize the results.

The candidate models that were attempted for each year were:

Model 1: Int BS1 BS2 MIN WIDE

Model 2: Int SL MIN WIDE

Model 3: Int VMT VMT^2 SL MIN WIDE

Model 4: Int LVMT SL MIN WIDE

The models that had significant coefficients were rerun and the final models are reported in the tables. If a model does not appear, the BEATS program failed to run.

Table 38. EB regression results for lowered speed limits before year 1.

Model		Estimate	Standard Error	T
Model 2	Int.	0.4166	1.174	0.355
	SL	-0.038	0.010	-3.689
	MIN	-0.416	0.093	-2.543
Model 4	Int.	1.653	0.506	3.277
	LVMT	0.883	0.066	13.402
	SL	-0.0384	0.010	-3.881
	MIN	-0.531	0.160	-3.317
Model 3	Int.	3.852	0.494	7.797
	VMT	0.021	0.002	8.540
	VMT^2	0.003	0.001	5.003
	MIN	-0.623	0.173	-3.601
	SL	-0.039	0.011	-3.684

Table 39. EB regression results for lowered speed limits before year 2.

Model		Estimate	Standard Error	T
Model 2	Int.	3.676	0.468	7.854
	VMT	0.0204	0.002	9.115
	VMT^2	-0.003	0.001	-5.365
	MIN	-0.43	0.163	-2.632

Table 40. EB regression results for lowered speed limits before year 3.

Model		Estimate	Standard Error	T
Year 3 Model 1	Int.	-0.282	0.129	-2.191
	BS1	1.093	0.393	2.779
	BS2	-0.418	0.158	-2.647
Year 3 Model 3	Int.	4.149	0.494	8.402
	VMT	0.022	0.002	9.489
	VMT^2	-0.003	-0.001	-5.647
	SL	-0.048	0.011	-4.514

Table 41. EB regression results for raised speed limits before year 1.

Model		Estimate	Standard Error	T
Model 1	Int.	-0.461	0.138	-3.344
	BS1	0.752	0.210	3.574
	BS2	-0.554	0.198	-2.805
Model 2	Int.	1.956	0.422	4.635
	SL	-0.064	0.011	-5.809

Table 42. EB regression results for raised speed limits before year 2.

Model		Estimate	Standard Error	T
Model 1	Int.	-0.398	0.129	-3.074
	BS1	0.714	0.198	3.612
	BS2	-0.526	0.185	-2.84
Model 2	Int.	1.96	0.387	5.057
	SL	-0.063	0.010	5.397
Model 3	Int.	4.783	0.552	8.669
	VMT	0.034	0.004	8.479
	VMT^2	-0.006	0.001	-6.16
	SL	-0.083	0.016	-5.209
Model 4	Int.	4.783	0.552	8.669
	LVMT	0.035	0.004	8.479
	SL	-0.006	0.001	-6.16
	MIN	0.083	0.0159	-5.209

Table 43. EB regression results for raised speed limits before year 3.

Model		Estimate	Standard Error	T
Year 3 Model 1	Int.	-0.225	0.136	-1.652
	BS1	0.644	0.21	3.067
	BS2	-0.572	0.195	-2.939
Year 3 Model 2	Int.	2.056	0.413	4.977
	SL	-0.062	0.011	-5.705
Year 3 Model 4	Int.	2.174	0.426	5.106
	LVMT	0.943	0.076	12.328

Table 44 identifies which models and variables were selected for the various years.

Table 44. Summary of annual EB regression results.

		Lowered Speeds			Raised Speeds		
		Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
Model 1	BS1	*	*	■	■	■	■
	BS2			■	■	■	■
	MIN						
	WID						
Model 2	SL	■	*	*	■	■	■
	MIN	■					
	WID						
Model 3	VMT	■	■	■	*	■	*
	VMT^2	■	■	■		■	
	SL	■		■		■	
	MIN	■	■				
	WID						
Model 4	LVMT	■	*	*	*	■	■
	SL	■				■	
	MIN	■				■	
	WID						

In the table, the darkened square indicates the variables that were significant in the models. The * indicates models that did not run on BEATS. Note that for Model 1, the indicator variables for initial posted speed were always significant when the BEATS program would run. Similarly, for Model 2, the significant variable when the program ran was SL or the continuous initial speed limit variable. Model 3 indicates the significant non-linearity of VMT with respect to accident count. The quadratic variable was significant whenever the computer program was able to run. Along with this fact, initial posted speed (SL) was also usually significant. Likewise, non-linearity was indicated by the significance of the log of VMT whenever the program was executable. The only other covariate that seemed to play a role was MIN — whether or not this was a major or minor arterial.

Before/After Without Comparison Group

The classical before/after without comparison described in chapter 3 was applied to these data using both the classical and EB estimate. (The EB estimate was only available for the raised speed sites since the BEATS program failed for the lowered speed limit sites. The results of these methods for the entire 3-year before and after periods are listed in table 45.

Table 45. Before/after analysis without comparison group.

	Lowered		Raised	
	Class	EBC	Class	EBC
Expected Before Accidents	2860	*	1753	1725.3
Estimated Treatment Effect	22%	*	-13.7%	-28.5%
T	10.01	*	4.52	10.7

The classical method concluded that there was a significant increase in accidents (22 percent) where speed limits were lowered, and both methods agreed that there was a significant decrease in accidents (13.7 and 28.5 percent, respectively) when speed limits were raised. Bear in mind that the before/after methodology **without covariates** is the weakest methodology.

Before/After With Test for Comparability

Since 3 years of data were available before the speed change, the test for comparability described in chapter 3 can be performed. Table 46 lists the observed accident counts for each before and after year for the treatment and reference sites. The test for comparability using the before periods for the treatment and reference sites is the chi-square test on the frequencies in tables 47 and 48.

Table 46. Accident frequencies by separate before and after periods for treatment and reference sites.

	Lowered		Raised	
	TRT	REF	TRT	REF
Before year 1	916	1909	535	723
Before year 2	911	1835	583	816
Before year 3	1033	2114	635	874
After year 1	1136	2421	679	1021
After year 2	1218	2620	666	988
After year 3	1315	2532	686	974

Table 47. Accident counts for test for comparability for raised speed limit sites.

	TRT	REF
Before year 1	535	723
Before year 2	583	816
Before year 3	686	974

Table 48. Accident counts for test for comparability for lowered speed limit sites.

	TRT	REF
Before year 1	916	1909
Before year 2	911	1835
Before year 3	1033	2114

Tables 49 and 50 list the estimates for all of the methodologies and models and the resulting likelihood ratio statistics. The EBC estimates were those from the EBC models that had the best fit. The L.R. statistics are to be compared with the value of 3.84. If the statistics exceed this value, there is a significant difference in the groups, i.e., they are not comparable and the collapsed analysis of the next section would be invalid.

Table 49. Expected before accidents for test for comparability analysis for raised speed limit sites.

Group	Time	Class		EBC	
		TRT	REF	TRT	REF
All	YR1 YR2 YR3	535 583 635	723 816 874	548.1 593.5 639.5	730.4 828.1 888.0
	L.R. χ^2	0.199		0.416	
≤ 30 mi/h	YR1 YR2 YR3	102 115 124	224 243 266	101.1 113.1 121.9	221.6 240.6 266.0
	L.R. χ^2	0.055		0.064	
35 and 40 mi/h	YR1 YR2 YR3	394 421 456	181 201 227	397.9 399.2 457.8	181.2 193.2 227.6
	L.R. χ^2	0.443		0.554	
≥ 45 mi/h	YR1 YR2 YR3	39 47 55	318 372 381	44.0 45.1 56.4	313.3 368.6 380.0
	L.R. χ^2	0.55		0.134	

1 mi/h = 1.61 km/h

Table 50. Expected before accidents for test for comparability analysis for lowered speed limit sites.

Group	Time	Class		EBC	
		TRT	REF	TRT	REF
All	YR1 YR2 YR3	916 911 1033	1909 1835 2114	844.1 810.5 975.5	2009.4 1926.0 2192.0
	L.R. χ^2	0.357		0.425	
< 45 mi/h	YR1 YR2 YR3	132 123 142	924 877 976	115.2 104.7 130.3	1021.6 924.5 998.3
	L.R. χ^2	0.078		0.066	
$\geq$ 45 mi/h	YR1 YR2 YR3	784 788 891	985 958 1138	744.0 745.1 856.4	987.8 1001.5 1193.7
	L.R. χ^2	0.56		0.14	

1 mi/h = 1.61 km/h

All tests for comparability were passed and the subsequent analysis was justified.

Before/After With Comparison Group

The classical analysis of before/after data with comparison group described in chapter 3 was applied to these data. Tables 51 and 52 list the observed frequencies for these data, along with the estimated treatment effects and tests of significance. The L.R. χ^2 in the table is the actual test statistic, not the p-value; hence, values less than 3.84 are not significant at the 0.05 level.

Table 51. Speed accident frequencies for raised speed limit sites.

Group	Time	Class		EBC	
		TRT	REF	TRT	REF
All	Before	1753	2413	1644.9	2463.5
	After	2031	2983	1994.7	2877.6
TRT Effect		-6.0%		1.2%	
L.R. χ^2		2.31		0.57	
$\leq$ 30 mi/h	Before	341	733	336.4	744.1
	After	488	852	497.3	863.6
TRT Effect		23.1%		27.4%	
L.R. χ^2		5.76		6.42	
35 and 40 mi/h	Before	1271	609	1144.3	657.8
	After	1312	710	1250.4	707.8
TRT Effect		-11.0%		1.2%	
L.R. χ^2		3.22		2.52	
$\geq$ 45 mi/h	Before	141	1071	164.2	1061.6
	After	231	1421	247.0	1306.2
TRT Effect		23.5%		22.3%	
L.R. χ^2		3.41		3.02	

1 mi/h = 1.61 km/h

Table 52. Speed accident frequencies for lowered speed limit sites.

Group	Time	Class		EBC	
		TRT	REF	TRT	REF
All	Before	2860	5858	2610.2	5922.4
	After	3387	7155	3552.3	7652.5
TRT Effect		-3.0%		5.3%	
L.R. χ^2		0.998		1.54	
< 45 mi/h	Before	397	2777	445.1	2988.3
	After	452	3200	532.3	3423.6
TRT Effect		-2.0%		-4.4%	
L.R. χ^2		0.027		2.17	
$\geq$ 45 mi/h	Before	2463	3081	2165.1	2777.0
	After	2935	3955	3020.0	4228.9
TRT Effect		-7.1%		-8.4%	
L.R. χ^2		3.18		3.77	

1 mi/h = 1.61 km/h

Note that none of the treatment effects was statistically significant, except for sites with initial posted speeds $\leq$ 30 mi/h (48.3 km/h) where speed limits were raised. Both the classical and EB methods agreed and estimated a 23.1 percent and 27.4 percent increase, respectively, in accidents where speed limits were raised at these very low posted speed sites. However, given the low frequencies and low number of sites in this category, this conclusion is not strongly supported. The treatment effect estimates varied dramatically, but since there was no statistical significance, these fluctuations can be considered as random fluctuations about zero — no significant treatment effect when speed limits are either raised or reduced. This is comparable to MDOT's findings and, hence, its recommendation that policy regarding changing speed limits should not be based on potential safety benefits alone is supported. Other factors, such as potential increased compliance, should be the primary motivation for speed limit changes.

Summary and Conclusions

In summation, the following issues emerged as a result of this case study analysis:

1. In applying the extended EB method using covariates, some sites may be dropped from the analysis if all covariates are not available for that site. In other words, the existing methodology cannot handle missing data. There are methods for incorporating such incomplete data in ordinary least-squares modeling and it may be possible to develop corresponding methods for the extended EB methodology. However, at this time, incomplete data result in a loss of information and the extent of that loss is not known. For this study, we found that a significant reduction in total accidents at the reference sites for the raised speed limit analyses resulted (as compared to the total accidents in MDOT's study). This loss of information could seriously skew conclusions in studies where the sample size is small to begin with. Although the classical method could have used data from the sites with missing covariates, the two methodologies would not have been comparable, and since the main purpose of this study was to compare methodologies, the classical method used the same data as the EB method.

2. In order to conduct the test for comparability, this study only included sites that had three complete years of before and after data. This resulted in further elimination of sites and reduction of total accidents. An analysis could have been preformed using only 2 years of before and after data that would have included these sites, but then it would have resulted in a loss of 2 years of information from the other sites. The decision was made to use the most data and that was the 3 years of information. This issue, however, will arise when analyzing multiple years of accident data, and the researcher will have to decide which way to partition the data.

3. There will be cases where a covariate can be viewed as either continuous or categorical, such as initial posted speed in this study. Generally speaking, the recommendation is to use the continuous format, as categorizing results in a loss of information. Both methods were modeled in this study. However, due to the few categories of initial posted speed, the speeds were divided into groups and they seemed to perform equally well. The interpretation of results was also clearer when initial posted speeds were categorized.

4. The before/after methodology without comparison, which is never the recommended methodology (especially when a comparison group is available), produced very different results and estimates than the comparison group analysis. Significant increases in accidents (22 percent) at sites where speed limits were lowered and significant decreases in accidents (14 percent to 29 percent) at sites where speed limits were raised were estimated by this methodology. This example points out how dangerous it would be to draw conclusions from this very weak methodology.

5. The sensitivity of the BEATS program was revealed in the analysis of these data. The program failed to run in many instances, and the reasons for this are not known. The program was fairly successful when individual years were analyzed, but would not run when all three before years were combined in a single data set. There may be a sample-size limitation in the program that needs further investigation.

6. The definition of the initial posted speed categories provided an internal comparison or test of comparability between the accident rates for the three categories of low, moderate, and high initial posted speeds. The accident rate for the low initial posted speed sites (before changing speeds) was higher than the accident rate for the moderate sites, which, in turn, was higher than the accident rate for the high posted speed sites. As the initial posted speed increased, the accident rates decreased.

7. All treated sites were comparable to their respective reference sites in the before period.

8. Initial posted speed was a significant variable in modeling accident rates in the before period, whether it was defined as a continuous variable or as a categorical variable. There is a strong indication of non-linearity of VMT with accident frequency that warrants further investigation.

9. There was no statistically significant change in accidents when speed limits were either raised or lowered. The only exception to this was for the low posted speed sites. A significant increase in accidents (23 percent to 27 percent) was found at these sites when the speed limits were raised. However, given the low accident frequencies at these sites and the low number of sites, this statistical significance is probably an anomaly due to the small amount of data.

CHAPTER 9: MINNESOTA SIGNALIZATION STUDY

The methodologies of this study were applied to Minnesota HSIS data by Jun Wang and were reported in an HSIS Summary Report contained in appendix A. The treatment group consisted of 13 intersections where new traffic signal controls were installed. The reference group consisted of 79 sites that were defined to include intersections that were comparable to the treatment group with respect to daily entering traffic, number of approach legs, intersection configuration, etc. All study sites were examined using Minnesota's video disc system to verify comparability of sites with respect to these roadway geometrics. The sites were examined using the *Manual on Uniform Traffic Control Devices* (MUTCD) warrant criteria and it was found that 38 percent of the reference sites met at least one of the volume warrants and 14 percent met the accident warrants for signalization, though no signals were installed. Accidents occurring within 250 ft (76.2 m) of the treatment and reference locations, on both major and minor crossroad locations, were considered intersection-related and were included in the study. Accidents occurring during the construction period were excluded. The before and after treatment periods varied from 21 to 31 months.

The treatment sites had a higher accident rate than the reference sites before treatment. This is indicative of regression-to-the-mean potential. There was a 30 percent reduction in accidents after new traffic signals were installed according to the classical methodology, and a 25 percent reduction with the EB method, both of which were statistically significant at the 0.05 level. For injury accidents only, the classical estimate was a 38 percent reduction and the EB estimate was a 23 percent reduction, again both statistically significant. Although both methods found significant reductions, the EB estimate was considerably lower (more conservative) than the classical.

A significant issue raised by this case study is the selection of reference sites. An intersection in the reference group that does not satisfy at least one of the MUTCD warrants is not a site that could have been selected for treatment, from a practical standpoint. However, does this have any statistical impact in violation of the exchangeability assumption? This issue warrants further investigation.

CHAPTER 10: CONCLUSIONS AND FUTURE RESEARCH

One of the strongest conclusions to emerge from this study revolves around the collection of reference group data. Considerable effort was made to define and identify reference groups that were comparable using video discs and large automated accident and roadway inventory data bases. Yet, in most cases, there was little, if any, statistical difference in the conclusions drawn by either method. The question emerges, then, is the cost and effort of collecting reference group data justified and, if so, under what conditions? That is, when is there sufficient RTM bias to warrant the additional effort and can a set of criteria be established to guide the researcher in this judgment before embarking upon this hefty data collection effort? It would be interesting to see a sensitivity analysis of just how sensitive (or robust) the classical methodology is to RTM biases.

Another area of concern that emerged in this study is the issue of the assumption of exchangeability. In many of the case studies, this assumption came into question either with regard to traffic volumes, section length, or reference sites that were candidates for treatment, such as the Minnesota signalization study. Again, a sensitivity analysis of the robustness of the EB methodology to departures from exchangeability might produce guidelines for the researcher prior to data collection to ascertain whether or not it is reasonable to assume that the reference group to be collected has any practical chance of being within acceptable ranges of exchangeability.

Several results indicated that there is some validity to the assumption of a non-linear relationship between accident frequency and VMT from the models examined in these case studies. Given that traditional accident analysis methods are based on accident rate, and that inherent in the calculation of rate is an assumed linear relationship between accident frequency and VMT, what is the impact of ignoring the non-linearity? Is there some range over which linearity can be assumed? If so, what is it? If there is a non-linear relationship, how does it impact traditional rate-based methodologies? This is a very important issue that has continuously surfaced in the literature, but which has never adequately been addressed. Given the results of these case studies and the strong support for non-linearity, it may be timely to conduct a thorough investigation of this relationship and its impact on the way things have always been done.

The issue of sample size once again emerged as an important and key consideration in accident analyses. There were some case studies where even significant treatment effects were questionable due to small sample sizes. But how small is too small? Can the sample size be too large so as to obscure practical significance? Again, this is an issue that has been raised and we now have some guidelines, but they are not in a readily usable form for the highway safety analyst. It would appear that a very useful tool would be a sample-size spreadsheet in which various entries could be changed (expected treatment effect, potential RTM bias, traffic volume ranges, etc.) and the table would automatically calculate recommended sample sizes. This effort would probably yield a low-cost, yet highly useful and visible, product.

The existing version of the BEATS program is not user-friendly and it requires additional effort to even be useable by the statistically sophisticated researcher. The issue of who should use this complex methodology requires careful consideration. Because of the assumptions of exchangeability, etc., the methodology in the hands of someone who does not thoroughly understand its requirements has the potential for misuse and misapplication. The workshop participants seemed to be in agreement that two versions of the computer program should be developed. One version would be very user-friendly, similar to the format of the original BEATS program, and would only preform the ranking analyses. This version could be distributed to practitioners for use in high hazard location identification, etc. The second version could be less user-friendly, but easily executed by the researcher who has a good understanding of the statistical methodology and the impacts of the assumptions. This version should also include some graphics, such as likelihood functions and confidence intervals. This, too, could be a very low-cost effort at this point, but is really needed if this methodology is to be used at all. Given the effort and resources devoted to the development of this methodology, it would seem that this additional effort is highly warranted.

In conclusion, this study has attempted to apply existing accident analysis methodologies to five carefully selected data sets. The effort expended in data collection was substantial, yet, from a practical standpoint, did not make a difference in the conclusions reached by the methodologies. Before such effort is expended, guidelines should be developed for the researcher that indicate whether or not the collection of a reference group is warranted. The most critical need for future research appears to be in the area of developing such guidelines through sensitivity analyses, sample-size spreadsheets, etc. Also, the computer program should be enhanced to be useable by practitioners who do have the ability and desire to engage in this data collection effort and to use the extended EB methodology.

Summary Report

HSIS

HIGHWAY SAFETY INFORMATION SYSTEM

The Highway Safety Information Systems (HSIS) is a multi-State safety data base that contains accident, roadway inventory, and traffic volume data for a select group of States. The participating States—Illinois, Maine, Michigan, Minnesota, and Utah—were selected based on the quality of their data, the range of data available, and their ability to merge data from the various files. The HSIS is used by FHWA staff, contractors, university researchers, and others to study current highway safety issues, direct research efforts, and evaluate the effectiveness of accident countermeasures.

U.S. Department of Transportation
Federal Highway Administration

Research and Development
Turner-Fairbank Highway
Research Center
6300 Georgetown Pike
McLean, Virginia 22101-2296

APPENDIX A

THE APPLICATION OF AN IMPROVED ACCIDENT ANALYSIS METHOD FOR HIGHWAY SAFETY EVALUATIONS

The evaluation of safety effects for various treatments has long been a subject of controversy within the transportation safety research community. Safety evaluations mostly rely on assessment of accident experience over before-and-after periods. One of the common pitfalls in the assessment methodologies is the failure to account for regression-to-the-mean (r-t-m) bias. The sampling bias due to the r-t-m phenomenon may seriously affect conclusions drawn in safety treatment evaluation studies. Safety treatment sites are generally selected because they have a high accident rate or accident count. If a site has an unusually high number of accidents occurring before the treatment, accident occurrence at that same site the following period would, in all probability, be lower even without any intervention at that site. This is the phenomenon known as r-t-m. Therefore, a simple before-and-after comparison for sites where the treatment is selected based on the accident experience is likely to result in an overestimation of the treatment's effect.

In a recent Federal Highway Administration (FHWA) study, a new method, titled Empirical Bayes Estimation of Safety and Transportation (EBEST), was developed for providing a better estimate of the expected accident experience for a treated site, adjusted for any r-t-m bias.[1] With this method, a microcomputer program was developed to allow easy application of the analysis technique. The methodology was developed and initially tested using simulation and hypothetical examples. This study was undertaken to apply the EBEST methodology to actual data from the Highway Safety Information System (HSIS). The installation of traffic signal controls at previously unsignalized intersections was selected as the treatment to be evaluated.

State Data Bases Used

This study employed data from only one HSIS State. Minnesota was selected because of the availability of sufficient safety treatment data, including the start and completion dates of roadway improvement and signal installation projects as well as a videodisc photolog system. The videodisc system allows users to access images of the State-maintained roadway network, collect additional information about study locations, and verify existing HSIS data.

Analysis Methods

A critical requirement of the EBEST methodology is the use of data from a reference group. The reference group is a sample of sites that are generally similar to the treatment sites with respect to roadway and traffic characteristics. The reference group and treatment group should represent the population of potential treatment sites. A recent enhancement made to the EBEST method includes a regression model to control for factors that may differ among treatment and reference sites.[2] The EBEST method also incorporates measures of exposure (traffic volume, section length, etc.) and can account for changes that occur over time. In many cases, the reference group can serve as a comparison group to account for potential time effects.

In addition to the EBEST method, the traditional before-and-after method with comparison group (i.e., the "classical" method) was also applied in this study. The purpose of applying both methods was to compare the results of a method that corrects for r-t-m bias to a method that does not. In this study, the reference group also served as the comparison group to adjust for time effects. Statistically, the reference group was found to be an appropriate comparison group.

The treatment group consisted of 13 intersections where new traffic signal controls were installed. Accident, traffic, and intersection configuration information for each treatment site were extracted from HSIS. The reference group was defined to include intersections that were comparable to the treatment group with respect to specific criteria (e.g., daily entering traffic, number of approach legs, intersection configuration, etc.). After applying these criteria, 79 sites were selected for the reference group. Depending upon the size of the treatment group, the reference group should be two to five times larger than the size of the treatment group. All study sites (treatment and reference sites) were examined using Minnesota's videodisc system to verify the site information and locations.

Because most traffic engineers in the United States determine the need for traffic signal control based on the warrants in the *Manual on Uniform Traffic Control Devices (MUTCD)*, a crude signal warrant analysis was attempted to determine how many of the reference group sites were candidates for signal installation. However, because detailed volume data were not available, average hourly approach volume were estimated and compared to the minimum vehicle volume warrant and the interruption of continuous traffic warrant. Relying on several assumptions, it was found that 38 percent of the reference sites met either of the volume warrants and 14 percent met the accident experience warrant. These results illustrate the practical difficulties encountered in selecting a reference group that represents the population of potential treatment sites. After further examination of the characteristics of the treatment and reference groups, it was concluded that the reference group was acceptable for this analysis.

All accidents that were reported between 1985 and 1990 that occurred within 76.2 m (250 ft) of the treatment and reference site intersections on both major and minor cross road approaches were retrieved from HSIS for the analysis. Accidents that occurred during the construction period were excluded from the analysis. Reference sites were matched with treatment sites and corresponding before-and-after time periods for the treatment site were used in the analysis. Consequently, the before-and-after treatment periods varied from 21 months to 31 months.

Results

Figure 1 shows that the treatment sites had a higher accident experience than the reference sites in the before period. This is an indication of a sampling bias with potential for r-t-m.

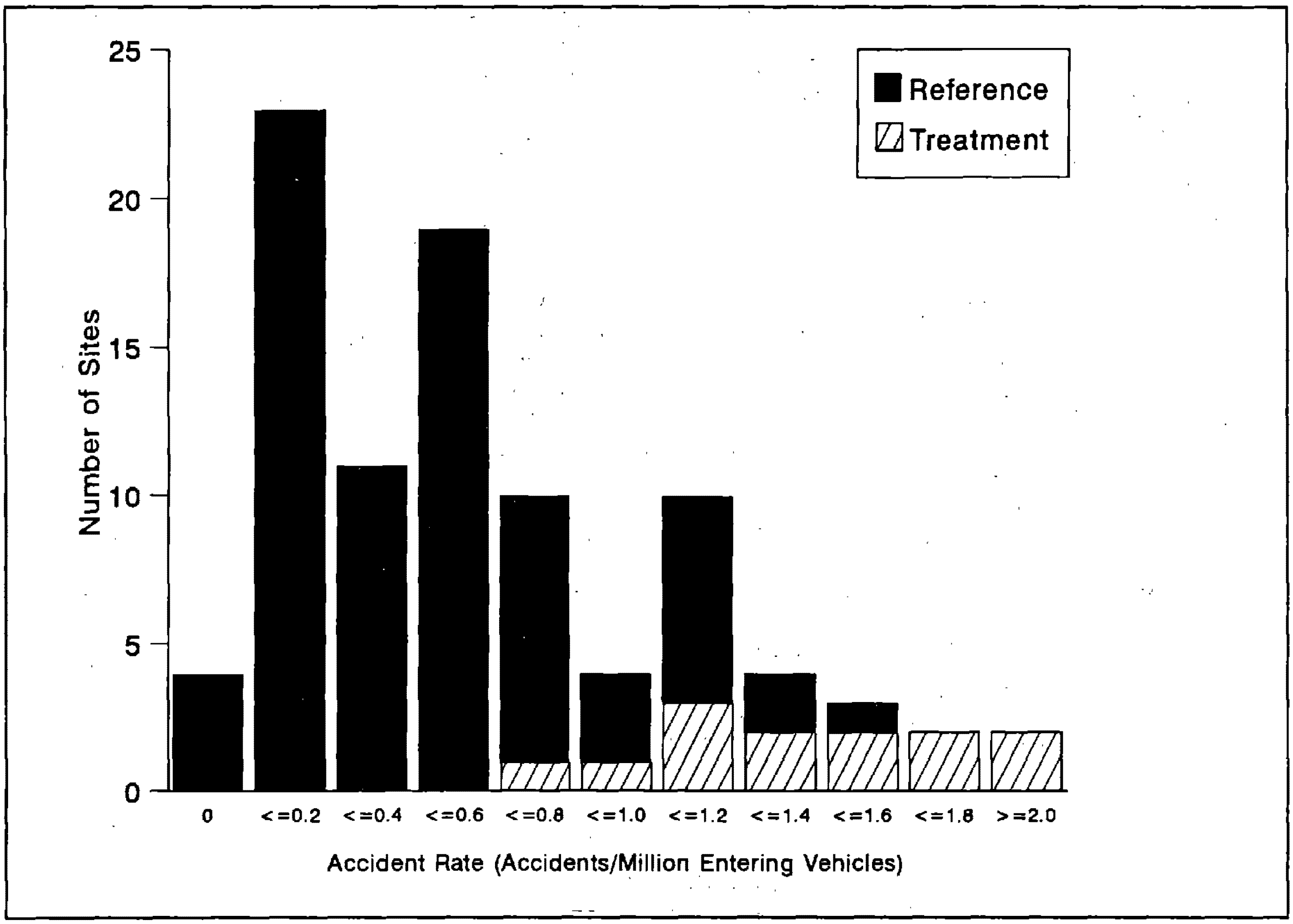

Figure 1. Before accident rate distribution for all sites.

Table 1. Results from EBEST method and classical method.

		EBEST	Classical
Total Accidents	Treatment Effect	-25%	-30%
	95% Interval	-38%, -10%	-42%, -15%
Injury Accidents	Treatment Effect	-23%	-38%
	95% Interval	-43%, +5%	-56%, -13%

Table 1 shows that for total accidents, the treatment effect is 30 percent (30 percent reduction in total accidents after new traffic signals were installed) with the classical method and 25 percent with the EBEST method. The 95 percent confidence intervals are -42 percent, -15 percent and -38 percent, -10 percent for the classical and EBEST methods, respectively. In other words, based on the classical method, there is a 95 percent probability that the true treatment effect is between a 15 percent and 42 percent reduction in total accidents. Based on the EBEST method, there is a 95 percent probability that the true effect is between a 10 percent and 38 percent reduction. For injury accidents, the treatment effect is 38 percent with the classical method and 23 percent with the EBEST method. The 95 percent confidence intervals are -56 percent, -13 percent and -43 percent, +5 percent for the classical and EBEST methods, respectively.

The study results show a significant difference between the EBEST method and the classical method in the evaluation of injury accidents. It appears that the classical method has overestimated the treatment effect. This overestimation can be attributed to the lack of adjustment for r-t-m bias.

Study Implications

This was the first study to apply the EBEST method to actual data to evaluate the safety effect of a treatment. Reference sites were identified using the roadway data commonly found in State data bases and were verified with videodisc photologs. A crude traffic signal warrant examination was conducted to determine how well the reference sites represented the population of potential treatment sites.

The findings of the signal warrant examination raise questions about the methods for selecting reference sites. To many traffic engineers, an intersection in the reference group that does not satisfy at least one of the *MUTCD* warrants is not a site that could potentially be "treated" with traffic signal control. Consequently, the question remains: What truly constitutes an appropriate reference group for the treatment being analyzed? This issue is being addressed in a current research study in which the EBEST method is being applied to several data sets. This research study will include several before-and-after studies, such as the one reported here. It will also examine identification and ranking of high-frequency accident locations using the EBEST method. The strengths and limitations of the EBEST method will be further identified and questions concerning the selection of the reference group will be studied.

For More Information

This research was conducted by Jun Wang, a former graduate research fellow at the Turner-Fairbank Highway Research Center (TFHRC), under the supervision of Michael Griffith of FHWA. The study was sponsored by the graduate research fellowship program of FHWA. For more information, contact Jeffrey F. Paniati, HSIS Program Manager, HSR-30, (703) 285-2568.

References

1. Pendleton, O.J. Application of New Accident Analysis Methodologies: Volume I, General Methodology, Publication No. FHWA-RD-90-091, Washington, DC, 1991.

2. Christiansen, C.L. and Morris, C.N., Empirical Bayes Analysis of Rates Using a Hierarchical Regression Model, Technical Report No. 83, Center for Statistical Sciences, University of Texas at Austin, 1990.

APPENDIX B: WORKSHOP REPORT

A 1½-day workshop was held May 23-24, 1995, at the Turner-Fairbank Highway Research Center to present results from the study "Evaluation of Accident Analysis Methodologies," sponsored by the FHWA, Contract No. DTFH61-92-C-00029. Five experts were in attendance and provided written reports after the workshop commenting on the study results and implications. A roster of the experts providing these comments appear at the end of this appendix. This report will summarize the comments and issues raised by the workshop participants.

GENERAL COMMENTS

1. For estimating treatment effect, the methodology would not be usable as a whole by practitioners who cannot understand the theory behind it. Perhaps a two-stage approach would be more feasible wherein:

   Stage 1: The practitioner determines whether the EB methodology is warranted (i.e., how serious is the regression-to-the-mean potential biases for this particular application).

   Stage 2: If the EB procedure is not warranted, the practitioner proceeds with the classical analysis. If the EB methodology is warranted, the practitioner is advised to consult suitable qualified analysts for assistance.

2. For the identification of high hazard locations, a "black box" user-friendly computer software application may be feasible for the lay practitioner to employ the EB method without the consultation of qualified analysts with an indepth understanding of the theory.

3. In many cases, the sample size required of the EB methodologies is prohibitive, both in terms of cost and effort in collecting. This should not be a deterrent to evaluation since there might be a tendency to administer treatments in small doses to avoid the evaluation issue.

4. When is a comparison group needed or warranted? If a suitable reference group is available, is a comparison group needed? If a good comparison group is available, can it serve as a reference group? What are the advantages of having both, and is the effort warranted in attaining both?

5. In studies where a comparison condition is warranted, should a comparison group even be considered? In some cases, there is no choice, e.g., the Michigan raised pavement marker study. However, the issue of exposure becomes more complicated with comparison conditions, i.e., are day/night, wet/dry exposure estimates available?

Preceding Page Blank

6. The exchangeability assumption creates considerable practical difficulties. This requirement may be unnecessarily restrictive in the case of the extended EB methodology (with covariates). Perhaps this assumption can be relaxed if appropriate significant covariates are identified and used in the adjustment of the estimates.

7. There would appear to be a danger that the reference group, if defined as those sections that were candidates for treatment but were not selected, may be biased in the opposite direction. Would too large a reference group tend to overcorrect for shrinkage?

8. What are the appropriate exposure measures? Whereas coarse measures such as AADT are generally available, are these always sufficient or appropriate? For an initial screening in high hazard identification, AADT may be adequate. For treatment evaluation, however, this is highly dependent on the application. The practitioner should be forewarned that AADT is not a universally adequate exposure indicator.

9. AADT appears to be better suited for use as a covariate than as a denominator in the EB modeling methodology. There also appears to be some support for a non-linear relationship between AADT (VMT) and accidents.

10. There should be a limit as to how much extra resources should be used in trying to develop the "perfect" model. The regression models should be recognized as being predictive, rather than causal, models. Practically speaking, predictive variables (urban/rural, roadway geometrics, etc.) are easier to obtain and are more reliable than causal ones.

11. How important is it to use a negative binomial as opposed to a Poisson or normal error structure? Few analysts are versed in modeling that is more sophisticated than ordinary least squares.

12. What are the appropriate relative numbers for the sample sizes of reference and treatment groups?

13. For future research, perhaps the EB and extended EB methods should be compared to other non-traditional methods such as Hauer's PIL methodology.

14. For future research, perhaps one should do more validation of model results, e.g., in the identification of high hazard locations, subsequent years should be revisited to see if, in fact, those sites targeted as hazardous were indeed hazardous.

15. For future research, this extended EB methodology should be applied to the accident migration issue. It is still of concern that the treatment effect may be affecting non-treated sites.

16. The case studies in this report will probably get cited and although the intent of these analyses was primarily for demonstration, strong warning labels should be placed on drawing any definitive conclusions from the results of these case studies.

17. For future research, some effort should be spent on standardizing the extended EBEST methodology by developing regression models for standard reference groups. These standard models can be transferred in time to other jurisdictions and the methodology can be structured to allow adjustments for spatial and temporal differences among reference groups. This would also get around the problem of having to use a biased reference population caused by a treatment group that is large and biased. This would also provide useful input for the Interactive Highway Design Models under current development.

18. This methodology inherently assumes that omitted variables have random effects. In reality, many important variables that are omitted are expected to have fixed effects.

19. Many variables (e.g., AADT) are subject to large sampling and non-sampling errors. The assumption of "error-free" covariates is generally highly questionable.

20. The validity of the Poisson assumption depends on the way the sites are defined in terms of level of aggregation, e.g., length of road sections, radius of intersections, length of time intervals, etc. In theory, the assumption is violated when a site is not totally "homogeneous."

21. The existing EB methodology is weak in distinguishing "treatable" and "untreatable" variables, which is a very desirable capability for design engineers to effectively rank the "unsafety" of various sites.

22. The combination of generalized linear modeling and the empirical Bayes viewpoint (the extended EBEST methodology) represents a genuine advance in the statistical methods available to highway safety analysts. The EB method offers a procedure for the rational combination of group and individual information with modest data sets.

23. The validity of any conclusions is conditional on the satisfaction of a set of assumptions. At present, there are no straightforward guidelines for assessing the validity of these assumptions, so it appears that the method could be easily "misused." If a potential user cannot explain how one might validate the exchangeability and distributional assumption for their particular application, they should seek consultation before applying the method. This, of course, applies to the application of any statistical methodology.

24. Application of EB methods are least controversial when it can be assumed that the observables are generated by a hierarchical random mechanism, e.g., where accident rates are assigned to sites as independent gamma random outcomes and then accident

counts are Poisson outcomes conditioned on the site-specific rates. The EB estimates can then be justified as large-sample approximations. Before the method receives wide distribution, some estimates of the accuracy as a function of sample size should be obtained.

25. There are basically two types of questions addressed in safety evaluations: (1) Has the treatment had an aggregate beneficial effect? and (2) Has the treatment had a beneficial effect at a particular site? EB methods are most effective in addressing the latter question. The former question can generally be addressed using classical methods conducted on well-designed "clinical" experiments. The role of the comparison group is to provide the necessary data for conducting these experiments, while the reference group provides the input necessary to address site-specific issues. The debate at the workshop proved that at present there is no common understanding of what counts as a good reference group. The two options, then, seem to be to abandon reference group designs in favor of true controlled experiments or to develop a commonly agreed-upon protocol for the construction of acceptable reference groups.

26. Collinearity among the extended EBEST methodology will be a problem, as well as outliers, etc. Ordinary least squares provide indicators of such problems in the modeling process. A potential extension of this research might be the development of model diagnostics and internal checks on the validity of model assumptions.

27. The argument that RTM is a result of poor design, but practitioners will not employ good design, in practice may be because they have no good models to follow. Future research should be focused on developing examples of well-conducted field research to serve as paradigms and for developing a scenario of potential "protocols" for designing good studies.

28. A user-friendly version of BEATS using the extended methodology is highly desirable and should include a profile of the likelihood graph for the zeta parameter. Perhaps two separate programs should be available — one for high hazard identification that could be very simple and distributed widely to non-sophisticated practitioners and one for safety evaluation that would be more complex and would require a sophisticated user who understands the methodology assumptions, etc.

29. The issue of uncertainty in exposure measures (AADT) is critical and essential to the application of all research methodologies.

30. Future research should be directed at conducting a sensitivity analysis on the reference group, both in terms of type and quantity of sites selected. This should also include a test of the "robustness" of the exchangeability assumption.

31. Develop a list of step-by-step instructions for engineers to follow in designing problem identification and before/after evaluation studies.

32. Explain the exchangeability assumption in "plain English."

33. For future research, allow the convolution parameter to be a function of exogenous variables and to vary by site.

34. For future research, allow fixed-effect models to be easily entertained and let the intercept term vary by individual sites.

35. Further work on the EB methodology is not warranted and should not have a high priority within the FHWA research program. This is not an indictment of the methodology, but the recognition that many research topics demand the attention and resources of the FHWA.

36. The FHWA should promote and encourage the use of the EB methodology by the States for problem identification. If States can see that their own predictions of future accidents at specific sites are "less wrong more often" when the EB methodology is used, they may begin to accept the methodology. A "hands-on" demonstration that the methodology works will be more effective than a tutorial on "why" the methodology works.

37. For future research, could protocols be developed whereby comparisons are run to ascertain whether or not the reference group is suspect? Simulations could be conducted whereby samples from populations of gamma distributions could be sampled and divided into treatment and reference groups in various ways based on the samples accident and exposure (and possibly covariate) characteristics. Then the consequences of selecting a "poor" reference group (too small, too biased, etc.) could be evaluated since the treatment effects and population characteristics are known.

38. How are rate-based analyses (i.e., in problem and high hazard location identification) impacted by the fact that accidents and VMT appear to be non-linearly related? How serious must the non-linearity be to affect such analyses?

WORKSHOP PANEL ROSTER

Dr. Gary A. Davis
Dept. of Civil and Mineral Engineering
500 Pillsbury Drive
Minneapolis, MN 55455-0220

Lindsay Griffin
Texas Transportation Institute
Texas A&M University
College Station, TX 77843-3135

Dr. Shaw-Pin Miaou
Center for Transportation Analysis
Energy Division
Oak Ridge National Laboratories
P.O. Box 2008 MS-6366, 5500A
Oak Ridge, TN 37831

Martin Parker
Martin R. Parker & Associates
38549 Laurenwood Drive
Wayne, MI 48184-1073

Dr. Bhagwant Persaud
Ryerson Polytechnical University
Department of Civil Engineering
350 Victoria Street
Toronto, Ontario, Canada M5B 2K3

REFERENCES

1. Griffin, L.I., III. *Three Procedures for Evaluating Highway Safety Improvement Programs*, TARE-51, Texas Transportation Institute, College Station, TX, 1982.

2. Pendleton, O.J. *Application of New Accident Analysis Methodologies: Volume I, General Methodology*, Publication No. FHWA-RD-90-091, FHWA, Washington, DC, 1991.

3. Christiansen, C.L. and Morris, C.N. *Hierarchical Poisson Regression Modeling*, Report No. HCP-1994-2, Harvard Medical School, Cambridge, MA, 1994.

4. Hauer, E. *Identification of Priority Investigation Locations in New York State*, Transportation Research Board, Washington, DC (publication pending).

5. Parker, M.R. *Comparison of Speed Zoning Procedures and Their Effectiveness*, No. 89-1204, Michigan Department of Transportation, 1992.

6. Wang, J. *The Application of an Improved Accident Analysis Method for Highway Safety Evaluations*, Publication No. FHWA-RD-94-082, FHWA, Washington, DC, 1994.

7. Michaels, R.M. "Two Simple Techniques for Determining the Significance of Accident-Reducing Measures," *Traffic Engineering*, September 1966, pp. 45-48.

8. Hauer, E. "Reflections on the Methods of Statistical Inference in Research on the Effect of Safety Countermeasures," *Accident Analysis and Prevention*, **15**, No. 4, 1983, pp. 275-285.

9. Arnold, S.F. and Antle, C.E. "An Empirical Bayes Solution for the Problem Considered by Williford and Murdock," *Accident Analysis and Prevention*, **20**, No. 4, 1982, pp. 229-301.

10. Davis, G.A. *A Statistical Method for Identifying Areas of High Crash Risk to Older Drivers*, Minnesota Department of Transportation, 1991.

11. Hauer, E. "Empirical Bayes Approach to the Estimation of 'Unsafety': The Multivariate Regression Method," *Accident Analysis and Prevention*, **24**, No. 5, 1992, pp. 457-477.