U.S. Department
of Transportation

**Federal Highway
Administration**

Publication No. FHWA-RD-90-091
September 1991

Application of New Accident Analysis Methodologies

Volume I: General Methodology

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

FOREWORD

This research was initiated by a request from the Federal Highway Administration's Office of Safety and Traffic Operations R&D. The study developed methodology to correct for one of the most serious problems in accident analysis -- the regression-to-the-mean bias. Regression-to-the-mean is the phenomenon where the number of accidents at a high-accident location decreases even if no safety improvements are made. In addition, a menu-driven microcomputer program was developed to allow easy application of this new analysis technique. The method developed in this study provides a better estimate of the expected safety for a site.

The report is in three volumes. Volume I presents an intuitive, non-technical explanation of the regression-to-the-mean methodology. The required assumptions and data requirements are defined in lay terms for ease of comprehension to the highway engineer. Technical, statistical explanations are relegated to Volume III. Volume II of the report briefly describes the computer program and presents examples of the computer output, focusing on interpretation of the output results. Parties interested in receiving the computer program should contact Michael S. Griffith of the Federal Highway Administration on (703) 285-2382.

Volumes I and II will be distributed with two copies to each Region and six copies to each Division Office. Four of the Division copies should be sent to the State. Volume III will be distributed on a limited basis, one copy to each Region and two copies to each Division Office. One of the Division copies should be sent to the State. All volumes of the report will be sent to the Transportation Research Information Service Network, Department of Transportation Library, and the National Technical Information Service (NTIS) in Springfield, Virginia, to be available for interested parties.

R. J. Betsold
Director, Office of Safety and Traffic Operations Research and Development

NOTICE

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-RD-90-091	PB92-115179	

4. Title and Subtitle
APPLICATION OF NEW ACCIDENT ANALYSIS METHODOLOGIES
Volume I - General Methodology

5. Report Date
September 1991

6. Performing Organization Code

7. Author(s)
Olga J. Pendleton

8. Performing Organization Report No.

9. Performing Organization Name and Address
Texas Transportation Institute
The Texas A&M University System
College Station, TX 77843-3135

10. Work Unit No.
2A3B0111

11. Contract or Grant No.
DTFH61-87-C-00015

12. Sponsoring Agency Name and Address
Office of Safety and Traffic Operations R & D
Federal Highway Administration
6300 Georgetown, Pike
McLean, VA 22101-2296

13. Type of Report and Period Covered
Final Report
June 1987 - May 1991

14. Sponsoring Agency Code

15. Supplementary Notes
Contracting Officer's Technical Representative:
Michael S. Griffith, HSR-30

16. Abstract

Researchers in the field of accident analysis have long been aware of the problems associated with drawing statistical inference on safety using accident data. Aside from the problems of accessibility and quality, accident data present a real challenge when it comes to statistical analysis. One of the most serious problems in accident analysis is the regression-to-the-mean bias which occurs due to the non-random site selection process in safety measure evaluation studies. This study presents a new empirical Bayes method (EBEST) which adjusts for regression-to-the-mean bias. Three typical applications in accident analysis are considered for regression-to-the-mean bias, namely: 1. the evaluation of safety treatments, 2. the identification of high hazard locations, and 3. the assimilation of information from multiple safety measure studies (meta-analysis). A computer program was developed to execute these analyses as a part of this study. This manuscript describes the EBEST (Empirical Bayes Estimation of Safety and Transportation) methodology and presents examples of how the method works for each of the three accident analysis applications. This report appears in three volumes. Volume I, General Methodology, FHWA-RD-90-091, is a non-statistical review of the study. Volume II, A Users Manual for BEATS, FHWA-RD-91-014, is a user's manual for BEATS computer program, and Volume III, Theoretical Development of New Accident Analysis Methodology, FHWA-RD-91-015, contains the theoretical development of the procedure.

17. Key Words	18. Distribution Statement
Regression-to-mean, accident analysis safety treatment evaluation, empirical Bayes, before-after studies, meta-analysis, synthesis studies, high hazard location ranking	No Restriction. This document is available to the public through the National Technical Service, 5285 Port Royal Rd., Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	76	

Form DOT F 1700.7 (8-69)

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimetres	mm
ft	feet	0.305	metres	m
yd	yards	0.914	metres	m
mi	miles	1.61	kilometres	km
		AREA		
in^2	square inches	645.2	millimetres squared	mm^2
ft^2	square feet	0.093	metres squared	m^2
yd^2	square yards	0.836	metres squared	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	kilometres squared	km^2
		VOLUME		
fl oz	fluid ounces	29.57	millilitres	mL
gal	gallons	3.785	litres	L
ft^3	cubic feet	0.028	metres cubed	m^3
yd^3	cubic yards	0.765	metres cubed	m^3

NOTE: Volumes greater than 1000 L shall be shown in m^3.

Symbol	When You Know	Multiply By	To Find	Symbol
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5(F-32)/9	Celcius temperature	°C
		Illumination		
fc	foot-candles	10.76	lux	lx
fL	foot-Lamberts	3.426	candela/m^2	cd/m^2

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimetres	0.039	inches	in
m	metres	3.28	feet	ft
m	metres	1.09	yards	yd
km	kilometres	0.621	miles	mi
		AREA		
mm^2	millimetres squared	0.0016	square inches	in^2
m^2	metres squared	10.764	square feet	ft^2
ha	hectares	2.47	acres	ac
km^2	kilometres squared	0.386	square miles	mi^2
		VOLUME		
mL	millilitres	0.034	fluid ounces	fl oz
L	litres	0.264	gallons	gal
m^3	metres cubed	35.315	cubic feet	ft^3
m^3	metres cubed	1.308	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams	1.102	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

* SI is the symbol for the International System of Measurement

(Revised July 1989)

Table of Contents

VOLUME I

VOLUME I - (Con't.)

VOLUME II

VOLUME III

List of Figures

VOLUME I

VOLUME II

VOLUME III

List of Tables

VOLUME I

VOLUME II

VOLUME II - (Cont'd.)

VOLUME II - (Cont'd.)

VOLUME II - (Cont'd.)

VOLUME III

Chapter 1 - Setting the Stage

1.1 Introduction

The statistical analysis of accident data has long been considered a difficult and sometimes seemingly intractable problem to those faced with the challenge of gleaning useful and accurate information about roadside safety. When existing statistical methods are used for such purposes as evaluating the effect of highway safety measures, the results are often disappointing and inconclusive. After devoting considerable effort and funding in collecting the appropriate accident data, a safety improvement which was thought to prove obviously effective turns out to be ineffective or, perhaps, one which is thought to have minor impact yields surprisingly large reductions in accidents. Why is this happening and what can be done about it? Is the problem with the data, the statistical method, the design or a combination of these? How does the transportation researcher recognize the problem and more importantly how is it remedied? These are the questions to be addressed in the following pages along with solutions which promise to be effective.

Three particular safety applications are featured, namely, the evaluation of safety treatments, the ranking and identification of high hazard locations, and the combining of safety information from multiple safety studies for drawing an overall assessment (e.g., synthesis study analyses). Each of these applications shares a common problem when accident data are used for these purposes. This problem involves the very nature of accident data, namely, that accident occurrence represents a rare, low probability random event. Unfortunately, most of the statistical methods which are commonly used in these applications do not recognize or utilize this very basic characteristic of accident data.

The reasons that the popular statistical methods are used in spite of this deficiency are obvious. To begin with, these methods are generally the only methods that are taught in basic statistical courses. Subsequently, they are the only methods for which user friendly computer software is readily available. Even accident data analysts, who are aware that they are using an ineffective tool continue to use the standard, normal distribution statistics such as the T-test, analysis of variance, or regression. As a result, a tremendous loss of statistical information has occurred.

All of these problems, pose a serious logistic impasse to the appropriate analysis of accident data. However, more recently, an even greater problem has surfaced. This problem is regression-to-the-mean sampling bias which inadvertently results from the usual treatment site selection process. Specifically, only high accident locations are generally selected for treatment whch is contradictory to the unbiased statistical sampling procedure of random sampling.

1.2 Problems in Accident Analysis

There are three major problems in finding the appropriate statistical method for analyzing accident data. The first is the one just described, namely, the non-Gaussian (non-normal) distribution property of accident counts and rates. The second problem is the potential confounding of a time effect with the treatment effect. However, the third problem is potentially more serious, namely the inability to design accident safety studies based on purely random, unbiased samples. Frequentist statistical techniques, even if based on the more appropriate discrete distribution, still require that samples (sites) be randomly selected.[8,18] Violation of this requirement results in a biased sample and a potentially serious confounding phenomenon known as regression-to-the-mean (r-t-m). In the text that follows, the traditional statistical methods will be reviewed and the need for new methods to adjust for r-t-m will be illustrated.

In the evaluation of a highway safety treatment, the most primitive statistical method is the two-sample T-test comparing before and after accidents at the treated sites. It was readily recognized that there was a serious problem with this method - namely, that other effects which were unrelated to the treatment, such as time, could be confounding the true effect of the treatment. To illustrate, suppose the treated sites were experiencing an annual increase in accidents due to increasing traffic volumes. Then, accidents in the year following treatment may have been expected to increase regardless of the treatment. Now further suppose the treatment does indeed reduce accidents. The true amount that accidents are reduced will be obscured by the increase caused by time. The treatment effect, in this case, would be diluted and the researchers might conclude that the treatment was ineffective when, in fact, it was *effective*.

This problem was recognized many years ago and the following solution proposed. To counter the confounding effects such as time, select a similar group of sites (control or comparison groups) and observe accidents at these sites over the same period as the treated sites. With this design, a statistic can be computed which will adjust for the time effect. This statistic is called the odds ratio or cross-product ratio and the design is called the before-after with comparison group design.

The confounding of a treatment effect with the time effect and the role of a comparison group in adjusting for this confounding will be illustrated in the following hypothetical example. At the same time, the importance of the log transformation when dealing with *relative* changes in accident counts will also be demonstrated. An excellent account of this phenomenon can be found in an article by Griffin.[9]

Suppose the data in table 1 reflects total accidents at 10 treated and 10 comparison sites for 1 year before and 1 year after some safety treatment was imposed at the 10 treated sites.

Table 1.
Comparison group adjustment for time effect confounding.

	Total Accidents Comparison	Treatment
Before	400	100
After without Trt	800	200
After with Trt	800	100

The data in table 1 assumes there is a doubling of accidents over time, as seen in the comparison group which increased from 400 to 800 accidents. This would mean that of the 100 accidents we observed in the treated sites, 200 accidents would have been expected due to the doubling over time. Now suppose the treatment was effective and reduced accidents by 50-percent. Of the 200 accidents we expected to see at the treated sites, only 100 would have occurred due to the treatment's effectiveness.

Analyzing this data without benefit of knowledge of the doubling effect due to time (which we gleaned from the comparison group data), would have resulted in the conclusion that there was no change in accidents due to the treatment, i.e., we observed 100 accidents before and 100 after. Using the cross-product ratio, however, the treatment effect is computed as the relative change of the odds of accidents before to after in the comparison group to the change in odds in the treatment group or

$$\frac{400/800 - 100/100}{100/100} = -.50 \tag{1}$$

The negative sign indicates a reduction and thus this quantity can be thought of as a measure of the percent reduction in accidents, or 50-percent.

Graphically, this can be seen in figure 1.

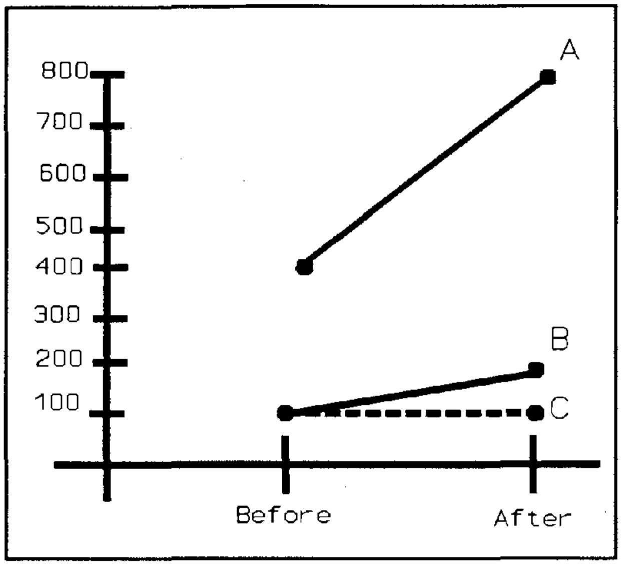


Figure 1. Hypothetical time effect.

Referring to the numbers in table 1, note that A = comparison sites, B = treatment sites without treatment, C = treatment sites with treatment. The distance between lines B and C in the after period for the treatment sites reflects the reduction actually due to treatment. Graphing this same data on the log scale, i.e., the log of total accidents as the ordinate scale would reflect a truer picture of the relative change in accidents relative to the accident magnitudes as shown in figure 2.

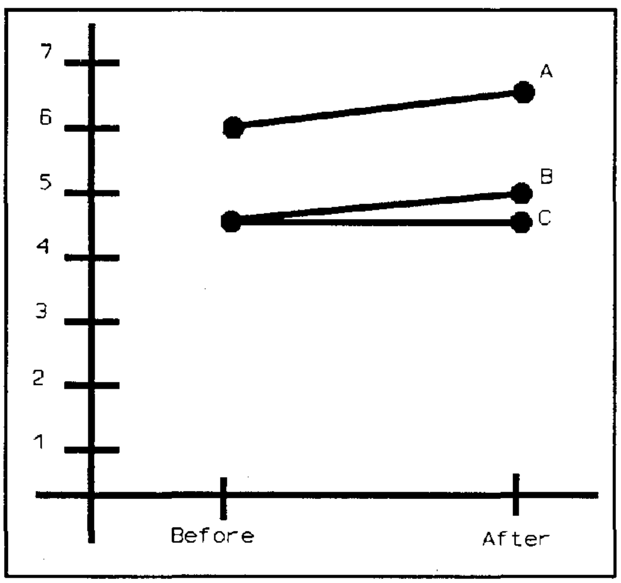


Figure 2. Hypothetical time effect - log scale.

Note that in figure 2, the lines depicting the time effect are now parallel for both groups as they should be since a doubling effect was experienced in both groups. The log transformation reflects this since it allows for a relative comparison of change not an absolute one. This is one important reason for analysing accident frequencies in the log scale. The test statistic for testing the statistical significance of the treatment effect takes this into account as:

$$T = \frac{\ln(O.R.)}{S.D.(O.R.)} \tag{2}$$

where the O.R. is the odds ratio and S.D. represents the standard deviation of the odds ratio. This test statistic has a standard normal distribution and is compared to normal z-values at a specified level of significance. Generically, the data of table 1 can be depicted as:

	Comparison	Treatment
Before	A	B
After	C	D

Then

$$O.R. = \frac{A/C}{B/D} = \frac{AB}{CD} \tag{3}$$

and

$$S.D.(O.R.) = \sqrt{\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}} \tag{4}$$

As an example, to test the significance of the treatment for the hypothetical example of table 1:

$$O.R. = \frac{(400)(100)}{(800)(100)} = .50 \tag{5}$$

The odds ratio minus 1 is used to determine the direction of the effect for tables set-up in the order of table 1, i.e., comparison on the left, treatment on the right, before above and after below. Note that changing this order changes the interpretation of the direction of the effect. In this example, the effect is

$$O.R. - 1 = -.5 \tag{6}$$

or a 50-percent reduction.

Now the standard deviation of the log odds ratio is

$$S.D.(O.R.) = \sqrt{\frac{1}{400} + \frac{1}{100} + \frac{1}{800} + \frac{1}{100}} = .154 \tag{7}$$

The test statistic is then,

$$T = \frac{\ln(.5)}{.154} = -4.50 \tag{8}$$

At the 5-percent level of significance, this number would be compared to the z-value of -1.960 for a two-tailed test and since the test statistic is less than this value, the conclusion would be that the 50-percent reduction in accidents due to this treatment is statistically significant, i.e., the treatment is significant. Without the comparison group, the test statistic for evaluating the treatment effect would have been

$$T = \frac{B-D}{\sqrt{B+D}} = \frac{100-100}{\sqrt{100+100}} = 0 \tag{9}$$

This test statistic would be compared to the same z - value at the 5-percent level of significance and the conclusion would have been that there was no significant change in accidents to due treatment.

The above examples serve to show how one very important potential confounding effect, time, can be adjusted for in safety treatment evaluations with the additional information (data) provided by the comparison group design. However, there is another very important confounding effect which is not adjusted for by this design and requires additional information and a new statistical methodology to account for it. The problem is regression-to-the-mean (r-t-m).

The sampling bias due to the r-t-m phenomenon can seriously affect conclusions drawn in safety treatment evaluation studies. Highway sections are generally selected for treatment because the number of accidents at the treatment sites is unusually high. Thus, these treatment sections represent a sample from the upper end of the population distribution of accidents from which it was drawn. Another sample drawn from this population, at some future time (after treatment), would be expected to be closer to the center of the distribution. Thus, if a site has an unusually high number of accidents occurring before treatment, accident occurrence at that same site the following year would, in all probability, be lower, apart from any intervention at that site. This is the very real phenomenon known as r-t-m. Figure 3 depicts various degrees of the r-t-m phenomenon.

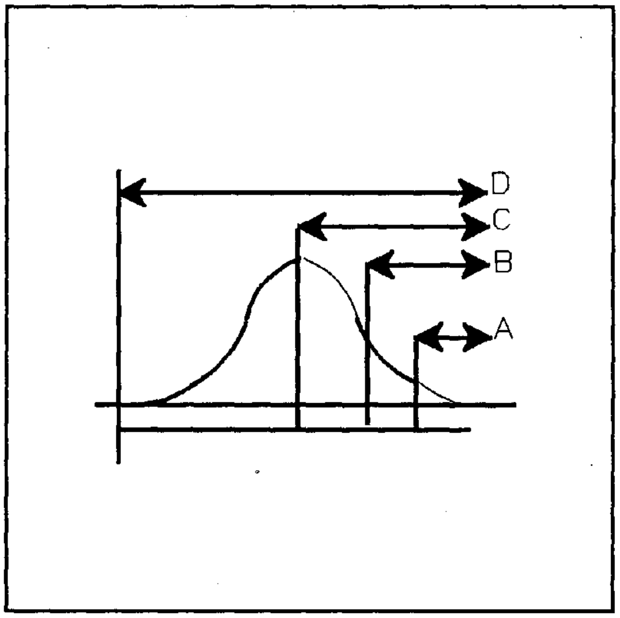


Figure 3. Regression-to-the-mean example.

Note: this figure is purely anecdotal and does not represent the real distribution of accident frequencies.

The areas of the curve labeled A through D represent varying degrees of r-t-m potential from most severe to none. That is, if the treatment sites represent a sample from Area A, r-t-m potential is high. A subsequent observation on this same site is likely to be smaller as this value regresses to its true mean. If treatment sites represent samples from area D, they do, in fact, reflect a random sample and r-t-m is not likely to be a problem. The way in which this phenomenon confounds the estimate of treatment effectiveness will be demonstrated in an example in a later section.

The problem of regression-to-the-mean has been recognized for quite some time but only in the past decade have solutions been proposed.[2,10] During this time, the empirical Bayes (EB) methodology has received increasing attention as a proposed solution for the r-t-m problem in accident analysis.[3,4,12,13,17] Although the concept of this methodology has been widely embraced, actual use of this method in accident analysis has been limited. There are several possible explanations for this. EB methods are not traditionally taught and hence researchers in the transportation field are not familiar with them. The subject matter is difficult and has, for the most part, been explained in terms which require a certain degree of statistical training to understand. The computer software for executing the computations is not readily accessible. And finally, but most importantly, the data required for this procedure have not routinely been used. Note that this does not mean that the data were not available. The necessary data are often available in some form but have been traditionally ignored in accident analysis. Specifics on this point will be addressed later.

The procedure developed in this study can adjust for this sampling bias. This procedure is called the EBEST procedure (Empirical Bayesian Estimation of Safety in Transportation). It is most beneficial and superior to classical methods when there is a high degree of sampling bias (r-t-m potential). For this reason, safety measure evaluation is the most critical application for which the EBEST procedure is suited. From a practical standpoint, it is logical to select sites for treatment not in a purely random (ignorant) fashion but according to their likelihood of benefitting from the treatment. When we do this, however, we violate our classical statistical assumption of random sampling.

Whereas the use of empirical Bayes in this application is not new, the EBEST methodology developed in this study differs from previously proposed empirical Bayes methods in two important ways: (1) The EBEST method uses the statistically superior method of maximum likelihood (as opposed to the approximate method of moments) to derive the estimates and (2) the EBEST procedure incorporates a measure of exposure (e.g. traffic volume, section length, etc.) in the prior distribution assumption which allows each site to be evaluated individually and weighted by its exposure.[10] Before developing the EBEST methodology, certain basic terms and concepts of the empirical Bayesian methodology will be established.

1.3 A Comparison of Frequentist and Bayesian Philosophies

Statistical methods can be classified into two broad areas which differ philosophically and mathematically. These areas are Bayesian methods and non-Bayesian (frequentist) methods. The frequentist methods are the ones most often taught and widely used.

Philosophically, Bayesians believe in the use of additional prior information in developing statistical methods. Frequentists methods require no prior information about the unknown parameters. Some view this as operating from a point of uniform ignorance. Others view frequentist methods as being more "fair" and unbiased. There is some feeling that, in the purely Bayesian strategies, such biases could dictate the method's outcome. At any rate, these philosophies have been widely debated for centuries and it is not the intent of this section to join in this debate but merely to point out the philosophical differences of the methods.

The Bayesian methods can be further classified into two areas - pure Bayesian and empirical Bayesian (EB) methods. The pure Bayesian assumes the prior information using no data whereas the empirical Bayesian attempts to obtain estimates of the prior unknown parameters using data. This later requirement has made the empirical Bayes methods much more appealing to some more classically minded researchers since it appears to be less biased than the pure Bayesian methods. In this study, we will present an EB method for adjusting for r-t-m bias in accident analysis.

For this application, focus is on knowledge of the true accident rate for the entire population represented by the potential treatment sites. The Bayesian philosophy embraces the concept that each site has a true accident rate, λ_i, and that a group of k such sites, i = 1,...,k, would have means that would have the same distribution, a gamma distribution, with some mean and variance. Figure 4 is an example of a gamma distribution with a mean of.5 (μ) and different variances, σ^2's.

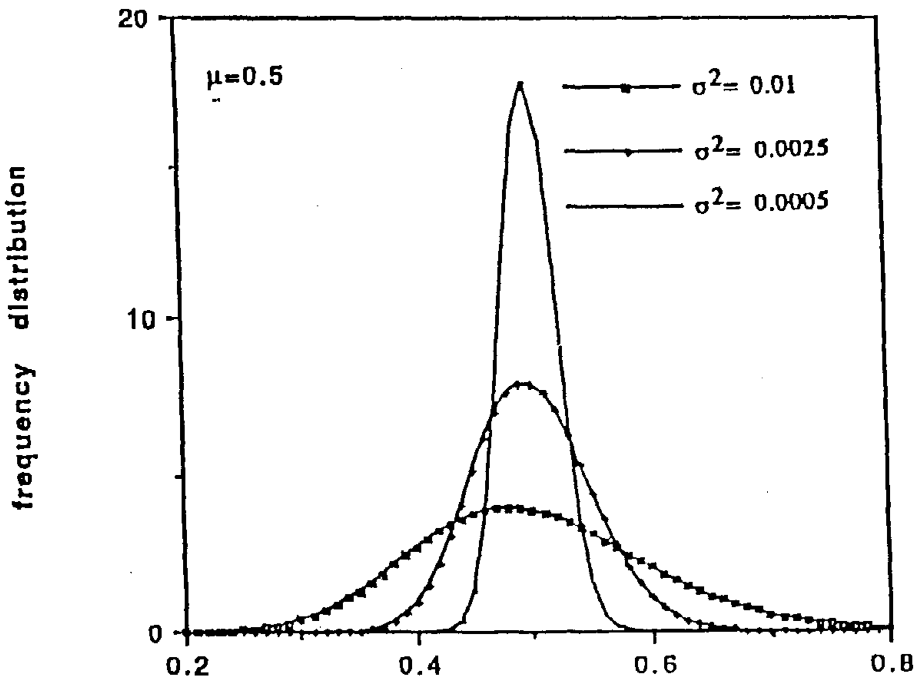


Figure 4. Example of a gamma distribution.

The variances of a gamma distribution representing accident rates is determined by the accident rate and the exposure measure. The exposure measure may be traffic volume, section length, time, etc., depending upon the application. The distribution in figure 4 could represent three highway types all of which have the same average accident rate of .5 accidents per 100,000 vehicle miles traveled, for example, but have different exposures which control the variability of the probability distribution of their mean rates. The variance of a gamma distribution is the mean divided by the exposure, hence, the highway type with the greatest exposure has the smallest variance. Exposures for this example would be 50 (.5/.01), 200(.5/.0025), and 1000(.5/.0005). Hypothetically, then, these distributions might represent highway types of 50,000 vehicle miles, 200,000 vehicle miles, and 1,000,000 vehicle miles per year but which all have true means fluctuating about an annual accident rate of .5 accidents per 100,000.

Knowledge about the gamma distribution allows the Bayesian to make probability statements about the true site means. The pure Bayesian assumes the parameters of this gamma distribution are known from some source (e.g. engineering knowledge) other than data. The EB methodology assumes the prior distribution is known (using engineering knowledge) but estimates the parameters of this distribution, the mean and variance, using data from sites similar to the treatment sites, collectively defined as the reference group.

A necessary prerequisite to obtaining good estimates is having a data set which is representative of the entire population of interest. The non-Bayesian has this same prerequisite and attempts to meet this by requiring a random sample. The empirical Bayesian attempts to meet this prerequisite by requiring a suitable reference group. If, in fact, the sample is truly random and representative of the entire population, results from the non-Bayesian and empirical Bayesian analysis are likely to be similar. The more biased the sample, the more the benefit to be gained from the EB analysis. This, however, depends upon the data. Inadequate or inappropriate data can lead to false conclusions using EB methods just as faulty data can cause problems in any type of analysis. It is essential, then, that the data requirements be clearly understood by those attempting to use this method, especially if the method is to be applied to data from a biased or non-random sample. (Such is generally the case in safety treatment evaluations.) One of the important data requirements is the availability of data on a reference group.

Basically, then EB is a statistical method which allows an adjustment for r-t-m bias in the treatment site selection process and thus provides a more realistic estimate of expected safety apart from any treatment effect. This is accomplished using additional data, or information, about the population from which the treatment sites were drawn. This additional information comes from reference group data.

2.1 Important Concepts and Assumptions

As described in chapter 1, the empirical Bayes philosophy has proven to be a promising solution to the many statistical problems with accident data. Both the problems encountered with the non-normal distribution of accidents and regression-to-the-mean sampling bias can be resolved using the appropriate prior distribution assumptions and the data to estimate the parameters of that distribution.

Hauer and others have suggested EB methods but the parameter estimating procedures recommended (method of moments) are not statistically optimal.[10] The reason for this was, no doubt, the computational complexity of the more optimal procedure, the method of maximum likelihood.[14] In this study, this computational problem has been resolved and a very accurate and efficient computer program developed for obtaining maximum likelihood estimates. Another critical issue which was not addressed in earlier studies in this area is the use of exposure measures in obtaining the estimates. With accident data, a dominant factor in reflecting a site's exposure to accident potential is traffic volume. The method developed in this study does, however, use measures of exposure, whether it be traffic volume, section length, or duration of the treatment period, in obtaining the estimates of expected safety for a given site.

The method developed herein, the EBEST method, provides a statistically optimal estimate of the expected safety of a site adjusting for both the site's exposure to accident potential and r-t-m sampling bias. Like any statistical method, the EBEST method requires certain assumptions to be met. Also, it is essential that the data satisfy certain standards of adequacy and quality. The assumptions and data requirements will be defined in this chapter.

2.2 The Reference Group

A critical data requirement that sets EB methods apart from frequentist methods is the use of data on a reference group. The reference group and treatment group, collectively, should represent the entire population of potential treatment sites. It is important that the reference group be carefully defined so that it satisfies this requirement.

This reference group data may be available either prospectively at the time of treatment site selection, or retrospectively, after treatment implementation. Collecting data retrospectively for the reference group may not be easy. Two critical considerations are (1) defining the appropriate reference group and (2) retrieving the accident data. Reference groups may not be difficult to define. For example, if the treated sites were a group of urban intersections, a reference group might be a large representative sample of all urban intersections with similar characteristics (roadway geometry, type of traffic, etc.) that potentially would have received the same treatment. Likewise, data retrieval may not be so difficult due to advances in data retrieval software. An example of such software is the

LANSER program (Local Area Network Safety Evaluation and Reporting System).[5] LANSER is an accident data retrieval program which allows the user to specify both highway and accident data descriptors and then create data subsets which meet these criteria. This is done using a PC network system which contains up to 4 years of Texas accident data. Hence, a main frame computer is not required and data retrieval is both inexpensive and expedient. Systems such as LANSER greatly facilitate retrospective accident data collection.

Another possibility is that the reference data may already exist. That is, when the treatment group was selected, it is possible that the entire population of potential treatment sites was available. In Texas, for example, highway departments have access to the accident histories for the entire population of intersections which are candidates for signalization. These candidate intersections are generally ranked according to accident count and/or rate and then those with highest values are selected for treatment. Therefore, the reference group was available at some point in time, but then ignored in the safety measure evaluation process. In this case, the reference group data was available prospectively, before treatment implementation.

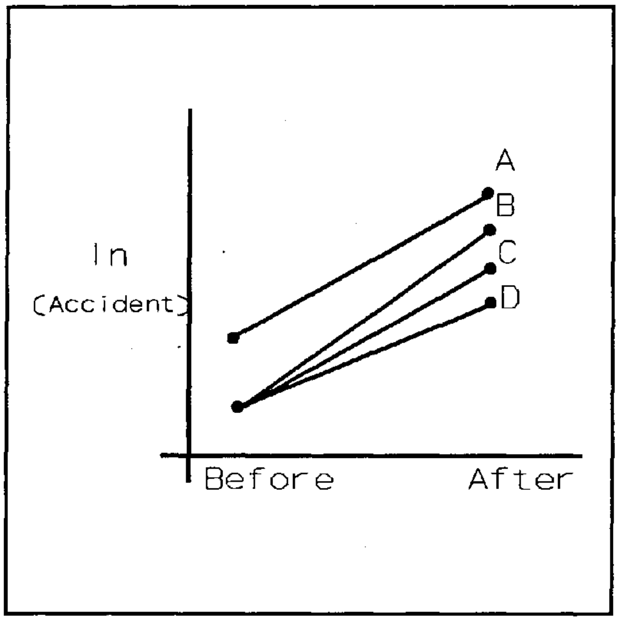
It is important, then, that researchers in this area recognize the need for reference group data and preserve and <u>use</u> this information in the safety evaluation process. If this data requirement is recognized at the time of treatment site selection, the more tedious task of retrospective reference data retrieval may be avoided altogether.

The adequacy of the reference group data, both in quality <u>and</u> quantity is critical to the success of EBEST or any EB method. The number of sites in the reference group should be large, much larger (say five times larger) than the number of treatment sites. This will often be a natural consequence if the data which were available when sites are selected is preserved for future analyses. Also, it is important that the reference group truly represent the potential treatment sites and that this group be fairly homogeneous. Again, these criteria are automatically met if the reference data is obtained from the original data used in the treatment site selection process and if this data represented the population of potentially treatable sites. When reference data is obtained retrospectively, it is essential that the quality of this data, i.e., the definition of the reference group site characteristics, and the quantity be critically assessed.

In some respects, the reference group may appear to be a comparison group. However, this is not the case. There is a very important, though subtle, difference. Comparison groups are selected for one primary purpose - to represent a time trend. There is no restriction that the comparison group must come from the same population as potential treatment sites. Indeed, sometimes the comparison group represents a completely different population - for example, when it represents a comparison condition like type of accident (dry versus wet) or time (day versus night). Clearly, daytime accidents do not have the same accident rate as nighttime accidents. Thus, daytime accidents do not constitute a suitable reference group for nighttime accidents but they may serve as a suitable comparison group by representing the trend in time from the before to after period.

reference group for nighttime accidents but they may serve as a suitable comparison group by representing the trend in time from the before to after period.

Conversely, a suitable reference group may not be a suitable comparison group. If the treated group is biasedly selected, then the reference group may also represent a biased sample, biased in the opposite direction. In this case, using the reference group as a comparison to adjust for the time trend could actually worsen the bias in the estimation of the treatment effect. A hypothetical example of this is illustrated in figure 5.



A = True Comparison Group
B = Observed Reference Group
C = True Reference and Treatment Group
D = Observed Treatment Group

Figure 5. Effect of regression-to-the-mean.

In figure 5, an example is presented representing the true and observed accident frequency (in the log scale) that might be expected when the treatment group is biased. The A and C lines represent the true values and the B and D lines represent the observed values. The treatment group represents a group of sites with unusually high accidents. Since the reference and treatment groups come from the same population, they are represented by the same solid line (line C). Now since the treatment group is biased on the side of high accidents, the number of accidents in the after period is likely to decrease apart from any treatment effect (line D). On the other hand, the reference group is likely to experience an <u>increase</u> in accidents since the part of the population that would have had the higher accidents would have been selected for treatment (line B). The relative distance of these lines measured at the before and after points reflects the amount of change one would see during these periods. The true value lines show no change but the observed value lines show that the difference between the treated sites and reference sites is much greater in the after period. If a treatment effect were to be evaluated, using the reference group as a comparison group, the conclusion might be drawn that the treatment was effective, when, in fact, no treatment was even administered.

To further emphasize this, a numerical example is shown in table 2.

Table 2. Hypothetical example illustrating reference, comparison and treatment groups.

	Accident Counts		
True	Reference	Treatment	Comparison
Before	25	25	46
After without treatment	50	50	92
After with treatment	50	25	92
OBSERVED			
Before	15	40	42
After without treatment	40	60	90
After with treatment	40	30	90

Hypothetical before and after accident counts are shown, (for this example we will assume exposures - traffic volume, section length, etc. - are equal and thus the counts would reflect the same trends as the rates). The true accident counts for the reference and treatment groups in the before period are both equal to 25 and the comparison group is assumed to be slightly higher at 46. Now consider the accident counts we would most likely observe during the before period for these groups (row 4). The reference group would be lower than the true (15) and the treatment group would be higher (40) since we have selected sites with high accidents for treatment. Similary, row 2 reflects the true counts assuming a doubling over time with no treatment effect and row 3 reflects the true counts with a 50-percent expected reduction due to treatment in the treatment group. Remember, these are the true values. Now, what are we likely to <u>observe</u> in these groups? Rows 5 and 6 represent likely observed values with and without treatment. The reference group would likely be lower than the true mean, 40 vs 50. The comparison group might remain close to its true mean, 90 vs 92. The treatment group without treatment would be greater than its true mean (50 to 60). Assuming a 50-percent reduction from this due to treatment, then, the treated sites after treatment would be expected to go from 60 to 30.

Now consider estimating treatment effectiveness using the odds ratio. With the comparison group, the odds ratio is:

$$O.R._c = \frac{42/90}{40/30} = .35 \tag{10}$$

Subtracting one from this value the treatment effect is estimated to be a reduction in accidents of 65-percent. Now suppose we use the reference group as a comparison group:

$$O.R._R = \frac{15/40}{40/30} = .28 \tag{11}$$

Using the reference group we would estimate an even greater treatment effect of 72-percent reduction. Recall the actual reduction was 50-percent.

Thus, using a reference group as a comparison group can cause an even greater distortion in the estimation of a treatment's effect, especially if the number of reference sites is small. Of course, this can occur if the comparison group is small and biasedly selected, as well. If the comparison group does not represent a truely random sample and is somehow selected biasedly according to high or low accident experience, distortion will occur. If the reference group is to be used in place of a comparison group, this should only be done if the number of reference group sites is extremely large (say five times) relative to the number of treatment sites to diminish the effect of any bias.

Examples of potentially appropriate reference groups for three safety treatments are given in table 3. In the first case, if the treatment was conversion of two-way stops to four-way at urban intersections, it would not be appropriate to consider all urban intersections regardless of signing as a reference group since these do not constitute a sample of potentially treatable intersections. Only two-way stop intersections would be candidates for treatment. Furthermore, additional restrictions such as traffic volumes might be considered since very low volumed two-way stops are also probably not potential candidates for conversion to four-way stops.

Table 3. Examples of reference groups for various safety treatments.

TREATMENT:	Conversion to 4-way Stop from 2-way for urban intersections
APPROPRIATE REF:	All 2-way Stop urban intersections which could have been converted to 4-way
INAPPROP. REF:	All urban intersections regardless of stop sign or signal type
TREATMENT:	Resurfacing of 2-lane roads that have not been resurfaced in two or more years
APPROPRIATE REF:	All 2-lane rural roads which could have been resurfaced and have not been resurfaced in the last two years
INAPPROP. REF:	All 2-lane rural roads, even those that were just resurfaced
TREATMENT:	Raised pavement marker installation on 2-lane unlit rural roads with curvature of more than some specified amount
APPROPRIATE REF:	2-lane unlit rural roads with the same degree of curvature which might have been selected for treatment
INAPPROP. REF:	All two-lane rural roads including lengthy sections with no curvature

2.3 The EBEST Methodology

In this section, an intuitive explanation of how the EBEST method works will be presented, in as non-technical terms as possible. Those interested in more technical details are referred to volume III.

The EBEST procedure uses the reference and treatment group together to estimate the unknown parameters of the assumed gamma distribution for the true site mean accident rates, λ_i. The sample data, accident counts (z_i), are assumed to represent a Poisson distribution about some true mean accident count ($\lambda_i e_i$, where e_i is the site's exposure). The estimate of these gamma distribution parameters is based upon the accident counts and exposures at all (treatment plus reference group) sites. Exposure may be traffic volume, section length, number of months, etc. The key to defining exposure for a given problem rests in the assumption of exchangeability which will be explained in the next section.

The EBEST estimate of the true site accident rate, $\hat{\lambda}_i$, is a function of both the estimated mean rate for all the sites, $\hat{\mu}$, and the site's observed accident rate, y_i, as follows:

$$\hat{\lambda}_i = B_i\hat{\mu} + (1 - B_i)y_i \tag{12}$$

where B_i, is:

$$B_i = \frac{\epsilon_i}{e_i + \epsilon_i} \tag{13}$$

e_i being the site's observed exposure (traffic volume, section, length, etc.) and ϵ_i the estimated value of the site's exposure using the EBEST estimate based on all sites. Basically, this value controls the variability of the gamma distribution for the site means.

The expected accident rate, $\hat{\lambda}_i$, then, is a value somewhere between the observed value for that site, y_i, and the estimated true accident rate for the population of potential treatment sites, $\hat{\mu}$. The amount that the observed value is adjusted by is called the shrinkage factor, B_i. If there is not much regression-to-the-mean, the value of the B_i will be small and the estimated rate for site i will be similar to the observed value. If the value of B_i is close to one, the estimated rate for site i will be closer to the estimated rate for the entire population of potential treatment sites. Thus, the shrinkage factor, B_i, indicates how much "weight" is given to the observed site information in estimating its true mean rate.

By contrast, the frequentist, non-Bayesian, methodology estimates the true accident rate for all sites using only the observed treated site data, y_i. No reference group data is ever required or used. In so doing, the r-t-m bias is confounded in the estimate and no prior knowledge about the distribution of site means is ever used (the true site means are treated as a fixed, not random, variable). The *pure* Bayesian methodology, on the other hand, would not use the data to estimate the parameters but rather, use known (or guessed-at) values. The EB method incorporates something of both the non-Bayesian and pure Bayesian methods by using data and prior knowledge simultaneously. In this way, information from all of the sites is used in the estimation of the rate of a particular site.

2.4 EBEST Assumptions

The assumptions on the probability distributions - namely, the Poisson for accident counts and the gamma for the accident rates, are reasonable and easily justified based on the nature of accident occurrence. However, there is one critical assumption in the EB methodology which requires careful scrutiny by the would-be user. This is the assumption of exchangeability.

Exchangeability is synonymous with assuming that the true site means are identically and independently distributed. In other words, our sample of potential treatment sites (reference sites) should be somewhat similar and homogeneous with regard to factors that influence their safety. Traffic volume is one obvious dramatic factor to be considered. Other more subtle factors that could affect exchangeability are highway geometrics or other site specific factors. If these are identified, sites can be divided into more homogeneous subgroups in the estimation process.

The key, then, to satisfying exchangeability is to identify all factors which might affect a site's safety and to ensure that either (1) all sites in the potential treatment site data set (reference group) are homogeneous with regard to these factors or (2) these factors are measured for each site and used in the estimation process. Given the critical nature of this assumption, additional examples will be provided to assist the researcher in ensuring that this assumption is met.

A check for exchangeability would be to ask, "Could I guess which site would be safer based on the information I have apriori, i.e., before examining the data?" If so, the true site means are not exchangeable. Suppose that the only data available for assessing safety for a group of sites is their accident counts. Without any other measures of exposure such as traffic volume, the accident count must serve as a surrogate measure for the site's safety. Suppose, now, that some very busy urban intersections are combined with low volume residential intersections in the treatment group. Before the data is even collected one could guess which intersection would have more accidents. In this example, the exchangeability assumption about the true site accident *counts* is violated. To satisfy it, traffic volume data is necessary. Now, consider the same question about a site's accident *rate*, defined as accidents per vehicle. Is it possible to guess which intersection would have the highest rate? If not, then we can consider accident rate, as defined here, as being exchangeable and implement the EBEST method.

Although traffic volume is the most obvious variable which affects exchangeability, other variables can be a factor. For example, section length may vary among sites for some safety treatment studies. Another factor might be time periods. For construction zone treatments, duration of the construction period may vary. Collectively, these factors can be termed <u>exposure</u> factors. If sites vary in their exposures, data representing these exposures is essential to the analysis in meeting the exchangeability assumption.

In general, any information about a particular site which would allow you, apriori, to guess at the relative magnitudes of the site means must be included in the analysis. In a way, this is just common sense. If you were asked to compare sites using some measure of their accident experience, you would naturally take these factors into consideration. Therefore, a way to assure that you have sufficient data to satisfy exchangeability is to define the most appropriate measure of accident occurrence that would allow you to compare sites on a fair basis. That is, you would not compare accident counts for a site with 10 MVM (million vehicle miles traveled or 16.1 million kilometers traveled) to a site with 1000 MVM (1610 million kilometers traveled). But you would compare accidents per MVM. You would not compare accidents per 10,000 ADT (average daily traffic) on a 5-mi (8 km) section to a 10-mi (16.1 km) section but you would compare accidents/ADT/mile. You would not compare total fatalities for 3 years at one site to 2 years at another, but you would compare fatal accidents per year. In each of the above cases, trying to estimate a site's expected safety not knowing a site's MVM, section length, or time period violates the exchangeability assumption. The assumption can be met by collecting data on these measures and using this as a measure of exposure in the EBEST procedure.

Exposure, then, is a measure of a sites' relative accident risk. If sites have varying exposures, then some measure of their exposure is essential to satisfy the assumption of exchangeability. Exchangeability is a critical assumption in EB analysis and significant violation of this assumption will invalidate any results and conclusions drawn from the analysis, just as with any statistical procedure.

3.1 Safety Treatment Evaluation

3.1.1 *Safety Treatment Evaluation Concerns*

Accident data often serve as a surrogate measure of safety and, as such, have been used in a variety of ways in researching transportation safety. Of all the uses, the one most common and often subject to much debate and criticism is the evaluation of a highway safety treatment.

The criticisms of accident analysis for highway safety evaluation are well-founded. These criticisms include:

1. Insufficient accident data on which to base conclusions either with respect to numbers of sites or low accident counts.

2. Lack of data on traffic volumes forcing conclusions to to be drawn on accident counts rather than rates.

3. Use of inappropriate statistical methods in the assessment of significance of the effect.

4. Lack of comparison group data, or inappropriate comparison group data to adjust for potential time effects, etc.

5. Non-random treatment site selection leading to potential r-t-m confounding effects.

Most safety measure evaluation studies can be found to fall victim to at least one, if not more, of these criticisms. It has been argued that better planning and experimental design considerations could alleviate many of these problems. That is, more sites or a longer time period could resolve sample size and low count problems. The second problem, traffic volume, is often avoidable as this data is frequently available but not retrieved or used in the analysis.

Comparison group data has been recommended to alleviate problems in safety treatment evaluations. While comparison group data is desirable and can account for changes attributable to other factors, like time, comparison groups are not a panacea for all problems in this area. Chief among these is regression-to-the-mean bias. Comparison groups do nothing for adjusting for r-t-m bias which can cause a treatment to falsely appear to affect accidents. The frequentist, non-Bayesian analysis for a before/after study using a comparison group, while seemingly addressing a similar problem, i.e., adjustment for certain "expected changes independent of the treatment", can only adjust for changes in time (from before to after periods) or differences in comparison and treatment groups (different volumes, section lengths, etc.).

Conversely, previously proposed methods which adjust for r-t-m generally do nothing to control for these other factors. If there is a change in accidents from the before to after period that is purely associated with a time trend or if there is a similar change in traffic volumes, these changes are confounded in the treatment evaluation and unaccounted for even in Hauer's "debiasing" technique.[10]

Finally, it is important to note that no statistical technique is robust enough to overcome some of the problems discussed here - specifically those caused by insufficient data, unavailability of essential data, or data of poor quality. Furthermore, assumptions are almost always required and violation of these assumptions can result in erroneous conclusions. The EBEST method proposed in this study for safety treatment evaluation is no exception. However, the EBEST method *can*, with the right data, alleviate the last three problems. The reference group data is used by the EBEST procedure to adjust for r-t-m bias (problem 5). This same group can be used to adjust for the same confounding effects (time) as the comparison group if the same sites are observed in the post treatment period *and* the number of reference group sites is very large relative to the number of treatment sites (problem 4). The EBEST method is also the most appropriate statistical technique since it uses the appropriate probability distribution assumptions (gamma and Poisson) rather than the normal distribution assumption (problem 3). The EBEST method is the only method proposed to date that will do this.

3.1.2 *The EBEST Method*

The EBEST procedure provides a better estimate of the expected accident experience for a treated site, adjusted for any r-t-m bias. In evaluating a safety treatment, this estimate is used to ask what effect did the treatment have on the safety of these sites, above and beyond any expected changes which are independent of the treatment. The EBEST procedure does this by requiring data on a reference group. The reference group and treatment group together, then, is used to provide an estimate of expected safety adjusted for r-t-m.

Similarly, the reference group data could be used to provide an estimate for the expected safety adjusted for time effects and other factors. That is, it can play the role of a comparison group provided data is available during the after period for the same reference sites. Given that the number of sites in the reference group will be large, this may or may not be a desireable alternative to collecting comparison group data depending on the amount of effort involved.

The following options are then available to the analysts:

1. Collect data on the reference group during the before treatment period <u>and</u> a comparison group before and after treatment.

2. Collect data on the reference group during both the before and after period.

If option 1 is used, the classical odds ratio (also called the cross product ratio) is used to assess a treatment's effect and is computed as:

$$O.R._{1} = \frac{A/B}{C/D} \qquad (14)$$

A = the number of accidents in the comparison group before.
B = the number of accidents in the comparison group after.
C = the EBEST estimate of the number of accidents in the treated group apart from any treatment effect.
D = the number of accidents observed in treated group after.

The quantity, C, is computed using the estimated accident rate for each site from the EBEST method, i.e.,

$$C = \sum_{i=1}^{k} \hat{\lambda}_i e_i \qquad (15)$$

where e_i is the observed site exposure and $\hat{\lambda}_i$ is the EBEST expected accident rate site i, i=1,..., k (12). In other words, C is the total expected accident counts (the product of the expected rate and exposure-traffic volume, section length, etc.)

If option 2 is used, the quantities A and B in equation 14 are defined as:

A = the EBEST estimate of the number of accidents in the reference group using the before data.
B = the observed number of accidents in the reference group after.

The quantities C and D remain unchanged. The quantity A is:

$$A = \sum_{i=1}^{m} \hat{\lambda}_i e_i \qquad (16)$$

where e_i is the observed site exposure and $\hat{\lambda}_i$ the EBEST expected accident rate for the m reference sites. If the reference group is indeed playing the role of a comparison group, the ratio of A/B will be estimating the change due to time alone apart from any other effects, just as the ratio of A/B does for the comparison group.

Note that although the measure of the treatment effect is based on the ratio of counts, exposure (traffic volume, etc.) is implicit in this estimate as it was used to determine the r-t-m adjustment. Also, the cross product ratio adjusts for any differences in exposures between the treatment and comparison group or reference group ratios as long as it is reasonable to assume that the before-to-after change in these ratios is equal for both groups. This is an inherent assumption in using the cross product ratio and is what "makes it work." If this assumption is questionable, so is the cross product ratio statistic.

Given either of the above estimates, statistical estimates and inferences can be drawn using classical methods of hypothesis testing and confidence interval estimation (although Bayesians typically do not embrace these procedures). Nevertheless, the analyst needs an answer to the question "is this safety treatment significantly effective?; yes or no and with what confidence?"

3.1.3 *When is EBEST* **the** *Best ?*

The EBEST method's primary advantage over traditional statistical methods is the ability of the procedure to adjust for r-t-m effects in safety treatment evaluation studies. The question that naturally arises is when and under what conditions is the EBEST method most superior to the simpler classical analyses?

Technically, the EBEST method is always superior to the classical in that it uses more information about the probability distribution of the site means. However, if there is little or no r-t-m bias, the shrinkage coefficients for each site will be close to zero giving nearly full weight to the observed site mean rather than the overall average site mean of all of the sites combined. Therefore, if both options are available, i.e. the EBEST method and the frequentist method, it would be prudent to take advantage of the superior EB method. On the other hand, without the available software for computing the EBEST method, the frequentist method is by far the easiest. Another advantage in using the frequentist method would be ease of explanation to a non-statistical audience. At any rate, the question of when to use EBEST was addressed via a simulation study. This section will summarize this study and its conclusions, again, in non-technical terms.

The objective of the simulation study was to determine the conditions under which the EBEST method would produce significantly superior results from the classical. Since we know the greatest advantage will occur under severe r-t-m sampling bias, this condition was imposed on the simulated data drawn from appropriate probability distributions. Another factor known to produce the greatest difference is the variability of the site means. Hence, the population parameters from which the sampled site means was drawn was selected with varying degrees of variability.

Specifically, samples of size n were drawn from a gamma distribution like that of figure 4 with specified means and variances. This population represents the population of the true mean accident rates for the potentially treatable sites depicted in the lower portion of figure 6. The sample of size n would reflect the sample of treatment and reference group true mean accident rates, with the variability of the distribution being controlled by the exposures ϵ_i, as explained in 1.3. Exposures are sampled for each site from a negative exponential distribution. The negative exponential distribution is a skewed, decaying exponential function which assumes most of the exposure values, e.g., traffic volumes, are clustered about the mean but a few sites may have extremely high values (volumes) as shown in the top portion of figure 6. The particular parameters used in this figure generated exposures with a mean of 14 units and a variance of 16 units. This would correspond to a case where most sites had the same traffic volumes but a few sites had extremely high volumes. The n observed accident counts are then sampled from Poisson distributions with means equal to the product of the site mean rates times the site exposures, i.e., the middle portion of figure 6. These n observed accident counts form the data which is typically available at the time of treatment site selection (i.e., the reference group). The observed site accident counts are ordered, and the highest n_t of them selected to represent the group selected for treatment, typical of the site selection process choosing the sites with highest accidents. The remaining $(n - n_t)$ accident counts represent the observed accident counts for the reference group sites. Table 4 gives an example of a sample of 10 simulated sites before and after treatment, assuming a 10-percent treatment reduction. The 90 reference sites are not listed; however, these 10 sites are the sites with the 10 highest accident counts from the sample of 100.

Table 4.
Sample simulation data.

True site mean rate = .05
Mean exposure = 500
Variance exposure = 300
Number of treated sites = 10
Number of reference sites = 90
Treatment effect = 10% reduction

	Before				**After**	
Site #	Accident Count	Exposures	Accident Rate	Accident Count	Exposure	Accident Rate
1	55	344	.160	59	344	.172
2	48	849	.056	44	849	.052
3	52	328	.158	46	328	.140
4	48	943	.051	41	943	.043
5	45	1202	.037	42	1202	.035
6	39	312	.125	18	312	.058
7	43	727	.059	39	727	.054
8	52	1612	.032	67	1612	.042
9	51	417	.122	46	417	.110
10	68	569	.120	49	569	.086
Total	501			451		

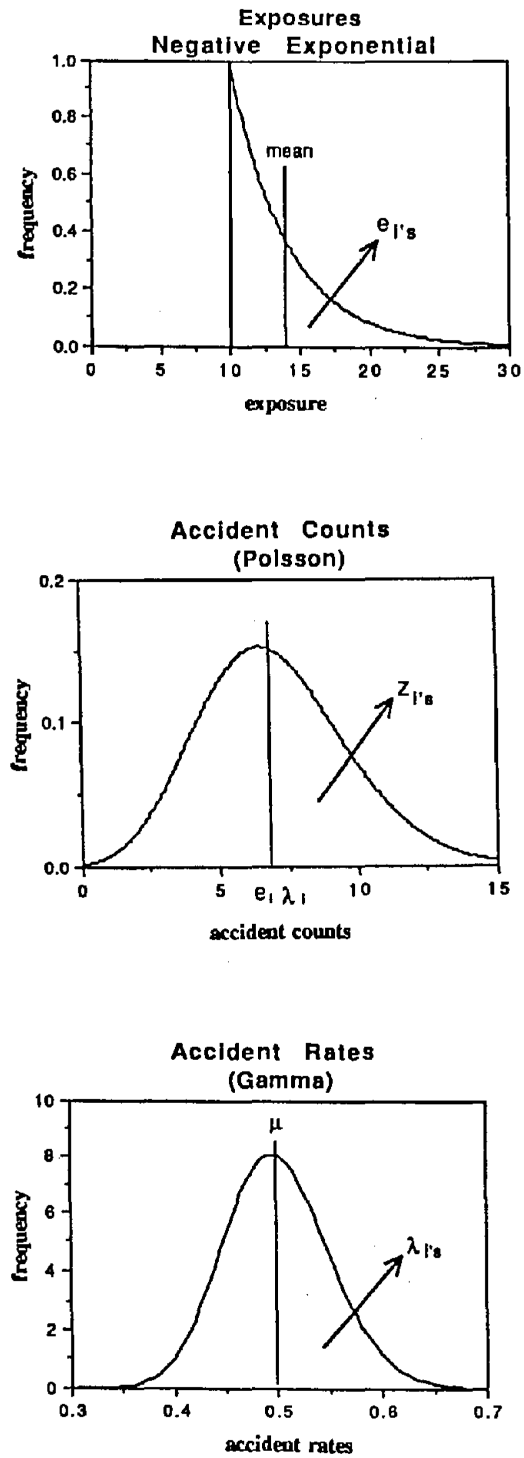


Figure 6. Graphical depiction of simulation process.

Once the simulated sample has been drawn, statistical estimates for the sample are computed and compared to the known population values. For a fixed set of parameter values, five replicate samples were drawn to provide estimates of the consistency and variability of deviations from the known values. The figure of five replicates was chosen based on preliminary results comparing five, ten and 20 replications. Very little differences were observed in increasing the replication process beyond five whereas a substantial increase in numerical computation time was experienced.

The statistics used to compare the estimates are:

Θ_{EB} - estimated EBEST odds ratio for treatment effect
Θ_F - estimated frequentist odds ratio for treatment effect
D_{EB} - difference in percent change from the true and the EB estimate of treatment effect
D_F - difference in percent change from the true and frequentist estimate of treatment effect

Other statistical estimates of differences between the procedures such as squared error loss, etc. were compared and described in volume III.

Selection of the gamma parameters was based on the following considerations. gamma distribution parameters were assumed to be $\mu = 0.5$ and $\epsilon = 1000$, 200, and 50, respectively. These correspond to scenarios where the true accident rate of all sites is the same, i.e. 0.5 accidents per unit exposure, but the exposures differ from a high exposure, 1000 to a low exposure, 50. This would result in high, medium and low accident counts, as might be expected for the situations of high, medium, and low volume roadways. This would be equivalent to considering the range of accident counts for highway sections to differ from 10 accidents per day to 2 accidents per day to .5 accidents per day yet holding their initial relative safety (as reflected by the rate) constant. Similarly, other exposure units are accommodated in these same ratios, i.e. 500 accidents per year, 100 accidents per year,and 25 accidents per year or 100 MVM, 20 MVM, and 5 MVM. In this way, the sensitivity of the EBEST and frequentist methods for estimating the change in safety can be compared independent of differences in initial safety thresholds. Any insensitivity in the methods in estimating this change due to treatment, then, is attributable to the differences in sample magnitudes, i.e. in accidents counts, for sites which had the same relative "safety" (rates) before treatment. Changing the level of safety threshold, i.e., fixing the rates at other values, was examined, namely .2 and 3.5, but this had no effect on the sensitivity of conclusions drawn about these selected ratios. Figure 4 showed the distribution of the three assumed gammas and how the assumed variability decreased as exposure increased.

Treatment sample sizes were varied but the sample of reference plus treatment sites was fixed at 100. The proportion of treated sites in the sample varied from 10, 20, and 50-percent. This resulted in a range of degree of r-t-m bias from high to low since the 10, 20, and 50 sites with highest accident counts were selected for treatment, respectively. The degree of r-t-m bias is greatest when the 10 highest are selected from a group of 100 than when the 50 highest are selected. This is equivalent to sampling from the extreme tail of the distribution when only 10 sites are treated and to selecting the largest 50 (half) when 100 sites are selected.

Three levels of treatment effect were selected - 50, 20, and 10-percent reductions due to treatment. The three treatment levels are reported on one table for each of the gamma parameter values corresponding to high, moderate and low exposure scenarios. These are reported in tables 5, 6 and 7. Figures 7, 8 and 9 show the percentage difference for each case. The results will be discussed within exposure levels, i.e., by table. *(Note: A treatment effect of 90-percent reduction replaced the 10-percent reduction for the low exposure case since, for low exposures, even a 50-percent reduction was not detectable. Thus, an extreme treatment effect was used to see if there was <u>any</u> treatment level that could be detected with this smaller sample size.)*

High Exposure (table 5)

EBEST estimates are closer to the true values than the frequentist estimates when the treatment results in a 50-percent reduction (Θ is .5) and there is r-t-m bias. Furthermore, this improvement is uniform regardless of the degree of r-t-m bias for the EBEST estimates with percent deviation from the true being 4, 7, and 5-percent for n_t's (number of sites) of 10, 20 and 50, respectively. The frequentist estimate was worse as r-t-m bias increased, 14, 15 and 9-percent. The difference between the EBEST and frequentist methods precision was smallest when there was least r-t-m bias, namely when 50-percent of the sites were treated. When the samples were drawn randomly, i.e., no r-t-m, there was almost no difference in the frequentist and EBEST methods regardless of sample size (percent treated).

In estimating a smaller treatment effect of 20-percent reduction, the EBEST estimates were considerably closer to the true effect with differences of 8,14, and 16-percent compared to the frequentist estimated differences of 25, 27, and 18-percent for the 10, 20, and 50 treated site samples, respectively. The no r-t-m case reflected little differences in the two estimates except for the frequentist estimate for 20 treatment sites which was greatly inflated, probably due to a sampling anomaly.

For the even smaller treatment effect of 10-percent reduction, neither method is very close to the true although the EBEST procedure is closer with differences of 12, 20, and 19-percent compared to the frequentist differences which all exceeded 20-percent (i.e., a 30-percent reduction would be estimated as opposed to the true 10-percent reduction using the frequentist method). The no r-t-m situation shows an underestimation of treatment effect for both methods and estimates of an increase due to treatment by as much as 131-percent! What this means is that with high exposure and high accident locations, these sample sizes are insufficient for detecting such a small treatment effect and a purely random sample does not alleviate this problem. The EBEST method actually improves the estimate even in this no r-t-m situation because it is based on information from the larger number of 100 sites total. Thus, even in no r-t-m situations, if the accident exposure is high and the number of treatment sites is low relative to that exposure and the treatment effect is small, (lower right hand portion of table 5) the EBEST method will provide a better estimate than the frequentist (3-percent vs 39-percent deviation from the true). Note that in figure 7, the EBEST estimate is uniformly better than the frequentist.

Moderate Exposure (table 6)

EBEST estimates are closer to the true values than the frequentist estimates with a 50-percent treatment effect and there is high r-t-m bias (n_t is small), however, neither estimates are very good. That is, when 10-percent of the sites are treated, both EBEST and frequentist estimates are inflated, 13-percent and 20-percent greater reductions than is true, respectively. When 20-percent of the sites are treated, both overestimate at 14-percent and 17-percent difference, respectively. The two methods are essentially equal (11-percent and 10-percent over) when 50-percent of the sites are treated. Both methods were much more effective in estimating when there was no r-t-m bias and were not much different, although both methods uniformly underestimated the treatment effect rather than overestimated it.

For 20-percent reductions due to treatment, neither EBEST nor the frequentist method provide very good estimates, overestimating the effect of the treatment to be around a 50-percent reduction. The EBEST method is superior to the frequentist, however, in detecting this smaller treatment effect when there is no r-t-m, though both methods tend to underestimate the treatment effect and even estimate it in the reverse direction.

For 10-percent reductions, the same conclusion is reached, i.e., neither method is good and in fact, the situation is much worse with both methods claiming an inflated treatment effect of about 40-percent. The increase in the number of treatment sites from 10 to 50 does little to improve on this. See figure 8 for a summation.

Low Exposures (table 7)

As one might imagine, the situation worsens as accident counts decrease. Even for detecting a reduction of as much as 50-percent with 50 treatment sites, both the EBEST method and the frequentist overstate the case claiming 81-percent and 65-percent reductions, respectively. This indicates to the would-be researcher that safety treatments are nearly impossible to estimate accurately when accident counts and sample size are small, a finding that should not be too surprising. The main point here is that neither random sampling nor EB methods can alleviate this problem and there appears to be no statistical solution to this dilemma. The only success was obtained at the 90-percent level of reduction due to treatment. However, both methods were equally effective at detecting this.

Intersection safety treatments nearly all fall in this low exposure category and hence, treatment evaluation for intersections is an extremely unprecise task. Conclusions should be drawn guardedly, if at all. Unless an extremely significant reduction is anticipated or a large number of sites are treated, treatment effects will, in general, be undetectable.

In summation, the above simulation results indicate that:

- EBEST is a superior method for evaluating safety treatments when exposure (and subsequently accident counts) are high even when no r-t-m bias is present. This is true even for detecting small treatment effects. Examination of table 5 reflects this as the EB estimates are uniformly closer to the true values with the exception of the no r-t-m with 50 sites at 50-percent treatment effectiveness.

- EBEST is best in moderate exposure situations even in no r-t-m situations for evaluating large treatment effects of 50-percent or more but neither method is effective for estimating treatment effects of 20-percent or less. Examination of table 6 reflects this.

- Neither statistical method is effective if exposure (and subsequently accident counts) and sample size are small for detecting a treatment effect of 50-percent or less. However, if the treatment is highly effective (90-percent reductions), both methods are equally effective at estimating it. This is evidenced in table 7 where both the EB and frequentist estimates are very far from the true values except for the extremely effective (90-percent) treatments.

Therefore, the critical issue here appears to still be the one of exposure or sample size and magnitude of the treatment effect. No statistical procedure will be effective at identifying small treatment effects at low exposure locations. The surprising thing that emerged from this simulation, however, is that EBEST is not only best in severe r-t-m situations, but it is also superior when there is no r-t-m and the exposure is high, especially with small sample sizes. This fact means that the EB method is recommended even if there is no r-t-m problem.

Table 5.

Simulation study results;
high exposures.

Known treatment effectiveness: -50%

	With r-t-m				no r-t-m			
n_t	EBEST	FREQ	D_{EB}	D_F	EBEST	FREQ	D_{EB}	DF
10	-54%	-64%	4%	14%	-55%	-57%	5%	7%
20	-57%	-65%	7%	15%	-53%	-55%	3%	5%
50	-55%	-59%	5%	9%	-46%	-50%	-4%	0%

Known treatment effectiveness: -20%

	With r-t-m				no r-t-m			
n_t	EBEST	FREQ	D_{EB}	D_F	EBEST	FREQ	D_{EB}	DF
10	-28%	-45%	8%	25%	-21%	-22%	1%	2%
20	-34%	-47%	14%	27%	-6%	57%	-14%	-77%
50	-36%	-38%	16%	18%	-27%	-7%	7%	-13%

Known treatment effectiveness: -10%

	With r-t-m				no r-t-m			
n_t	EBEST	FREQ	D_{EB}	D_F	EBEST	FREQ	D_{EB}	DF
10	-22%	-40%	12%	30%	3%	39%	-13%	-49%
20	-30%	-43%	20%	33%	18%	121%	-28%	-131%
50	-29%	-31%	19%	21%	-2%	28%	-8%	-38%

Table 6.

**Simulation study results;
moderate exposures.**

Known treatment effectiveness: -50%

	With r-t-m				no r-t-m			
n_t	EBEST	FREQ	D_{EB}	D_F	EBEST	FREQ	D_{EB}	DF
10	-63%	-70%	13%	20%	-46%	-44%	-4%	-6%
20	-64%	-67%	14%	17%	-46%	-47%	-4%	-3%
50	-61%	-60%	11%	10%	-49%	-41%	-1%	-9%

Known treatment effectiveness: -20%

	With r-t-m				no r-t-m			
n_t	EBEST	FREQ	D_{EB}	D_F	EBEST	FREQ	D_{EB}	DF
10	-48%	-56%	28%	36%	-13%	22%	-7%	-42%
20	-39%	-46%	19%	26%	-17%	-1%	-3%	-19%
50	-45%	-39%	25%	19%	17%	87%	-37%	-107%

Known treatment effectiveness: -10%

	With r-t-m				no r-t-m			
n_t	EBEST	FREQ	D_{EB}	D_F	EBEST	FREQ	D_{EB}	DF
10	-39%	-49%	29%	39%	10%	65%	-20%	-75%
20	-33%	-39%	23%	29%	13%	59%	-23%	-69%
50	-36%	-31%	26%	21%	52%	142%	-62%	-152%

Table 7.

Simulation study results;
low exposures.

Known Treatment effectiveness: -50%

	With r-t-m				no r-t-m			
n_t	EBEST	FREQ	D_{EB}	D_F	EBEST	FREQ	D_{EB}	DF
10	-72%	-75%	22%	25%	-32%	7%	-18%	-57%
20	-82%	-72%	32%	22%	-51%	-48%	1%	-2%
50	-81%	-65%	31%	15%	-41%	-14%	-9%	-36%

Known Treatment effectiveness: -20%

	With r-t-m				no r-t-m			
n_t	EBEST	FREQ	D_{EB}	D_F	EBEST	FREQ	D_{EB}	DF
10	-58%	-61%	38%	41%	23%	93%	-43%	-113%
20	-56%	-56%	36%	36%	-13%	6%	-7%	-26%
50	-52%	-41%	32%	21%	-38%	-31%	18%	11%

Known Treatment effectiveness: -90%

	With r-t-m				no r-t-m			
n_t	EBEST	FREQ	D_{EB}	D_F	EBEST	FREQ	D_{EB}	DF
10	-95%	-96%	5%	6%	-79%	-61%	-11%	-29%
20	-94%	-94%	4%	4%	-91%	-92%	1%	2%
50	-95%	-94%	5%	4%	-92%	-92%	2%	2%

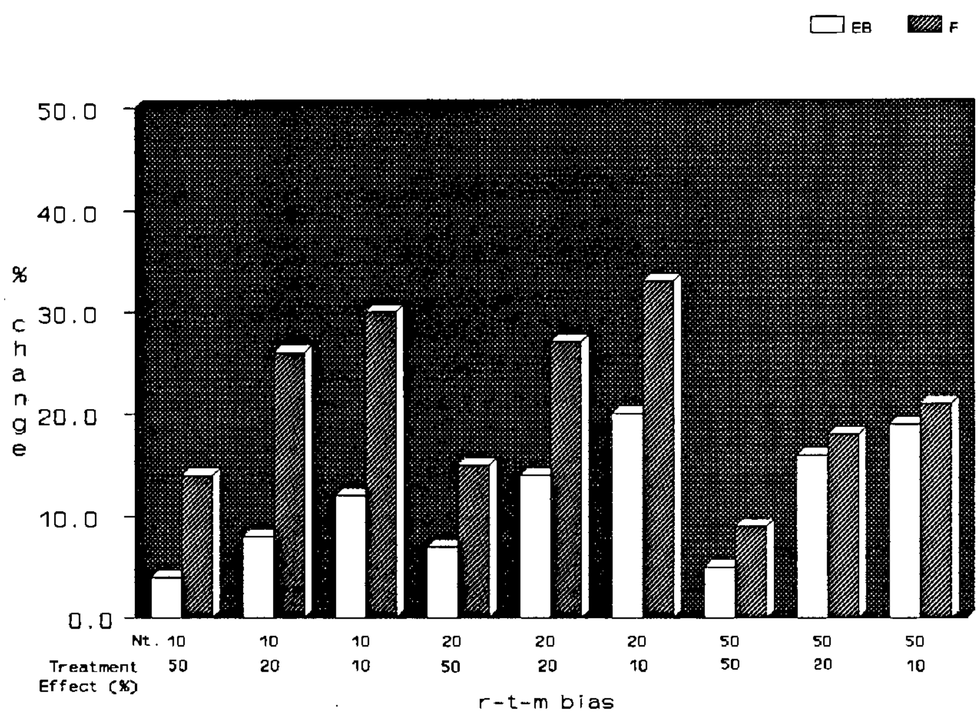


Figure 7. Differences in treatment effect estimates for EBEST and frequentist methods; high exposure.

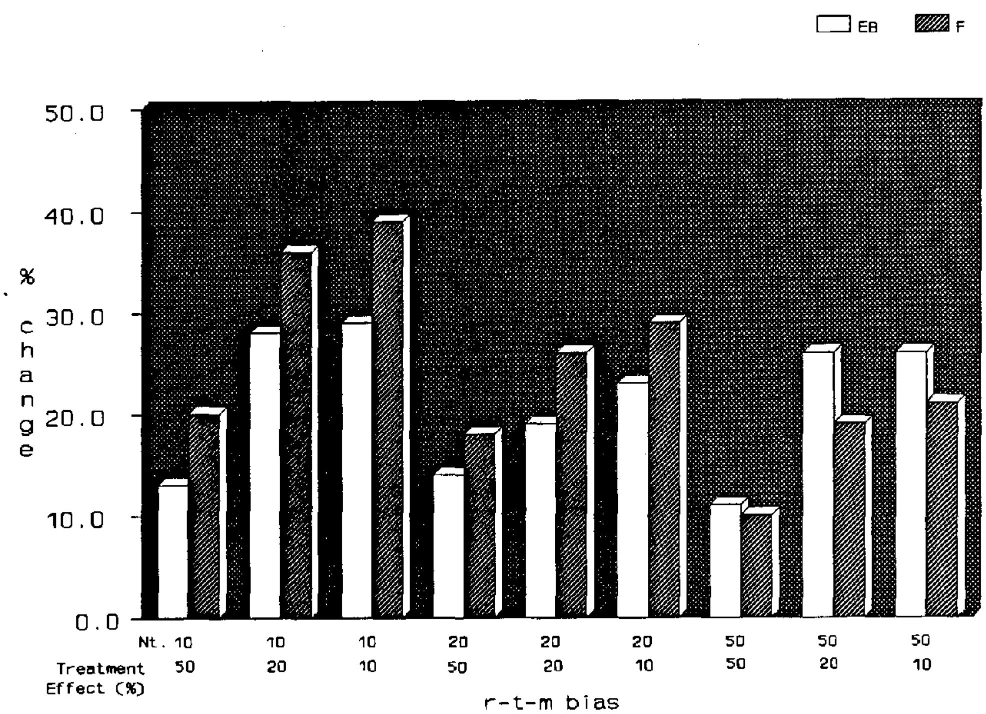
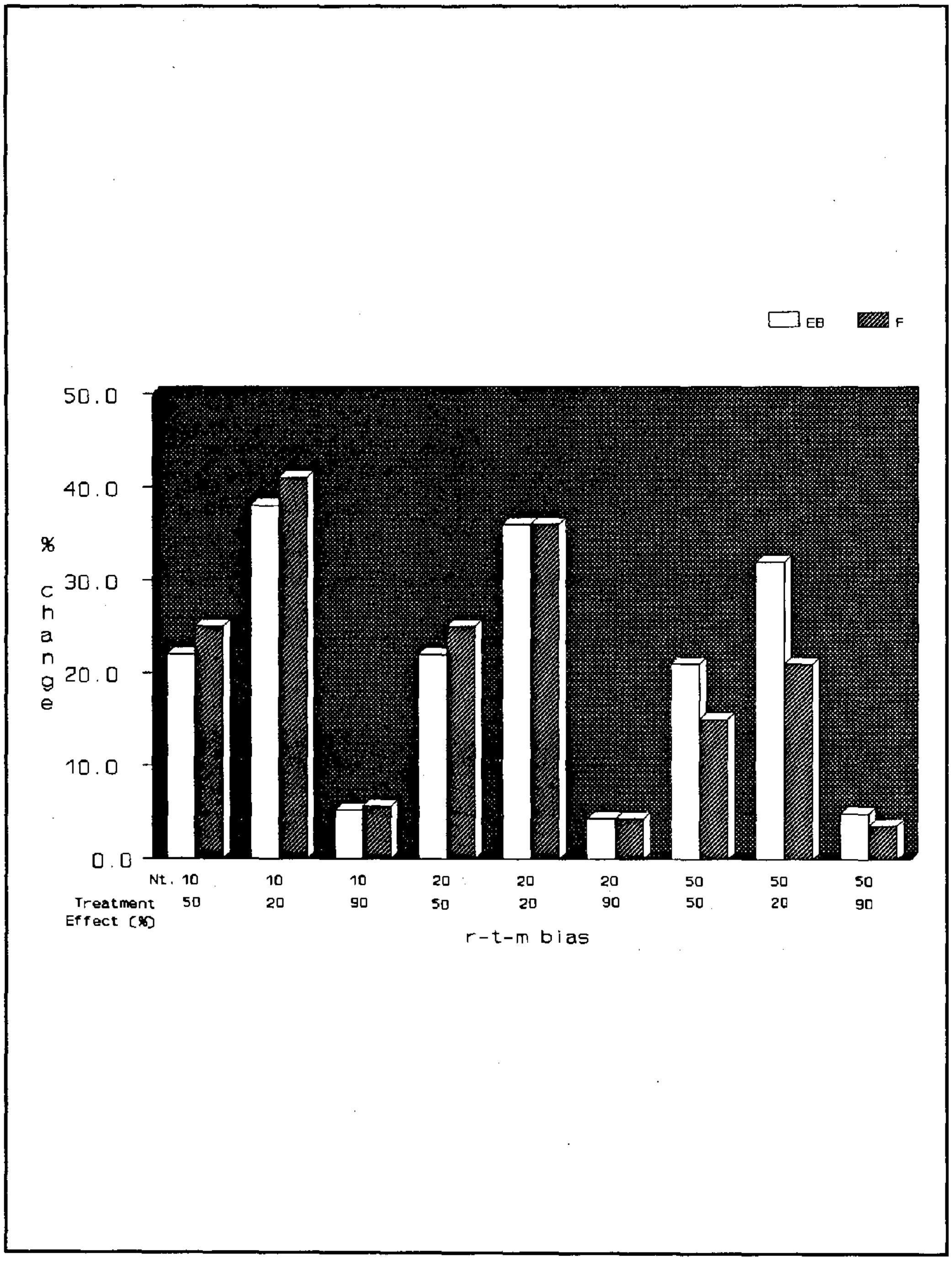


Figure 8. Differences in treatment effect estimates for EBEST and frequentist methods; moderate exposure.



Note: Treatment effect of 90-percent reduction rather than 10-percent was used for low exposures.

Figure 9. Differences in treatment effect estimates for EBEST and frequentist methods; low exposure.

3.1.4 *Data Needs*

The data requirements for evaluating a safety treatment using the EBEST method will vary with the validity of certain assumptions. The more you are able or willing to assume about the sites, the less data will be required of the method. The following is a listing of the critical data elements and assumptions which are required if these data are not collected.

Treatment group data

<u>*accident counts*</u> - before and after accident counts for all treatment sites. The durations need not be equal for all sites of after data, however, if they differ, the durations must be known, and incorporated as an exposure variable. The duration should be long enough to allow for reasonably large counts of accidents yet not so long that other non-treatment related factors might influence the analysis. For example, to evaluate a safety law, two years of before data is generally desireable and at least one year of after data. Three years of after data might be excessive and increase the chances of other confounding factors like changes in posted speeds or construction to obscure the conclusions.

<u>*exposure measures*</u> - any measure of a site's exposure to accident potential should be available for the before period, at a minimum. These exposure variables might include traffic volumes, section length, number of lanes, duration of time, etc. **If no exposure data is used, it must be justifiable and reasonable to assume all treatment sites have identical exposures.** Whereas this may be reasonable for some specific studies, it is a very unique condition and will generally not be acceptable. Omission of this measure simply because of convenience is never justified. Omission of this measure of exposure often results in violation of the very critical assumption of exchangeability as explained in 2.1, and could negate or invalidate study conclusions based on this method. Measures of exposure for the after period are desireable but may be omitted if it is justifiable to assume the after exposures would not change substantially from the before to after periods. In this case, the before exposures are entered with the after data counts into the computer program.

Reference Group data

<u>*accident counts*</u> - before accident counts for all reference group sites. The size of the reference group should be much larger (say five times larger) than the size of the treatment group so that the reference and treatment group together can be considered to be a representative sample of the population of potential treatment sites. Preferably, the reference group data should be available at the time of treatment site selection. The duration of the before periods should correspond to those for the treatment sites. After accident counts for the reference group, if available, can be used in lieu of a comparison group to adjust for time changes from before to after.

exposure - before exposures measure for the reference group sites is essential and should be the same measures identified for the treatment group. After exposures can be assumed to be the same when it is reasonable to make this assumption.

Comparison Group

accident counts - before and after accident counts for a comparison group. A comparison group (or condition) should provide a reasonable estimate of the expected change in accidents at the treated sites that is attributable to time alone, independent of treatment. The only justification for omitting comparison group data is if no change over time can be assumed or, if the reference group data is available for the after period and can provide this time adjustment. The size of the comparison group is generally the same as the size of the treatment group.

exposure - exposure measures for the comparison group are not required by this method. The cross product ratio method assumes that any change in before to after exposures for the comparison group is equal to that of the treatment group. That is, if traffic volume is expected to double for the treatment group, it should be expected to double for the comparison group. When the comparison group is appropriately selected, this assumption is automatically justified. That is, an appropriate comparison site for a given treatment site should be similar to that treatment site to begin with so any changes in exposure, should also be similar. This assumption is an inherent part of the cross product ratio methodology and if this is violated, this method is inappropriate.

3.2 Ranking High Hazard Locations

Sites are typically ranked according to some safety measure, either accident counts or rates, using only information about their safety measured at one point in time. The frequentist ranking methods typically rank on accident count, or rate, or both, but do not use any prior distribution assumptions in this process.(1)

Just as the EBEST procedure provided a more realistic estimate of a site's safety for safety measure evaluation by using information on all of the sites, the same concept applies in the ranking procedure. That is, suppose a collection of sites is to be ranked according to some measure of their "safety." Typically, either total accidents or accident rates are used for this purpose. The classical procedure assumes that the observed safety measure is a fixed and true estimate of that site's safety. In reality, we know this is not true. That is, there is some variability about this observed value for a given site. Furthermore, these site estimates vary in their *reliability*. That is, we have more confidence in some site estimates than others. For example, sites with higher exposures are more likely to experience accidents than those with low exposure. So if we observe x number of accidents at a high exposure site we have more *confidence* that x is close to its true value than y accidents observed for a site with low exposure. The EBEST procedure takes these concepts into account and provides a site estimate that incorporates the concept of the variability about the observed estimate and weights the estimates according to their exposure.

The EB methods described in 2.3 provide expected site rates using the estimated shrinkage factor (weights) for that site. These shrinkage factors, themselves, can provide valuable information on a site's ranking since it reflects the amount of "faith" we can put in the number of accidents observed at that site for this time period. Thus, even though a site experiences an unusually large number of accidents, if the shrinkage factor is large, one might assume that this is an atypical occurrence and perhaps that site should not be considered as "hazardous" as it appears. Rankings based on EBEST estimates for this site will tend to give such a site a lower rank than the frequentist methods.

Higle and Witkowski first proposed the use of the EB estimated procedure for ranking locations.[12] The EBEST procedure differs from their method primarily by being a maximum likelihood procedure rather than method of moments. A comparison of the relative behavior of the EBEST method and Higle and Witkowski's method can be found in volume III.

It is extremely critical that the exchangeability assumption be carefully scrutinized in this application of EB. Satisfaction of this assumption may be more difficult in the ranking application than in safety treatment evaluation. The reason for this is that ranking tasks tend to involve large numbers of sites which may not be very homogeneous with respect to exposure factors, etc. The EB assumption of exchangeability will be questionable in these cases. Also extreme differences in traffic volumes will produce small values of shrinkage coefficients, B_i's, and thus yield similar estimates to the non-EB methods.

In a ranking procedure, the entire group (or population) of interest is generally available as it is this "group" that is being ranked. Thus, the reference group is always available by the very nature of the task. Sampling bias, then, is not generally of concern. The EBEST method is used not so much as a sampling bias adjustor, but rather, as a statistical tool which uses prior information which will result in giving greater weight to sites with greater exposure. Traditional non-EB methods would give equal weight to accident information from each site. Examples where this ranking can produce substantially different results can be found in Efron and Morris, a *Scientific American* article.[6]

The EBEST ranking procedure is straightforward. The EBEST estimate for each site's true accident rate is computed using the data on all sites in the ranking pool (accident data and exposure data). Then the EBEST estimate is ranked, as opposed to the observed data. Either the estimated rate or estimated count may be ranked. The issue of which to rank, rate or count, will depend on the application.

3.3 Combining Information from Multiple Studies (MACEST)

Meta-analysis is the science of combining information from multiple sources in a quantitative fashion. It is used when the information to be combined represents some summary of the raw data in multiple studies, as is often the case in published research studies. Decisions are often made using a subjective type of meta-analysis by administrators and policy makers. More recently attempts have been made to develop objective, quantitative measures on which to base these global decisions.[11]

In the area of highway safety, attempts at doing meta-analyses of a sort have, in the past, appeared in the form of synthesis reports on specific safety treatments.[7] However, these reports often fail to provide any statistically justifiable estimate for the effect of the treatment over multiple studies.

In this study an empirical Bayes method for a Meta-Analysis Combining Estimates of Safety Treatments (MACEST) is developed. Since, from a practical standpoint, only classical safety treatment effects are available in the literature to date, MACEST uses the classical cross-product ratio estimate for before/after studies with a comparison group. It is recognized that these are not the best estimates (EBEST estimates!) and that they suffer from r-t-m potential. Yet, these are the only estimates of safety currently available for a meta-analysis application. Future research might consider the development of a MACEST procedure for EBEST measures of safety treatments. Mathematically, this estimation procedure differs slightly from the EBEST procedure. Details of this were developed in earlier research and will be given in volume III.[13]

Although the individual safety measure effects are non-Bayesian, the meta-analysis procedure uses an EB method to combine these estimates. Based on the assumption that the log of the cross product ratio has a standard normal distribution and variance equal to the sum of the reciprocals of the number of accidents in each group (before-after-treatment-comparison), the EB procedure computes an estimate of the combined cross product ratios weighting by the variance of the cross product ratios. The assumed prior distribution is normal and maximum likelihood estimates are computed on the marginal distribution as usual. This EB procedure is standard and has been used for numerous applications.[15]

Any statistical tool, no matter how effective, will not receive general usage unless it is easily implemented. This is one of the reasons that less effective methods, like T-tests, analysis of variance, and regression are the most popular methods for analyzing accident data even though these methods require assumptions that are not justifiable. Many computer programs exist for these techniques. Even the cross product ratio, though easily computed by a hand calculator, is not used as often as it should be or <u>would</u> be if computer programs were available to compute it.

For this reason, a user-friendly computer program for implementing the EBEST methodology, was developed. The computer program is called BEATS -Bayesian Estimation of Accidents in Transportation Safety. This section will briefly describe BEATS. Complete documentation and instructions are in volume II.

In addition, examples are presented in this section, from both simulated and actual data sets, for each of three safety applications.

4.1 The Computer Program - BEATS

BEATS is a pre-compiled menu driven program for the personal computer. The only requirements are a DOS-driven IBM-compatible PC. Memory requirements depend only on the size of the data files as BEATS can be executed on a computer without a hard disk drive, e.g., on a laptop portable. One need only insert a floppy diskette containing the BEATS program and enter the title "BEATS" to begin execution.

The menu screens direct the user from this point with complete instructions. Options are available for the three specific analyses:

- Safety Treatment Evaluation.
- Ranking Location.
- Meta-Analysis.

For each of these methods, the user may request more information in the way of tutorials or the user may opt to bypass the tutorials and go directly to data analysis. Since the tutorials contain information vital to the appropriateness of the EBEST method, i.e., assumptions and data requirements, first time users are encouraged to go through them. The by-pass option is designed primarily for frequent users. An example of the menu screens and tutorial information is presented in volume II.

The user must refer to an existing ASCII file which contains the data in the required format. The format is a free-format , i.e., data elements need not follow an exact format, but all data elements must be numeric and variables separated by at least one space. Missing data, such as no exposure for a site if the exposure option was selected or no accident information for one site, is not accepted by this program. These sites must be edited from the data file prior to program execution.

Data is required for all variables which might cause sites to vary in their exposure or accident risk potential. Since sites rarely have the same traffic volumes, it is highly recommended that data always be available on traffic volumes and used in this procedure. Rarely will omission of traffic volume be justified, however, this program allows the user the option of not entering an exposure variable. Emphatic warnings are expressed if this option is selected.

The availability of comparison group data is also optional, if reference data is available during the after period to adjust for time change. If both are available, the program will estimate the treatment effect both ways.

In addition to the problems that could occur when EBEST assumptions are violated or certain data elements are not available, there may be times when the BEATS program cannot find a numerical solution. This happens when the data for whatever reasons, seem to represent an unusual likelihood function where a unique maximum cannot be found. When this occurs, the algorithm will not converge to a solution. A limitation of 10 iterations is built into the program to prevent it from executing continuously. After 10 iterations, if convergence is not obtained, a message to this effect will be printed.

Unfortunately, there is no simple solution to this problem. There is some incompatibility between the data and the assumed models, and only with the assistance of a statistician can it sometimes be resolved. Experience to date indicates that this phenomenon is rare and not likely to occur. It did occur occasionally during the simulation study. These instances seemed to occur when the number of sites was small and the variability in exposures was large or when exposures were large relative to the number of accidents.

Failure to converge occurred with one of the data sets examined in this study. The task was to rank Texas cities, similar to the county ranking example which is presented in 4.3. However, the only exposure measure available was city population - city traffic volumes was not available. City population, alone, is not a good surrogate for measuring a city's potential safety as cities with low populations near large cities were less safe than cities with the similar low populations in remote rural areas. For example, a city of only 5,000 may be located near a city of 100,000 (e.g. suburbs) and be less safe than a city of 5,000 in a remote rural area. Using the population of 5,000 as a measure of exposure (accident risk) is obviously not appropriate. Traffic volume, however, would be. In this case, the BEATS program did not converge as the likelihood was representing more than one population and a unique maximum could not be found. The exchangeability assumption had been violated and the assumed model did not represent the data. Without valid exposure data, the EBEST method had to be abandoned for this task.

The user should be alerted to the possibility of this problem. When it occurs, the data should be scrutinized for violations of assumptions of the model. If, as in the city ranking example, the problem cannot be corrected, the EBEST method cannot be used. This can occur with any statistical method when the data is not compatible with the assumed model. Whereas most other statistical methods will still numerically provide a solution, though incorrect, without warning, the EBEST method will simply fail and no solution will result. Perhaps this is preferable to providing a solution but a wrong one.

4.2 Safety Measure Evaluation Examples

Two examples are presented in this section, one from simulated data and one from a safety treatment study. The purpose of this section is to illustrate the execution and interpretation of the EBEST method using the BEATS computer program.

A simulated example is presented here which illustrates the extreme differences that can occur between the EBEST and the frequentist procedure. For the sake of interpretation, a hypothetical scenario has been created to represent this simulated data.

Suppose 20 previously non-signalized intersections were selected to receive signalization. They were selected from a group of 100 similar intersections which were candidates for treatment. The selection was made based on their accident counts during the preceding year. The 20 intersections with the highest number of accidents were selected. The intersections had AADT's of around 15,000 with the maximum for a given site being 25,700 and the minimum, 10,200. *The mean accident rate of 0.2 was assumed with a variance of .0002 for the gamma distribution. Also, a 50-percent reduction due to treatment was assumed. The 100 sampled observations were ordered by count, and the 20 highest selected as the treatment group. The remaining 80 formed the reference group. Another sample of 20 was drawn to represent the comparison group before and still another group of 20 from the same population was drawn to represent the comparison group after. The treatment group <u>after</u> was drawn from a distribution assuming a 50-percent reduction in the population means, and the reference group 80 sites <u>after</u> drawn from the same population as the before, again assuming no change due to time.*

The descriptive statistics are given in table 8. The 20 treatment sites had 111 accidents before and only 35 accidents after signalization. The total AADT for the 20 sites was 280,500 (exposure was entered per 1000 AADT), and exposures ranged from 24,600 to 10,200. The average number of accidents per 1000 AADT was 0.40 for the selected treated sites (i.e., the 20 sites with highest accident counts).

Using the classical estimate of treatment effectiveness with the comparison group, a 64-percent reduction in accidents was estimated which is statistically significant at the 5-percent level (z-value of -4.09 which is less than -1.96). By contrast, the EBEST estimate yields a 44-percent reduction which is also statistically significant with a z-value of -3.38. Recall that this data was simulated assuming a true reduction effect of 50-percent. The classical method overestimates this value by 14-percent whereas the EBEST estimate underestimates it by 6-percent. The EBEST estimate is clearly closer to the true value. This is because while the frequentist method uses the before accident count of 111 as the expected count after, the EBEST procedure estimates the expected after count to be 65, adjusting for the r-t-m bias induced by selecting the intersections with the 20 highest accident counts.

Table 8.
Hypothetical safety treatment study descriptive statistics.

Group	n	AADT/ 1000	Before # acc.	Before rate	After # acc.	After rate
Treatment	20	280.5	111	0.40	35	.12
Comparison	20	281.0	55	0.20	53	.19
Reference	80	1034.7	181	0.17	225	.22

4.3 Ranking Texas Counties

An example was run for this portion of the study using Texas accident data. Data for this example consisted of accident histories for 254 Texas counties from 1982 through 87. All hazardous moving violation accidents and the total county-wide average daily traffic was input. The years 1982 through 1986 were used to determine the EBEST estimates. The county rankings were then compared based on their expected ranking for accident count and rate and their 1987 observed ranking.

The user has the option to rank by accident count, rate, or both. Both were done in this example. Table 9 lists the top 25 counties by their 1987 EBEST accidents (EBEST ACC. COUNT) and table 10 lists the top 25 counties by their 1987 EBEST accident rate (EBEST RATE). Table 11, ranking on EBEST accident counts, provides potential accident reductions for each county using the EBEST estimates for various possible percent reductions, 10, 20, 30, 40, 50-percent, and no reduction. One of the ways in which this output might be helpful would be the following. Suppose a particular type of safety treatment could be expected to result in a 10-percent reduction in accidents. If this treatment were to be applied in the 3 counties with the highest accident counts in table 11, a reduction of 2293 accidents (1004 + 776 + 513) could be expected. Table 12 gives these same expected numbers of accident reductions but for the 25 counties with the highest EBEST rates.

In table 13, the counties that differed most in their EBEST rank ratings are listed. While this example is not dramatic, it does show that there can be considerable discrepancy between the two rankings methods. The counties are listed by county numbers, 47, 25, and 252. These counties had exactly the same numbers of accidents, namely 34. Thus their county rankings were equivalent using the frequentist method, 165, and the EBEST count ranks were also similar, (161, 163 and 175th, respectively.) The difference in rankings occurred when accident rates were ranked. Whereas the frequentist method ranked county 25 as 28th, the EBEST method ranked it as 18th (table 10). If the decision on which counties were to receive supplemental funding was based on such ranking and if only the 20 counties with the highest accident rates were to receive the funding, county 25 would not have received funding using the frequentist method.

To further understand the differences in these methods, the EBEST likelihood distributions are examined in figure 10. These plots show the estimated distribution of the site accident rates. The counties which have similar distributions are counties 47 and 252. These counties not only had similar mean rates but similar and smaller variances than county 25. Thus, they would not have been expected to vary much in the rankings by either procedure. County 25, on the otherhand, not only had a higher mean rate but its rate could be expected to vary considerably more given the wider variability in the likelihood distribution as seen in figure 10. This is the basic difference in the two methods. Whereas the frequentist method assumes that a site's true mean accident rate is a fixed quantity with no variability, the EBEST procedure allows for the variability and takes it into account in estimating that site's expected accident rate.

Figures 11 and 12 further depict how different the EBEST likelihoods can be. Figure 11 shows these likelihoods for the counties ranked high, middle, and low on accident rates and figure 12 depicts this for accident counts.

Table 9.
Counties ranked by 1987 EBEST accident counts.

OBS	RANK	CNTY.	ACC. COUNT	VMT	OBSERVED RATE	EBEST RATE	EBEST COUNT
1	1	101	10038	223908.84	0.044831	0.044824	10036.46
2	2	57	7759	172025.67	0.045104	0.045094	7757.41
3	3	220	5128	100089.15	0.051234	0.051207	5125.22
4	4	15	4623	87800.56	0.052653	0.052619	4619.94
5	5	227	2648	51047.79	0.051873	0.051816	2645.10
6	6	71	1994	31668.85	0.062964	0.062805	1988.96
7	7	178	1140	25760.16	0.044254	0.044199	1138.42
8	8	108	1062	19129.53	0.055516	0.055329	1058.42
9	9	61	948	17547.24	0.054056	0.053868	944.70
10	10	155	932	18730.68	0.049758	0.049626	929.53
11	11	152	914	19551.74	0.046748	0.046651	912.11
12	12	123	893	21954.28	0.040677	0.040642	892.28
13	13	84	884	15250.32	0.057966	0.057701	879.95
14	14	43	866	16605.82	0.052150	0.051974	863.07
15	15	31	786	13659.29	0.057543	0.057253	782.04
16	16	212	747	16547.25	0.45143	0.045048	745.42
17	17	20	717	13786.03	0.052009	0.051799	714.10
18	18	14	685	14717.57	0.046543	0.046417	683.15
19	19	188	642	12653.44	0.050737	0.050528	639.35
20	20	170	601	14800.35	0.040607	0.040559	600.29
21	21	221	583	10570.01	0.055156	0.054826	579.51
22	22	92	545	11044.71	0.049345	0.049130	542.62
23	23	68	501	9756.93	0.051348	0.051066	498.25
24	24	21	481	8295.13	0.057986	0.057503	476.99
25	25	79	472	13286.78	0.035524	0.035544	472.27

Table 10.
Counties ranked by 1987 EBEST accident rates.

OBS	RANK	CNTY.	ACC. COUNT	VMT	OBSERVED RATE	EBEST RATE	EBEST COUNT
1	1	32	69	736.98	0.093626	0.081775	60.27
2	2	71	1994	31668.85	0.062964	0.062805	1988.96
3	3	240	386	6111.43	0.063160	0.062350	381.05
4	4	10	50	738.73	0.067684	0.061267	45.26
5	5	226	414	6733.41	0.061484	0.060794	409.35
6	6	214	107	1689.63	0.063327	0.060599	102.39
7	7	4	73	1185.63	0.061571	0.058094	68.88
8	8	126	424	7246.53	0.058511	0.057946	419.91
9	9	84	884	15250.32	0.057966	0.057701	879.95
10	10	21	481	8295.13	0.057986	0.057503	476.99
11	11	31	786	13659.29	0.057543	0.057253	782.04
12	12	111	126	2185.96	0.057641	0.055946	122.30
13	13	108	1062	19129.53	0.055516	0.055329	1058.42
14	14	221	583	10570.01	0.055156	0.054826	579.51
15	15	61	948	17537.24	0.054056	0.053868	944.70
16	16	37	205	3806.22	0.053859	0.053035	201.86
17	17	3	332	6220.84	0.053369	0.052870	328.89
18	18	25	150	2780.69	0.046543	0.052830	146.90
19	19	15	4623	87800.56	0.050737	0.052619	4619.94
20	20	43	866	16605.82	0.040607	0.051974	863.07
21	21	227	2648	51047.79	0.055156	0.051816	2645.10
22	22	20	717	13786.03	0.049345	0.051799	714.10
23	23	220	5128	100089.15	0.051348	0.051207	5125.22
24	24	68	501	9756.93	0.057986	0.051066	498.25
25	25	107	241	13286.78	0.035524	0.050976	238.26

Table 11.
Counties ranked by EBEST accident counts.

			PERCENT REDUCTIONS					
OBS	RANK	COUNTY	10%	20%	30%	40%	50%	0%
1	1	101	1003.65	2007.29	3010.94	4014.58	5018.23	10036.46
2	2	57	775.74	1551.48	2327.22	3102.96	3878.70	7757.41
3	3	220	512.52	1025.04	1537.57	2050.09	2562.61	5125.22
4	4	15	461.99	923.99	1385.98	1847.98	2309.97	4619.94
5	5	227	264.51	529.02	793.53	1058.04	1322.55	2645.10
6	6	71	198.90	397.79	596.69	795.58	994.48	1988.96
7	7	178	113.86	227.72	341.57	455.43	569.29	1138.58
8	8	108	105.84	211.68	317.52	423.37	529.21	1058.42
9	9	61	94.47	188.94	283.41	377.88	472.35	944.70
10	10	155	92.95	185.91	278.86	371.81	464.76	929.53
11	11	152	91.21	182.42	273.63	364.84	456.05	912.11
12	12	123	89.23	178.46	267.68	356.91	466.13	892.28
13	13	84	88.00	175.99	263.99	351.98	439.98	879.95
14	14	43	86.31	172.61	258.92	345.23	431.53	863.07
15	15	31	78.20	156.41	234.61	312.82	391.02	782.04
16	16	212	74.54	149.08	223.63	298.17	372.71	745.42
17	17	20	71.41	142.82	214.23	285.64	357.05	714.10
18	18	14	68.32	136.63	204.95	273.26	341.58	683.15
19	19	188	63.94	127.87	191.81	255.74	319.68	639.35
20	20	170	60.03	120.06	180.09	240.12	300.15	600.29
21	21	221	57.95	115.90	173.85	231.81	289.76	579.51
22	22	92	54.26	108.52	162.79	217.05	271.31	542.62
23	23	68	49.82	99.65	149.47	199.30	249.12	498.25
24	24	21	47.70	95.40	143.10	190.80	238.50	476.99
25	25	79	47.23	94.45	141.68	188.91	236.13	472.27

Table 12.
Counties ranked by EBEST accident rates.

			PERCENT REDUCTIONS					
OBS	RANK	COUNTY	10%	20%	30%	40%	50%	0%
1	1	32	6.027	12.05	18.08	24.11	30.13	60.27
2	2	71	198.896	397.79	596.69	795.58	994.48	1988.96
3	3	240	38.105	76.21	114.31	152.42	190.52	381.05
4	4	10	4.526	9.05	13.58	18.10	22.63	45.26
5	5	226	40.935	81.87	122.81	163.74	204.68	409.35
6	6	214	10.239	20.48	30.72	40.96	51.19	102.39
7	7	4	6.888	13.78	20.66	27.55	34.44	68.88
8	8	126	41.991	83.98	125.97	167.96	209.95	419.91
9	9	84	87.995	175.99	263.99	351.98	439.98	879.95
10	10	21	47.699	95.40	143.10	190.80	238.50	476.99
11	11	31	78.204	156.41	234.61	312.82	391.02	782.04
12	12	111	12.230	24.46	36.69	48.92	61.15	122.30
13	13	108	105.842	211.68	317.52	423.37	529.21	1058.42
14	14	221	57.951	115.90	173.85	231.81	289.76	579.51
15	15	61	94.470	188.94	283.41	377.88	472.35	944.70
16	16	37	20.186	40.37	60.56	80.75	100.93	201.86
17	17	3	32.889	65.78	98.67	131.56	164.45	328.89
18	18	25	14.690	29.38	44.07	58.76	73.45	146.90
19	19	15	461.994	923.99	1385.98	1847.98	2309.97	4619.94
20	20	43	86.307	172.61	258.92	345.23	431.53	863.07
21	21	227	264.510	529.02	793.53	1058.04	1322.55	2645.10
22	22	20	71.410	142.82	214.23	285.64	357.05	714.10
23	23	220	512.522	1025.04	1537.57	2050.09	2562.61	5125.22
24	24	68	49.825	99.65	149.47	199.30	249.12	498.25
25	25	107	23.826	47.65	71.48	95.31	119.13	238.26

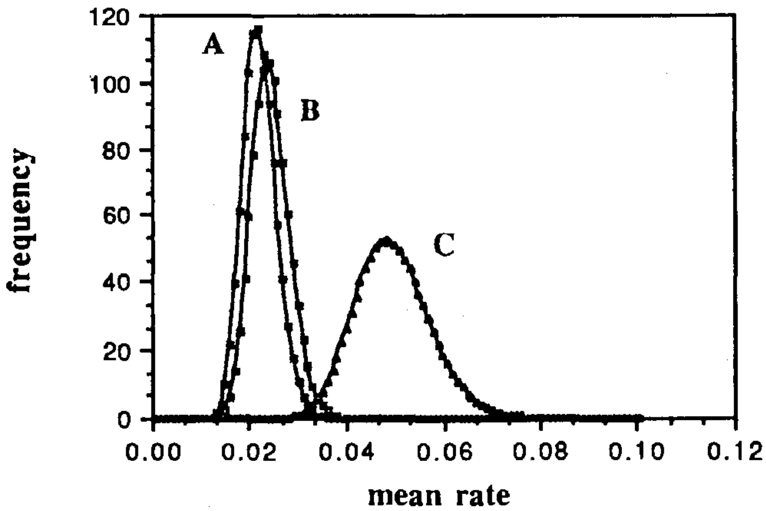


Figure 10. Likelihood distributions for counties where EBEST ranks differed A = #47, B = #252, C = #25.

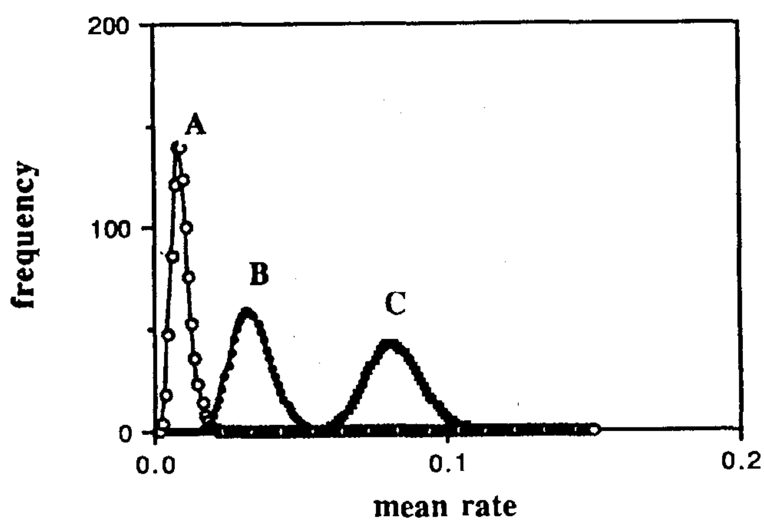


Figure 11. Likelihood distributions for counties with highest, middle and lowest EBEST ranks on rates A = low, B = middle, C = high.

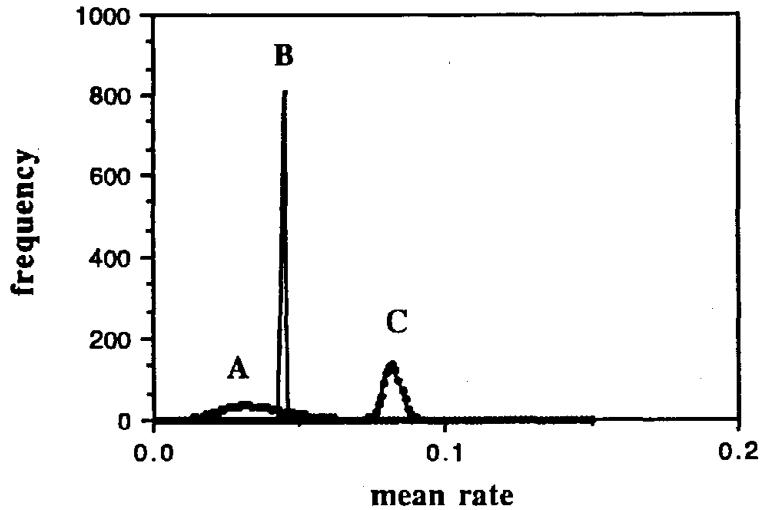


**Figure 12. Likelihood distributions for counties with highest, middle and lowest EBEST ranks on counts
A = low, B = middle, C = high.**

Table 13.
Counties which differed in rankings.

County	frequentist rate rank	EBEST rate rank	frequentist count rank	EBEST count rank
47	212	210	165	161
25	28	18	165	163
252	193	198	165	175

4.4 Combining Safety Information from Multiple Sources

Although an exhaustive search was conducted including *numerous* synthesis reports, no actual data was found which provided all of the ingredients for a proper meta-analysis. This may be indicative of a more serious problem than the development of a statistical method for this task: namely, is data readily available for it's implementation?

"Availability" is the key word here - not existence. The necessary data generally existed and was available at some point in the studies. However, the data never made it into the synthesis reports or even, perhaps, into the original reports on the studies investigated. Hence, the first important step in being able to use this method is to educate researchers in **what** to report from their studies so that the literature will be complete enough for future use in a meta-analysis.

The essential data elements needed for a meta-analysis combining treatment effects from multiple before/after safety treatment studies using the cross-product ratio with a comparison group are estimates of the:

- Treatment effect.
- Variability of the treatment effect including any relevant weighting factors.

The estimate of the variability is the key issue here and generally the limiting factor and missing element in most research reports. In order that meta-analysis be meaningful, a good estimate of the variability of the statistic being combined - in this case, the cross product ratio - is essential. Without this, a simple unweighted average of the treatment effects is probably the best one can do.

The variability of the cross-product ratio is the sum of the reciprocals of the accident counts in each of the Before/After-Comparison/Treatment cells as shown in equation 4. This estimate must be weighted by any other factors which are different among the studies and critical to the magnitude of the accident counts. Examples of such factors are:

- Traffic volumes.
- Number of sites in the study.
- Duration of before/after time periods.
- Section length.

By multiplying the estimated variability by the factors, the amount of information from each study is weighted to account for them, the factors also help to satisfy the exchangeability assumption which is no less critical to MACEST than EBEST. For example, if studies had different numbers of sites, the estimated variance should be multiplied by the number of sites. This product is then entered in the BEATS program. Just as with the exposure variable for the safety treatment program, BEATS will not compute this product for you and assumes the data entered has accounted for all exposure factors.

An example illustrating the use of meta-analysis will be presented using a hypothetical set of data. This data is used for illustration only. Suppose several studies were available on the safety effect of left-turn phasing. (Note: the appropriate analysis would be to use each study's raw data whenever available. Meta-analysis is only proposed as an alternative when the raw data are no longer available and only the summary data exist. Assume treatment/comparison group accident counts were given in the study reports along with the cross-product ratio estimate of the treatment effect. Table 14 lists data for 5 studies. Each study had different numbers of sites on which the treatment effects were based. Table 15 gives the MACEST results for combining information on these studies. The number of sites, n, was used as a weighting factor.

Table 14.
Hypothetical multiple study data on treatment evaluations.

	No.	Treatment		Comparison	
Study	sites	Before	After	Before	After
1	11	31	43	109	128
2	26	458	492	356	528
3	10	66	95	22	29
4	15	179	194	63	44
5	5	32	43	17	7

Table 15.
MACEST for hypothetical data.

Study	Treatment Effect	n	Weighted MACEST	Unweighted MACEST
1	+18%	11	+32%	+22%
2	-28%	26	0%	-27%
3	+9%	10	+42%	+15%
4	+55%	15	+14%	+58%
5	+226%	5	-15%	+220%
Overall	+37.8%		+20%	+63%

The MACEST estimates are different from the uncombined estimates. The weighted MACEST estimate says there was a 20-percent increase in accidents due to treatment while the frequentist overall estimate yields a 37.8-percent increase. Note that the effect of the last study, study 5, which showed a tremendous increase was only based on five sites. This study had a strong effect on the overall frequentist estimate and the unweighted MACEST (i.e., not weighting by the number of sites) which gave a 63-percent increase. This example shows the importance of weighting factors such as the number of study sites.

Chapter 5 - Conclusion

5.1 Summary

The empirical Bayes methodology developed and presented in this study promises to resolve many of the problems encountered in accident analysis for (1) safety treatment evaluation, (2) ranking high hazard locations, and (3) combining estimates from multiple safety treatment studies. The methodology developed for the first two tasks, EBEST, adjusts for r-t-m sampling bias and weights information from each site according to that site's exposure or accident potential. The methodology developed for the third task, MACEST, provides an estimate of the combined effect of a safety treatment for multiple studies when only summary statistics are available, as is often the case in the literature. The MACEST procedure also weights the amount of information each study contributes by the site's accident potential and variability.

Frequentist statistical methods for these accident analysis applications have been ineffective due to problems such as r-t-m and varying exposure or accident potential among sites or studies. The methods developed in this research overcome these problems.

Simulation studies showed that using the EBEST procedure yields significant improvements over the classical methods when:

- There is considerable r-t-m bias for high and moderate exposures.
- The number of sites is low.
- There is a small, marginal treatment effect.
- There is high variability among sites, as reflected by high exposures.

EB methods are superior, but closer to frequentist methods when either:

- There is a large treatment effect.
- Exposures are low.

Neither the EBEST method nor the frequentist method are successful for the combination of low exposures, <u>and</u> a small (10%) treatment effect.

A computer program, BEATS, was developed for applying these methods to all three accident analysis tasks. Tutorials explaining the methodology, assumptions, and data requirements are available for first-time users. BEATS is user-friendly and operates on any DOS-driven PC from a floppy diskette. Memory requirements are minimal and restricted only by the size of the data to be analyzed.

Although the data requirements for using either EBEST or MACEST are reasonable, data for use as examples was extremely difficult to find. This is probably due to a lack of awareness on the part of researchers as to what data elements are important both to collect and to report, rather than severity of the data requirements.

For the EBEST procedure, these data requirements include a reference group and measurement of site exposure. Reference group data is often available at the time of site selection (prospectively), but then discarded and ignored in the evaluation process. Even in cases where the reference group was never available or lost, it can often be retrieved or recreated after the fact (retrospectively) using advanced data retrieval software such as LANSER.[5] In retrospective reference group retrieval, care must be taken in its description and sample size to ensure that the reference and treatment sites collectively, before treatment, do form a good representative sample of the population of potential treatment sites.

Exposure data is another issue. Again, however, the problem is not one of inability to collect the data or lack of existing data on site exposures (e.g., traffic volumes) but rather the negligence linking exposure data to accident data files. Traffic records systems must be modified to include this extremely valuable information not only to facilitate use of the methods developed here but to improve accident analysis study results, in general. The accident analysis researchers are becoming increasingly aware of this and advances are being made in developing more complete traffic record data bases so that, hopefully, exposure data will become an easily attainable data element. The EBEST method will accommodate analyses without exposures. However, it does so by assuming equal exposures (i.e., ignoring the problem) as do all existing methodologies.

The MACEST procedure requires data on safety treatment estimates in before/after studies with a comparison group, the number of accidents for each of these categories, and exposure measures if studies differ on them. Again these data were generally available at one time. However, researchers neglect to include these data in their research reports and publications. In this case, the information is permanently lost to the research public and can never be retrieved. Changing reporting practices seems a small price to pay for the opportunity of potentially contributing to further meta-analyses combining information from multiple studies. MACEST is to be used only when raw data are not available and inferences need to be drawn on summary statistics alone.

The methods developed here, EBEST and MACEST, offer accident analysts a potentially powerful tool for extracting meaningful conclusions for improving public safety. Hopefully, the transportation research audience will embrace them and use them to their fullest potential. But first this audience must be made aware of them and educated about their data requirements so that these concerns can be addressed at the study planning stage.

5.2 A Look to the Future

Although the new methods developed in this study resolve many of the problems facing the accident analyst, other problems still haunt us and extensions to these methods are easily identifiable. One such extension is the incorporation of other variables such as roadway geometrics, weather, time, etc.

Often times sites are known to differ in certain aspects and it would be of interest to consider adjusting for these differences in the treatment evaluation process. For example, treatment sections might include curved and straight roadways. Degree of curvature would be a variable of interest in the safety evaluation of the treatment. Examples of other such variables are number of lanes, intersection type, highway functional class, etc. The current BEATS computer program would require that the data be analyzed in subgroups, divided according to these variables, if these variables might be causing a problem with the exchangeability assumption. That is, if you have two-lane and four-lane sections and you know the accident rate is higher on two-lane sections without even looking at the data, you do not have exchangeability and must analyze two-lane sections separately from four-lane ones. This limitation can be overcome by incorporating these variables in the estimation of the true site means. This would require modeling the true site means as a regression function dependent upon these other variables in the prior gamma distribution. The theoretical work for this task has been done by Morris and Christiansen, though not yet published. The next step would be to modify the BEATS program to perform the analysis.

Another valuable extension to the EBEST methodology and BEATS would be the ability to handle multiple time periods to test for time trends. This would allow testing for such things as comparability of the treatment and comparison group before treatment. The frequentist method for doing these does not adjust for r-t-m bias in the treatment group. The frequentist method for handling multiple before periods is known as the before/after design with comparison group and comparability check.[9] In this procedure, the before data for both the treated and comparison group is statistically tested for comparability. This is a check to see if the comparison group really provides "comparable" information on safety, i.e., similar to the treatment group. If it does not, then the assumption that this comparison group is providing us with good estimates for what would have happened without treatment is invalid and the estimate of treatment effectiveness will not be reliable. This issue is separate from the issue of r-t-m, but, nonetheless, important. Using the classical method when r-t-m bias is present, still risks yielding a poor estimate of treatment effectiveness. Thus, a very useful (and important) extension of the EBEST methodology would be to incorporate multiple time periods and checks for comparability which also adjust for r-t-m bias.

Extensions of the MACEST procedure are numerous. In addition to combining safety treatment estimates from multiple studies, other applications of meta-analysis are the combining of:

- Test statistics (t-values, F-values) from multiple studies to obtain a single test statistic for testing some hypothesis.

- P-values (levels of significance) from multiple studies to obtain an overall significance level.

In general, any statistic can be combined using a MACEST procedure provided that the probability distribution of that test statistic is known and some estimate of the variability of that test statistic is available from the data.

Further extensions of meta-analysis may, however, be premature. Unless the research audience is educated on the important information to report in studies -until publications and reports are required to follow certain guidelines in consistency and completeness - the data for doing a meta-analysis will not be available.

On the other hand, if we begin to develop synthesis - type data bases which contain the essential pieces of data for a specific safety treatment evaluation, for example, this file could be continually updated whenever a new study was completed and MACEST would provide a continuous, updated estimate of the effectiveness of that safety treatment.

For example, consider intersection conversion from left-turn permissive to protected. Many studies have been done and continue to be done, at all levels, to address this safety treatment. Suppose a literature search on this topic revealed a number of studies which provided sufficient information - treatment effect, number of sites, etc. - to use MACEST. A data base could be constructed to create and maintain such data and, as future left turn conversion studies evolved, they could be added to this data base. The result would be a sort of "intelligent system" for evaluating left-turn conversions at intersections.

In addition to extensions in the statistical methodology for accident analysis, future R&D efforts need to be directed toward unifying the advances currently being made in

- Improved accident data systems
  (e.g., HSIS - Highway Safety Information System, FHWA's new highway safety database).

- Sophisticated data retrieval software
  (e.g., LANSER - Local Area Network Safety Evaluation and Reporting System, a Texas Transportation Institute data retrieval program).

- Powerful statistical methodologies
  (e.g., EBEST).

Research efforts in all of these areas have, independently, been striving to solve some aspect of the accident analysis problem. However, these efforts have been achieved independently.

The time has come to merge them - to put the pieces of the puzzle together - so that insight may be gleaned into the true safety picture which accurately and honestly reflects the status of safety on our highways.

When this study was intially undertaken, it was assumed that data on which to apply the methodology would be available. As the methodology unfolded, it was discovered that either the needed data on safety evaluation studies was not preserved (i.e., the reference group information was available at the time of treatment site selection but then discarded) or there was no way to retrieve this information retrospectively in an automated way (i.e., via computer). Thus, simulations and hypothetical examples were the only sources for examining the effectiveness of the EBEST methodology.

Since this study's inception, however, advances have been made both in the existence of multi-state automated accident data bases merged with roadway inventory information, like HSIS, and in data retrieval software, like LANSER. The next step in the research process, then, should be to implement the EBEST methodology on actual safety treatment studies. This could be done by:

1. Contacting the States in the HSIS data base and requesting identification of sites and a before-and-after time period wherein certain safety treatments were implemented. Comparison sites would also need to be identified.

2. Retrieving the accident data and exposure variable from HSIS.

3. Applying the EBEST and frequentist methods to these data to compare results.

It would be desireable to solicit more than one kind of safety treatment so that examples similar to the simulation study parameters would result, i.e., high, medium and low exposures, high, medium, and low treatment effectiveness and high, medium, and low numbers of treated sites.

The result of such a study would have significant impact on the useage of this methodology by the research community. Simulations and hypothetical examples are of academic interest, but in order for a methdology as demanding data-wise and statistically complex as this to be accepted by those who would use it, the practical aspects must be evidenced. Real world examples of highway safety treatments are the most convincing marketing tools for implementation and acceptance of such new methodologies. It is hoped that future studies will be encourged to achieve these objectives.

BEATS	-	a menu driven, precompiled PC DOS computer program for computing the EBEST estimate for safety treatment evaluation and ranking and the MACEST estimate for meta-analysis of safety treatment estimated effects.
frequentist	-	non-Bayesian statistics.
comparison group	-	a group of sites which is independent of the treatment and provides an estimate of the expected change in accidents from before to after independent of the treatment. The comparison group need not be a separate group of sites but may be the same sites under a different condition which are not affected by the treatment (weather, time of day, etc.); the comparison group may be the reference group if post treatment data is available.
cross product ratio	-	the ratio of the odds of one event as compared to another; in this case, the odds of an accident in the treatment group before-to-after compared to the odds in the comparison or reference group.
EBEST	-	the empirical Bayes estimation procedure for estimating the expected accident rate apart from any treatment effect and adjusted for r-t-m sampling bias; empirical Bayes Estimation of Safety in Transportation.
empirical Bayes	-	a statistical methodology which assumes a prior information about the probability distribution on the true population parameters and estimates these parameters using data.
exchangeability	-	the assumption that the true site means are identically and independently distributed as gamma random variables; this assumption is violated if there is any way to tell, in advance of data collection, which sites would have higher true means.
negative exponential distribution	-	the assumed continuous probability distribution for traffic volumes which assumes most volumes are clustered about the mean but a few sites may have extremely high volumes.

exposure - some quantity or group of quantities which reflect a site's safety potential such as traffic volume, section length, or time.

gamma distribution - a continuous probability distribution assumed as the prior distribution for the true site mean accident rates

HSIS - Highway Safety Information System, a new highway safety data base developed by FHWA and HSRC. The HSIS is designed to provide a detailed system linking accident, roadway and traffic data for problem analysis.

hyperparameters - unknown true values of the gamma distribution, i.e. the distribution of the population representing the reference and treatment group collectively, namely, μ_o and ϵ.

LANSER - **L**ocal **A**rea **N**etwork **S**afety **E**valuation and **R**eporting system, a data retrieval program for Texas accident data.

MACEST - **M**eta-**A**nalysis for **C**ombining **E**stimates of **S**afety **T**reatments, the empirical Bayes estimate for combining safety treatment estimates from multiple studies; the estimates are the frequentist cross-product ratios and the assumed prior distribution is the log normal probability distribution.

maximum likelihood - the statistical method used by the EBEST methodology for estimating the hyperparameters using the negative binomial distribution; the computational procedure requires numerical iteration but produces a statistically optimal estimator to that of the method of moments.

meta-analysis - the methodology of combining statistical estimates when only summary data is available, e.g., combining estimates from multiple studies in the literature, when the raw data is unavailable.

method of moments - a statistical method for deriving population parameter estimates which is computationally easier than maximum likelihood estimation for some probability distributions, in this case, the negative binomial, but which does not produce estimates with statistically optimal properties.

negative binomial	-	the resulting discrete probability distribution of a Poisson distributed random variable, accident counts, given the gamma prior distribution hyperparameters.
prior distribution	-	the probability distribution on the true population parameters - in this case, the gamma distribution of the true site mean accident rates, λ_i.
reference group	-	a group of sites similar to the treatment site such that the reference group and treatment group collectively represent the entire population of potential treatment sites.
regression-to-the mean	-	the phenomenon where the observed value will shrink toward the population's true mean during the post treatment period independent of treatment; abbreviated r-t-m.
relative squared error loss function	-	the squared error loss divided by the true value of the estimate, i.e., a measure of the deviation from the true relative to the magnitude of the true; used in the simulation study to measure the relative error of EBEST as compared to frequentist estimates, relative to the magnitude of the known accident rate.
shrinkage coefficient		a weighting factor which adjusts each site's observed rate for r-t-m sampling bias; the closer this value is to 1, the more the weight toward the overall true site mean of all sites; the closer to o, the more the weight toward the observed site accident rate.
treatment group	-	a collection of locations which have received some safety treatment.

Notation

μ_O = mean accident rate for the prior gamma distribution of all true site means, λ_i's.

ϵ = mean for all site exposures for the prior gamma distribution

λ_i = true mean accident rate for each site

y_i = site i's observed accident rate

e_i = site i's exposure

z_i = site i's observed accident count, $z_i = e_i y_i$

$\hat{\lambda}_i$ = EBEST estimate of the accident rate for site i

$\hat{\mu}_o$ = EBEST estimate or the accident rate for all sites

$\hat{\epsilon}$ = EBEST estimate of the mean exposure for all sites

$\hat{B}_i$ = shrinkage factor for site i

Θ = the true safety treatment effect

$(\Theta - 1)$ = % change in accidents expected for the safety treatment

$\hat{\Theta}_F$ = non-Bayesian estimates of Θ; cross product ratio

$$\hat{\Theta}_F = \frac{\sum_{i=1}^{n_c} z_{icB} / \sum_{i=1}^{n_c} z_{icA}}{\sum_{i=1}^{n_T} z_{iTB} / \sum_{i=1}^{n_T} z_{iTA}}$$

$\hat{\Theta}_{EBc}$ = EBEST estimate of Θ using the comparison group

$$\hat{\Theta}_{EBc} = \frac{\sum_{i=1}^{n_c} z_{icB} / \sum_{i=1}^{n_c} z_{icA}}{\sum_{i=1}^{n_T} e_i\hat{\lambda}_i / \sum_{i=1}^{n_T} z_{iTA}}$$

Notation (continued)

$\hat{\Theta}_{EBR}$ = EBEST estimate of Θ using the reference group.

$$\hat{\Theta}_{EBR} = \frac{\sum_{i=1}^{n_R} e_i\hat{\lambda}_i/(\sum_{i=1}^{n_R} z_{iRA})}{\sum_{i=1}^{n_T} e_i\hat{\lambda}_i/(\sum_{i=1}^{n_T} z_{iTA})}$$

n_c = number of sites in comparison group
n_T = number of sites in treatment group
n_R = number of sites in reference group

subscripts c on z_i's refer to comparison group

subscripts T on z_i's refer to treatment group

subscripts B on z_i's refer to Before group

subscripts A on z_i's refer to After group

u = true mean accident rate for all sites = μ_o

r = true mean exposure

1. Acuff, A., O.J. Pendleton, and B.H. DeLucia. 1990. "Save City/Save County Technical Report." Texas State Department of Highways and Public Transportation, Safety and Traffic Operations Section, Texas Transportation Institute, TARE-84.

2. Abbess, C., D. Jarrett, and C.C. Wright. 1981. "Accidents at Blackspots: Estimating the Effectiveness of Remedial Treatment, with Special Reference to the *Regression-to-the-Mean* Effect." *Traffic Engineering and Control*, Vol. 22, No. 10, pp. 535-542.

3. Ben-Akiva, Mosche and Avishai Cedar. 1988. "Estimation of Improved Intersection Safety." Massachusetts Department of Public Works Report, pp. 121.

4. Brinkman, C.P. 1986. "Improved Highway Safety Analysis." *Public Roads*.September, Vol. 51, No. 6, Federal Highway Administration, Washington, D.C.

5. DeLucia, B.H., M.R. Crouse, D. Young, and A. Acuff. 1987. "Local Area Network Safety Evaluation and Reporting System." For the Texas State Department of Highways and Public Transportation.

6. Efron, B., and C. Morris. 1977. "Stein's Paradox in Statistics." *Scientific American*, Vol. 235, no. 5.

7. Rosenbaum, Merton and Charles Pinnell. 1982. "Synthesis of Safety Research Related to Traffic Control and Roadway Elements. Volume I, FHWA-TS-82-232, and Volume II, FHWA-TS-82-233." Federal Highway Administration, Washington, D.C.

8. Fleiss, J.L. 1981. *Statistical Methods for Rates and Proportion*. Wiley & Sons. New York, pp. 321.

9. Griffin, L.I., III. 1982. "Three Procedures for Evaluating Highway Safety Improvement Program." Texas Transportation Institute, TARE - 51, pp. 49.

10. Hauer, E. 1980. "Bias-by-Selection." *Accident Analysis and Prevention*, Vol. 12, pp. 113-117.

11. Hedges, L. V. and Olkin, I. 1985. *Statistical Methods for Meta-Analysis*. Academic Press., pp. 369.

12. Higle, J. and Witkowski, J. 1988. "Bayesian Identification of Hazard Locations." *Transportation Research Record* 1185, pp. 24-36.

13. Leotsarakoas, Christos. 1988. "A Comparative Analysis of Traffic Accidents in Highway Construction Zones." Ph.D. Dissertation, Ohio State University.

14. Mood, A.M., F. A. Graybill, and D. C. Boes. 1963. *Introduction to the Theory of Statistics*. McGraw-Hill. pp. 564.

15. Morris, Carl N. 1983. "Parametric empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association*, Vol. 78 no. 381, pp. 47-55.

16. Morris, C. N., O. J. Pendleton, M. K. Bishop, and C. L. Scaff. 1989 "Empirical Methodology in Traffic Accident Analysis." University of Texas Center for Statistical Science Technical Report #75.

17. Robbins, H. 1956. *An empirical Bayes Approach to Statistics*. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, 157-163, University of California Press.

18. Weed, R.M. 1986. "Revised Decision Criteria for Before-and-After Analyses." *Transportation Research Record* No. 1068, pp. 8 - 17.