Publication No. FHWA- RD-92-049
October 1992

A Systemwide Methodology for Evaluating Highway Safety Studies

U.S. Department of Transportation
Federal Highway Administration

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 a new, better procedure for assessing widespread geographic effects resulting from the use of safety treatments. Most currently used statistical methods for countermeasure evaluation assess the spot or local changes that result from the new treatment. The effects of the treatment on the entire "system" or nearby contiguous links are seldom evaluated. The new approach is ideal for applications where large quantities of data reside in a PC computer format. The research findings identify how much treatment and comparison site data are required to conduct a statistically valid analysis.

Incorporated in the methodology are known procedures for determining if regression-to-the-mean (RTM) problems exist and how those problems can be minimized when they are present. A previous research study developed a computer program (BEATS) which identifies and corrects for regression-to-the-mean bias in highway safety studies. Those interested in receiving the computer program should contact Michael S. Griffith of my staff on (703) 285-2382.

This report will be distributed with one copy to each Region and two copies to each Division Office. One of the Division copies should be sent to the State. 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.

Lawrence Powers

for Lyle Saxton
Director, Office of Safety and Traffic
Operations Research and Development

NOTICE

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof. The contents of this report reflect the views of the contractor who is responsible for the accuracy of the data presented herein. The contents do not necessarily reflect the official policy of the Department of Transportation. This report does not constitute a standard, specification, or regulation.

The United States Government does not endorse productsor manufacturers. Trade or manufacturers' names appear herein only because they are considered essential to the objective of this document.

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2.	3. Recipient's Catalog No.
FHWA-RD-92-049	PB93-127504	

4. Title and Subtitle	5. Report Date
A SYSTEMWIDE METHODOLOGY FOR EVALUATING HIGHWAY SAFETY STUDIES	OCTOBER 1992
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Olga J. Pendleton	

9. Performing Organization Name and Address	10. Work Unit No.
Texas Transportation Institute College Station, TX 77843-3135	2A3B0111
	11. Contract or Grant No. DTFH61-88-C-00022

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Office of Safety and Traffic Operations R & D Information & Behavioral Systems Division Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296	FINAL REPORT JULY 1988 - APRIL 1992
	14. Sponsoring Agency Code

15. Supplementary Notes

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

16. Abstract

This study developed a new, better procedure for assessing widespread geographic effects resulting from the use of safety treatments. Most currently used statistical methods for countermeasure evaluation assess the spot or local changes that result from the new treatment. The effects of the treatment on the entire system or near-by contiguous links are seldom evaluated. The new approach is ideal for applications where large quantities of data reside in a PC computer format. Incorporated in the methodology are known procedures for determining if regression-to-the-mean (RTM) problems exist and how those problems can be minimized when they are present. A previous research study developed a computer program - Bayesian Estimation of Accidents in Transportation Studies (BEATS) - that identifies and corrects regression-to-the-mean bias in highway safety studies.

Step-by-step guidelines were developed on how to plan and use the new procedure to evaluate highway safety studies. The guidelines describe what are adequate sample sizes, how to identify RTM problems, and how computerized data analysis procedures can be used.

17. Key Words	18. Distribution Statement
Systemwide methodology, accident analysis, highway safety studies, before-after studies, regression-to-the-mean, accident migration	No Restriction. This document is available to the public through the National Technical Information 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
Unclassifed	Unclassified	113	

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	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
AREA				
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
VOLUME				
fl oz	fluid ounces	29.57	milliliters	ml
gal	gallons	3.785	liters	l
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3

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

MASS				
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
TEMPERATURE (exact)				
°F	Fahrenheit temperature	5(F-32)/9 or (F-32)/1.8	Celcius temperature	°C
ILLUMINATION				
fc	foot-candles	10.76	lux	l
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
FORCE and PRESSURE or STRESS				
lbf	poundforce	4.45	newtons	N
psi	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

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

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

(Revised August 1992)

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

CHAPTER 1. INTRODUCTION

WHAT IS ACCIDENT MIGRATION?

The evaluation of highway safety countermeasures is probably one of the most difficult and challenging aspects of accident analysis. It is a task that has been performed for decades and yet, today, there is still no general consensus on what is the optimal methodology for evaluating the safety effectiveness of a treatment. Progress is rapidly being made in this area. However, specific problems such as the accident migration phenomenon have only begun to be explored.

The concept of "accident migration" is often first met with raised brows and skepticism. The belief that a predetermined number of accidents are "just waiting to happen" in a system so that reducing accidents in one area will result in an increase in another is counterintuitive and unacceptable to the average researcher in this field. Such statements conjur up visions of accidents posing as "little gremlins" along the roadway. If they are chased away at one location by some safety measure they just move on down the road and cause trouble there. Yet, restated, this concept is quite believable. Examples of specific countermeasures are readily available which, if applied to one road section, could result in an increase (or decrease) in accidents on a related (intersecting or contiguous) road section. Consider the application of selective traffic enforcement at preannounced road sections. Traffic which normally flows through enforced sections may divert to adjacent sections to avoid detection. The vehicles most likely to change course could be those most likely to have an accident (i.e., habitual speeders, unlicensed drivers, or drivers under the influence of drugs or alcohol). Hence, accidents may increase at the adjacent sections due to selective traffic enforcement at selected sections. This would be called accident migration due to implementation of a safety measure.

Most currently used statistical methods for countermeasure evaluation do not consider the accident migration phenomenon since the focus is on treated site measures of effectiveness rather than more global measures. That is, rather than considering the entire system of road sections in a study area and measuring the safety of the "system", methods of before/after analysis measure the safety locally, i.e., at the treated site. Even before/after designs with comparison or control groups may exclude significant portions of the system and hence ignore the fact that the safety of the system may have been affected by the countermeasure.

A key issue in the implementation of a more global safety measure, however, is data availability and cost. Frequently, the cost of obtaining additional data makes even comparison or control section data unattainable. Therefore, any solution or recommendation for an improved methodology for countermeasure evaluation must consider the cost and feasibility of obtaining additional data if the methodology is to be widely accepted and adopted. The statistical concepts of experimental design, sampling, and modeling may provide the necessary information for being able to selectively acquire a minimal (cost-

effective) amount of data for projecting treatment effectiveness estimates for an entire system of road sections with various types of exposures or accident "potential".

In this study, statistical concepts were explored and applied to several simulated accident scenarios. The objectives of this study were:

> To develop a methodology for determining system-wide safety effects of improvements made at specific local sites and assess under what statistical conditions these effects can be accurately evaluated.

One of the first tasks in this study was to arrive at some definitions and concepts on what constitutes accident migration and what are its various forms. Having done this, the task of determining what types of accident migration are relevant to this study and the scope of the study could become more focused.

The literature refers to two types of accident migration--geographic and non-geographic. Geographic migration is the type of migration which generally comes to mind, namely:

> A change in accident rate or frequency at a particular location in a roadway system resulting from treatment at another location.

Non-geographic migration, according to the literature, would consist of examples such as a shift across severity levels due to a treatment effect (seat belts) or a change in accident type due to a treatment effect (more rear-end collisions after posting a stop sign).

Early on in this study it was jointly decided that this study would focus on geographic migration. The reason for this is that methods already exist for assessing non-geographic types of migration, whereas, for geographic migration, the main issues are how to design an experiment to detect it, as well as to define the conditions under which it is detectable.

To distinguish between the terms "geographic" and "non-geographic" migration, consider the establishment of a new terminology. If the change in accidents (type or severity) occurs within the treated domain, define this as accident metamorphosis. In this way, the term accident migration can be reserved for changes in accidents outside the treated domain. Table 1 characterizes this definitional concept.

Table 1. Accident metamorphosis and migration.

Change in Accident Type or Location	Treatment Domain: Spot	Treatment Domain: Area
Inside Treatment Domain	(1) Metamorphosis	(2) Metamorphosis or Migration
Outside Treatment Domain	(3) Migration	(4) Migration

To better understand table 1, consider an example from cell 1. Imagine that a traffic signal is installed at a given intersection. By the implementation of this treatment, the total number of accidents at the intersection may decrease and the distribution of accident types may metamorphose into a new distribution (e.g., fewer angular collisions, but more front-to-rear collisions). It is not practical to consider changes in accident location within an intersection (or any other spot that has undergone treatment) and, therefore, accident migration is not defined in cell 1.

Regarding cell 2, imagine that 10 mi (16.1 km) of highway are widened while existing structures (e.g. bridges and culverts) are left unchanged. Under these circumstances it is quite conceivable that accidents within these 10 mi (16.1 km) of highway will both metamorphose and migrate after widening: fewer run-off-road accidents may be recorded, but more bridge accidents may occur (metamorphosis); fewer accidents may occur along curves and tangents, more may occur at bridges (migration). This example serves to illustrate that in cell 2 accident metamorphosis and migration are not completely independent and, in fact, may be highly interrelated.

In cells 3 and 4 accident migration involves the increase in accident rate or frequency at locations outside of the treatment domain. If a particular curve is improved, the accidents at that curve may decrease, but accidents at the next curve may increase (cell 3 conditions). If 10 mi of highway are resurfaced, wet weather accidents throughout those 10 mi may go down while wet weather accidents throughout the next 10 mi downstream may go up (cell 4 conditions). Examples of countermeasure studies which would correspond to the above cells are:

1. Traffic signalization.
2. Lane widening but bridges left same width.
3. Bridge widening, signing.
4. Resurfacing, raised pavement markers, 65 mi/h (105 km/h) speed study.

The purpose of this project is to develop a methodology or statistical procedure to evaluate accident migration--a formidable task. First, consider the simpler task of evaluating accident metamorphosis. To evaluate accident metamorphosis in cells 1 or 2 (or theoretically in cells 3 or 4) of table 1, the most obvious analysis is a standard chi-square (x^2) or likelihood chi-square (G^2) test, although it should be pointed out that such tests, like other tests used in before/after analyses, are susceptible to regressional artifact. Imagine, for example, that a given intersection is upgraded to reduce angular collisions. Three years of data are available for the intersection before upgrading and for one year after, as shown in table 2. The headings -3, -2, and -1 represent 3 years, 2 years and 1 year before and +1 represents 1 year after, respectively.

Table 2. Hypothetical intersection accident data.

	Before			After	
Accident Type	-3	-2	-1	+1	Total
Angular	28	32	29	17	106
Front-to-rear	17	15	12	20	64
Head on	6	6	4	4	20
Side swipe	6	9	7	7	31
Other	5	8	6	10	29
	62	70	58	60	250

The standard G^2 before/after statistic will be used to analyze this data. This statistic is similar to the chi-square statistic with the exception that it can be partitioned into sub-levels, similar to analysis of variance sums of squares, since it is based on logarithms and is therefore additive. The general form of the equation is:

$$G^2 = -2 \sum m_{ij} \ln \left(\frac{\hat{m}_{ij}}{m_{ij}} \right) \quad (1)$$

where $\hat{m}_{ij}$ is the expected cell count, the same expected count as used in the chi-square test, and m_{ij} is the observed cell count. This statistic has a chi-square distribution under the null hypothesis that the observed cell frequency equals the expected with (r-1) (c-1) degrees of freedom where r is the number of rows and c the number of columns in the table. A more detailed account of the G^2 analysis can be found in Griffin.[1]

Two sources of G^2's are computed--one for the before data only (3 by 5 table or 8 degrees of freedom) and one for collapsing all 3 years of before data into one column with a single column of after data (2 by 5 or 4 degrees of freedom). The total degrees of freedom for the 4 by 5 table is 12. Table 3 lists these statistics.

Table 3. G^2 analysis.

Source	G^2	df
Before	1.876	8
Treatment	7.968	4
	9.844	12

The G^2 before of table 3 indicates that there is no significant difference in the types of accidents in the before period at these sites. The critical χ^2 with 8 degrees of freedom at $\alpha = .05$ is 15.5. The computed G^2 value would have to exceed this in order that the hypothesis that there is no difference in accident types during the before period be rejected. Therefore, the relative proportions of different accident types were consistent in the before period.

This conclusion supports collapsing over the three before time periods to assess the treatment effect. Had this not been the case, the treatment effect could not be assessed in this problem and the source of the inconsistency would need to be determined. The G^2 treatment is not significant either at $\alpha = .05$ (critical $\chi^2 = 9.49$) and hence the conclusion is that the treatment had no effect in changing the distribution of accident types.

However, had the data in table 2 been regrouped by accident type, the conclusion would have been different. Consider the regrouped numbers in table 4 and the G^2 analysis of table 5.

Table 4. Hypothetical intersection accident data (regrouped).

	Before			After	
Accident Type	-3	-2	-1	+1	Total
Angular	28	32	29	17	106
Other	34	38	29	43	144
	62	70	58	60	250

Table 5. G^2 analysis of regrouped data.

Source	G^2	df
Before	0.338	2
Treatment	6.609	1
	6.947	3

The before G^2 is not significant, again indicating a consistency in the proportion of angular and non-angular accidents in the before period, but the treatment G^2 is significant (critical $\chi^2 = 3.84$), indicating that the proportion of angular accidents significantly decreased after treatment.

The point of this example is two-fold; 1. statistical methods already exist to analyze accident metamorphosis, and 2. a major concern and issue if accident metamorphosis is to be assessed, is of how to partition the accident data. The latter depends largely on the question to be addressed. In this example, obviously, the second partition is more meaningful in that the question is--does this safety measure affect angular accidents? Partitioning the data further only diluted the effect and addressed a more global issue of whether the measure affected any accident types.

Having addressed the issue of accident metamorphosis to this extent, the focus shall now primarily be on accident migration as this is the objective of this study. There are four key issues to consider in assessing migration. (It will be assumed that the term migration refers to "geographic" migration throughout this report.) They are the "what, where, when and how" of accident migration. The what has already been addressed in the definition of geographic migration. The following are issues for the other three.

Where does accident migration occur?

In other words, where is the geographical region of concern for potential accident migration? In order to answer this, a clear definition of the geographical treatment area is necessary i.e., the area wherein the treatment is being evaluated. Highway treatments are implemented on two general types of areas, 1. spot locations (e.g., intersections and bridges), and 2. segment locations (e.g., roadway sections). It is possible for the treatment area to be a spot and the migration area to be surrounding spots (intersections), or for the treatment area to be a spot and the migration area to be segments. The point is that in the definition of accident migration for a particular highway treatment, the geographical domain of the migration plays an important role.

When does accident migration occur?

That is, for what types of safety treatments is accident migration likely to be a real and critical factor? Assessing and measuring the effect of migration will necessarily add time and cost to any evaluation study. When is this additional effort warranted?

How does accident migration occur?

What are the mechanisms of accident migration? Is it likely to occur because of physical or operational changes in the system (narrow view) or changes in driver behavior (broad view)? Also, how does accident migration manifest itself? Is it causing a change in accident type, accident severity, or total accidents?

Narrow view of accident migration

The narrow view of accident migration explains the phenomenon of accident migration in purely physical or operational terms. Imagine, for example, that there are two overpasses on a low volume east-west rural road, one mi (1.6 km) apart. The western overpass is 14½ ft (174 in or 426.7 cm) high. The other overpass one mi (1.6 km) to the east is 14 ft (168 in or 442.0 cm) high. Last year, 100,000 vehicles passed through this section of highway from west to east. Ten of these vehicles were over 174 in high (442.0 cm), 5 were between 168 and 174 in (426.7 and 442.0 cm) and the rest (99,985) were less than 168 in (426.7 cm) high. Obviously, 10 vehicles struck the western overpass and 5 struck the eastern overpass.

To reduce the number of accidents at the first (western) overpass, the highway department increased the height (clearance) of the structure. In the year following treatment to the overpass, another 100,000 vehicles (with the same distribution of vehicle height observed the previous year) traveled down the road from west to east. This year 5 vehicles struck the western overpass (a reduction of 50-percent) and 10 struck the eastern, unmodified overpass (an increase of 100-percent). In this example five accidents have simply been displaced to a second overpass. This displacement of accidents, this change in accident location, takes place without any change in driver behavior.

Consider a second example. On another low volume, east-west rural road there are two sharp horizontal curves, one mi (1.6 km) apart. The more westerly curve is a 10° curve; the curve one mi (1.6 km) to the east is a 12° curve. Each year 100,000 vehicles travel this road from west to east. Some of the drivers of these vehicles drive very cautiously (i.e., slowly, attentively), others drive more carelessly (i.e., faster, less attentively). For the sake of this example, let us assume that the caution or carelessness with which drivers operate their vehicles can be measured along a dimension called driver

intensity (I). Further assume that driver intensity is normally distributed with a mean of 50 and a standard deviation of 10. And, finally assume that at driver intensities (I) up to 88.24 (i.e., z = 3.824), 10° horizontal curves can be successfully negotiated without crashing. At driver intensities up to 87.58 (i.e., z = 3.758), 12° horizontal curves can be successfully negotiated without crashing. See figure 1.

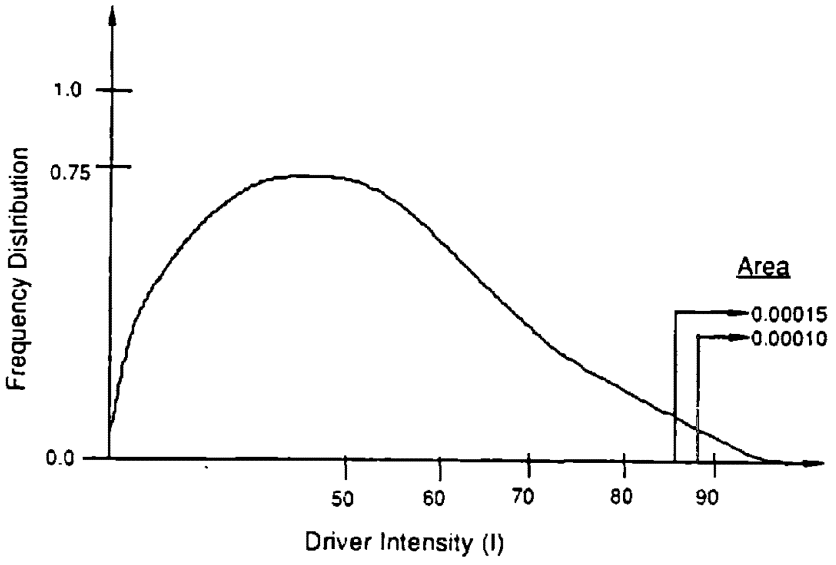

Figure 1. Hypothetical driver intensity distribution.

Of the 100,000 vehicles driven from west to east down our hypothetical highway, 99,985 are driven at intensities below 87.58, 5 are driven at intensities between 87.58 and 88.24, and 10 are driven at intensities beyond 88.24. Obviously, 10 vehicles crash at the westerly (10°) curve. Five more crash one mi to the east at the 12° curve. The highway department alters the geometry on the western (10°) curve to such an extent that 5 of the 10 vehicles that would otherwise have crashed at the first curve are now able to successfully negotiate it. But, note that the five vehicles which did not crash due to the efforts of the highway department are being driven with an intensity in excess of 87.58. When they encounter the 12° curve one mi to the east, they will crash. Or, in other words, by correcting the first (10°) curve, accidents were reduced from 10 to 5 (50-percent), but accidents at the second (12°) curve were increased from 5 to 10 (100-percent).

Again, five accidents have been displaced to a location one mi to the east. And, again, this displacement of accidents, this change in accident location, has taken place without any change in driver behavior. Two axioms follow from the narrow view of accident migration:

1. If the site (e.g., the overpass or the horizontal curve) that is improved is not the most severe site in the system (i.e., not the lowest overpass or the sharpest horizontal curve), then at equilibrium, the accidents reduced at the improved site will be exactly offset by increases in accidents at the remaining sites in the system. Systemwide, accidents remain the same.

2. If the site that is improved is the most severe site in the system (and no other site is equally severe), then at equilibrium, accidents reduced at the improved site will more than offset increases in accidents at all remaining sites in the system. Systemwide, accidents decrease.

Each of these axioms should be carefully considered by the researcher and well defined before embarking on assessment of accident migration.

Broad view of accident migration

The broad view of accident migration acknowledges that treatments to the roadway environment can result in the displacement or migration of accidents as previously discussed. In some circumstances (e.g., the overpass example) all, or nearly all, of accident migration can be viewed as a simple, physical displacement of accidents. But, the broad view of accident migration also assumes that driving behavior interacts with the environment. By this view, driver intensity is not an immutable distribution, as implied in the horizontal curve example, but a distribution that changes over time as a function (among other things) of current environmental (i.e., roadway, traffic, and weather) conditions and recently encountered conditions.

Drivers who have just encountered a severe horizontal curve will drive with less intensity. Conversely, drivers who have just encountered a relatively benign horizontal curve may drive with more intensity. Or, referring once again to figure 1, the distribution of driver intensity can be altered (shifted to the left or right, made more or less variable, or skewed) by recently encountered conditions. Unfortunately, no theory of human behavior can quantitatively describe how driver intensity would be affected by recently encountered environmental conditions. In the absence of such information, and assuming that the behavioral aspect of accident migration can be simply added to the more physical or operational aspect, the net gain or loss in accidents to the system due to a safety treatment is not well defined.

A HISTORICAL OVERVIEW

Although the potential existence of the accident migration phenomenon was apparently recognized for some time, Boyle and Wright appear to be the first to dare to acknowledge this phenomenon in the literature and to attempt to quantify the safety effect for a specific data set.[2] Their methodology is fraught with design and data limitations, as well as statistical inadequacies, which they readily acknowledge. These limitations are further illuminated by Huddart's eloquent discussion.[3] Nonetheless, Boyle and Wright are to be commended for their pioneering efforts and the attention which their article generated on this controversial topic.[2]

There are several controversial aspects to this concept of accident migration, the most obvious one being a direct result of the terminology. As Boyle and Wright state, this terminology implies that there is some physical movement of accidents.[2] Some of the arguments they use to support the existence of the accident migration phenomenon, like the "near-miss" hypothesis, seem to border on acceptance of the physical movement concept. Although it is quite plausible that, for some drivers, keeping an intersection dangerous so that they will experience "near-misses" will make them more cautious drivers at other intersections in the system, these near miss events are surely rare occurrences and would likely not have a dramatic influence on another rare occurrence, accidents, as Huddart points out.[3]

A more realistic interpretation of the accident migration phenomenon is that there are some safety measures which affect not only the sites where they are implemented, but an entire system or network of sites within some surrounding region. This effect can be reflected in the change in the correlational structure of the accident experience among the sites before and after implementation of the countermeasure. Thus, if this correlational structure can be quantified, a system-wide effect of the countermeasure would be readily available in the form of a measure of change in the before and after correlational structure.

Although the concept of accident migration has been recognized for quite some time by the researchers in the field of countermeasure evaluation, most of the literature on this subject appears to be rather current. In one of the most recent articles on this subject Persaud recommends an empirical Bayesian approach first suggested by Abbess et al. and more recently formalized by Hauer.[4,5,6] This recommendation is based on the fact that the regression-to-the-mean phenomenon is obviously a factor to be considered in assessing the extent of accident migration. In this study, we reviewed and critically assessed the limited literature on the subject of accident migration, including the articles of Abbess et al., Boyle and Wright and related discussions (Huddart, Stein, McGuigan, Boyle and Wright), and the articles of Persaud. (See References 2, 3, 5, 7, 8, 9, 10, 11, 12, 13.) These articles provide a foundation for the research conducted in this study.

SAFETY MEASURE EXAMPLES REQUIRING SYSTEM-WIDE ASSESSMENT

There are numerous examples of countermeasures wherein a system-wide analysis could be an important factor. The following is a list of some of these examples:

- Increased stop sign size to better alert driver. Accidents may increase at surrounding locations with standard size stop sign.
- Bypass construction around an urban area. Accidents may decrease within the urban area but migrate to the perimeter (bypass) area.
- Left-turn protected signalization. Accidents may increase at left-turn permissive locations due to driver expectancy.
- Widening of roadway width without corresponding widening of bridge structure width. Accidents rates may migrate from normal roadway sections to bridge structures with restricted width.
- Use of reduced speed zones on short sections of highway. Accidents may reduce in these reduced speed zones but increase farther down the road.
- Increasing the speed limit from 55 to 65 mi/h on rural interstates. Accident rates may not significantly change on rural interstates but increase on contiguous and intersecting sections.
- Use of speed bumps to control speed. Traffic may divert to alternate routes, and correspondingly, accident rates may increase.
- Use of selective enforcement to control speed or DUI. Traffic may divert to alternate routes, and correspondingly, accident rates may increase.
- Left turn prohibited at some intersections. Accident rate may increase at other intersections without left turn restrictions.
- Traffic control devices (innovative flagging, signing, etc.) in construction zone taper areas. Some types of traffic control devices may cause erratic lane-changing in the pre- or postconstruction zone tapers so that accidents are reduced within the taper but are increased on contiguous road sections.

The extent to which the accident migration phenomenon poses a problem will vary according to the type of countermeasure. A list of potential countermeasures or safety improvements was developed in consultation with experienced traffic engineers in the respective safety areas. These countermeasures were evaluated as to the probable impact of accident migration to sites away from the sites of actual improvement. Safety improvements were classified into three groups, (1) potentially highly affected by accident migration, (2) possibly affected by accident migration, (3) not likely to be affected by accident migration. This list is presented in Table 6.

Table 6. List of safety improvements and assessment of the need for a system-wide evaluation.

Safety Improvement	Need for System-wide evaluation
Intersections	
Signalization	Highly Probable
Conversion to 4-Way Stop	Highly Probable
Left-Turn Treatment	Highly Probable
Installation of Zebra Crossings	Moderately Probable
Road Sections	
Raised Pavement Markers	Moderately Probable
Increasing Lane Width	Moderately Probable
Adding or Removing Shoulders	Highly Probable
Increasing Number of Lanes	Moderately Probable
Resurfacing	Moderately Probable
Modification of Roadside Appurtenances	Unlikely
Installation of Guardrail	Unlikely
Construction Zones	
Reflective Pavement Markings	Highly Probable
Traffic Control Device	Highly Probable
Signings and Warnings	Highly Probable
Selective Enforcement	
Speed	Highly Probable
DWI	Highly Probable

Additional specific examples of countermeasures in actual previously conducted studies are presented below to illustrate where accident migration could have been a probable factor.

Attempts to alleviate or reduce certain types of accidents occurring at work zones may result, in some cases, in the "accident migration" phenomenon. For example, there has been research conducted to improve flagger safety at work zones. For certain situations, portable traffic signals have been suggested for control of alternating two-way traffic over a one-lane section of road instead of flaggers. The potential for flagger-vehicle accidents are obviously removed with the removal of the flagger from the work site. However, it has been noted that driver noncompliance with the portable signals may result in head-on vehicle collisions in the middle of the work zone. Complete analysis of the accident potential for flaggers versus portable traffic signal control was not performed as part of that research.

Another example consisted of studies on the use of radar transmitting devices at work zones to reduce traffic speeds. The devices emitted a microwave signal which activated commercially available radar detectors approximately 0.5 mi in advance of the work zone. The research indicated that the transmitter did result in a slight decrease in speeds approaching the work zone, and had the biggest effect upon the higher (i.e., 60 mi/h (95.2 km/h) and above) speed vehicles. It was hypothesized that the transmitter could result in safer traffic operations through the work zone. However, it was not possible to determine if accidents might increase upstream of the work zone where the radar signal was first detected. Since not all vehicles are equipped with radar detectors, the transmitter may generate speed differentials as some vehicles (with radar detectors) slow down upon receiving the signal while other vehicles do not. Again, full-scale investigation of the upstream accident implications of the transmitter was beyond the scope of the study.

Another study was conducted to specifically evaluate accident migration as a result of modifying existing shoulders to add lanes. No significant effects were found; however, modifications could be made to the experimental design and statistical methodologies to reevaluate this study. Specifically, no suitable locations could be found to evaluate geographical migration, however, it was suspected that even though there was no change in total accidents, median barrier accidents increased due to lane narrowing. This would be a prime example of accident migration across accident type.

STATISTICAL MEASURES OF EFFECTIVENESS

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? The evaluation of treatment effects at migration sites adds a new dimension to this problem, but, basically, the statistical issues are the same.

The basic problem in accident analysis 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 are inappropriate and do not recognize or utilize this very basic characteristic of accident data. For these reasons, two statistical methods for evaluating treatment and migration effects are used in this study. Both methods are based on the before/after with comparison group design. That is, it is assumed that comparison group data is available for both the treated and migratory sites and that accident counts and volumes (i.e., accident rates) are available for a before and after time period for all groups.

With this experimental design, two statistics are available to estimate a treatment effect and test its significance. They are the classical or frequentist cross product (odds ratio) estimate and the recently developed EBEST (Empirical Bayes Estimate of Safety in Transportation) estimate. Both of these estimates and their statistical tests will be reviewed in this section.[1,14]

The odds ratio

Assume the data is oriented in the following two-by-two tabular format with accident counts as data entries:

Table 7. Tabular format for accident data.

	Comparison	Treatment
Before	A	B
After	C	D

The odds ratio (cross product, frequentist) estimate of treatment effectiveness is:

$$O.R. = \frac{A/C}{B/D} = \frac{AD}{BC} \tag{2}$$

The statistical test of significance for this estimate is:

$$T = \frac{\ln(O.R.)}{S.D.(O.R.)} \tag{3}$$

where the O.R. is the odds ratio and S.D. represents the standard deviation of the odds ratio, as follows:

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

This test statistic has a standard normal distribution and is compared to normal z-values at a specified level of significance. The estimate is interpreted as a percent reduction by computing:

$$x = (O.R. - 1)\ 100 \tag{5}$$

Then the treatment is said to significantly increase (if x is positive) or decrease (if x is negative) accidents by x - percent if T, in absolute value exceeds 1.96 (5-percent level of significance). If T in absolute value does not exceed 1.96, the safety treatment is said to have caused no change in accidents. This same statistic and test can be used with sites where potential migratory treatment effects could occur.

This is the standard method used in most before/after studies with a comparison group. The comparison group is necessary to control for other potential confounding factors independent of the treatment such as time, roadway geometrics, etc. The comparison group cannot, however, adjust for the very serious potential confounding effect of regression-to-the-mean.

Regression-to-the-mean adjustment

Sampling bias due to the regression-to-the-mean (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 2 depicts various degrees of the r-t-m phenomenon. (Note: this figure is purely anecdotal and does not represent the real distribution of accident frequencies.)

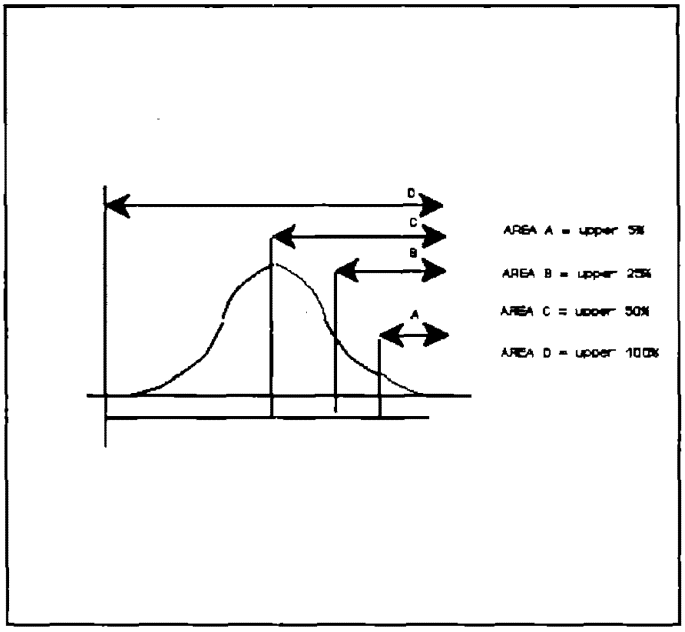

Figure 2. Regression-to-the-mean simulation.

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.[5,6] 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. (See references 14, 17, 18, 19, 20, 21.) Although the concept of this methodology has been widely embraced, actual use of this method in accident analysis has been limited. The procedure developed in an earlier study can adjust for this sampling bias.[14,15,16] This procedure is called the EBEST procedure. 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 the classical statistical assumption of random sampling. Similarly, migration sites may necessarily be selected in a biased rather than random fashion.

Whereas the use of empirical Bayes in this application is not new, the EBEST methodology 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.

The EBEST procedure uses a 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 reference group is a sample representing the pool of potential treatment (or migration) sites. 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 subsequently.

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, $\hat{}$, and the site's observed accident rate, y_i, as follows:

$$\hat{\lambda}_i = B_i \hat{} + (1 - B_i) y_i \tag{6}$$

where B_i, is:

$$B_i = \frac{\epsilon_i}{e_i + \epsilon_i} \tag{7}$$

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, $\hat{}$. 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 estimated 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).

In order to use the EBEST method, certain critical assumptions must be met by both the treated and migration sites. One such assumption is called 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 or potential migration sites (i.e., the reference group for the treatment sites and the reference group for the migration 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. For more details on understanding the concept of exchangeability, the reader is referred to Pendleton et al.[16]

The EBEST procedure, then, 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, if r-t-m is a potential problem for the migration sites, a reference group for these sites will also be required to adjust for the r-t-m effect.

To compute the EBEST estimate for evaluating a safety treatment, the odds ratio estimate is modified by computing the quantity, C, using the estimated accident rate for each site, i.e.,

$$C = \sum_{i=1}^{K} \hat{\lambda}_i e_i \tag{8}$$

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

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. Statistical estimates and inferences can be drawn using classical methods of hypothesis testing and confidence interval estimation.

The treatment and migration effects using both methods were computed and tested in this study. These were labeled as:

$\hat{\theta}_{TF}$ = treatment effect at treated sites using frequentist method.

$\hat{\theta}_{MF}$ = treatment effect at migration sites using frequentist method.

$\hat{\theta}_{TEB}$ = treatment effect at treated sites using the EBEST method.

$\hat{\theta}_{MEB}$ = treatment effect at migration sites using the EBEST method.

Data needs

The data needs in assessing accident migration depend upon whether or not r-t-m is a potential problem. The degree of r-t-m is also a factor, and guidelines as to how much r-t-m must be present before a problem exists will be given in chapter 3. If there is no reason to believe that the migration sites that were selected were either the most hazardous or least hazardous sites, that is, if they represent a random sample from the population of potential migration sites, then there is no need for a reference group for the migration sites. Similarly, though less likely, if the treatment sites were not selected as the most hazardous sites, there is no need for a treatment reference group. Specific data requirement for these groups are listed below.

Treatment and migration group data

accident counts - Before and after accident counts for all treatment and migration sites are needed. 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 desirable and at least one year of after data. Three years of after data might be excessive and increase the chances of other confounding factors such as changes in posted speeds or construction to obscure the conclusions.

exposure measures - 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 and migration 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, and could negate or invalidate study conclusions based on this method. Measures of exposure for the after period are desirable but may be omitted if it is justifiable to assume the after exposures would not change substantially from the before to after periods.

Reference group data for treated sites

accident counts - Before accident counts for all reference group sites are needed. The size of the reference group should be much larger (say 5 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 durations 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 exposure measures for the reference group sites are 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.

Reference group data for migration sites

Reference group data for migration sites will only be necessary if there is reason to suspect that the migration sites were selected with an r-t-m bias. It is feasible that this could occur in either direction. That is, the migration sites could be the most hazardous sites due to their proximity to the treatment sites. Hence, all factors contributing to the degree of hazard of the treatment sites would be present and affect the migration sites. In this case, the migration sites would have a high (positive) r-t-m bias similar to the treatment sites. On the other hand, the migration sites could have come from the same selection pool as the treatment sites (i.e., they could have been candidates for the treatment but were not hazardous enough to be selected). In this case, the migration sites would be experiencing less accidents than usual and have a lower or negative r-t-m bias. Criteria for determining whether or not bias is a problem were developed in this study and provided in chapter 3.

Comparison Group

accident counts - Before and after accident counts for a comparison group for both treatment and migration sites should provide a reasonable estimate of the expected change in accidents at the 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. The size of the comparison group is generally the same as the size of the treatment group and migration group, respectively. It may be reasonable to use the same comparison group for the migration site as the treated site depending on the situation.

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 that 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.

Correlation

Maher first addressed the issue of correlation between the migration and treatment sites and the likely effect such correlation might have on assessing accident migration.[22] Maher contended that not only do we have a potential "reverse regression-to-the-mean" phenomenon confounding the true accident migration effect, but in the fact that a site "neighbors" a treated site, there is an underlying correlation between these sites which can confound the migration effect. He demonstrates this correlational effect via simulation studies using two different procedures. In one, the moving average method, a moving average was computed using a lognormal distribution assumption on the mean accident rates. In the second, the clustering method, a number of different "centers" was selected on a grid and then the surrounding means were generated such that they decreased as distance from these centers increased. The results of these two procedures, he claims, were similar.

The statistics used to compare before/after reductions was the usual percent reduction in before to after. This was computed for:

1. Adjacent neighbor sites from "treated" sites, P_{mig} .
2. "Treated" sites, P_{reg} .
3. All untreated sites whether adjacent to the treated sites or not, P_{unt} .

"Treated" sites were not actually treated, of course, but selected based on accidents greater than some selected threshold criterion, k. Since there was no actual treatment effect, change before-to-after in the treated sites was attributable to r-t-m, P_{reg}, reverse regression-to-the-mean, P_{unt}, or accident migration, P_{mig} .

As the site selection threshold increases, the size of reverse regression-to-the-mean effect falls steadily to zero, supporting McGuigan's observation.[9] The change in accidents at the neighboring sites (the "migration" effect) did not fall nearly as quickly and tended to remain at about the 10-percent level. The conclusion: "the prime determinant of the magnitude of the migration effect is the amount of spatial correlation between the site means." As this correlation approaches zero, the "migration" effect approaches the reverse regression-to-the-mean effect.

Maher's results were based on a limited number of simulations and choices of k values (two examples with 10 simulations and k varying from 7 to 18 with mean and shape parameters of 4.74 and 1.72 and 3.53 and 1.78, respectively). In summation, he concludes that any model which tries to account for the effect of accident migration must incorporate some spatial correlation among neighboring sites as well as adjusting for the regression-to-the-mean effect. He proposes "more mathematical and less simulation based research" to develop "formulae, graphs or tables which would allow the expected size of the migration effect to be determined."

The basic discovery in his study, that of spatial correlation, appears to be a vital one to the problem of isolating true accident migration. However, his paper just touches the subject in a small and limited way. To truly grasp this phenomenon, a much larger simulation study with more choices in means and site selection criteria is in order. Also, the author uses two methods to "impose" a type of spatial correlation, but neither allows for the control of this spatial correlation by varying levels, high to low. For example, the moving average process (the author does not state the choice of lag so it is assumed to be one) will result in some value of spatial correlation--but what is it for these simulations? In the second procedure, it would appear that some type of correlation is imposed (as Maher claims the site means decrease as distance from the center increases), but how much correlation is imposed and how was this done? It would appear that a better way to conduct a simulation study would be to impose varying degrees of spatial correlation on the structure and control for the amount of correlation. A further important limitation of this study, which is imbedded in the spatial correlation concept, is exposure (traffic flow). Also, the study neglects to propose a statistical method for measuring accident migration that would incorporate both the concept of spatial correlation and regression-to-the-mean.

Because of Maher's conclusion that correlation would be an important factor, this factor was included in the simulations of this study. However, the correlational structure was imposed in such a way as to allow control of the degree of correlation. This was done by selecting the parameters of the gamma distributions. First, two variables must be defined to represent accident rates at the treatment and migration sites in such a way so that they have a common element. Let y_T and y_M represent the random variable, accident counts, for the treated and migration sites, respectively. Define:

$$y_T = x + w \tag{9}$$

$$y_M = z + w \tag{10}$$

where x, z, and w are all independent Poisson random variables with means λ_x, λ_z, and λ_w, respectively. Then, the mean (and variances) of the y's will be:

$$E(y_M) = \lambda_w + \lambda_z \tag{11}$$

$$E(y_T) = \lambda_w + \lambda_x \tag{12}$$

and the y's will be correlated with covariance:

$$COV(y_T\, y_M) = \lambda_w \tag{13}$$

and correlation:

$$\rho_{TM} = \frac{\lambda_w}{\sqrt{(\lambda_w+\lambda_x)(\lambda_w+\lambda_z)}} \tag{14}$$

In this simulation study, it is assumed that treatment and migration sites come from the same population and thus have the same expected accident rate in the before period. With this assumption (14) simplified to:

$$\rho_{TM} = \frac{\lambda_w}{\lambda_w+\lambda_x} \tag{15}$$

Now by specifying a correlation (ρ_{TM}) and a true accident rate at both the treated and migration sites, λ_x, the common element, λ_w, could be determined.

That is, let

$$\lambda_{TB} \overset{d}{=} Gam(\lambda_w + \lambda_x\ ,\ \epsilon) \tag{16}$$

and

$$\lambda_{TM} \overset{d}{=} Gam(\lambda_w + \lambda_z\ ,\ \epsilon) \tag{17}$$

where λ_{TB} = true mean of treated sites before.

λ_{TM} = true mean of migration sites before.

$\overset{d}{=}$ mean "distributed as".

ϵ = true exposure.

Assume the true mean rate for both the migration and treatment sites is .09 and the desired correlation between treated and migration accident rates is .80. Substituting into (15) we can solve for λ_w:

$$\lambda_w = (.09)(.80) = .072 \tag{18}$$

Now, random variables can be drawn from the three separate distributions:

λ_{xi} and λ_{zi} come from gammas with mean rates of .018 (.09 - .072) and λ_{wi} from a gamma with mean .072. Then, by adding these rates accordingly:

$$\lambda_{TBi} = \lambda_{wi} + \lambda_{xi} \tag{19}$$

and

$$\lambda_{MBi} = \lambda_{wi} + \lambda_{zi} \tag{20}$$

Thus, treatment and migration site mean rates have been generated which are correlated.

MEETING THE STUDY OBJECTIVES

It was initially felt that a retrospective, historical approach using real data from safety measure studies would be the most effective method for accomplishing the objectives of this study. That is, according to the extensive literature on before/after countermeasure evaluations, it appeared that a bountiful supply of accident data existed for accomplishing the goals and objectives of this project. A retrospective study would necessarily entail the identification of data sources, preferably automated, which could provide accident histories not only for the treated or local sites which received the safety improvement, and for some non-local sites, such as control and comparison groups, but for a complete system of sites which could be affected by accident migration. In addition, to address the regression-to-the-mean issue, reference group data or data on a sample of potential treatment and migration sites would be necessary.

Anyone familiar with accident data would recognize the difficulty of finding data for this task. The task of supplementing data on completed studies with data not

originally collected in that study would be formidable. Still, an earnest attempt was made to retrieve such data. Several potential accident data sources were considered for inclusion in this study. This required considerable expenditure of time and resources. In all cases, one or more of the following difficulties was encountered:

1. The data did not reside in a location oriented format, i.e., it was not possible to retrieve accidents by specifying the roadway section or location where it occurred.

2. Accident migration locations could not be identified.

3. Reference groups could not be identified.

4. Treatment sites could not be isolated from the population of sites from which they were drawn as that "piece" of information was no longer in the database.

5. The data were politically sensitive.

For these reasons, the retrospective approach using real data had to be abandoned. Instead, simulated data was used. In this way, known treatment and migration effects could be imposed on the data for comparison with estimated values. Also, all of the necessary data groups were easily produced and regression-to-the-mean could be imposed on both treatment and migration groups. The accident counts and volumes were selected to represent real data by examining data from the Texas state accident and roadway inventory files. Interstate, two-lane rural roads, intersections, etc. were examined to determine realistic accident and volume data for these roadway types.

CHAPTER 2. ASSESSING THE CONDITIONS WARRANTING A SYSTEM-WIDE EVALUATION

In this chapter the results of the three stages of this study will be presented. Stage 1 assesses the ability to detect treatment effects at both treated and migration sites when there is no regression-to-the-mean bias in the selection of the migration sites. In stage 2, r-t-m bias was imposed on the migration sites in both directions to assess the ability to detect accident migration. Stage 3 further refines these results based on factors identified in stage 2 as being important.

SIMULATION PROCEDURES AND PARAMETERS--STAGE 1

For this set of simulations, it was assumed that the treated sites had severe r-t-m bias, i.e. that the most hazardous sites were selected for treatment. The migration sites, on the other hand, were randomly drawn so that no r-t-m bias was expected. In this way, the statistical methods for evaluating treatment effect at the migration sites were unconfounded by any sampling bias. Inability to detect a system effect under this "ideal" scenario would put serious doubt on the ability to detect this effect under more complicated conditions.

The following parameters were considered to be most important in providing a complete analysis of the accident migration problem. These parameters are defined and reasons provided to justify their relevance to the study.

accidents - Both accident counts and accident rates are important variables to consider in the ability to detect a system effect. Obviously, this effect is going to be more difficult to measure in low accident count situations. Both accident counts and accident rates are varied to reflect the entire range of possible countermeasure situations wherein it is highly probable that a system-wide evaluation is warranted.

traffic volumes - Traffic volumes tend to follow different distributions depending upon highway or intersection types. Traffic volumes are simulated from distributions similar to those in real world situations where accident migration is a potential problem.

number of sites - The ability of the proposed statistical methods to be able to detect migration depends upon the amount of information available, specifically the number of sites measured. The number of sites provides varying degrees of accuracy in estimating both within and between site variability among the treated and migration sites as well as the correlation among the treated and migration site pairs.

treatment effect and migration effect - The smaller the treatment or system effect the more difficult it is to detect by any statistical method. Thus, by varying this parameter it is possible to define the minimum magnitude of either effect which would be reasonable to detect using the proposed statistical methods.

correlation - The correlation among the treatment sites and their corresponding potential migration sites provides an indication of how the migration is occurring on a site-by-site basis, i.e. systematic versus global migration. The degree of correlation is varied in this study to compare the ability to measure both of these types of migrations. More details are provided in the statistical methods section.

Ideal conditions

In order to be able to test the sensitivity of the statistical methods in assessing treatment effects, the first step was to control for outside factors that might confound the issue. That is, if the methods were not successful in assessing a treatment effect under ideal conditions, there would have been no point in pursuing lesser than ideal conditions and the study would have terminated. The following is a list of simulation parameters used in this stage.

Number of sites

Three values were studied to reflect low, moderate and large number of sites in a safety measure study. Specifically, groups of 10, 25 and 100 sites were used. Ten sites are typical numbers in intersection or construction zone safety measure studies and in any situation where the cost of evaluation per site is a limiting factor. Twenty-five reflects the "average" number of sites based on past study experience. The upper figure of 100 sites reflects situations where multiple site information may be easy to obtain at low cost, such as in the evaluation of the 65 mi/h (104 km/h) speed limit.

Treatment effect

Three values were studied for varying treatment effects at both the migration and treatment sites. These values were no effect, 10-percent reduction, 30-percent reduction and 50-percent reduction for treatment sites and none, 10-percent increase, 30-percent increase, and 50-percent increase at the migration sites. These parameters were paired as follows:

0-percent at treated sites and 0-percent at migration sites - The purpose of this combination is to test the "false-positive" error rate of the statistical method for evaluating a safety treatment's effects. That is, in actuality there is no effect due to treatment at either the treated or migration sites and any effect is purely due to random noise or regression-to-the-mean. An effective statistical method should have a small false-positive error rate if it is to be proposed for assessing migration and treatment effects. Of course, this will depend upon the variability in the data. Thus, it is anticipated that upper bounds on the variability tolerance will result from this simulation scenario.

10-percent reduction at treated sites and 10-percent increase at migration sites - This combination determines the sensitivity of the statistical method to identify minimal treatment effectiveness at the treated sites with a parallel minimal harmful effect on migration sites. It was anticipated that this would probably be the most difficult effect to measure and would be very dependent on the number of sites and magnitude of the accident rates. Again, bounds on these parameters were determined so that statements like - "It will not be possible to detect minimal treatment effects if accident rates are (below some threshold)" or "there are fewer than (?) sites available for study" could be determined.

30-percent reduction at treated sites and 30-percent increase at migration sites - This reflects a moderate effect at both sites.

50-percent reduction at treated sites and 50-percent increase at migration sites - This was a critical combination as any statistical method which cannot detect this amount of effect at reasonable levels of accident frequencies and numbers of sites would not be of practical value for any safety evaluation study. This simulation was conducted first.

Correlation

The degree of correlation in accident rates among treatment and migration pairs of sites would provide a measure of the degree to which migration was systematically occurring in the system. No correlation coupled with a large shift in mean rates at both sites would signify a global effect i.e., an effect that was not occurring systematically on a site by site basis. On the other hand, a strong correlation both before and after coupled with a large shift in mean rates would indicate the strongest possible migration effect--systematically at all pairs of sites. This is explained in further detail in the statistical methods section. The correlation parameters for this study were .10, .80, and .95 (for r^2's of .01, .64, and .90, respectively).

Accident rates, counts, and volume

Four accident rates were selected to represent a range of roadway accident rates. Accident analyses conducted on Texas accident records supplied the accident and volume parameters for this study. The combinations of accident count, traffic volume, and accident rates selected for this study represented typical highway roadways where countermeasure evaluations would likely be implemented. From a statistical standpoint, it was imperative that the entire range of feasible values of accident counts, rates, and traffic volumes on these roadways be covered in the simulation process. Based on the analysis of 1989 Texas accident records, various accident rates and VMT's were computed for various roadway sections. The data are listed in table 8. From this list, parameters were selected for the

Table 8. Accident data for 1989 Texas roadways.

Road Class	POP	VMT (100,000)	TOT ACC	RATE* (MVMT)	FI	F	Length
Interstate	≤100,000	1047	52	0.572	20.0	0.764	12.146
	≤ 25,000	1023	43	0.523	16.0	0.698	12.185
	≤ 10,000	218	21	1.121	8.0	0.259	11.205
	>100,000	2857	355	0.927	115.0	2.135	12.081
2-Lane US & State	≤10,000	116	10	1.247	4.0	0.234	11.252
	>10,000	220	21	1.700	7.0	0.1264	9.591
2-Lane	≤10,000	40	6	2.286	2.5	0.1320	9.343
	>10,000	140	19	1.761	7.0	0.1543	8.730

*Note: These are averaged rates averaged over all roadway sections of this type (control section rates).

simulation study representing the following road types:

1. Urban interstates (population > 100,000).
2. Rural interstates (population ≤ 100,000).
3. Urban two-lane US & state roads (population ≥ 10,000).
4. Rural two-lane county roads (population ≤ 10,000).

From table 8, these four groups represent rates ranging from 2.286 to 0.572 accidents per million vehicle miles traveled (per 1.6 million km) and annual accident frequencies from 355 to 6. This range should encompass the ranges of most roadways that might be considered in a safety measure evaluation study. Although these numbers were based on the four roadway types listed, they also likely represent other types. For example, the rural two-lane county road parameters may also reflect low volume intersections, etc. The intent in this selection was not to specifically analyze all possible combinations but to select a representative range of parameters.

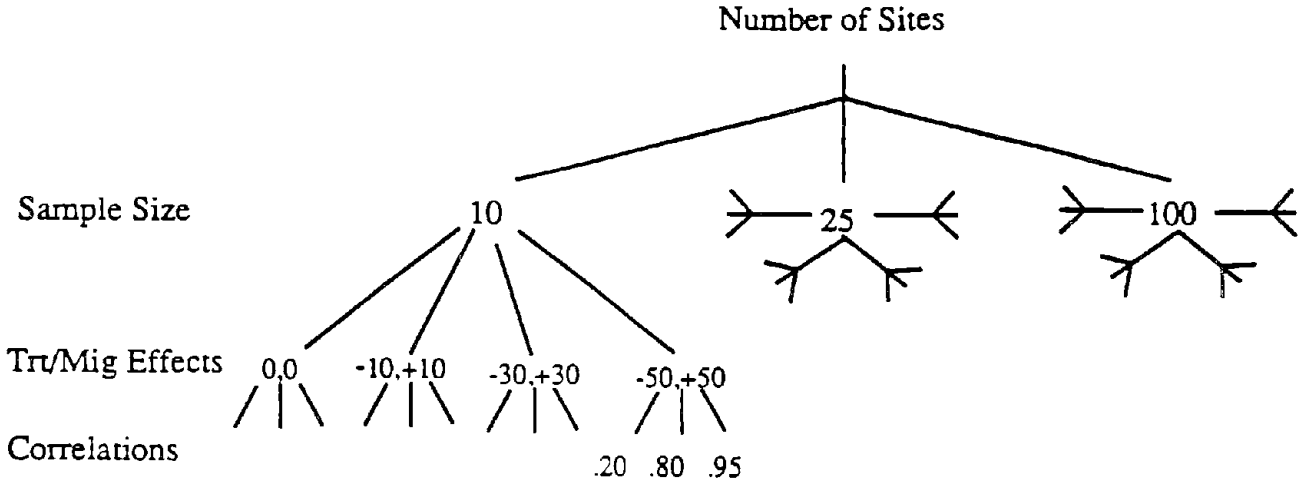

Figure 3. Stage 1 simulation study schematic.

A schematic showing these simulation parameters is shown in figure 3. Five replicate runs were made for each combination for a total of 720 computer runs (3 correlations x 3 sample sizes x 4 trt/mig effects x 4 road types x 5 replicates).

Simulation methodology

Data for before and after treatment at treatment and migration sites reflecting the parameter ranges just described were generated as follows.

The data generated for the before period for i=1, ..., n (10, 25, or 100) treatment sites consisted of:

1. True treatment site mean, λ_{TBi}, where λ_{TBi} comes from a gamma distribution whose parameters are defined by the four rate/traffic volume combinations as noted in the previous section.
2. Site exposure values, e_i, (traffic volumes) assuming these follow a negative exponential distribution whose parameters are defined by the four traffic volumes.
3. Accident counts, z_{Bi}, which are assumed to come from a Poisson distribution with mean $\lambda_{TBi}e_i$ from the parameter selections of 1 and 2.

This same sampling procedure is followed for the corresponding n migration sites in such a way that the migration and treatment true accident rates (λ_{MBi} and λ_{TBi}) are correlated using .95, .80 or 20.

For the after period, the sampling process is modified as follows. For the treatment sites, it is assumed that the overall mean rate of all the sites (i.e., λ_{TB}, the mean of the gamma distribution from which the λ_{TBi}'s were drawn), has shifted by the amount of the assumed treatment effect, i.e., one of the following

$$\lambda_{TA} = .5\lambda_{TB},\ .7\lambda_{TB},\ .9\lambda_{TB} \text{ or } \lambda_{TB} \tag{21}$$

where λ_{TA} is the rate after treatment and λ_{TB} is the rate before treatment. This is equivalent to assuming that the treatment site means have been changed by 50 percent, 70 percent, 90 percent or 100 percent from what they were before treatment. These correspond to a 50 percent, 30 percent, 10 percent and 0 percent reduction due to the treatment.

Similarly, for the migration sites the true mean is shifted by the corresponding quantities to produce a 50 percent, 30 percent, 10 percent and 0 percent increase in mean accident rate in the after treatment period.

$$\lambda_{MA} = 1.5\lambda_{MB},\ 1.3\lambda_{MB},\ 1.1\lambda_{MB} \text{ or } \lambda_{MB} \tag{22}$$

The traffic volumes are held constant for the before to after period for all stages of the study. This allows for the evaluation of the effect of the treatment and migration apart from any effect on the traffic volumes. It is important in assessing migration that changes in accidents be assessed separately from changes in volume and we do this in this phase of the simulation study by holding the traffic volumes constant. This insures that at 50-percent reduction in accident rate corresponds to a 50-percent reduction in accident count since the denominator for rate (traffic volume) is unchanged. In practice, traffic volumes should be tested separately for significant changes over time as a change in traffic volume is not, definitionally accident migration but rather traffic migration. Figure 4 provides a graphical depiction of this simulation process.

Tables 9-11 list the relative deviations of the estimated effect from the true. The relative deviations, not absolute deviations, are the relevant criteria to measure since the measurement of interest, treatment effectiveness, is a ratio estimate. Relative deviation is defined as

$$1 - \frac{\hat{\theta}}{\theta} \tag{23}$$

where $\hat{\theta}$ is the estimated treatment effect and θ is the true effect. For example n=100 and z=355 in table 9, 50-percent true reduction, the estimated reduction was 49.9 percent so the estimated $\hat{\theta}$ was 50.1 and the true effect 50.0 yielding a relative deviation of:

$$1 - \frac{50.1}{50} = -.002 \tag{24}$$

That is, the estimate is within 0.2-percent of the true value. Since this number is negative, it is an overestimate of the true θ. Positive numbers represent underestimates. For the 50-percent increase at migration sites for this same set of conditions, the relative percent deviation is:

$$1 - \frac{153.8}{150.0} = -.025 \tag{25}$$

or within 2.5-percent of the true θ.

The columns of tables 9-11 represent the sample size (n), the accident frequency for each of the four accident rate parameter choices (z), and the true treatment effects used in the simulation for both the treated and migration sites, e.g. 50-percent reduction of the treated sites (-50 TRT), 50-percent increase at migration sites (+50 MIG), etc. The entries in the table are deviations of the estimated treatment effects in percent from the true effects as listed in the column headings.

The number of sites, n, are listed in decreasing order, hence, one might expect the deviations to increase going down the table, i.e., the estimates, based on 100 sites, would logically do a better job of estimating the true effect than the ones based on 10 sites. The z's, accident frequencies, are ordered from highest to lowest and similarly, within a particular n grouping, one would expect to do a better job of predicting the true effect at sites which had more accidents. That is, there would be a better chance of seeing treatment effect at sites which had 355 accidents per year than ones that only had 6 per year before treatment.

Finally, treatment effects are ordered across the table from highest to lowest. Here one would expect a statistically significant treatment effect to be found where the amount of reduction or increase is greatest. Hence, progressing from left to right across these tables one might expect to see more statistical non-significance on the right side of the table. Statistically *significant* effects are denoted by *, i.e. these are effects which are statistically different from zero.

Table 9 contains simulation results for the high correlation situations. The following points can be noted:

1. The treatment and system effects are estimated within $\pm$ 10 percent except for the case where n = 10.

2. Most treatment and system effects are not statistically significant when there was no real effect except (0 percent change) for n = 25 and z = 6.

3. Even a small (10-percent) change is detectable and statistically significant except at n = 10 for both treatment and migration sites.

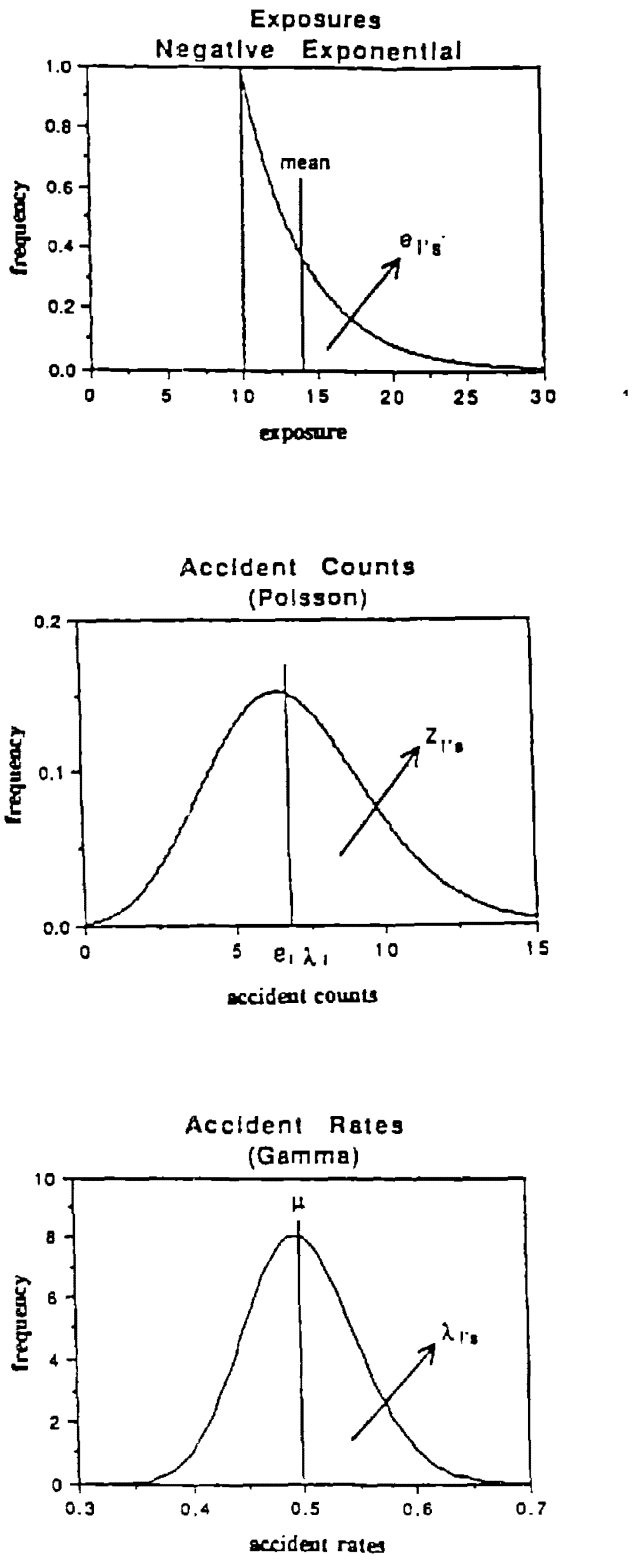

Figure 4. Graphical depiction of simulation process.

Table 9. Stage 1 relative percent deviations of estimates: r = .95.
(* indicates statistical significance)

n	z	-50 TRT	+50 MIG	-30 TRT	+30 MIG	-10 TRT	+10 MIG	0 TRT	0 MIG
100	355	-0.2*	-2.5*	-1.3*	0.2*	0.0*	0.6*	0.5	0.6
	52	-1.6*	-3.0*	-0.7*	-0.3*	-1.1*	-3.0*	-2.1	3.7*
	21	-0.0*	1.4*	-1.3*	-3.9*	-0.2*	1.3*	2.2	3.0
	6	7.0*	3.2*	10.9*	3.1*	-4.6*	-2.4*	-3.3	2.3
25	355	0.4*	-0.2*	0.4*	-1.0*	-0.5*	1.9*	2.6	1.2
	52	-3.2*	-0.1*	0.3*	-2.7*	1.0*	1.3*	1.2	5.7
	21	0.2*	3.7*	-1.3*	-2.9*	-4.6*	-2.9	1.5	3.9
	6	-3.6*	0.5*	10.1*	10.1*	-2.7*	-3.7	-10.6*	9.8*
10	355	1.6*	-3.5*	-3.0*	-7.9*	-0.6*	1.2*	15.6*	24.2*
	52	4.0*	-0.4*	-1.1*	5.0*	-15.6*	-14.0*	18.3*	17.3*
	21	5.8*	-1.9*	18.7*	-20.0*	-17.1*	-9.8*	16.5*	29.3*
	6	-9.0*	3.4*	22.6*	-16.3*	-1.0*	-5.6*	17.4*	19.4*

Table 10 reflects results when there is moderate correlation between the migration and treatment sites before treatment (r^2 = .64). The following points are noted:

1. Again, as with the high correlation case, the true treatment effect at the treated sites and migration sites is estimated within $\pm$ 10 percent and even closer for the 50-percent reductions except at low accident frequency sites (for migration sites n = 25 and z = 6) and where only 10 sites are used.

2. The treatment and system effects are generally statistically non-significant for the no effect case but there is an exception for n = 25 and z = 6.

3. The treatment and system effects are generally deemed significant when there was only a 10-percent change.

Table 10. Stage 1 relative percent deviations of estimates: r = .80.
(* indicates statistical significance)

n	z	-50 TRT	+50 MIG	-30 TRT	+30 MIG	-10 TRT	+10 MIG	0 TRT	0 MIG
100	355	-0.4*	-0.1*	1.0*	-2.1*	0.1*	-0.2*	-0.3	0.7
	52	0.2*	1.2*	-0.7*	-1.1*	2.7*	3.9*	-1.8	2.3
	21	1.0*	-1.7*	-0.4*	4.0*	0.2*	-2.5*	2.9	3.3
	6	3.4*	3.6*	13.7*	3.5*	0.1*	-1.3*	1.4	-3.3
25	355	1.2*	-4.3*	4.6*	5.8*	0.2*	1.9	3.5	1.3
	52	-0.4*	-0.1*	12.9*	-2.6*	2.3*	1.9	2.0	1.2
	21	-4.0*	5.5*	-0.1*	3.3*	-1.3*	-7.5*	-5.5	-1.9
	6	-0.2*	6.3*	0.9*	12.4*	4.0*	-9.5*	22.5*	9.8*
10	355	-2.6*	9.5*	-2.7*	0.6*	3.7	-2.0	-2.2	12.1*
	52	-10.6*	4.9*	-4.1*	-2.8*	-8.4*	-4.7	3.8	1.8
	21	-24.4*	80.5*	1.0*	17.5*	-7.0*	14.5*	-5.2	-19.8*
	6	0.0	53.5*	25.7*	2.9*	-16.2*	43.6*	0.8	-8.0

Table 11 reflects results when there is little or no correlation between accident rates for the treatment and migration sites before treatment. The following points are noted:

1. Both the true treatment effect at treated sites and at migration sites is estimated within $\pm$ 10 percent (even closer for 50-percent reductions) except at low accident frequency sites (z = 6) and low sample site (n = 10).

2. Most of the treatment and system effects are statistically non-significant for the no change cases.

TABLE 11. Stage 1 relative percent deviations of estimates: r = .20. (* indicates statistical significance)

n	z	-50 TRT	+50 MIG	-30 TRT	+30 MIG	-10 TRT	+10 MIG	0 TRT	0 MIG
100	355	0.6*	-1.7*	-2.7*	-1.2*	3.5*	-0.8*	-1.9	0.6
	52	-1.4*	0.5*	-3.4*	-4.8*	0.6*	-0.9*	0.6	-1.4
	21	3.6*	2.9*	4.4*	1.7*	-1.3*	2.0*	-2.4	-1.1
	6	7.2*	4.9*	-1.4*	3.7*	2.4*	-5.4	-2.3	2.4
25	355	4.6*	0.9*	-0.4*	2.0*	-0.1*	-0.7*	-2.3	1.0
	52	3.4*	2.5*	6.8*	5.5*	3.6*	4.3*	7.2	0.3
	21	-2.0*	4.8*	9.1*	4.0*	-2.7*	-0.5*	-7.1	0.7
	6	-5.8*	4.5*	-2.6*	4.9*	0.4*	2.3*	-8.4*	-2.0
10	355	-5.0*	2.7*	3.6*	-8.4*	2.4	-5.1	3.3	5.5
	52	-11.0*	0.1*	2.7*	-6.0*	-12.1*	2.5*	3.5	-5.2
	21	-7.4*	30.6*	15.1*	-6.7*	-23.0*	5.3*	1.0	-12.1*
	6	-42.60*	30.9*	31.7*	-9.5*	-41.6*	5.8*	-1.9	2.6

An example of more details from the simulations is presented in table 12. The rest of the detailed tables can be obtained from the FHWA. Table 12 corresponds to the correlation condition of .95 and n = 100. The variance of the 5 replicate effects is reported in parenthesis. Thus, .5011 was the average treatment effect $\hat{\theta}_T$ of the 5 replicates which had a variability of .0002. The treatment effect estimate of .5011 corresponds to -49.89 percent change due to treatment, $(\hat{\theta}_T - 1)\ 100$. A 95-percent confidence interval about this estimate would be:

$$.5011 \pm 1.96 \sqrt{(.0002)/4} =$$

$$.5011 \pm .0139 = (.4872, .5150)$$

In terms of the percent reduction, these limits would be 51.3-percent and 48.5-percent reductions (.4872-1) and (.5150-1) times 100, respectively. This means we can expect our estimate of the true treatment effect of a 50-percent reduction to be between 51.3-percent and 48.5-percent reductions with 95-percent confidence.

The last two columns of table 12 represents the number of replicates that were statistically significant for testing the following hypothesis:

$(T/M)_1$ Is the effect significantly different from zero?

$(T/M)_2$ Is the effect significantly different from the true effect?

The ratios T/M represent the number of the five replicates that were statistically significant for the treatment sites/migration sites. Thus, for the 50-percent treatment/system true effect and z = 355 from table 10, all 5 replicate estimates were significantly different from zero for both the treatment and migration sites, (T/M) = 5/5. This means that in all cases a 50-percent reduction at treatment sites and a 50-percent increase at migration sites was statistically detectable. Similarly, in all but 2 of the treated site replicates, the estimate was not significantly different from the true and for the migration sites only 1 was not significantly different from the true--4 were significantly different $(T/M)_2$ = 2/4. Whereas, ideally we would like this ratio to be 0/0 we must realize that with such large n and small variability, statistical significance is going to be fairly easy to get but should not be confused with practical significance. For example, if we examine the five replicates that gave rise to the 4 significant migration sites it can be noted that a difference of 4.07 percent ($\hat{\theta}_m$ = 1.5407 and the true is 1.50) is deemed statistically significant with an n of 100 sites. Column $(T/M)_1$ is the more important column because very small numbers in this column mean that the treatment or system effect is not statistically detectable.

TABLE 12. Stage 1 simulation results: r = .95, n = 100.

Label	$\bar{z}_T$	TRT%	VAR	Mig %	VAR	$(T/M)_1$	$(T/M)_2$
.5 P1.1	355	-49.89	(.0002)	53.79	(.0052)	5/5	2/4
P2.1	52	-49.16	(.0005)	54.47	(.0127)	5/5	3/2
P3.1	21	-49.91	(.0015)	47.85	(.0306)	5/5	2/3
P4.1	6	-53.54	(.0021)	45.19	(.0319)	5/5	1/3
.7 P1.1	355	-29.05	(.0004)	29.75	(.0036)	5/5	2/4
P2.1	52	-30.47	(.0018)	30.31	(.0113)	5/5	2/3
P3.1	21	-30.87	(.0004)	35.08	(.0176)	5/5	0/2
P4.1	6	-22.43	(.0292)	25.96	(.0136)	4/5	3/3
.9 P1.1	355	-10.02	(.0005)	9.44	(.0011)	5/2	3/3
P2.1	52	-11.04	(.0029)	13.26	(.0047)	5/5	3/3
P3.1	21	-9.84	(.0053)	8.64	(.0042)	4/2	2/1
P4.1	6	-14.06	(.0108)	12.60	(.0065)	4/4	3/2
1.0 P1.1	355	0.49	(.0028)	0.61	(.0067)	5/5	5/5
P2.1	52	-2.14	(.0040)	3.68	(.0021)	3/2	3/2
P3.1	21	2.19	(.0046)	3.02	(.0146)	1/2	1/2
P4.1	6	-3.32	(.0070)	2.23	(.0074)	1/3	1/3

General conclusions for stage 1

1. Treatment effect at both treated and migration sites is generally detectable and accurate within $\pm$ 10 percent even for the minimal effect of 10 percent for n = 25 or 100.

2. Treatment effect cannot be accurately detected for 10 or fewer sites under any conditions and fluctuates dramatically. There is a tendency to find a spurious treatment and migration effect when such few sites are used.

3. These results are consistent across all correlation combinations.

Thus, it can be concluded from this phase that system effects due to treatment can be adequately assessed under these ideal conditions and that at least 25 sites will be necessary to assess it. The same is true for evaluating treatment effect at the treated sites.

SIMULATION PROCEDURES AND PARAMETERS--STAGE 2

The objective of the second phase of the accident system simulation study was to determine the effect of r-t-m on the assessment of accident migration. From stage 1 it was concluded that under ideal conditions of no r-t-m problem among migration sites, safety treatment effects could successfully be identified at both the treated and migration sites if at least 25 sites were treated and 25 sites identified as potential migration sites. This will be further discussed in the next chapter. Since it was not possible to adequately assess treatment effects at either treated or migration sites for only 10 sites, the n=10 situation was dropped from consideration in the second phase of this study. Also, since there was little difference in the r=.80 (moderate correlation) and r=.95 (high correlation) cases, only the two ranges in correlation, high and low, were investigated for this phase. Three r-t-m potential scenarios were examined. Treated sites were ordered so that the n sites with the highest number of accidents were selected for treatment and the remaining sites were assigned to the reference group. The reference and treatment group together was set at 10 times the number of treated sites (1000 and 250, respectively, for the n=100 and n=25 cases). For the migration sites, the following three conditions were examined:

1. No r-t-m problem among migration sites.

2. Severe r-t-m at the migration sites.

3. Reverse r-t-m at the migration sites.

For the no r-t-m situation, the n migration sites were selected from the middle of the distribution of accident counts, i.e. the 1000 samples of accident counts from the Poisson distribution were ordered and the 50 counts above the mean and 50 counts below the mean were selected as representing the accident counts at migration sites before treatment and the other 900 formed the reference group. For severe r-t-m, the 100 highest accident counts were selected as representing the migration sites and for reverse r-t-m the lowest 100 were selected. Severe r-t-m corresponded to situations where the potential migration sites were also high hazard locations, perhaps because of their proximity to the treated sites, but were not candidates for the treatment. This might occur, for example, if the treatment were adding left-turn lanes at intersections which tended to have a high number of rear-end collisions and the migration sites were sites close to the treated sites which also had a high number of rear-end collisions but were not treated because there was only a limited amount of funding in the project. Reverse r-t-m could occur if the migration sites had an unusually low accident count. This could occur if both the treatment and migration sites come from the same population and, given that the treated sites were selected from that population because they had high accidents, the sites remaining in the population had low accident counts. As an example, consider the treatment of signalization of intersections with non-signalized intersections serving as the potential migration group. The very fact that these intersections are not signalized may be because they don't have enough accidents to warrant a signal. Exposures were simulated as in stage 1, i.e. there was no change in exposures from the before and after period.

The total number of computer runs for this study phase was 1920 (2 correlations times 2 groups (treated and migration) times 3 r-t-m situations times 2 sample sizes time 4 road types times 4 trt/mig effects times 5 replicates). For each of these 1920 data sets, the BEATS (Bayesian Estimation of Accidents in Transportation Studies) program was executed which was a massive computational undertaking, especially considering the total sample sizes of 1000 or 250 in each case.[15] The following is an attempt to summarize these results. Figure 5 represents the schematic for the stage 2 simulations.

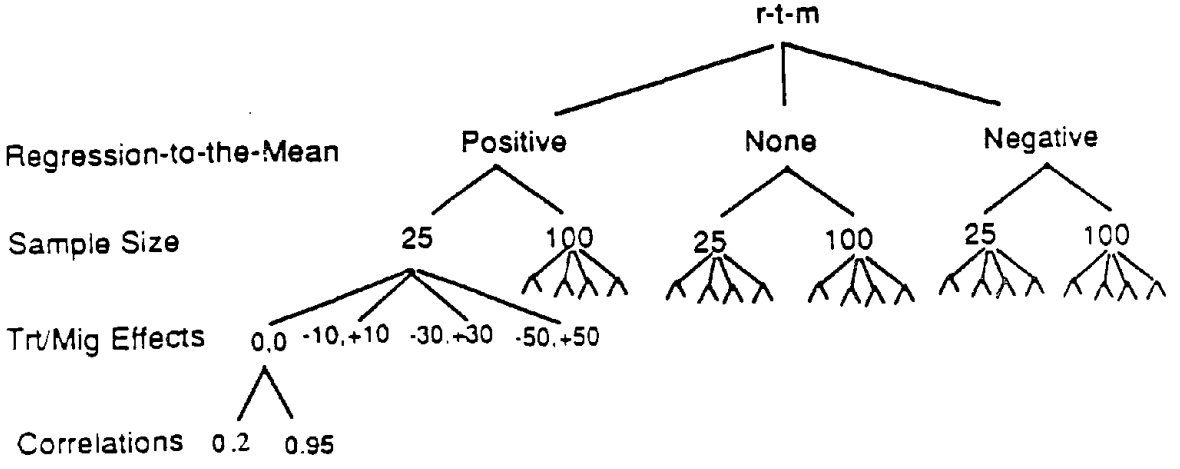

Figure 5. Stage 2 simulation study schematic.

The data were generated similarly to the procedure for stage 1. In addition to drawing reference group data for the treatment group, reference group data was also drawn for the migration group to provide the necessary data for the EBEST methodology. The same four treatment/migration effects were used, namely 50 percent, 30 percent, 10 percent, and 0 percent reductions at the treated sites and 50-percent, 30 percent, 10 percent, and 0 percent increases in accidents at the migration sites. No reference group data was drawn or needed for the after period. Tables 13-18 summarize the deviations of effectiveness for all the conditions. More details of the simulation studies are available from the FHWA. Table 19 is an example of the more detailed tables. These tables contain the average of the five replicate treatment effects estimates in percent change, the variance of the five replicate estimates (VAR), the average shrinkage for the five estimates, β_i, and the resulting significance test results. The significance tests results are labeled as:

$$(n_1/n_2)(n_3/n_4)$$

where n_1 is the number of replicates where the frequentist estimates were significantly different from zero, n_2 is the number of replicates where the EBEST estimates were significantly different from zero, n_3 is the number of replicates where the frequentist estimate was different from the true value, and n_4 is the number of replicates where the EBEST estimate was differed from the true.

As in stage 1, the interpretation of n_3 and n_4 is of little practical importance since the large accident frequencies caused nearly all of these tests to be "statistically" significant. The interpretation of n_1 and n_2 is of more practical importance as this indicates when a treatment effect would be deemed statistically significant. These results show that in nearly all cases, the treatment is found to be significant even when there was no true effect. This is purely an artifact of the r-t-m phenomenon. A rather surprising and puzzling result emerges from these estimates, namely, that the r-t-m problem is not satisfactorily alleviated by the EBEST methodology. Although the EBEST estimate is always closer to the true value than the frequentist estimate, the improvement is marginal. Based on all the simulation the average shrinkage value was around .10 with a maximum value of .37. This means that on the average 90 percent of our observed count is used and only 10 percent of our estimate is based on the combined treatment-reference group information.

The amount of adjustment is dependent primarily on the exposure information--especially on the amount of variability among the exposures. The exposures for this data were assumed to follow a negative exponential distribution and the mean exposure values were selected from typical highway type VMT's. The negative exponential has been a proposed distribution based on past literature, however, it is not known if this assumption has ever been tested or validated using actual data. The following are possible reasons for the minuscule improvement in the EBEST procedure:

1. The distributional assumption on the VMT's is not valid yielding simulated VMT's that are unrealistically highly variable.

2. The real world situation is such that there really is this much variability in VMT's and, if so, no statistical procedure exists for adjusting for this degree of r-t-m given the amount of varibility in exposures.

3. The amount of r-t-m introduced in these scenarios is unrealistic and lesser degrees of r-t-m might show more improvement. Recall we are assuming that out of 1000 sites the top 100 are selected for treatment, which is pretty severe r-t-m.

The original objective of this study was not the evaluation of the EBEST methodology. However, if r-t-m cannot be satisfactorily adjusted for in evaluating treatment effects at treated sites, it cannot be expected to be adjusted for at migration sites. In fact, in the no r-t-m cases for migration sites, system effects are satisfactorily identified in tables 13 and 14. Therefore, if one can be sure that the migration sites are not subject to any r-t-m bias, one can be equally assured that with 25 or more sites, treatment effects at the migration sites are detectable. The exceptions to this broad statement is with the lower accident count sections, especially 2-lane rural roads, where, as in stage 1 it was found that the estimates were not within the arbitrary 10-percent criteria for estimating the effect most of the time. There are a few other minor exceptions, however, if the criteria is expanded to being within 15-percent of the true estimate, most of these exceptions disappear. This can be explained by examining the variability in the estimates based on the detailed tables. These exceptions are within the expected variability.

For the high r-t-m and reverse r-t-m conditions, the migration assessment attempts are futile! Note that in tables 15 and 16 corresponding to high r-t-m we almost consistently conclude there is no effect or even a significant **reduction** in accidents at the migration site when there was actually a 30-percent or even 50-percent **increase.** And when there is little or no change in accidents, we estimate that there is a significant reduction in accidents at the migration sites due to the treatment. This conclusion is logical since high r-t-m at the migration sites indicates that we selected high accident locations before treatment and would expect to see a subsequent reduction in the after period regardless of treatment. However, given that we know there really was a significant increase and finding our estimate is so far from the true leaves little doubt that attempting to assess migration in these cases in not only foolish but dangerous and could lead to a conclusion of unwarranted safety at migration sites which were actually made more dangerous by the treatment.

The reverse r-t-m situation is equally bad in the other direction. Tables 17 and 18 summarize these results and what we see is that we uniformly estimate a tremendous increase in accidents at the migration sites due to the treatment and many of these increases are ridiculously large, most exceeding 100 percent! Thus, if the migration sites happen to have extremely low accidents, one can almost be completely assured that one could find a detrimental effect due to the treatment at these sites even when there was actually no change.

Table 13. Stage 2 deviations of estimates: r = .95, no r-t-m at migration sites. (* indicates statistical significance)

n	z	-50 TRT_F	+50 MIG_F	-50 TRT_E	+50 MIG_E	-30 TRT_F	+30 MIG_F	-30 TRT_E	+30 MIG_E
100	355	9.3*	2.4*	9.0*	2.4*	10.1*	4.0*	9.7*	4.0*
	52	17.1*	10.9*	15.7*	10.6*	22.9*	5.9*	21.0*	5.8*
	21	20.9*	12.8*	18.8*	11.4*	27.7*	15.9*	24.7*	14.6*
	6	25.3*	22.4*	23.2*	19.7*	33.5*	22.3*	30.7*	25.5*
25	355	8.1*	3.4*	7.9*	3.5*	6.9*	2.9*	6.6*	3.0*
	52	15.5*	2.2*	14.1*	2.5*	25.2*	1.1*	23.4*	0.9*
	21	19.1*	4.8*	17.0*	3.4*	28.0*	6.2*	25.1*	5.3*
	6	26.9*	38.4*	25.1*	34.0*	33.4*	28.5*	30.4*	25.3*
n	z	-10 TRT_F	+10 MIG_F	-10 TRT_E	+10 MIG_E	0 TRT_F	0 MIG_F	0 TRT_E	0 MIG_E
100	355	11.8*	1.3*	11.3*	1.4*	-15.9*	0.1*	-15.4*	0.0
	52	30.8*	3.3*	28.4*	3.3*	-34.3*	1.1*	-31.6*	1.1
	21	39.0*	16.2*	35.3*	15.3*	-43.4*	9.4*	-39.2*	8.4*
	6	42.9*	14.3*	39.3*	11.7*	-51.8*	22.4*	-48.0*	19.8*
25	355	11.5*	3.6*	11.0*	3.5*	-16.6*	-3.2*	-16.0*	-3.2*
	52	31.3*	1.7*	28.8*	2.0*	-39.9*	0.1	-37.5*	-0.3
	21	39.1*	11.0*	35.6*	9.8*	-42.7*	11.9*	-38.4*	10.7*
	6	43.0*	10.1*	39.2*	12.0*	-51.2*	16.3*	-47.4*	14.1*

Table 14. Stage 2 deviations of estimates: r = .20, no r-t-m at migration sites.

(* indicates statistical significance)

n	z	-50 TRT_F	+50 MIG_F	-50 TRT_E	+50 MIG_E	-30 TRT_F	+30 MIG_F	-30 TRT_E	+30 MIG_E
100	355	6.7*	-0.7*	6.4*	-0.6*	9.3*	0.9*	8.9*	0.9*
	52	15.8*	1.0*	14.3*	0.8*	22.6*	-1.1*	20.9*	-1.5*
	21	20.1*	4.1*	17.8*	3.3*	26.0*	15.2*	26.0*	14.3*
	6	25.2*	18.3*	27.0*	11.2*	35.2*	20.7*	29.3*	15.9*
25	355	10.4*	2.6*	10.3*	2.7*	10.4*	0.7*	10.0*	0.8*
	52	14.5*	14.0*	-0.2*	13.1*	21.7*	3.5*	19.8*	2.8*
	21	20.4*	10.0*	2.4*	8.0*	26.8*	11.9*	22.3*	10.7*
	6	22.1*	18.4*	9.3*	11.2*	35.2*	20.7*	29.3*	15.9*
n	**z**	**-10 TRT_F**	**+10 MIG_F**	**-10 TRT_E**	**+10 MIG_E**	**0 TRT_F**	**0 MIG_F**	**0 TRT_E**	**0 MIG_E**
100	355	11.7*	1.9*	11.1*	1.9*	-15.2*	0.2	-14.6*	0.3
	52	30.3*	7.9*	27.8*	7.6*	-32.9*	-0.3	-30.0*	-0.5
	21	34.8*	7.5*	30.6*	7.2*	-40.3*	0.9	-35.9*	0.9
	6	43.2*	19.0*	35.2*	13.6*	-47.7*	1.7*	-39.0*	1.2
25	355	12.6*	0.2*	11.0*	0.3*	-15.5*	1.4	-14.9*	1.4
	52	25.5*	-1.5*	22.8*	-1.4*	-31.6*	3.3	-28.9*	3.0
	21	31.6*	4.1*	27.2*	3.7*	-34.2*	3.2*	-29.4*	2.4
	6	43.2*	18.9*	35.2*	13.6*	-47.7*	16.6*	-39.0*	12.2*

Table 15. Stage 2 deviations of estimates: r = .95, positive r-t-m at migration sites. (* indicates statistical significance)

n	z	-50 TRT_F	+50 MIG_F	-50 TRT_E	+50 MIG_E	-30 TRT_F	+30 MIG_F	-30 TRT_E	+30 MIG_E
100	355	9.0*	-22.9*	8.7*	-22.0*	40.1*	-18.3*	9.7*	-17.6*
	52	17.1*	-45.5*	15.8*	-41.2*	10.7*	-43.4*	21.0*	-39.7*
	21	20.0*	-60.7*	17.7*	-54.4*	28.0*	-52.4*	9.1*	-46.6*
	6	25.0*	-77.6*	23.1*	-72.1*	33.9*	-66.4*	31.0*	-61.4*
25	355	8.5*	-24.8*	8.2*	-23.9*	6.7*	-21.5*	6.4*	-20.8*
	52	15.0*	-49.7*	13.2*	-45.8*	24.8*	-42.2*	23.0*	-42.2*
	21	18.1*	-64.4*	15.8*	-58.4*	30.9*	-54.9*	20.9*	-49.5*
	6	26.5*	-78.7*	24.7*	-69.8*	35.5*	-58.7*	32.7*	-53.4*
n	z	-10 TRT_F	+10 MIG_F	-10 TRT_E	+10 MIG_E	0 TRT_F	0 MIG_F	0 TRT_E	0 MIG_E
100	355	11.6*	-16.2*	11.0*	-15.5*	-14.5*	-13.9*	-13.9*	-13.9*
	52	30.7*	-38.6*	28.2*	-35.8*	-34.1*	-32.1*	-31.5*	-32.1*
	21	38.9*	-43.1*	35.2*	-50.9*	-44.1*	-36.9*	-39.9*	-36.9*
	6	44.2*	-60.6*	40.6*	-59.2*	-51.3*	-49.2*	-47.4*	-45.2*
25	355	11.9*	-14.7*	-31.4*	-14.0*	-17.1*	-19.4*	-16.5*	-18.9*
	52	31.2*	-36.7	-48.7*	-33.7*	-41.0*	-38.6*	-37.8*	-32.7*
	21	41.0*	-45.8*	-57.5*	-41.4*	-43.6*	-40.0*	-39.5*	-35.6*
	6	46.6*	-56.2*	-61.4*	-51.8*	-47.6*	-48.2*	-43.4*	-47.9*

Table 16. Stage 2 deviations of estimates: r = .20, positive r-t-m at migration sites. (* indicates statistical significance)

n	z	-50 TRT_F	+50 MIG_F	-50 TRT_E	+50 MIG_E	-30 TRT_F	+30 MIG_F	-30 TRT_E	+30 MIG_E
100	355	6.5*	-22.8*	-6.2*	-21.9*	9.2*	-19.7*	-68.8*	-29.9*
	52	16.4*	-52.1*	14.9*	-52.3*	23.1*	-45.0*	-81.2*	-41.4*
	21	20.0*	-60.4*	17.8*	-53.7*	25.6*	-44.9*	-82.4*	-38.7*
	6	24.8*	-71.8*	20.5*	-58.3*	35.4*	-61.2*	-89.6*	-50.2*
25	355	10.4*	-18.0*	10.0*	-17.2*	10.1*	-15.4*	10.0*	-14.9*
	52	14.3*	-42.8*	12.8*	-38.5*	21.0*	-43.6*	19.1*	-40.2*
	21	20.0*	-56.7*	17.6*	-49.1*	28.0*	-49.9*	-25.0*	-43.6*
	6	23.0*	-78.3*	18.5*	-55.4*	43.5*	-66.5*	-27.4*	-55.9*
n	z	-10 TRT_F	+10 MIG_F	-10 TRT_E	+10 MIG_E	0 TRT_F	0 MIG_F	0 TRT_E	0 MIG_E
100	355	11.7*	-17.6*	11.1*	-17.0*	-15.2*	-15.4*	-15.5*	-15.0*
	52	10.2*	-32.7*	17.7*	-29.5*	-32.8*	-34.5*	-30.0*	-31.8*
	21	15.3*	-41.2*	31.2*	-36.0*	-39.7*	-36.6*	-35.2*	-31.8*
	6	24.5*	-54.8*	16.8*	-46.0*	-48.4*	-50.3*	-40.1*	-42.1*
25	355	12.7*	-19.4*	12.2*	-18.7*	-15.0*	-15.5*	-14.4*	-15.0*
	52	27.0*	-30.6*	24.3*	-27.2*	-31.8*	-27.2*	-29.0*	-24.1*
	21	31.6*	-40.9*	27.3*	-35.8*	-35.6*	-37.9*	-31.0*	-33.2*
	6	44.8*	-55.6*	37.2*	-46.3*	-48.6*	-39.1*	-40.5*	-28.8*

Table 17. Stage 2 deviations of estimates: r = .95, negative r-t-m at migration sites. (* indicates statistical significance)

n	z	-50 TRT_F	+50 MIG_F	-50 TRT_E	+50 MIG_E	-30 TRT_F	+30 MIG_F	-30 TRT_E	+30 MIG_E
100	355	9.5*	62.7*	9.2*	57.1*	10.2*	74.7*	9.8*	68.8*
	52	17.2*	202.2*	15.8*	146.6*	22.7*	162.1*	20.4*	118.6*
	21	21.0*	356.8*	18.9*	181.8*	28.4*	279.6*	25.4*	153.9*
	6	25.1*	856.5*	23.0*	353.6*	34.5*	897.1*	31.8*	309.3
25	355	8.9*	75.2*	8.6*	69.2*	9.1*	70.3*	6.6*	65.2*
	52	15.3*	99.4*	13.9*	142.0*	25.1*	164.4*	23.3*	118.2*
	21	18.4*	349.5*	16.1*	175.4*	29.9*	257.3*	27.3*	137.3*
	6	25.9*	909.1*	24.1*	356.3*	35.0*	718.1*	32.2*	251.2*
n	z	-10 TRT_F	+10 MIG_F	-10 TRT_E	+10 MIG_E	0 TRT_F	0 MIG_F	0 TRT_E	0 MIG_E
100	355	11.5*	149.5*	10.9*	45.0*	-16.4*	41.7*	-15.8*	37.8*
	52	31.7*	143.4*	50.7*	104.8*	-34.7*	125.6*	-34.7*	90.2*
	21	38.3*	255.0*	34.7*	140.2*	-43.4*	231.6*	-39.3*	126.1*
	6	43.8*	649.4*	40.0*	138.4*	-51.3*	583.3*	-47.5*	228.6*
25	355	11.4*	51.1*	10.9*	46.7*	-17.4*	47.1*	-16.8*	42.9*
	52	30.2*	147.0*	27.6*	105.2*	-38.8*	133.6*	-36.4*	96.0*
	21	41.9*	305.0*	38.3*	164.8*	-43.4*	204.1*	-39.2*	110.8*
	6	44.6*	788.3*	40.7*	120.1*	-49.6*	661.0*	-45.6*	24.08*

Table 18. Stage 2 deviations of estimates: r = .20, negative r-t-m at migration sites. (* indicates statistical significance)

n	z	-50 TRT_F	+50 MIG_F	-50 TRT_E	+50 MIG_E	-30 TRT_F	+30 MIG_F	-30 TRT_E	+30 MIG_E
100	355	6.7*	65.3*	6.4*	59.0*	9.1*	55.4*	8.7*	50.3*
	52	16.5*	147.4*	15.1*	103.6*	23.0*	117.4*	21.1*	81.8*
	21	20.0*	248.9*	17.8*	127.7*	25.7*	269.6*	17.6*	145.9*
	6	25.0*	247.4*	20.7*	192.3*	25.1*	231.6*	29.3*	178.2*
25	355	10.4*	68.4*	10.1*	61.4*	10.4*	54.1*	10.0*	48.8*
	52	14.1*	150.0*	12.6*	105.7*	21.3*	129.5*	19.4*	92.7*
	21	10.6*	269.6*	18.3*	137.8*	28.1*	224.3*	25.0*	120.3*
	6	23.2*	511.2*	19.3*	188.6*	32.2*	891.4*	25.9*	136.7
n	z	-10 TRT_F	+10 MIG_F	-10 TRT_E	+10 MIG_E	0 TRT_F	0 MIG_F	0 TRT_E	0 MIG_E
100	355	11.5*	45.3*	11.0*	41.0*	-15.2*	42.5*	-14.5*	38.6*
	52	30.3*	122.8*	27.7*	87.5*	-32.7*	99.3*	-29.9*	70.2*
	21	33.9*	201.6*	29.7*	110.3*	-40.1*	180.8*	-35.1*	98.5*
	6	43.7*	942.8*	35.0*	140.1*	-46.5*	841.8*	-38.1*	129.4*
25	355	12.9*	30.7*	12.9*	46.2*	-15.3*	37.9*	-14.7*	34.0*
	52	23.3*	104.2*	23.6*	88.7*	-32.1*	117.2*	-23.5*	85.1*
	21	30.4*	172.1*	25.9*	99.1*	-33.1*	156.2*	-28.2*	80.4*
	6	49.0*	119.0*	36.1*	63.2*	-46.8*	892.0*	-38.2*	133.1*

Table 19. Stage 2 simulation results-treatment effect: r=.95, n=100, no r-t-m at migration sites.

Label	$\bar{z}_T$	TRT_F%	(VAR)	TRT_E%	(VAR)	$\bar{B}_{IF}$	Sig
.5MP1.1	355	-59.3	(.0002)	-59.0	(.0002)	.0303	(5/5)(5/5)
MP2.1	52	-67.1	(.0001)	-65.9	(.0001)	.1037	(5/5)(5/5)
MP3.1	21	-70.9	(.0007)	-68.8	(.0007)	.1536	(5/5)(5/5)
MP4.1	6	-75.3	(.0006)	-73.2	(.0007)	.1470	(5/5)(5/5)
.7MP1.1	355	-40.1	(.0007)	-39.7	(.0007)	.0308	(5/5)(5/5)
MP2.1	52	-52.9	(.0008)	-51.0	(.0009)	.1024	(5/5)(5/5)
MP3.1	21	-57.7	(.0009)	-54.7	(.0011)	.1514	(5/5)(5/5)
MP4.1	6	-63.5	(.0009)	-60.7	(.0010)	.1451	(5/5)(5/5)
.9MP1.1	355	-21.8	(.0007)	-21.3	(.0007)	.0354	(5/5)(5/5)
MP2.1	52	-40.8	(.0013)	-38.4	(.0013)	.1033	(5/5)(5/5)
MP3.1	21	-49.0	(.0011)	-45.3	(.0013)	.1475	(5/5)(5/5)
MP4.1	6	-52.9	(.0011)	-49.3	(.0013)	.1389	(5/5)(5/5)
1.0MP1.1	355	-15.9	(.0001)	-15.4	(.0001)	.0313	(5/5)(5/5)
MP2.1	52	-34.3	(.0010)	-31.6	(.0012)	.0993	(5/5)(5/5)
MP3.1	21	-43.4	(.0012)	-39.2	(.0012)	.1531	(5/5)(5/5)
MP4.1	6	-51.8	(.0012)	-48.0	(.0014)	.1422	(5/5)(5/5)

SIMULATION PROCEDURES AND PARAMETERS--STAGE 3

At this point, it was deemed advisable to reconsider the next stage in this study. Stage 3 was originally designed to assess the conditions under which a system effect could be detected when yet another confounding factor is introduced--namely, a time effect. What had been done up to this point in stage 2 was to assume the time effect was adequately adjusted for by a comparison group so that any treatment effects which were observed were due to treatment alone. Whereas this is still an important consideration, it now appeared that more pressing unanticipated concerns needed to be addressed in order to meet the objectives of this contract.

The primary objective in this contract was to determine under what conditions, if any, could a system effect adequately be assessed. At this point, we can only say the following:

1. It can be adequately assessed for roadways with accident rates typical of interstate highways and urban 2-lane roads if there is no r-t-m bias at the migration sites, at least 25 sites are available, and the time effect has been adjusted for by a comparison group (or there is no significant change over time.)

2. A system effect cannot be assessed if there is severe r-t-m bias in either direction, i.e. if either the most dangerous or the least dangerous sites were selected.

With regard to the migration problem we have not yet been able to say if there are some lesser r-t-m situations wherein both the treatment and migration sites could be better evaluated and, if so, how the highway safety analyst might determine a priori the degree of r-t-m for a particular study. That is, suppose we find that as long as the treatment and migration site accident rates are within a given range of the estimated mean accident rate using the EBEST procedure the treatment effect can adequately be assessed at both treatment and migration sites. Then, guidelines could be specified for the analyst to follow to determine ahead of time whether the sites meet this criteria before doing the study. Conversely, a study that has already been done using the EBEST procedure could be checked to see if it followed the guidelines and, if not, the validity of the studies conclusions could come into question.

Due to the findings in stage 2, stage 3 was modified from the original proposal. The following questions were addressed in stage 3:

1. How does the ratio of the size of the treatment sites to the reference sites affect the results?

2. What effect would result if, rather than choosing the exposures randomly, a correlation was forced between the accident count and exposure?

3. Do the exposures from the data pulled on the Texas roadways follow a negative exponential and how does the variability in the real data compare to what is simulated?

4. What effect will decreasing the variability of exposures by truncating the ranges have on the results?

5. What effect will varying the degree of r-t-m have on the results?

The following reflects the results of stage 3 for each of these questions.

Question 1. How does the ratio of the size of the treatment sites to the reference sites affect the results?

The parameters selected for this simulation were the 50% reduction, 50% increase in treatment/migration sites respectively before and after treatment for low correlation between the two groups, no r-t-m in the selection of migration sites and for rural interstates (z = 52, VMT = 1047). Again, 5 replicates were run for each simulation condition. It was felt that this choice of parameters reflected the more moderate choices unconfounded by migration r-t-m so that we could focus on the effect of the differing the ratio of treatment to reference group size on the shrinkage coefficient, b_i. In stage 2 we found that there was very little difference in the EB adjustment due to low shrinkage coefficient values. The number of sites used were: 100/200, 25/250, 25/125, and 10/20, where these numbers refer to treatment/treatment plus reference group sizes, respectively.

Table 20 depicts the descriptive data from these simulations and table 21 summarizes the data and estimates. It can be noted that reducing the size of the reference group has little if any effect on increasing the b_i's. Examining the variability of the exposures explains this. In other words, the variability in the exposures is far more influential in determining the b_i's than the degree of r-t-m as reflected by the ratio of the treatment to reference group. We do however note that both the treatment and migration effects are deemed statistically significant in all cases except when the sample size is extremely small, i.e. 10/20. Therefore, it would appear that for n=100, another 100 is a suitable size for the reference group and for n=25, 4 times the number of treatment sites (100 reference sites) is sufficient to detect both treatment and system effects.

Question 2. What effect would result if, rather than choosing the exposures randomly, a correlation was forced between accident count and exposure?

In the real world setting, we know that accident counts and exposure are positively correlated. Yet, in our sampling scheme, we randomly generate exposures, rates and counts. What would be the effect of imposing a high correlation restriction between counts and exposures in the simulation? Would this result in a more realistic distribution of exposures and hence affect the shrinkage coefficient?

A relationship was forced between the e_i's and z_i's as follows:

$$e_i = z_i \, \bar{\epsilon}/\bar{z} + \delta_i \qquad (26)$$

where δ_i is a Poisson random variable with $\lambda = 1$, $\bar{\epsilon}$ is the mean of the exponential distribution representing exposures, and $\bar{z}$ is the mean of the Poisson distribution representing the counts.

Table 20. Comparing treatment/reference sample size ratios: descriptive statistics.

		Acc. Count				Acc. Rate			
Ratio	**Time Period**	**Trt**	**Mig**	**Tref**	**Mref**	**Trt**	**Mig**	**Tref**	**Mref**
100/200	B	7082	4802	3471	5673	.0630	.0519	.0401	.0534
	A	3012	7403			.0268	.0800		
25/125	B	2293	1191	4274	5461	.0725	.0520	.0458	.0536
	A	855	1822			.0270	.0746		
25/250	B	2670	1177	10554	9476	.0762	.0493	.0492	.0424
	A	943	1932			.0269	.0808		
10/20	B	668	469	340	561	.0695	.0492	.0388	.0566
	A	287	690			.0259	.0724		

A sample of 100 e_i's were generated assuming this relationship (correlated) and compared to the e_i's simulated randomly from the simulation program. Table 22 lists the descriptive statistics along with the chi-square goodness-of-fit test statistics testing that the distribution of the resulting accident rates from these two procedures follows a Poisson distribution. That is, we want to ensure that forcing this relationship has not resulted in accident rates that violate the Poisson assumption. From table 22 the correlation between exposure and accident count increased from .563 to .916 using the relationship in (26).

The χ^2 statistics show that the resulting accident rates follow a Poisson distribution. Rejection of this hypothesis would require a χ^2 value greater than the critical value of 123.2 at the .05 level of significance. Thus, the assumed relationship between exposure and accident count increases the correlation between these variables without distorting the Poisson nature of the rates.

Table 21. Comparing treatment/reference sample size ratios on r-t-m adjustment.

		Exposure		Trt Effect (%)		Avg.
		Avg	**Var(x10^3)**	**Freq**	**EB**	**Shrinkage b_i**
100/200	Trt	1124.1	122.02	-57.4	-56.4	.145
	Mig	925.0	36.88	+37.9	+54.1	.170
	Tref	866.2	25.34			
	Mref	1061.4	129.7			
25/125	Trt	1265.0	154.8	-62.6	-61.1	.131
	Mig	916.0	28.6	53.2	53.1	.159
	Tref	932.0	44.9			
	Mref	1018.5	120.2			
25/250	Trt	1401.8	205.4	-64.5	-63.1	.118
	Mig	956.8	57.4	+64.1	+63.1	.162
	Tref	952.6	44.6			
	Mref	993.7	94.3			
10/20	Trt	1105.0	105.3	-67.3	-65.1	.161
	Mig	953.6	62.7	+44.7	+43.3	.245
	Tref	875.7	26.1			
	Mref	990.8	96.6			

Table 22. Chi-square goodness-of-fit tests on exposures.

	no correlation		**X^2**	**with correlation**		**X^2**
	z	**e**		**z**	**e**	
Mean	66	1072	0.47	81	1228	0.37
Std. dev.	25.1	321.3		59.2	485.4	
correlation	.563			.916		

The next thing to examine is how this forced correlation affected the r-t-m adjustment. Table 23 shows the results with and without correlated exposures. As the tables show, forcing this correlation does not noticeably change any of the results. In fact, it would appear for the five replications in this case that the b_i's are actually somewhat smaller, on average, and the estimates of treatment effect are worse (farther from the true value).

Table 23. Comparing treatment effects at treatment sites with and without correlated exposures.

	no correlation			with correlation		
Sample Size	Freq	EB	Avg Shrinkage	Freq	EB	Avg Shrinkage
100/200	-57.4	-56.4	.145	-61.0	-59.7	.131
25/125	-62.6	-61.1	.131	-67.2	-66.2	.100

Question 3. Do the exposures from the data pulled on the Texas roadways follow a negative exponential and how does the variability in the real data compare to what is simulated?

Figure 6 is an exponential distribution and Figures 7-10 depict the frequency distributions for the Texas data. From these frequency histograms the distributions would appear to follow a negative exponential. Chi-square goodness-of-fit tests on these data confirmed this conclusion.

Table 24 lists some comparison statistics between the assumed and actual distribution parameters. From this table, it appears that our simulated data actually had less variability among exposures than in reality. Recall that the greater the variability in exposure, the less r-t-m adjustment to be expected. Hence, had the simulated exposures had variabilities closer to the observed, the r-t-m adjustment would have been even smaller. In conclusion, then, the simulated data appears to yield a conservative estimate on the variability of the exposures as compared with real data.

Question 4. What effect will decreasing the variability of exposures by truncating the ranges have on the results?

Table 25 reflects two levels of exposure truncation, 500/1500 and 800/1200. These correspond to only selecting exposures within specified restricted ranges, i.e no exposure less than 500 or greater than 1500 was selected to represent rural interstates with mean VMT's of 1047, etc. The truncated exponential appears to have little if any effect on the shrinkage coefficients if exposures are not correlated with accidents (recall question 2), yet, there was a sizeable increase in the average shrinkage if we force the correlation between exposure and

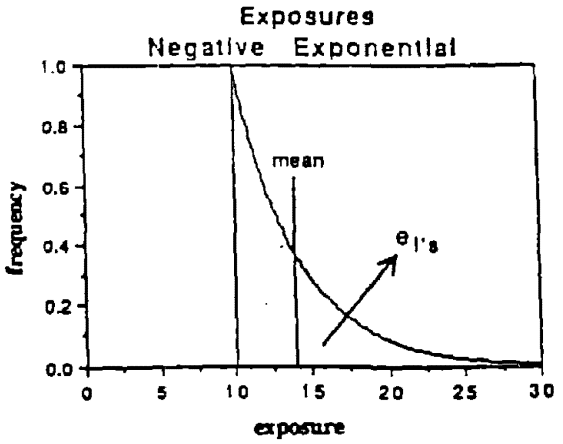

Figure 6. Theoretical Negative exponential distribution.

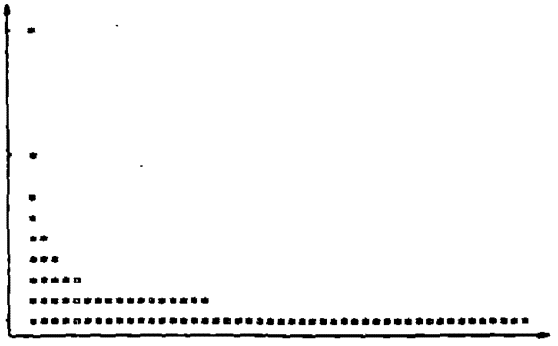

Figure 7. Frequency distribution of VMT for 1989 Texas rural interstates.

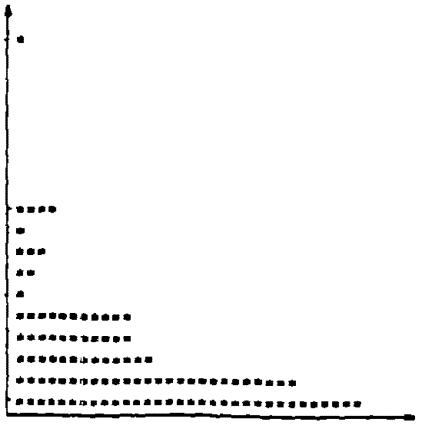

Figure 8. Frequency distribution of VMT for 1989 Texas urban interstates.

Figure 9. Frequency distribution of VMT for 1989 Texas rural 2-lane roads.

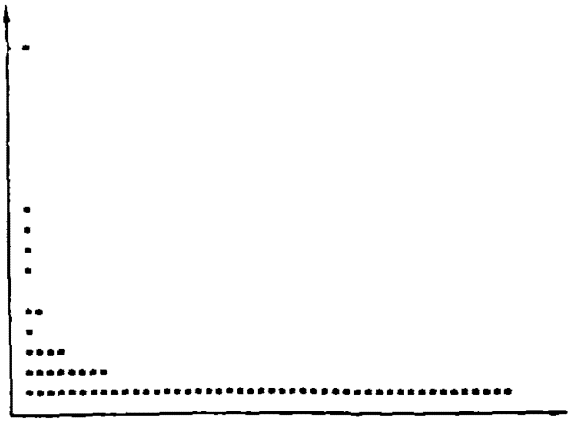

Figure 10. Frequency distribution of VMT for 1989 Texas urban 2-lane roads.

Table 24. Comparing Texas exposure data to simulated data.

	Mean	Var.	Median
Rural Int			
OBS	1046.5	2979.1	531.2
SIM	1047.0	900.0	699.9
Urban Int			
OBS	2587	7388.0	1782.2
SIM	2850	1000.0	1850.0
Rural 2-lane			
OBS	39.7	6599	17.8
SIM	40.0	2250	25.0
Urban 2-lane			
OBS	220.1	71.3	144
SIM	218.0	49.0	140

accidents. Recall in question 2 we concluded that correlating the exposures alone did not change either the distribution of the exposures or the r-t-m adjustment using the non-truncated exposures. Now, however, it would appear that the combination of truncating exposures and forcing a correlation will affect both the degree of r-t-m adjustment (though still not by an impressive amount) and, in the majority of the cases, improve the estimate of effectiveness. Thus, the combination of assuming exposures are correlated with accidents and limiting the range of exposures appears to both increase the improvement of the estimate and the degree of r-t-m adjustment.

Question 5. What effect will varying the degree of r-t-m have on the results?

The degree of r-t-m was varied to determine if this would have any effect in improving the EB adjustment or in assessing accident migration.

Table 25. Truncated exposures: treatment effectiveness (%).

		Trt		Mig		Avg. Shrinkage	
Exp ratio	**Sampling ratio**	**Freq**	**EB**	**Freq**	**EB**	**Trt**	**Mig**
500/1500	100/200						
	no corr.	-59.5	-58.4	+51.9	+51.7	.143	.176
	corr.	-55.9	-54.3	+58.6	+57.5	.236	.292
	25/125						
	no corr.	-66.4	-64.7	+71.6	+70.4	.143	.153
	corr.	-62.9	-60.8	+74.0	+71.9	.206	.288
800/1200	100/200						
	no corr.	-57.8	-56.5	+60.8	+61.8	.150	.162
	corr.	-57.9	-55.8	+52.2	+51.0	.330	.334
	25/125						
	no corr.	-68.5	-66.7	+55.5	+54.1	.149	.166
	corr.	-61.5	-58.9	+55.6	+55.1	.253	.330

The degree of r-t-m is being varied as follows:

High r-t-m - Choose 75 percent of the total sample to be the highest accident location letting the remaining 25 percent be randomly selected for accident occurrence.

Moderate r-t-m - Choose 50 percent of the total sample to be the highest accident locations letting the other 50 percent be random.

Low r-t-m - Choose 25 percent of the total sample to be the highest accident locations letting the other 75 percent be random.

Table 26 reflects the sampling combinations used to impose high, low, or moderate r-t-m bias in the treatment group and high, low or moderate positive and negative r-t-m bias on the migration group. For example, high positive r-t-m bias was defined as 75 percent of the sites in the sample being drawn from populations with the highest accident rates and 25 percent being drawn at random. Low positive r-t-m reversed the 75 percent/25 percent combination and moderate positive r-t-m imposed a 50/50 split. For high negative (reverse)

Table 26. Sampling combinations: varying r-t-m.

		Trt/Ref			Migration		mig/ref	
Label	Sample ratio	High	ran	ran	High	ran	low	ran
HiTrt/Hi+Mig	100/200	75	25	100	75	25	-	100
	25/125	19	6	100	19	6	-	100
LoTrt/Lo+Mig	100/200	25	75	100	25	75	-	100
	25/125	6	19	100	6	19	-	100
HiTrt/Hi-Mig	100/200	75	25	100	-	25	75	100
	25/125	19	6	100	-	6	19	100
LoTrt/Lo-Mig	100/200	25	75	100	-	75	25	100
	25/125	6	19	100	-	19	6	100
ModTrt/Mod+Mig	100/200	50	50	100	50	50	-	100
	25/125	13	12	100	13	12	-	100
ModTrt/Mod-Mig	100/200	50	50	100	-	50	50	100
	25/125	13	12	100	-	12	13	100

r-t-m, 75 percent of the sites in the sample were drawn from populations with the lowest accident rates, etc. Table 27 lists the descriptive statistics for these runs and Table 28 gives the treatment effectiveness.

Table 27. Varying r-t-m: descriptive statistics.

			Acc. Count				r-tm bias	
Label	Sample ratio	Time period	Trt	Mig	Tref	Mref	Trt	Mig
HiTrt/Hi+Mig	100/200	B	61.9	61.8	41.1	41.2	19.0	18.8
		A	26.7	80.7				
LoTrt/Lo+Mig		B	57.2	57.1	45.7	46.5	10.0	9.8
		A	27.1	80.1				
HiTrt/Hi-Mig		B	62.4	41.8	41.5	67.7	20.0	-19.6
		A	27.0	74.1				
LoTrt/Lo-Mig		B	56.3	46.6	45.9	56.2	8.3	-10.4
		A	25.8	76.3				
ModTrt/Mod+Mig		B	61.2	60.5	44.2	43.9	17.7	16.3
		A	26.8	62.2				
ModTrt/Mod-Mig		B	59.3	59.5	42.9	43.0	14.0	14.4
		A	26.6	79.8				
HiTrt/Hi+Mig	25/125	B	73.5	70.2	47.7	46.7	41.3	35.0
		A	29.2	84.4				
LoTrt/Lo+Mig		B	62.0	62.3	49.6	51.3	19.2	19.8
		A	29.7	79.9				
HiTrt/Hi-Mig		B	73.5	36.1	47.7	55.3	41.3	-30.6
		A	29.2	72.4				
LoTrt/Lo-Mig		B	61.1	46.7	50.2	54.7	17.5	-10.2
		A	29.2	74.4				
ModTrt/Mod+Mig		B	68.7	68.9	48.0	49.4	32.1	32.7
		A	30.6	83.4				
ModTrt/Mod-Mig		B	67.3	68.9	48.9	48.2	29.4	32.5
		A	26.5	80.1				

Table 28. Varying r-t-m: treatment effectiveness (%).

		Trt		Mig		Avg. Shrinkage	
Label	Sample ratio	Freq	EB	Freq	EB	Trt	Mig
HiTrt/Hi+Mig	100/200	56.9	54.4	30.6	36.0	.318	.313
	25/125	60.3	57.7	20.0	29.0	.283	.325
LoTrt/Lo+Mig	100/200	52.7	51.5	40.9	43.7	.320	.329
	25/125	52.0	50.3	28.9	33.0	.271	.304
HiTrt/Hi-Mig	100/200	56.7	55.0	77.0	67.0	.296	.327
	25/125	60.3	57.7	101.4	80.0	.383	.360
LoTrt/Lo-Mig	100/200	54.3	53.3	63.7	59.3	.325	.330
	25/125	52.2	50.7	60.1	53.6	.281	.351
ModTrt/Mod+Mig	100/200	56.1	54.7	28.4	33.3	.296	.347
	25/125	55.4	53.1	21.5	28.1	.267	.312
ModTrt/Mod-Mig	100/200	55.1	53.6	34.2	38.9	.389	.329
	25/125	60.6	58.2	16.5	23.8	.337	.336

R-T-M bias is defined as the relative percent change between the observed and the true mean accident count, i.e.

$$100\left(\frac{\bar{z}-\delta}{\delta}\right) \tag{27}$$

where δ is the true mean count and $\bar{z}$ is the observed mean count based on 5 replications. For the highway type used in this example, i.e. rural interstates, $\delta = 52$, VMT = 1047. Thus, for the high treatment/high positive migration combination for the 100/200 sample ratio for the treatment group, r-t-m bias was:

$$100\left(\frac{61.9-52}{52}\right)=19\% \tag{28}$$

This means the treatment group sample mean represented a 19-percent increase over the true mean accident count that would be expected on two-lane rural roads. Note that for negative (reverse) r-t-m, the bias is positive.

In table 28, we can see the relationship between the degree of r-t-m and the accuracy of our estimates. The relative deviations from the true effect is computed as:

$$1 - \frac{\hat{\theta}}{\theta} \tag{29}$$

where $\hat{\theta}$ is the estimated treatment effect and θ is the true effect, namely, .50 (50-percent reduction) for the treated sites and 1.50 (50-percent increase) for the migration sites. For example, the relative deviation from the true for the Hi/Hi + 100 combination for the frequentist treatment and migration estimates, respectively, are:

$$1 - \frac{.431}{.500} = .138 \tag{30}$$

and

$$1 - \frac{1.31}{1.50} = .127 \tag{31}$$

Positive values indicate an underestimate of the true theta, and negative values indicate an overestimate.

Table 29. Relative deviations from true and r-t-m bias.

	r-t-m bias		Deviations from true Trt		Mig	
Label	**Trt**	**Mig**	**Freq**	**EB**	**Freq**	**EB**
Hi/Hi+100	19.0	18.8	.124	.088	-.126	-.093
25	41.3	35.0	.206	.154	-.200	-.140
Lo/Lo+100	10.0	9.8	.054	.030	-.064	-.042
25	19.2	19.8	.040	.006	-.141	-.113
Hi/Hi-100	20.0	-19.6	.134	.100	.180	.113
25	41.3	-30.9	.206	.154	.343	.200
Lo/Hi-100	8.3	-10.4	.086	.066	.091	.062
25	17.5	-10.2	.014	.014	.067	.024
MOD/MOD+100	17.7	16.3	.122	.094	-.144	-.111
25	32.1	32.7	.108	.062	-.190	-.146
MOD/MOD-100	14.0	14.4	.102	.072	-.105	-.074
25	29.4	32.5	.212	.164	-.223	-.175

In all treatment effects, there is a tendency to underestimate the true theta (i.e. overestimate the amount of the reduction) which is due to r-t-m. In the migration group all estimates are underestimated except in the high reverse r-t-m cases where there is a tendency to overestimate θ (underestimate the migration effect). The closer this number is to zero, the better the estimate. Consider using the criteria that the relative deviation should be within $\pm$.10 (10 percent of the true value). Given this criteria, the treatment effect at both the migration and treatment sites is adequately assessed using the EB procedure for the sample size of 100 (two exceptions to this are extremely close (-.113 and .111)). For the sample size of 25, the treatment effect is only assessable for the moderate and low r-t-m situations. The EBEST estimate produces a modest improvement though often bringing the estimate to within the 10-percent criteria.

The case that a sample size of 25 may be adequate if r-t-m is not extreme suggests an interesting experimental design planning criteria. That is for sample sizes of 25 or less, the treatment effect cannot be adequately assessed if the degree of r-t-m bias is such that the observed accident counts represent more than a 30-percent increase relative to the expected accident counts for sites similar to those being treated. Researchers can determine this prior to implementing the treatment if they have a reference group sample. That is, they can compute the mean accident count for the reference group plus the candidate treatment sites, then compute the accident counts for the candidate sites. If the r-t-m bias exceeds 30 percent, then more than 25 sites will be required to adequately assess treatment. This type of a priori planning could be very cost beneficial.

Given the fact that the combination of correlated accidents and exposures and truncated exposures appeared to produce the best results (and is more reflective of reality), additional simulations were conducted to:

1. Determine what truncation ratio would result in significant r-t-m adjustment.

2. Determine what degree of r-t-m is allowable in order to adequately assess accident treatment effectiveness at both treated and migration sites.

All simulations were conducted for the 50-percent effect on all four roadtypes for the 100/200 and 25/125 sampling combinations. These results will be discussed in Chapter 3.

Additional simulations were conducted to examine the effect of truncated exposures on the r-t-m adjustment. For these simulations, r-t-m bias was imposed on the treatment group at three levels:

1. Top 100 percent of the sample was drawn for treatment.

2. Top 80 percent of the sample was drawn for treatment.

3. Top 70 percent of the sample was drawn for treatment.

Exposures were correlated with accidents as in stage 3. Sampling ratios of 100/200, 50/100, and 25/125 were imposed (number of treatment sites/treatment plus reference sites). Two exposure truncations were imposed: 800/1200, and 1000/1100. Treatment effects of 50-percent and 30-percent reductions were assumed.

Tables 29 and 30 reflect the relative deviations from the true and average shrinkage for these scenarios. From these tables it can be noted that the average shrinkage remains about 30 percent in all cases. The EB adjustment brings the deviations from true within acceptable limits using the 20-percent criteria in all cases. This adjustment is especially significant in the 25/125 sampling ratios where the frequentist method is outside of the limits in all cases. Parentheses denote cases beyond these limits. This means that if the sample size is 25 sites or less, it is imperative that the EBEST estimate be used to adjust for r-t-m bias. To detect a 30-percent treatment effect as opposed to a 50-percent effect, there is some indication that an even larger sample size demands the EB adjustment. For example, in the case of the 1000/1100 truncation for the severe r-t-m case where all treatment sites were the most hazardous sites, even the sample sizes of 100 and 50 treatment sites were not adequately assessed using the frequentist method. The following chapter will summarize the conclusions of these simulation studies.

Table 30. Relative deviations from true and average shrinkage on truncated exposures: 30% effect. (denotes outside of acceptable limits.)

				Deviations from true	
Exposure truncation	**Degree of r-t-m**	**Sampling ratios**	**Avg. Shrinkage**	**Freq**	**EB**
800/1200	Top 100%	100/200	.34	.149	.108
		50/100	.37	.175	.132
		25/125	.33	(.257)	.183
	Top 80%	100/200	.34	.145	.106
		50/100	.39	.161	.121
		25/125	.32	(.250)	.190
	Top 70%	100/200	.34	.135	.098
		50/100	.39	.157	.120
		25/125	.34	(.219)	.161
1000/1100	Top 100%	100/200	.35	(.201)	.157
		50/100	.34	(.200)	.155
		25/125	.30	(.268)	.185
	Top 80%	100/200	.34	.176	.137
		50/100	.34	.174	.134
		25/125	.30	(.238)	.180
	Top 70%	100/200	.34	.151	.111
		50/100	.34	.165	.127
		25/125	.30	(.229)	.174

Table 31. Relative deviations from true and average shrinkage on truncated exposures: 50% effect. (denotes outside of acceptable limits.)

				Deviations from true	
Exposure truncation	**Degree of r-t-m**	**Sampling ratios**	**Avg. Shrinkage**	**Freq**	**EB**
800/1200	Top 100%	100/200	.34	.144	.101
		50/100	.35	(.202)	.162
		25/125	.34	(.265)	.102
	Top 80%	100/200	.36	.126	.085
		50/100	.36	.170	.134
		25/125	.34	(.231)	.169
	Top 70%	100/200	.36	.123	.086
		50/100	.36	.168	.135
		25/125	.34	(.228)	.169
1000/1100	Top 100%	100/200	.34	.162	.117
		50/100	.36	.191	.149
		25/125	.30	(.242)	.179
	Top 80%	100/200	.35	.149	.106
		50/100	.35	.166	.124
		25/125	.31	(.215)	.157
	Top 70%	100/200	.34	.164	.123
		50/100	.35	.152	.112
		25/125	.35	(.203)	.149

CHAPTER 3. SUMMARY AND CONCLUSIONS

This chapter will attempt to furnish guidelines for researchers conducting safety measure evaluations. These guidelines were derived from the simulation results of this study. The following questions will be addressed:

1. How many sites are required to adequately assess a treatment effect at the treated sites?

2. How many total accidents must be observed to declare a specified percent effect to be statistically significant?

3. Are reference sites necessary to adjust for r-t-m bias at the treated sites? If so, how many?

4. Is a comparison group necessary?

5. Is a system effect of potential concern? If so, how many sites must be specified to declare a treatment effect to be statistically significant at the migration sites?

6. Is r-t-m a potential confounding effect among the migration sites? If so, how large a reference group will be necessary to adequately adjust for this bias?

7. What statistics should be reported in order to provide sufficient information on which to base an evaluation of the quality of the study and conclusions drawn?

These questions will be posed and answered in the form of a protocol. A hypothetical numerical example will be presented to illustrate the procedure.

SUMMARY OF FINDINGS

A brief review of the major findings of the three stages of this study are presented here. For more detailed findings, the reader is referred to the summary findings at the conclusion of each stage as reported in chapter 2.

The main objective of this study was to determine under which conditions could a system effect or the effect of a treatment on locations besides the treated area be assessed. In order to do this it was also essential to identify the conditions under which the treatment effect could be assessed at the treated locations. Previous studies developed the methodologies for assessing a treatment effect but no specific criteria defining when this assessment is feasible was ever developed.

Through extensive data simulation, such criteria was established in this study. Detailed guidelines for designing safety evaluation studies based on these results are presented in the next section. The major findings of this study were:

Treatment effects can adequately be assessed at both treated and migration sites if there is little or no r-t-m bias in the site selection process using standard statistical techniques. This assessment is accurate for as few as 10 sites to detect a change in accidents of 50 percent or more or to detect a change of up to 10 percent with 25 or more sites for typical urban and rural interstates or two-lane roads. This is, of course, is an unrealistic ideal set of conditions.

In the presence of as little as 70 percent r-t-m bias, (i.e. 70 percent of the sites selected for treatment or potential migration had the highest (or lowest) accidents with the other 30 percent being randomly selected), the following conditions are necessary in order for the EBEST procedure to provide an adequate adjustment for the r-t-m bias:

1. The ratio of the maximum to minimum exposure (VMT) for the group of sites being evaluated must not exceed 2.0.

2. At least 25 sites are being evaluated (for the typical highway sections simulated in this study) or enough sites to yield an adequate accident count in the before period (details follow in the next section).

3. The number of sites in the reference group must be at least four times as great as the group being evaluated for 25 sites, but need only be the same size for 100 or more treatment sites. More insight is needed as to how big the reference group needs to be for treatment groups between 25 and 100.

If there is little or no r-t-m, there is no need to use the EBEST adjustment although it will always provide a better estimate of the true effect than the classical procedure. Guidelines for determining the potential degree of r-t-m bias for a given study is provided in the next section. Similarly, if the exposures are highly variable, the EBEST adjustment will not differ substantially from the classical. In this case, no existing statistical methodology can provide an adequate assessment of the treatment effect, and conclusions based on such a study are highly suspect.

Correlation between migration sites and treated sites appeared to have little if any effect on successfully evaluating the treatment at either treated or migration sites. However, examining this correlation may be useful in identifying a systematic (site by site) migratory effect. There appears to be some indication that a high correlation between accident count and VMT improves the degree of adjustment using the EBEST estimation procedure.

PROTOCOL FOR A GLOBAL ASSESSMENT OF A HIGHWAY SAFETY MEASURE

The following step-by-step guide will provide a systematic method for planning and evaluating a highway safety study. These steps include the determination of adequate sample sizes, identifying r-t-m potential, identifying accident migration potential, and data analysis procedures for a sound safety measure evaluation. A numerical example follows in the next section. It may be useful to follow along in the example as each step is being discussed.

I. Determine the number of sites and accident frequencies necessary to be able to:

1. Evaluate the treatment effect at the treated sites.

2. Evaluate the treatment effect at the migration sites if migration is found to be a factor.

3. Adjust for r-t-m bias in either the treatment or migration site groups.

Table 32 gives the minimal total accident frequencies necessary to declare a given percent change in accident occurrence statistically significant. These figures were developed using the statistical formula for testing significance of the odds ratio at the 5-percent level of significance. Basically, equation 5 was used and the question posed--what is the fewest number of accidents required so that this test statistic would be greater than 1.96 (the normal probability value at the 5-percent level of significance), in absolute value? It is assumed that the treatment and comparison group are equal and that the treatment group change in accidents is adjusted for any change in accidents in the control group from before to after.

In other words, to declare a 10-percent reduction in accidents due to treatment, at least 1423 total accidents must be observed in both the treatment and comparison group before treatment (2846 accidents in both groups together in the before period). The number of sites or time period required to attain this level of accident frequency will depend upon the type of roadway and traffic volume. For example, in defining the parameters for this simulation study, urban interstates were found to have an average of 52 accidents per mile per year and 2-lane rural roads an average of 6 accidents per mile per year. In order to detect a 10-percent reduction in accidents for a safety measure for urban interstate sections, 27 sites would be required over a one year period or 14 sites over a two year period would be needed to obtain the necessary 1423 total accidents. If

Table 32. Total accidents required in the treatment group to declare specified treatment effects statistically significant.

Treatment effect (percent change)	Total accidents in Treatment Group before treatment
10	1423
15	607
20	328
25	201
30	133
35	94
40	69
45	52
50	40
55	32
60	25
65	20
70	17

the safety measure is imposed on two-lane rural roads, many more sites would be needed (1423/6=237 sites for one year). Of course, a 10-percent reduction is very minimal and it is reasonable to expect that the sample size would be large to be able to detect such a small change.

These sample sizes were determined using the odds ratio test statistic in equation 3 and setting T = 1.96, the critical value at the .05 level of significance. In other words, we are asking what the minimum sample size must be so that a given percent effect (O.R.-1) will be statistically signifcant. We assume that the treatment and comparison group are equal (A = B in equation 2) and that there is no change in the comparison group (A = C). In order for the treatment group to have the given effect, the relationship of A to D is known and there remains but one unknown to solve for, namely A. For example, consider the 10-percent reduction. The O.R. = .90 and equation 2 becomes:

$$1.96 = \frac{\ln(.9)}{S.D.(O.R.)} \tag{32}$$

where

$$S.D.(O.R.) = \sqrt{\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}} \tag{33}$$

but

$$O.R. = \frac{A/C}{B/D} = \frac{A/A}{A/D} = .9 \tag{34}$$

and

$$D = .9A \tag{35}$$

Now, returning to (33),

$$S.D.(O.R.) = \sqrt{\frac{3}{A} + \frac{1}{.9A}} \tag{36}$$

and solving for A in (32) we get

$$A = \frac{(3 + \frac{1}{.9})}{(\frac{\ln(.9)}{1.96})^2} = 1423 \tag{37}$$

The point is that this sample size figure can provide guidelines in designing a traffic safety study and in determining whether or not a safety measure can adequately be assessed. Similarly, if accident migration is to be assessed, the same numbers of migration sites will be required to detect given amounts of change. Recall that one of the findings from the simulation study was that accident migration was not detectable for fewer than 25 sites. Even for the urban interstates with 52 accidents per year the above table indicates that 27 sites would be required to detect a 10-percent change. Thus, it is not surprising that for fewer than 25 sites, unless a sufficiently long time period were observed, there will not be enough total accidents to detect a treatment effect at migration sites. Note that although these sample sizes were computed based on the ability to detect change at the .05 level of significance, these can be adjusted for any level of significance by multiplying by $(z/1.96)^2$ where z is the normal probability value for another level of significance.

II. Identify r-t-m potential.

In order to determine whether or not the sites selected for treatment have a potential r-t-m bias, the following steps should be taken. Recall that r-t-m bias occurs when either the most hazardous or least hazardous sites have been selected for treatment evaluation. If r-t-m bias is not a factor, the data needs will be reduced and classical statistical methods can be used to evaluate the treatment effect. If r-t-m bias is a factor, then additional information on the accident rate for the population of all potentially treatable sites is needed.

1. Define the population of potential treatment sites for the particular safety measure to be implemented.

2. Define the site selection criteria. How were the sites selected for treatment, i.e. all sites that had more than a specified number of accidents?

3. Compute the accident rate for the selected treatment sites before treatment.

4. Either:

   a. Select a random sample of sites from the defined population of potentially treatable sites defined in step 1. This sample is called the treatment reference group, that is, sites which are similar to the treated sites and could have been selected for treatment had they met the treatment selection criteria from step 2. The number of sites needed for this sample will depend on the number of treatment sites as illustrated in table 33:

Table 33. Sample size requirements for reference group.

Treated Group	Reference Group
n = 25	4 times greater
n = 26-99	2-3 times greater
n = 100	Same size

That is, if 25 sites are treated, the reference group ideally should consist of 100 sites.

or

b. Estimate the accident rate you would anticipate for the population of potentially treatable (i.e., the treatment plus reference group) sites defined in step 1. This may be obtained from past studies or literature or simply be an educated guess on the part of the researcher.

5. Using the estimates computed in II.3 and II.4, determine the amount of potential r-t-m bias among the treated sites.

   a. Compute or estimate the average exposure, $\bar{e}$, for the population of potentially treated sites (e.g. average VMT).

   b. Compute an approximate 95% confidence interval about the mean rate for the potentially treated sites (step I.4) as:

$$\hat{\lambda} \pm 1.96 \sqrt{\frac{\hat{\lambda}}{\bar{e}}}$$

   where $\hat{\lambda}$ is the accident rate from II.4 and $\bar{e}$ is the average exposure of II.5.a. This ratio is an estimate of the variance of the assumed Gamma distribution of accident rates.

   c. Compute the ratio of the largest site exposure to the smallest site exposure in the treated group before treatment (i.e. largest VMT in the sample to smallest VMT in the sample.

   d. If either:

      the treated site accident rate is within of the limits computed

      or

      the ratio in II.5.c exceeds 2.0,

      then no r-t-m adjustment will be necessary. Note: the reason no r-t-m adjustment is recommended for the second criteria is not

that regression-to-the-mean is not a factor but rather because the exposure variable is so variable that the EBEST procedure will yield estimates similar to the classical procedure (i.e. there will be little shrinkage toward the overall mean rate) and hence the effort to adjust for this bias is not warranted. If this is the case and the treated site rate is considerably outside the acceptable limits, one might reconsider the entire evaluation effort as it will, in all probability, be confounded by the r-t-m bias.

It should be noted that the ratio of max to min exposure exceeding 2.0 could be due to either of the following:

1. The reference group is not homogeneous enough to adequately represent the pool of potentially treatable sites or

2. The highway type is such that this much variability is inherent for these road volumes.

If the conclusion from these calculations is that r-t-m bias is a serious factor and that the EBEST procedure will adjust for this bias, then a reference group sample of the appropriate size as specified in step 4 should be specified and data on both accident frequencies and exposures for these sites should be collected.

III. Identify accident migration potential for this study.

A critical factor at this stage of a system-wide analysis is defining and identifying potential migration sites or neighboring sites which could be impacted by the treatment. The criteria for defining such sites will depend on the particular study. It will require sound engineering judgement and a thorough understanding of the highway system involved in the safety treatment study. This step should not be taken lightly and should be carefully planned in advance.

1. If there are fewer than 25 treatment sites, accident migration will be nearly impossible to detect and not worth the effort. If there are 25 or more sites in the treatment group and there is reason to suspect that the treatment could affect accidents at locations other than the treated spots, proceed to design an accident migration study.

2. Define the population of potential migration sites for the particular safety measure to be implemented.

3. Select migration sites for each treatment site in the sample. There may be more than one potential migration site for each treated site. If so, these may be grouped as primary, secondary, tertiary, etc. and the migration assessment conducted separately for each group.

4. Define the type of accident that would represent a migration due to the treatment (e.g. if the treatment were to change four-way stop intersections to two-way, the succeeding four-way intersection on the road that was changed might experience more rear-end collisions).

IV. Identify r-t-m bias potential among the migration sites.

Basically this involves asking whether or not there is reason to believe that the migration sites that would be affected by the treatment were experiencing accident rates either above or below the rate one would expect for similar sites during the period before treatment. The following steps will provide guidance in addressing this issue.

1. From the sample of migration sites selected in III.3, compute the accident rate for the migration sites before treatment.

2. Either:

   a. Select a random sample of sites from the defined population of potential migration sites defined in step III.2. This sample is called the migration reference group, that is, sites which are similar to the migration sites that might be affected by the treatment. The number of sites needed for this sample will again depend on the number of treatment sites and should be at least the same size as the treatment reference group;

   or

   b. Estimate the accident rate you would anticipate for the population of potential migration sites defined in step III.2. This may be obtained from past studies or literature or simply be an educated guess on the part of the researcher.

3. Using the estimates computed in IV.1 and IV.2, follow the same procedure as outlined in II.5 using the data from the migration sites to determine if r-t-m bias is a potential problem in assessing the treatment effect on the migration

sites or if the exposures for the migration sites are suitably stable to provide a reasonable adjustment for the bias.

V. Selecting the comparison groups.

Comparison group data is almost always an essential component in a valid before/after study of a treatment effect. The role of the comparison group data is to adjust for other confounding effects which could affect accidents from the before to after period. Frequently such an adjustment is needed because there is a time trend in accident occurrence, i.e. accident rate is increasing over time. In other instances, the comparison group adjusts for differences in road geometry, weather, traffic volume or driving population. Only if it can be assumed that none of these factors are likely among the treatment sites can an analysis without a comparison group be justified.

For these same reasons, a comparison group is essential in the assessment of the treatment effect on the migration sites. The migration sites may, however, often share the same comparison group as the treatment sites and hence, only the treatment sites' comparison group may be necessary to assess the treatment effect at both the treatment and migration sites. Since the comparison group is generally selected based on criteria other the accident history, r-t-m bias is seldom a problem and therefore reference group data is not needed for the comparison sites. The reference group data for the treatment and migration sites may qualify as a comparison group provided that the accident history is available for the reference group sites during the post-treatment period.

VI. Treatment Evaluation

1. Estimate the change in accidents at the treatment sites using the odds ratio procedure with or without the EBEST adjustment depending upon the degree of r-t-m bias and availability of a treatment reference group.

2. Estimate the change in accidents at the migration sites using the odds ratio procedure with or without the EBEST adjustment depending upon the degree of r-t-m bias and the availability of a migration reference group.

3. Report, at a minimum, the following statistics. (Reference group statistics are needed only when r-t-m bias is a factor.)

For the following groups:

treatment before
migration before
comparison before
treatment reference before
migration reference before
treatment after
migration after
comparison after

report:

number of sites
average exposure
exposure variance
total accidents
accident rate
correlation between exposure and accident count.

Also report:

- percent-change in accidents at treated sites with statistical test of significance
- percent-change in accidents at migration sites with statistical test of significance
- average shrinkage if EBEST is used for both treatment and migration sites
- EBEST estimate for expected accident count and rate for both treatment and migration sites

VII. Assess the implications of accident migration for this study.

Is it of serious consequence? Should treatment implementation be reconsidered due to this effect? Do you recommend that future evaluations of this safety measure assess the effects at areas other than the treated locations? What experimental design considerations would you recommend if this evaluation were to be repeated? At this stage, it will be important to compare the treatment and migration effect. Has a significant reduction in accidents at the treated sites been offset by an increase in accidents elsewhere in the system? Is the system effect, in fact, a true effect or just an anomaly due to some other factor? These and similar questions must be addressed in completing this system-wide evaluation.

HYPOTHETICAL EXAMPLE TO ILLUSTRATE PROTOCOL

To illustrate the use of the step-by-step process of designing a treatment evaluation study the following hypothetical example will be given. Assume that the safety measure of interest is the installation of raised pavement markers on two-lane rural road sections.

I. To determine the number of sites necessary to evaluate the treatment effect we must know the average number of accidents per year at the potential treatment sites . Suppose the average number of accidents is 10 per mi per year. Assume we would like to be able to detect a reduction in accidents of at least 20 percent. From table 31, we know that we will need to observe a total of 328 accidents at the treatment sites during the before period. Therefore, we must treat at least 33 sites if only one year of data is to be used. This requires that the after period also covers the same time frame. Furthermore, we will also need to identify and observe the same number of comparison sites. There has also been some concern that the installation of pavement markers on one section may increase accidents on the adjoining sections because drivers will increase their speed due to better visibility provided by the markers. To be able to detect an increase in accidents of 30 percent, the migration sites must experience a total of 133 accidents in the period before treatment. Assuming the adjacent sections of road experience the same number of accidents per mile per year as all two-lane rural road sections, i.e. 6 per mi per year, then 23 migration sections would be required. Of course, if possible we would like to define a migration section for each treatment section but if this is not possible we can at least determine how large an increase could be detected given the sites we have.

II. 1. The population of potential treatable sites for this study are two-lane rural roads with traffic volumes of at least 40 MVMT.

2. The sites selected for treatment will be two-lane rural roads which experienced 10 or more accidents per year per mi.

3. The average accident rate among the sites selected for treatment is 5.0 accidents per MVMT.

4. The average accident rate for all two-lane rural roads in the state with greater than 40 MVMT had been previously determined to be 2.286 accidents per MVMT.

5. a. The average exposure for all two-lane rural roads with at least 40 MVMT is 60 MVMT.

b. The 95-percent confidence interval about the accident rate for all two-lane rural roads of greater than 40 MVMT is:

$$2.286 \pm 1.96 \sqrt{(2.286/60)}$$

for an upper limit of 2.669 and a lower limit of 1.9034.

c. The ratio of the maximum to minimum MVMT's at the treated sites was 1.4.

d. Since the accident rate at the treated site, 5.0, exceeds the upper limit of the confidence interval, we suspect that there is significant r-t-m bias in the treatment sites and that the sites receiving treatment are experiencing more accidents than the average. Also, since the ratio of maximum to minimum exposure is less than 2, we can anticipate that the EBEST procedure will result in a significant adjustment for the r-t-m bias.

III. 1. Since a total of 33 sites will be treated there should be sufficient data to assess the effect of the marker installation at migration sites.

2. The population of potential migration sites will be all 2-mi sections with similar roadway geometry, traffic volume, etc. as the migration sections.

3. Since we are interested in the potential effect of the markers at adjoining highway sections.migration sites will be defined as the 2-mi segment following the marked segments in either direction.

4. For this study we will define all accidents as potentially caused by the treatment.

IV. 1. The accident rate for the migration sites before treatment was 2.146 accidents per MVMT.

2. The accident rate for sites similar to the migration sites is thought to be 2.25 and their average MVMT to be about 45. This produces upper and lower 95-percent confidence limits of 2.69 and 1.812, respectively. Since the migration sites' accident rate is within these limits we can assume there is no r-t-m problem in the migration site selection and no reference group data will be required for the migration sites.

FUTURE DIRECTIONS

Measuring and evaluating the effect of accident migration is the first step in identifying and quantifying the problem. However, the next logical step is to offer solutions. How can the safety of the system be preserved, given that we now know the impact of the safety measure? Should the safety improvement be modified? Should portions of the system be modified? Below is an example of a countermeasure which could be impacted by accident migration because it reduces driver expectancy. A possible solution to this type of safety impact follows.

It is assumed that drivers become acclimated to a certain level of expectancy while driving. This level of expectancy is a function of the information load presented to the driver in which he must perceive the information for its source, process it, and respond accordingly. If the rate of information presented to the driver remains relatively constant over a period of time, it can be hypothesized that the driver will settle into a "state of expectancy" that is appropriate for that situation.

For example, on a given rural interstate highway, the driving task is relatively simple and the driver becomes accustomed to a more "relaxed" state of awareness or expectancy. On an urban freeway with heavy traffic, the driving task obviously becomes more complex. The driver becomes adjusted to a heightened state of awareness as required by the driving task.

It seems logical that when the driving task becomes less complex over a relatively short transition period, there will be no adverse effect on the driver due to a decreased state of awareness or expectancy. However, it is hypothesized that a sudden increase in

the state of expectancy required by the driving task will have an adverse effect on the driver.

A logical conclusion can be drawn that situations requiring a sudden increased state of expectancy by the driver will be associated with high accident rates. Safety improvements such as pavement overlays, striping and bridge widening may reduce accidents within a particular location or section of roadway, but may increase accidents at downstream locations that require a higher state of expectancy. A driver traveling on a two-lane rural highway that comes upon a newly paved section of roadway may choose to drive a little faster than he was previously driving. However, when he leaves the improved section further down the roadway, he may be traveling too fast for these conditions.

If it can be determined that it is important to evaluate a system-wide effect for these countermeasures which reduce driver expectancy, what can be done to improve the safety of the system? A possible solution might be to provide a smooth transition from a reduced state of driver expectancy to a heightened level by way of other highway modifications in a transition zone.

A disappointing aspect of this study was the inability to use actual data in using the accident migration methodology. A very useful effort would be to use the methodology developed in this study to actually design and conduct a global safety measure evaluation study. An alternative approach would be to identify data bases from which a retrospective accident migration study could be conducted. Ideally, both low accident frequency and high frequency locations should be included. It would also be of interest to design and evaluate an intersection study where more than one migration site could be identified per treatment site.

In order for the results of this study to be used, highway safety analysts must be made aware of the type of data needed in assessing a safety measure both at the treated spots and more globally. The data requirements for this are not excessive and, if known a priori to conducting the study, could be collected with very little additional cost or effort. Since there are so many examples of safety measures which could be compromising safety at areas other than where they are being evaluated, it seems essential that this methodology be implemented immediately.

Safety is not a "spot" issue--it's a global issue. We cannot afford the cost in public safety that we incur when we assess safety only at the point of implementation of a treatment. Now that the methodology has been developed for handling the statistically difficult problems such as regression-to-the-mean bias and guidelines have been established to design such global safety studies, it behooves the research community to employ these procedures to their fullest capacities. It is the hope of this effort that future studies will embrace the concepts developed here toward the betterment of highway safety.

APPENDIX A. LITERATURE REVIEW

A more detailed account of the literature reviewed for this study is provided here.

One of the first papers to address accident migration proposes an empirical Bayes methodology for evaluating treatment effectiveness using accident data.[5] Assuming a gamma prior distribution on mean accident rates, parameters for the negative binomial are estimated from the data using the method of maximum likelihood. Although the author clearly state that this is the preferred method of estimation, they propose and give formulas for the method of moments which are identical to Hauer's.

This manuscript carefully and in non-technical terms explains the Empirical Bayes methodology. The authors then state that even if an effect were found to be significant (based on 90 percent confidence interval estimates) the source of the change cannot be attributed to the treatment unless information is available on "similar, untreated sites" (comparison group). The authors then acknowledge that they do not know how to use this comparison group information in the Empirical Bayes Methodology and state that this methodology "would be more powerful...if the problem of incorporating control data can be satisfactorily resolved.

The data used in the example consists of accidents from 297 potential blacknodes in Hertfordshire form 1975 to 1979. The authors acknowledge Hauer's 1980 paper, but claim their method differs because it allows evaluation of individual sites and Hauer's method is restricted to evaluating "groups of sites". The authors indirectly address an issue of concern in accident migration by noting that if the reference sites are selected to be very similar to the treated sites, reverse regression to the mean is possible since the more hazardous sites were selected for the treatment.

The methodology in this manuscript does not differ from Hauer's. The main new contributions are:

1. Showing that the negative binomial distribution is a reasonable assumption and fits this data set.

2. Acknowledging the superiority of maximum likelihood estimation over the method of moments and using maximum likelihood in fitting the negative binomial to this data.

3. Proposing and computing confidence intervals for statistical inference.

The deficiencies in the methodology are:

1. No means of using comparison group data.

2. No means of using exposure data.

3. Lack of stressing the importance of the reference group or clear guidelines as to how the reference group should be selected.

This manuscript may be the first to bring up the reverse regression-to-the-mean concept which could affect surrounding sites (accident migration) and the reference group.

The next paper of importance in this area explains an empirical Bayes methodology for estimating the expected frequency at a site after treatment based on assuming that the true mean of the site is also a random variable distributed as a gamma distribution.[23] Based on additional assumptions that, for multiple sites, these true means fluctuate about the same constant mean (site means come from the same population), and that accidents are independent Poisson distributions, the authors show the following:

1. If you observe 200 accidents over 5 years and then observe x accidents in the 1 year prior to treatment, the estimate of the true mean accident rate for that site is $\hat{m} = \frac{200 + x}{5 + 1}$; i.e. the average number of accidents after the 6 year period.

2. If you have a control and treatment site and they both have the same true mean before treatment, and the after accidents at the treated site is θ times the before accidents at that site (i.e. θ is the treatment effect), then that treatment effect's mode (best value) will be estimated by:

$$\hat{\theta} = \frac{x_2}{\hat{m}}$$

where x_2 is the number of accidents at the treated site after and $\hat{m}$ is the average expected accidents at the *control* site before treatment.

The authors then give a hypothetical example wherein the control and treatment site have the same means before treatment (3.6) and the control site decreases to 3 while the treatment site increases to 5. The best estimate of treatment effect is .8 implying a 20-percent reduction in treatment. However, the probability of observing this pair (3,5) given the prior distribution assumptions is only .015, highly unlikely. Thus, given this observation, the probability that the true treatment effect is 20-percent

reduction is very small. The authors conclude that the evidence *does not contradict* the use of the Bayesian method and that this needs further research.

The following are specific claims and critiques of this paper:

Claim: Local government officers are disposed to take action about blacknodes before consensus about the true accident rate at a particular site has had time to stabilize.

Response: How do we know the site's true accident rate has not stabilized? Can we use multiple observations in time at the site to estimate retrospectively what that true mean is?

Claim: The advantage of the Bayesian approach is that the uncertainty about the true accident rate is demonstrated by the density function that is assigned to it's value.

Response: Yes, but how is "it's value" determined. This is the part which requires careful consideration (how good is our reference group?)

Claim: One of the assumptions in this application is the constancy of the true mean accident rate at a particular blackspot.

Response: I interpret this as the assumption that all sites in the data set have some true mean value which is distributed about the same, constant, true mean. If so, this is the weakspot of the method. When combining sites to get the estimate for our shrinkage, we must be sure these sites represent similar conditions (exposure, road geometry, etc.) or their true mean may be fluctuating about means which vary (are not constant) and which depend on some other measure, like exposure. That is, we must question the assumption of exchangeability for this data set.

Claim: When considering a control group, mean, m_1 and treatment, mean, m_2 with m_2 after = Θm_1, the assumption is made that $m_1 = m_2 = m$, i.e. That the true mean of the control site is equal to the true mean of the treated.

Response: If we assume the true site means are random variables about some mean, wouldn't it be highly unlikely that these two random variables come from the same distribution? Also, it implies the control site is also experiencing regression-to-the-mean bias. The hypothetical example in this paper assumed the same true mean accident rate for both the treatment and control and then observed that the treated

site had a reduction in accidents relative to this mean where as the control site had an increase above the mean. If they had the same true mean before, then either both treatment and control sites had a regression to-the-mean problem or neither did. Suppose neither did. Then the effect we see should be believable. Suppose both did. Then it is highly unlikely that the *control* site would have regressed even farther away from it's mean the next year in the same direction.

Figure 11 pictorially describes this process where m_{2A} is the treated site after accident rate and m_{1A} is the control site after frequency rate. A regression-to-the-mean situation would mean both sites are in the shaded area of the distribution since they have the same accident rate before treatment. A more realistic example might be more appropriate here.

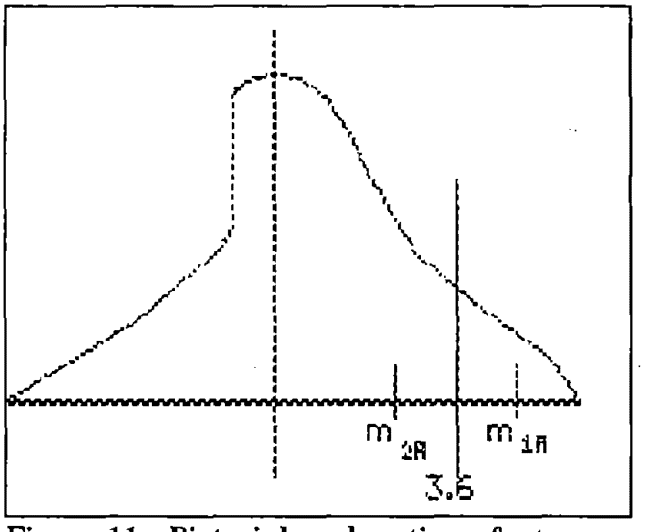

Figure 11. Pictorial explanation of r-t-m.

These authors have spearheaded the concept of accident migration. The method they propose is a candidate method for extending to the assessment of accident migration, and thus worthy of review. In practice, however, it would appear that their methodology is too restrictive for general application to the accident migration study.

Another migration example is presented in the following paper.[2] Between 1975 and 1978, inclusive, some 402 "blackspots" were treated throughout a 16 borough area in London, England. Each of these treated blackspots was at a roadway node or link. Actually, 8 of these 402 treated blackspots were not located at a link or node, however, these blackspots were eliminated before the first of three analyses, as explained later. Nodes are typically intersections; links are segments of roadway between intersections.

Accident data were collected at each blackspot for a period of three years before and after treatment. Three years of before and after data were also collected at all areas surrounding the treated blackspots. The surrounding areas were defined as follows:

1) For treated blackspots on links, the surrounding area included all of the link not receiving treatment and the two nodes at either end of the link.
2) For treated blackspots on nodes, the surrounding area included all of the links coming into the treated node, and the nodes at the ends of those links.

In addition to the accidents recorded at all 402 treated blackspots and surrounding areas, total accidents throughout the 16 boroughs being studied were recorded. The mean number of accidents throughout the 16 boroughs during a three year before period was 65,695, and 64,714 during a three year after period--a decrease in accidents of 1.49 percent. When accidents that occurred at treated blackspots are eliminated from these totals, 51,864 accidents were recorded for the before period, and 52,166 for the after period--an increase of 0.58 percent.

The accident recorded throughout the 16 boroughs (and outside the treated blackspots) formed a comparison group against which to access the effect of accident migration. As will be seen later, this comparison group is only weakly used in the analysis.

Some 366 treated blackspots are included in the first analysis. For these 366 blackspots and their surrounding areas the following accidents were recorded:

Table 34. Summary data for analysis I.

Time	Areas	Treated Blackspots
Before (3 yrs)	31,208	8,139
After (3 yrs)	30,632	6,410
Percent Change	-1.8	-21.2

During the before period, on average, there were 7.41 accident per blackspot per year. During the after period, the rate was 5.84 accidents per blackspot per year.

In the second analysis, the number of treated blackspots studied was further constrained. This analysis was performed on accidents collected at 219 treated blackspots and surrounding areas.

Table 35. Summary data for analysis II.

Time	Areas	Treated Blackspots
Before (3 yrs)	8,956	5,025
After (3 yrs)	9,269	3,921
Percent Change	+3.5	-22.0

There were 7.65 accidents per blackspot per year during the before period and 5.97 accidents per blackspot per year during the after period. Let's look a little closer at the 147 blackspots and surrounding areas that were eliminated between analyses I and II. By subtracting the accidents recorded in table 35 from those recorded in table 34, we can see the accidents that were eliminated between analyses I and II. These accidents are shown in table 36.

Table 36. Summary data for accidents eliminated between analyses I and II.

Time	Areas	Treated Blackspots
Before (3 yrs)	22,252	3,114
After (3 yrs)	21,363	2,489
Percent Change	-4.0	-20.1

There may have been good substantive reasons for eliminating the 147 treated blackspots and surrounding areas represented in table 36 before performing the second analysis. Nevertheless, it should be noted that accidents were decreasing by 4 percent in those surrounding areas that were eliminated, i.e., the elimination of the 147 treated blackspots and their surrounding areas biased the second analysis toward a finding of support for accident migration.

In the final analysis, the data set was constricted once more to contain 133 treated blackspots and surrounding areas.

Table 37. Summary data for analysis III.

Time	Areas	Treated Blackspots
Before (3 yrs)	5,386	2,935
After (3 yrs)	5,929	2,281
Percent Change	+10.0	-22.3

For the 133 treated blackspots in this last analysis there were, on average, 7.36 accidents per blackspot per year during the before period and 5.72 during the after period. The authors then argue that although accidents in the comparison group increased only 0.58 percent from before to after, for the sake of argument, let us assume that they may have increased by as much as 5 percent. Accordingly, accidents might have increased by as much as 5 percent in the areas surrounding the treated blackspots if those blackspots had not been treated. However, even allowing for an expected 5 percent increase in accidents in the "surrounding areas," the observed 10 percent increase in accidents in the "surrounding areas," the observed 10 percent increase is significantly by standard statistical procedures. On the evidence provided in this third analysis, the authors claim to have demonstrated accident migration.

Before leaving this third analysis, let us look at the 86 treated blackspots and surrounding areas that were eliminated between analyses II and III. Subtracting the accidents recorded in table 37 from those recorded in table 35 yields the following (table 38):

Table 38. Summary data for accidents eliminated between Analyses II and II.

Time	Areas	Treated Blackspots
Before (3 yrs)	3,570	2,090
After (3 yrs)	3,340	1,630
Percent Change	-6.4	-22.0

Accidents were decreasing by over 6 percent in those surrounding areas that were eliminated. Again, there may have been good substantive reasons for eliminating these 86 treated blackspots and their surrounding areas, nevertheless, the result of eliminating these data was to further bias the analyses toward sustaining the hypothesis that accident migration is a real phenomenon.

There are several troubling aspects to this study. First, the authors conducted three analyses before demonstrating accident migration. The data set for the third analysis was a subset of the second; the data set for the second analysis was a subset of the first. The question might honestly be asked: If the authors had not found a significant accident migration effect in the third analysis, would they have proceeded on to a fourth or fifth subsetting of the data and a fourth or fifth analysis?

Second, although the comparison group defined for purposes of assessing accident migration was, in fact, of little consequence in the analyses presented, the comparison group was, nevertheless, poorly defined. It is not clear how the authors defined the accident group. However, the same comparison group (i.e., the same comparison ratio) was used for all 133 surrounding areas in the third analysis. And, some of those surrounding areas may have started collecting "after" accident data as early as January 1975, while others may not have begun collecting "after" accident data until December 1978. How could one comparison group, one comparison ratio, possibly serve for all 133 surrounding areas at which accident migration was alleged to have occurred?

Third, during the mid-1970's the western world was undergoing an Arab oil embargo. The effect this embargo had on traffic accidents in greater London is not known. However, if it had some effect, this effect, as a minimum, makes longitudinal analyses more difficult to conduct and comparison groups and comparison ratios more difficult to define.

And fourth, the statistical analysis conducted to demonstrate accident migration was a fairly standard analysis. But, a statistical analysis conducted on an arbitrary and capricious data set is of little consequence, regardless of significance.

This paper is of monumental importance to this study of accident migration as it was this paper which actually spearheaded the concept of accident migration. In fact, the authors state that "a search of the literature has revealed no previous studies on the subject." It stirred up considerable controversy, as evidenced by the letters to the editor which followed. These letters will be summarized below:

The discussion paper which immediately follows the Boyle and Wright article, criticizes the paper as follows:

1. Huddart claims the findings are "unrealistic" - since high-intensity traffic signals have been the normal standard for some time, why should their replacement of old signals such a dramatic effect?

2. The statistical analysis is lacking and not robust.

3. The data is biased and thus the apparent migration effect is not real.[3]

Huddart makes valid points, especially with regard to the statistical analysis. It seems puzzling that they did not use the already proposed Bayesian methodology for handling regression-to-the-mean bias especially since one of the authors proposed this method in a much earlier paper.[2]

The authors respond to Huddart's discussion and specifically address his three criticisms as follows:

1. The term "realistic" is questioned. The authors claim to have reported facts and the observance of a phenomenon, whether their hypothesis explaining the cause of this phenomenon is accepted or not.

2. The authors stand by their statistical treatment. They claim their significance levels are "conservative".

3. The authors do not feel there is any bias in the data.[10]

Obviously, the authors intend to stand by their original claims and analyses. They do not adequately repute Huddart's criticisms.

Basically, Huddart dismisses most of Boyle and Wright's letter and says it is now up to the research audience to decide for themselves.[7] He does, however, respond to Boyle and Wright's claim that he has "misconceptions" about the "near-miss" theory and proceeds to logically explain through a series of equations, how risk compensation could not have caused the 10 percent increase in accidents observed by Boyle and Wright.

It seems there's confusion regarding the terms "near-miss". Huddart interprets this as an extremely rare event and I share Huddart's interpretation. In the next letter to the editor, however, we see that Boyle and Wright have a different interpretation.

The authors respond to Huddart's agruements in the previous letter by stating that, whereas they accept Huddart's formulas and mathematics they do not accept the assumptions on which they are based.[11] They define "near-misses" as a more common event including relatively mild forms of conflict. They also claim "near-misses" would affect other drivers, not just the one who performed the maneuver and that Huddart's formulas are not dynamic applying to a fixed point in time. Basically, this is a question on semantics and interpretation and a side-issue in the accident migration theory.

Stein correctly points out two serious flaws in the accident analysis of Boyle and Wright.[24]

1. The lack of a method for accounting for regression-to-the-mean bias and

2. The time effect over the long period of time covered.

Basically, we made the same criticisms as Stein in our literature review and concur on both issues.

The authors reply to Stein acknowledges the potential for regression-to-mean but claim they made an assessment of this and did not find it to be a problem in their data since the numbers of accidents is low and the distribution skewed close to zero.[25] Thus, reverse regression-to-the-mean for the migration sites is not going to be severe since they can't get lower than zero. They do not respond to the regression-to-the-mean bias in their treatment sites. The authors acknowledged a problem in the way the time window was defined, but feel this would not affect their results. Again, it is not clear why the authors do not entertain a better statistical method for handling regression-to-the-mean or time and I support Stein's criticisms.

Another manuscript presents the analysis of a data set to show that reverse regression-to-the-mean is a possible explanation for accident migration.[9] The data used in this study consisted of four years of accident data (1979-1982) from the Lothian Road Accident Statistics System. This system consists of a network of 3250 nodes (intersections) and 2600 links (highway sections greater than 40 m in length). Both accident totals and annual average traffic flow data were available. Although no "treatment" was actually administered to these sections, the four year time period was divided into 2 years before and 2 years after.

"Blackspots" were identified according to three different criteria--total accidents, accidents per million vehicle kilometers, and the difference in before/after totals. Varying levels of these criteria were selected, e.g., total accidents greater than 6, 5, 4, 3, 2, 1, etc. Then the before/after accident totals were compared for both the selected "blackspot" areas and the non-blackspot areas. There was a systematic increase in accidents at the non-blackspot areas and a systematic decrease in accidents in the blackspot areas. The increases in the non-blackspot areas were of the same magnitude as that observed by Boyle and Wright and increased in magnitude as the "blackspot" selection criteria weakened. The authors interpret this, then, as strong evidence that the "accident migration" phenomenon could well be attributed to this reverse regression-to-the mean phenomenon.

The methodology used to analyze the data was empirical bayes, EB, method of moments estimation on the negative binomial distribution. Using this method, McGuigan showed that if all sites with 7 or more accidents had been treated, "an increase in accidents of about 10 percent could be expected at the untreated sites on the basis of regression-to-the-mean alone, notwithstanding any migration effect".

A note follows regarding Steins criticism of Boyle and Wright which supports Stein's claims and further alleges that Boyle and Wright data set was biased. In conclusion, McGuigan warns that any accident migration effect must be measured relative to the change in accidents expected based on regression-to-the-mean bias.

This is an excellent article and clearly identifies a potentially serious problem in trying to measure accident migration. The only criticism is that exposure data, though available, was only used to select "blackspot" areas and never used in the analysis. Given the diversity of sites which were combined, it is highly likely that the EB assumption of exchangeability did not hold. Also, he uses the Abbess, Jarrett, Wright, and Hauer method of moments as opposed to maximum likelihood.

This study is extremely important to our study as it identifies a potentially serious problem in assessing accident migration. Regression-to-the-mean adjustments appear to be as critical a part of measuring accident migration as in measuring a treatment's effect.

In a letter to the editor, Boyle and Wright support their conclusion that there was a true accident migration effect which could not be accounted for by regression-to-the-mean.[12] They attack McGuigan's analysis of the Lothan data stating that it is unreasonable to assume all blackspot areas would have been treated. Nonetheless, whether all areas were treated or not, the regression-to-the-mean phenomenon was evident in the non-blackspot areas and this is McGuigan's point.

McGuigan responds to Boyle and Wright stating that he recognized the fact that all blackspot areas would not be treated, but nonetheless, wanted to point out the possibility of regression-to-the-mean as a factor in accident migration, not "refute the accident migration theory".[26] He further adds that, in personal communication, he has been told that the mean number of accidents, using the method of moments, for Boyle and Wright's data set was very small, i.e., two, and in this case would support the existence of a true accident migration effect, with little regression-to-the-mean aberration. He does raise the issue that it would be interesting to compute the maximum likelihood estimate of this parameter to get a more exact estimate. McGuigan adequately responded to Boyle and Wright.

This paper examines the effects accident migration versus reverse regression-to-the-mean using a data set from Ebbeck, 1976.[13] Ebbeck noted that whereas accidents were reduced by 50 percent at the treated sites, "the total area accidents are not being reduced, they are just being rearranged". The data set consisted of two year before and after accident data for 222 intersections of one-way streets converted from one-street stopped to all-way stop control in the four-year period, 1970-1973. Similar accident data was also available for all other intersections in the study area.

The methodology used for analysis was the empirical Bayes procedure using method of moment negative binomial parameter estimates. Using this methodology, Persaud found accident migration to be a plausible, realistic effect for the intersections immediately surrounding the converted streets. This was even more dramatic when right angle accidents were extracted. Thus, using a methodology which accounted for potential regression-to-the-mean, this study found that, "the all-way stop conversion programme . . . is causing the accident increases at the remaining one-street-stopped

intersection." The author recommends that this be explored further, perhaps through a planned study.

The primary criticisms are that no exposure data is used in this study bringing into question the EB assumption of exchangeability and that method of moments rather than maximum likelihood was used to estimate the parameters. The findings are, nonetheless, intriguing and it would be most interesting to subject this data to a more rigorous analysis. This study is extremely relevant to the accident migration study as it is the first in the literature to suggest that accident migration is a real phenomenon apart from regression-to-the-mean.

In an excellent review article, the authors state that two commonly assumed facts regarding evaluating a remedial treatment are now being challenged:

1. The remedial treatment can be regarded as a self-contained process and
2. The overall effectiveness can be measured in terms of accident total before and after treatment [27]

The authors claim that four processes need to be considered:

1. Regression-to-the-mean
2. Risk compensation
3. Accident migration
4. Interaction between subjective and objective risk

The paper then expands on each of these processes giving an excellent historical literature accounting of each including proposed methodologies. The authors recommend "controlled experiments" to further explore these four processes.

The paper is very comprehensive and well-written, however, it is not without bias, especially in its reporting of the series of dialogues on the author's 1984 paper. They even credit McGuigan as acknowledging that "the regression-to-the-mean effect has since been ruled out as a factor in accident migration" - a statement which McGuigan reacts to strongly in a later letter.

Maher shows, in another manuscript, that yet another very important and real statistical phenomenon must be taken into consideration in evaluating the "accident migration" effect--namely, the spatial correlation among neighboring sites[22]. Thus, not only do we have a potential "reverse regression-to-the-mean phenomenon confounding the true accident migration effect, but the fact that because a site "neighbors" a treated site, there is an underlying correlation between these sites which can confound the migration effect.

The author demonstrates this correlational effect by simulation studies using two different procedures. In one, the moving average method, a moving average was computed using a lognomial distribution assumption on the mean accident rates. In the second, the clustering method, a number of different "centers" was selected on a grid and then the surrounding means generated such that they decreased as distance from these centers increased. The results of these two procedures, he claims, was similar.

The statistics used to compare before/after reductions was the usual percent reduction in before to after. This was computed for:

1. Adjacent neighbor sites from "treated" sites, P_{mig},

2. "Treated" sites, P_{reg},

3. All untreated sites whether adjacent to the treated sites or not, P_{unt}.

"Treated" sites were not actually treated, of course, but selected based on accidents greater than some selected threshold criterion, k. Since there was no actual treatment, and change before-to-after in the treated sites was attributable to regression-to-the-mean, P_{reg}, reverse regression-to-the-mean, P_{unt}, or accident migration, P_{mig}.

As the site selection threshold k, increased, the size of reverse regression-to-the-mean effect falls steadily to zero, supporting McGuigan's observation. The change in accident at the neighboring sites (the "migration" effect) did not fall nearly as quickly and tended to remain at about the 10-percent level. The conclusion: "the prime determinant of the magnitude of "the migration effect "is the amount of spatial correlation between the" site means. As this correlation approaches zero, the "migration" effect approaches the reverse regression-to-the-mean effect.

These results were based on a limited number of simulations and choices of k values (2 examples with 10 simulations and k varying from 7 to 18 with mean and shape parameters of 4.74 and 1.72 and 3.53 and 1.78, respectively).

In summation, the author notes that any model which tries to account for the effect of accident migration must incorporate some spatial correlation among neighboring sites as well as adjusting for regression-to-the-mean effect. He proposes more mathematical and less simulation based research to develop formulae, graphs or tables which would allow the expected size of the 'migration' effect to be determined from a knowledge of the values of the important parameters in the accident parameters in the accident frequency model.

The basic discovery in this manuscript, that of spatial correlation, is a vital one to the problem of isolating true accident migration. However, as the author notes, this paper just touches the subject showing the effect of spatial correlation in a small and

limited simulation study. To truly grasp this phenomenon, a much larger simulation study with more choices in means and site selection criteria is in order. Also, the author uses two methods to "impose" a type of spatial correlation, but neither allows for the control of this spatial correlation of varying levels, high to low. For example, the moving average process (the author does not state the choice of lag so I assume it is one) will result in some value of spatial correlation, but what is it for these simulations? In the second procedure, it would appear that some type of correlation is imposed as the author say the site means decrease as distance from the center increases, but how much correlation is imposed and how was this done. It would appear that a better way to conduct a simulation study would be to impose varying degrees of spatial correlation on the structure and control for the amount of correlation.

A further important limitation of this study, which is imbedded in the spatial correlation concept, is exposure (traffic flow). Also, the study neglects to propose a statistical method for measuring accident migration which would incorporate both the concept of spatial correlation and regression-to-the-mean. This manuscript is extremely relevant to our study as it is the first to recognize that another process, a part from regression-to-the-mean, is an integral part of the accident migration phenomenon-namely, spatial correlation among neighboring site mean accidents.

In a letter to the editor, McGuigan reacts with surprise and alarm at Wright and Boyles claim, in the previous paper, that McGuigan supported the ruling out of regression-to-the-mean effect in accident migration.[28] McGuigan, in a clever, third person sentence, says he feels "rather put out at having been so substantially misrepresented"!

In another letter to the editor, the author applauds the strides made in the recent statistical method of Wright and Boyle but warns against losing site of the practical engineering interpretations of studies.[29] He specifically sites the pedestrian guard rail example stating that a better analysis would have been to compare numbers of adult and child pedestrian casualties before and after erection of the guardrails because previous studies have shown an interaction in these two types of accidents, namely, adult pedestrian accidents decreased while children pedestrian accidents decreased while children pedestrian accidents increased. The author urges that what is needed is "not so much for better statistical methods as for better engineering". I agree that sound engineering judgement must go hand-in-hand with statistical methodology development, but *both* are needed, not one at the expense of, or in void of, the other.

The authors reply to both of the proceeding letters.[30] However, their reply to McGuigan is more of a restatement of their previous claim that they do not feel regression-to-the-mean is a factor *in their particular study*. They promise, however, to examine the correlational structure of the data set and report on this later.

In another paper, Adams summarizes the state-of-the-art in accident analysis, to-date.[31] He claims there is uncertainty in this field on both how to measure safety effectiveness and how to recognize it when it occurs. Two factors which seriously obscure these results are: (1) poor statistical methods and (2) "terrible" data. He presents examples which show how injury data and fatality data frequently lead to opposite conclusions.

Adams addresses the issue of accident migration focusing on risk compensation and human behavior as the primary explanations of this phenomenon. He also defines migration as changes in types of accidents rather than strict geographical migration. In conclusion, he recommends that other indirect measures be used to evaluate safety, other than accidents, and that more sophisticated statistical techniques be developed to address regression-to-the-mean and accident migration. Although this is a good "opinion" and "summary" paper, it does little to offer solutions to the problem and sheds a very pessimistic light on the hope of ever finding a solution. In that sense, it is not very motivational for the potential researcher in this field but is, I'm afraid, very realistic.

In a more recent paper, two evaluations of lane additions to interstate highways are presented.[32] In both cases the additional "non-standard, mixed-flow" lane was created by restriping the highway to use part of the interior shoulder as a traffic lane. Changes in accidents from before to after were recorded throughout the section where the widening took place, and for several miles upstream and downstream from the modification. Accidents were adjusted by means of a control (comparison) group.

In the first evaluation, accident migration was convincingly demonstrated. In the second evaluation there was no evidence of accident migration. It should be noted that of the three conditions we are considering for the evaluation of accident migration, the two case studies presented here have <u>few</u> treatment sites but <u>many</u> accidents per site. In both of these studies only one treatment site is considered, but the number of accidents recorded at that site is in the hundreds.

<u>Case 1</u>

Approximately three miles of an interstate highway (I-405) in Los Angeles County were restriped to add an additional lane of traffic. Accident data were collected for these three miles of highway and for six miles of highway upstream and six miles of highway downstream from the project (treatment) area, for four years prior to treatment and two years following treatment. Further, all accidents on I-405 in Los Angeles County outside the study area (i.e., outside the treatment area and the six miles up and downstream from the treatment area) were collected for the same time periods, i.e., four years before and two years after treatment. The accident data collected, and the analyses based upon the data are shown below:

Table 39. Study area section I.

Control Time	Upstream	Project	Downstream	Study Area	Total Area
Before (4 years)	1,814	719	1,180	3,713	4,541
After (2 years)	903	497	971	2,371	2,944
Percent Change	23.22	-6.62	-26.93	1.50	
Chi-square (1 df)	31.55	1.04	23.38	0.18	

Positive changes in accidents represent reductions; negative changes in accidents represent increases. The percent changes in accidents and the resultant chi-squares are based upon cross-product ratios (i.e., odds ratio) formed with the accidents that occurred in the control area. Clearly, accidents decreased upstream and increased downstream following treatment. Throughout the three miles that were treated (i.e., the project area) accidents did not change significantly. These results are interpreted to indicate that there was a traffic "bottle neck" upstream from the project area which was displaced downstream following treatment, with a concomitant migration of accidents (particularly rear-end accidents) downstream.

Further analysis of the accidents in case 1 revealed that following treatment injury accidents and rear-end accidents decreased upstream (18 and 36 percent, respectively) and increased downstream (25 and 37 percent, respectively). Sideswipe accidents increased in the project area by 40 percent. All of these percentage changes were significant at $\alpha = 0.05$.

Case 2

At a second treatment location on I-405 in Orange County approximately two miles of highway received an added lane by restriping, on the northbound side of the Interstate. Accident data were collected for the northbound lanes of traffic (three years prior to treatment and two years after treatment) along six miles of highway upstream and six miles highway downstream from the project (treatment) area, as well as throughout the two miles of highway that received an added traffic lane. The control area in this second case study consisted of the adjacent, southbound lanes of traffic across from the study area. Accident data collected in the study area and the control area, and the accompanying analyses, were as follows:

Table 40. Study area section II.

Control Time	Upstream	Project	Downstream	Study Area	Total Area
Before (3 years)	282	213	253	748	969
After (2 years)	211	99	203	513	663
Percent Change	-9.36	32.07	-17.27	-0.24	
Chi-square (1 df)	0.74	8.69	2.22	0.0009	

In this second case study, using the same statistical procedures employed in the first case study, accidents increased in the project area by 32 percent, a significant increase at $\alpha = 0.05$. However, upstream and downstream from the project area, no significant change in accidents was observed. Further analysis of the data suggested that rear-end accidents were significantly reduced in the project area following treatment.

The statistical procedures used by the authors are neither well explained, nor well documented. The percent changes in accidents, as well as several of the chi-squares shown above, are our own calculations. The most conspicuous statistical shortcoming of this paper is the reuse of the same comparison data to carry out four inferential tests of significance as if they were independent, i.e., without any adjustment in α level. Although this is "poor form," the findings of significance are so strong that this shortcoming is of relatively little consequence.

In the final paper, reviewed for this study, Wright et. al. summarize recent developments in tackling two major problems in accident analysis--namely, regression-to-the-mean and accident migration[33].

The authors contend that there are 3 basic methods for evaluating accidents:

1. Matched control methods
2. Hauer's method
3. Their method

By Hauer's method it appears they are referring to his very early papers, not his later ones. I can make no distinction between methods 2 and 3 as these authors currently describe them and it would appear that the true originator of these methods is Robbins.

The authors then identify four problem areas:

1. How the treated sites are "defined"
2. How the reference sites are defined
3. The assumption of a gamma prior - i.e., is it reasonable?

4. Is their some other casual factor that has not been accounted for that would violate the assumption that the site means all fluctuate about the same overall accident frequency (exchangeability).

In summation, the authors note:

1. An objective definition of what constitutes the population of sites (reference plus treatment group) is essential.
2. The accuracy of estimates can be improved by dividing sites into homogeneous subgroups (to satisfy exchangeability).
3. Dividing type of accidents into "treatable" and "untreatable" groups will not necessarily improve precision but may remove bias.
4. The EB methods do not appear to be sensitive to the assumed form of the prior distribution.

Finally, the authors close with acknowledging that "none of the methods...have been tested against real data in a controlled experiment". The "controlled" experiment they propose is to select sites for treatment and compute a predicted regression-to-the-mean effect. Then, withhold treatment. They acknowledge that few, it any, "local authorities" would consider such an extreme experiment.

There appears to be some undercurrent about whose responsible for what method. There is little, if any, difference between the authors proposed method and Hauer's and both methods really have their roots in the statistical literature of the 50's. At any rate, it was interesting to note that, without explicitly using those terms, the authors show concern for the crucial assumptions of exchangeability and identification of an appropriate reference group. This appears to be the first acknowledgement of these important concerns in all the literature to date on empirical Bayes application in accident analysis.

Although the authors mention accident migration and promise to summarize the recent developments in this area as well as regression-to-the-mean, they focus on the latter and never again mention the former. Still, the paper is relevant since regression-to-the-mean is an obvious factor in assessing accident migration.

REFERENCES

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

2. A.J. Boyle and C.C. Wright, 1984. "Accident Migration After Remedial Treatment at Accident Blackspots." *Traffic Engineering and Control*, 25 (5), pp. 206-267.

3. K.W. Huddart, 1984. "Accident Migration - True or False." *Traffic Engineering and Control*, 25 (5), p. 267.

4. Bhagwant Persaud, 1986. "Relating the Effect of Safety Measures to Expected Number of Accidents." *Accident Analysis and Prevention*, 18 (1), February, pp. 63-70.

5. C. Abbess, D.F. 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*, 22 (10), October, pp. 535-542.

6. E. Hauer, 1986. "On the Estimation of the Expected Number of Accidents." *Accident Analysis and Prevention*, 18 (1), pp. 1-12.

7. K.W. Huddart, 1984. Reply to letter by Boyle and Wright. *Traffic Engineering and Control*, 25 (718), July/August, p. 392.

8. H.S. Stein, 1984. "Accident Migration after Remedial Treatment of Accident Blackspots." *Traffic Engineering and Control*, 25 (12) December, p. 618.

9. David McGuigan, 1985. "Accident Migration or a Flight of Fancy." *Traffic Engineering and Control*, 26 (4), April, pp. 229-233.

10. A.J. Boyle and C.C. Wright, 1984. "Accident Migration - a Rejoinder." *Traffic Engineering and Control*, 25 (7/8), p. 372.

11. A.J. Boyle and C.C. Wright, 1984. "Risk Compensation as a Possible Cause of Accident Migration: a Reply." *Traffic Engineering and Control*, 25 (9), September, p. 455.

12. A.J. Boyle and C.C. Wright, 1985. "Accident Migration - or a Flight of Fancy." (Letter to the Editor). *Traffic Engineering and Control*, 26 (7/8), p. 389.

13. Bhagwant Persaud, 1987. "Migration of Accident Risk after Remedial Blackspot Treatment." *Traffic Engineering and Control*, 28 (1), pp. 23-26.

14. Olga J. Pendleton, 1991. "APPLICATION OF NEW ACCIDENT ANALYSIS METHODOLOGIES Volume I - General Methodology." FHWA Report FHWA-RD-90-091. Federal Highway Administration, Washington, D.C., p. 76.

15. Olga J. Pendleton, Ofelia Gonzalez, and Horacio Duarte, 1991. "APPLICATION OF NEW ACCIDENT ANALYSIS METHODOLOGIES Volume II - A Users Manual for BEATS." FHWA Report FHWA-RD-91-014. Federal Highway Administration, Washington, D.C.

16. Carl N. Morris, Cindy L. Christiansen, Olga J. Pendleton, 1991. "APPLICATION OF NEW ACCIDENT ANALYSIS METHODOLOGIES Volume III - Theoretical Development of New Accident Methodology." FHWA Report FHWA-RD-91-015. Federal Highway Administration, Washington, D.C.

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

18. C.P. Brinkman, 1986. "Improved Highway Safety Analysis." *Public Roads*. September, 50 (2), pp. 49-54, Federal Highway Administration, Washington, D.C.

19. J. Higle and J. Witkowski, 1988. "Bayesian Identification of Hazard Locations." Transportation Research Record 1185, pp. 24-36, Transportation Research Board, Washington, D.C.

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

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

22. M.J. Maher, 1987. "Accident Migration - a Statistical Explanation?" *Traffic Engineering and Control*, 28 (9), September, pp. 480-483.

23. C. Abbess, D.F. Jarrett, and C.C. Wright, 1983. "Bayesian Methods Applied to Road Accident Blackspot Studies." *The Statistician*, 32.

24. H.S. Stein, 1984. "Accident Migration after Remedial Treatment of Accident Blackspots." *Traffic Engineering and Control*, 25(12), p. 618.

25. C.C. Wright and A.J. Boyle, 1984. "The Authors Reply to Stein's Letter." *Traffic Engineering and Control*, 25 (12), pp. 618, 619.

26. D.R.D. McGuigan, 1985. "Reply to letter by Boyle and Wright." *Traffic Engineering and Control*, 26 (7/8), pp. 390-391.

27. C.C. Wright and A.J. Boyle, 1987. "Road Accident Concision and Engineering Treatment. A Review of Some Current Issues." *Traffic Engineering and Control*, September, pp. 475-479.

28. D.R.D. McGuigan, 1987. "Letter to the Editor." *Traffic Engineering and Control.* 28(11), p. 648.

29. D. Stewart, 1987. "Letter to the Editor." *Traffic Engineering and Control.* 28(11), p. 648.

30. C.C. Wright and A.J. Boyle, 1987. "Letter to the Editor." *Traffic Engineering and Control.* 28(11), p. 648.

31. John G.N. Adams, 1988. "Evaluating the Effectiveness of Road Safety." July/August, p. 344-352.

32. Douglas Levine, W. Thomas, F. Golob, and Wilfred W. Recker. 1988. "Accident Migration." *Traffic Engineering and Control.* December, pp. 624-629.

33. C.C. Wright, C.R. Abbess, and D.F. Jarrett, 1988. "Estimating the Regression-to-the-Mean Effect Associated with Road Accident Blackspot Treatment." Accident Analysis and Prevention. 20(3), pp. 199-214.

&U.S. GOVERNMENT PRINTING OFFICE: 1992-625-979-No Program