PB97-116198

U.S. Department
of Transportation

Federal Highway
Administration

Publication No. FHWA-RD-96-141
November 1996

Indirect Methods to Account for Exposure in Highway Safety Studies

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

FOREWORD

One of the more complex issues in highway safety evaluation is how to incorporate some measure of exposure, some numerical measure of the opportunity for an accident to occur in the analysis. On a roadway section, for example, exposure might be vehicle-miles traveled (VMT). At fixed points, such as intersections, exposure might be some function of the annual average daily traffic (AADT) of the intersecting roads. However, the appropriate measure of exposure may not be at all clear, such as in driver- or vehicle-based analyses. In assessing the involvement of older drivers in left-turn accidents, for example, how does one measure the exposure of drivers by age at a specific site? Because of the difficulty of obtaining numerical exposure estimates, indirect methods have evolved for treating the exposure issue in a surrogate manner.

This report is intended to be an instructional text on these methods for use by the highway safety research community. The indirect methodologies presented here include before/after evaluation designs, case-control studies, and the induced-exposure methodologies. The numerical computations for each methodology are presented in simple, step-by-step formulas, followed by a numerical example. A discussion of both the advantages and disadvantages and appropriate and inappropriate applications are also presented. It is hoped that this report will serve as a comprehensive guide for the highway safety researcher, both in designing and evaluating highway safety studies.

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

NOTICE

BIBLIOGRAPHIC INFORMATION

PB97-116198

Report Nos:

Title: Indirect Methods to Account for Exposure in Highway Safety Studies.

Date: Nov 96

Authors: O. J. Pendleton.

Performing Organization: Texas Transportation Inst., College Station.

Performing Organization Report Nos: FHWA/RD-96/141

Sponsoring Organization: *Federal Highway Administration, McLean, VA. Office of Safety and Traffic Operations Research and Development.

Contract Nos: DTFH61-93-R-00123

Type of Report and Period Covered: Final rept. Oct 93-May 96.

NTIS Field/Group Codes: 85D (Transportation Safety), 85H (Road Transportation)

Price: PC A05/MF A01

Availability: Available from the National Technical Information Service, Springfield, VA. 22161

Number of Pages: 69p

Keywords: *Highway accident potential, *Accident risks, Highway safety, Accident analysis, Comparison, Case studies.

Abstract: One of the more complex issues in highway safety evaluation is how to incorporate some measure of exposure, some numerical measure of the opportunity for an accident to occur in the analysis. On a roadway section, for example, exposure might be vehicle-miles traveled (VMT). At fixed points, such as intersecting, exposure might be some function of the annual average daily traffic (AADT) of the intersecting roads. However, the appropriate measure of exposure may not be at all clear, such as in driver-or vehicle-based analyses. In assessing the involvement of older drivers in left-turn accidents, for example, how does one measure the exposure of drivers by age at a specific site. Because of the difficulty of obtaining numerical exposure estimates, indirect methods have evolved for treating the exposure issue in a surrogate manner. This report is intended to be an instructional text on these methods for use by the highway safety research community. The indirect methodologies presented here include before/after evaluation designs, case-control studies, and the induced-exposure methodologies. The numerical computations for each methodology are presented in simple, step-by-step formulas, followed by a numerical example. A discussion of both the advantages and disadvantages and appropriate and inappropriate applications are also presented. It is hoped that this report will serve as a comprehensive guide for the highway safety researcher, both in designing and in evaluating highway safety studies.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-RD-96-141		
4. Title and Subtitle INDIRECT METHODS TO ACCOUNT FOR EXPOSURE IN HIGHWAY SAFETY STUDIES		5. Report Date November 1996 6. Performing Organization Code
7. Author(s) Olga J. Pendleton		8. Performing Organization Report No.
9. Performing Organization Name and Address Texas Transportation Institute College Station, TX 77843-3135		10. Work Unit No. (TRAIS) A3b 11. Contract or Grant No. DTFH61-93-R-00123
12. Sponsoring Agency Name and Address Office of Safety and Traffic Operations R&D Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296		13. Type of Report and Period Covered FINAL REPORT Oct. 1993 - May 1996 14. Sponsoring Agency Code
15. Supplementary Notes Contracting Officer's Technical Representative: Michael S. Griffith, HSR-30		

16. Abstract

One of the more complex issues in highway safety evaluation is how to incorporate some measure of exposure, some numerical measure of the opportunity for an accident to occur in the analysis. On a roadway section, for example, exposure might be vehicle-miles traveled (VMT). At fixed points, such as intersections, exposure might be some function of the annual average daily traffic (AADT) of the intersecting roads. However, the appropriate measure of exposure may not be at all clear, such as in driver- or vehicle-based analyses. In assessing the involvement of older drivers in left-turn accidents, for example, how does one measure the exposure of drivers by age at a specific site? Because of the difficulty of obtaining numerical exposure estimates, indirect methods have evolved for treating the exposure issue in a surrogate manner.

This report is intended to be an instructional text on these methods for use by the highway safety research community. The indirect methodologies presented here include before/after evaluation designs, case-control studies, and the induced-exposure methodologies. The numerical computations for each methodology are presented in simple, step-by-step formulas, followed by a numerical example. A discussion of both the advantages and disadvantages and appropriate and inappropriate applications are also presented. It is hoped that this report will serve as a comprehensive guide for the highway safety researcher, both in designing and in evaluating highway safety studies.

17. Key Words	18. Distribution Statement
Safety evaluation, case-control, induced exposure, exposure, indirect exposure, comparison group, control group, matched comparison, yoked comparison, test for comparability.	No restrictions. This document is available to the public through the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	71	

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

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

APPROXIMATE CONVERSIONS FROM SI UNITS

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

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

(Revised September 1993)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Measurement of risk potential in accident analyses requires reliable measurements of the exposure of the at-risk entity (the motor vehicle or driver). Estimation of the probability of the occurrence of an accident necessarily hinges on the ability to estimate this risk potential. In the medical field, the analogy can be drawn to the probability of acquiring some disease. This probability is dependent on an individual's "exposure" to the disease. It is from this clinical use of the term *exposure* that the concept of exposure in accident analysis evolved. This concept, as related to motor vehicle accidents, however, does not always have a unique and precise definition. Depending on the application, e.g., the highway safety issue, exposure may be an absolute per unit measure, a relative measure, or an indirect measure.

Because of the difficulty in defining exposure for certain highway safety applications, there have been numerous attempts to perform accident analyses without using a direct measure of exposure. Exposure in the accident analysis field is defined as some numerical measure of the opportunity for an accident to occur. On a roadway section, for example, exposure might be vehicle-miles traveled (VMT). VMT is a function of the average annual daily traffic (AADT) and section length. At fixed points, such as intersections, exposure might be some function of the average annual daily traffic on the intersecting roads. However, it is not always clear what the exposure measure should be and hence there is a need for indirect methods to account for them.

Indirect methods are generally used when exposure data are either unavailable, unreliable, or undefinable. When these measures are appropriate, available, and reliable, rate-based accident analyses are possible (although rate-based methods assume that accidents and exposure are linearly related). Otherwise, it is necessary to use some methodology that accounts for exposure in an indirect manner. This report will present the most common methods used in before/after treatment evaluation studies. Each methodology will be defined and worked examples will be presented. A critique of each methodology, including the pros and cons, will be included. The reader of this report should be able to determine, from this presentation, which methods would be most suitable for their particular application and follow the step-by-step computational methods for applying the methodology to their research.

The methodologies contained in this report are:

1. Before/after treatment evaluation studies using accident counts.

   a. Without a comparison group.

   b. With comparison groups or conditions and only one before and one after time measurement.

   c. Case b above, but with multiple measurements in time to check for comparability of the comparison groups or conditions.

   d. With comparison groups or conditions and only one before and one after time measurement, but with replications.

2. Case-control methods.

    a. Classical.

    b. Double-pair comparison.

3. Induced-exposure methodologies in before/after treatment evaluation studies.

Table 1 summarizes these methods according to the amount of data required. The two-letter, one-number nomenclature for the rows (i.e., BY1) represent the time period of accident counts required, as a minimum, to apply the procedure (columns). B represents the before-treatment time period; A represents the after-treatment time period. Y1 and Y2 represent the first and second years of data, sequentially. If more data than the minimum are available, extensions of the methodologies can be performed, as will be shown.

Table 1. Minimum data requirements for indirect-exposure methodologies.

Groups	Time	1a	1b	1c	1d	2a	2b	3
Treatment	BY1				X			
	BY2	X	X	X	X	X	X	X
	AY1	X	X	X	X	X	X	X
	AY2							
Matched (Yoked) Comparison	BY1				*			
	BY2		*	X	*	*	*	*
	AY1		*	X	*	*	*	*
	AY2							
Unmatched (Reference) Comparison	BY1				*			
	BY2		*		*	*	*	*
	AY1		*		*	*	*	*
	AY2							

Note: The *'s indicate that these methodologies can be applied to data that either have a matched comparison group or that have an unmatched comparison group. It does not mean that both types of comparison groups are required by the methodology.

The definition of these methods will become clearer in the text. However, most of these methods are very commonly used and will be immediately familiar to the readers of this report. Only the terms assigned to these methods are relatively new and may be unfamiliar to many. The following topics will be presented and discussed for each methodology:

- Introduction and historical application of the methodology.
- Description, numerical computations, and examples.
- Critique.

CHAPTER 2: BEFORE/AFTER EVALUATION DESIGNS

INTRODUCTION AND HISTORICAL APPLICATION OF THE METHODOLOGIES

One of the earliest methods used to evaluate a treatment effect is based on only two measurements — one before treatment and one for some period after treatment.[1,2,3] These measurements are frequently annual accident counts. Some time period is usually left between the before and after periods to allow for driver adjustment to the change, construction period, etc. This is by far the simplest, but weakest, methodology and has serious deficiencies. For example, there is no "control" for any changes that might have occurred in the two time periods that were independent of the treatment, but which could affect the accident counts, such as an increase in traffic, differences in weather, etc. If a severe snowstorm occurred in the after period, accidents could have increased for this reason alone, and a treatment that really improved safety would be masked by this unrelated event. There is also the issue of the regression to the mean (rtm) phenomenon. Since it is highly likely that these sites were selected for treatment because they had an unusually high number of accidents before the treatment, accidents might decrease in the after period even without treatment simply because these sites are "regressing" or returning to the true number of accidents. In these cases, a treatment that is really ineffective might be classified as effective because the accidents decreased — the decrease being merely an artifact of rtm.

Historically, although the rtm phenomenon was recognized for some time, it was the lack of control issue that first received attention.[4] The methodology and inferences that can be drawn depend upon how the comparison group was selected. A comparison *group* is actually a group of sites selected as being similar enough to the treated sites to adequately reflect what would have happened to the treated sites had they not been treated. There are typically two ways in which the comparison group is selected. If the comparison group sites were selected to be matched and paired (either by proximity or roadway geometrics), this is referred to as "yoked" (like oxen) comparison groups. These are also sometimes referred to as "matched" comparison sites, although the term *matched* connotes a much more rigorous matching than what actually occurs in practice.

Another way to select the comparison group is from the pool of potentially treated sites that were not selected for treatment. Empirical Bayes methods refer to this type of comparison group as the reference group. That is, the treatment group and comparison group together represent a sample from the population of sites that would be candidates for the treatment. Since there is no one-to-one matching in the selection process, it is not possible to estimate site-specific treatment effects. On the other hand, by assessing the effectiveness of a treatment imposed at several different sites, there is some degree of protection against overestimating or underestimating treatment effectiveness due to the uniqueness of a particular treatment site. There is also the potential for increasing sample size, since comparison sites selected in this manner are far easier to find than by the yoked method. Increasing sample sites increases the statistical power of the test.

A comparison *condition* is an identified factor on which the accidents at the same sites can be partitioned so that one level of the partitioning reflects the treatment effect and the other represents conditions where the treatment would not be expected to have any effect. For example, in the evaluation of raised pavement markers, since they are thought to be more effective in reducing accidents at night than during the day, daytime accidents at the treated sites can be considered a comparison condition for nighttime accidents at those same sites. The daytime accident frequencies then serve as "controls" for the nighttime condition. Another often-used comparison condition is wet/dry, although it is more difficult to accurately define a "wet" accident than a "dry" accident.

When more than one point is available for the before period, the methodology can be extended to provide a check on the validity of the comparison group or condition.[3,6] Generally, this is done on yearly units. If 2 or more years of accident data can be retrieved, the before data can be subjected to a "test for comparability." If this test is failed, the safety treatment evaluation is questionable because the comparison group or condition is not behaving like the treatment group and may not be a valid control for before/after differences. If more than 1 year of after data is available, one can test "return-to-baseline" theories, although this is seldom done in practice.

The methodology of empirical Bayes was applied to accident analyses to adjust for regression to the mean. Although there are several variations on the application of this methodology, the basic philosophy is the same.[7-11] The above methodologies are all based on the premise that the number of accidents observed in the before period is a good "estimate" of what would be expected in the after period without the treatment. If regression to the mean is a problem, i.e., if the most hazardous sites are treated and these sites were experiencing an unusually high number of accidents, then the before accidents do not represent good estimates of what to expect apart from treatment. The empirical Bayes methods all attempt to adjust for this by providing better estimates of what to realistically expect. These estimates adjust for regression to the mean by providing some number between that observed and an estimate of the "true" mean for the treated sites. If the accidents are unrealistically high, the empirical Bayes estimate will "shrink" the observed mean closer to the "true" mean and hence the estimate of what to expect will be smaller than the observed before-accident count. This methodology is computationally quite complex and requires sophisticated software to compute.

1.a. Before/after without comparison group.

Description and numerical computations

Although there are several equivalent variations of methods for performing these before/after tests, the most common ones will be presented here. Highway safety studies are typically based on accident frequencies before and after a highway treatment. The most elementary method (1.a.) is the simple before/after test. The estimate of the treatment's "effect" is:

$$E = \frac{(D - C)}{C} 100 \qquad (1)$$

where E represents the percent change in accidents after treatment, positive values imply an increase in accidents and negative values imply a decrease in accidents after treatment; D is the number of accidents at treatment sites after treatment; and C is the number of accidents at the treatment sites before treatment.

A test statistic for testing the statistical significance of this estimate is:

$$T = \frac{D - C}{\sqrt{D + C}} \qquad (2)$$

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

Numerical example

The data for this numerical example will also be used with other methodologies and will be referred to as the Michigan raised pavement marker study. These data were obtained from the Michigan data in the Highway Safety Information System (HSIS) data base maintained by the University of North Carolina Highway Safety Research Center. A total of 17 locations, comprising 90.30 km (56.11 mi) where raised pavement markers were installed between August 1 and September 30, 1989, represent the treatment sites for this analysis.

The total accident counts in these data, which reflect the generic values used in equation 1, are:

Table 2. Total before/after accidents for raised pavement marker example.

Time	Treatment Sites
Before	567
After	564

For the before/after analysis without a comparison group, the estimate in equation 1 would be computed as:

$$E = \frac{(564 - 567)}{567}100 = -0.53 \qquad (3)$$

This computation is interpreted as a 0.53 percent **reduction** (due to the negative sign) in accidents due to raised pavement markers. To determine if this reduction is statistically significant, equation 2 is used to compute the test statistic as:

$$T = \frac{564 - 567}{\sqrt{564 + 567}} = -0.0892 \qquad (4)$$

Since the Z-value at the 0.05 level is 1.96 and the absolute value of T is less than 1.96, the conclusion would be that raised pavement markers did not significantly change the number of accidents at these sites. In fact, there was actually an increase in accidents, but since this is not statistically significant, it is just considered a random fluctuation about zero — no change.

1.b. Before/after with comparison groups or conditions: one before/after measurement.

Description and numerical computations

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

Table 3. Generic before/after with comparison group.

Time	Comparison	Treatment
Before	A	C
After	B	D

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

$$O.R. = \frac{A/B}{C/D} \qquad (5)$$

where A/B are the odds of the before-to-after accidents in the comparison group (condition) and C/D are the same odds for the treatment group. Any differences in the before-to-after period will cancel out, assuming that the same differences occurred in both groups. If the comparison group is a valid comparison group, this will be the case. The treatment effect is estimated as this ratio minus 1 (times 100 as a percent change) and negative values imply a decrease in accidents due to the treatment.

There are several candidate test statistics for testing the statistical significance of the treatment effect, but the one that will be chosen here is the likelihood ratio chi-square. The formula for the likelihood ratio chi-square is:

$$L.R\chi^2 = 2\left(\sum\sum o_{ij} \ln \left(\frac{o_{ij}}{e_{ij}}\right)\right) \qquad (6)$$

where o_{ij} are the observed frequencies and e_{ij} are the expected frequencies. For the classical method, the expected frequencies are the same ones used in the Pearson chi-square, i.e., the product of the cells corresponding marginal frequencies divided by the total frequency. For example, the expected value for the first cell of the table, A, would be:

$$e_{11} = \frac{(A + C)(A + B)}{(A + B + C + D)} \qquad (7)$$

The test statistic in equation 6 is to be compared with the chi-square value with one degree of freedom or 3.84. If this calculated value exceeds 3.84, there is a significant change in accidents due to the treatment. If the odds ratio of equation 5 is negative, there has been a significant reduction; and if it is positive, there has been a significant increase.

This test is only justified, however, if the treatment and comparison groups pass the comparability test. In cases where only 1 year of before and 1 year of after data are available, this check for comparability cannot be done, and the reliability of the conclusions depends upon the validity of the assumption that these groups are comparable.

Numerical example

The data for the Michigan raised pavement marker study will be used in this example. Accident data were collected for 2 years before and 2 years after the treatment. Because of the design of this study, both a comparison *group* and a comparison *condition* are available for this example. A total of 42 sites, comprising 235.41 km (146.28 mi) where raised pavement markers were not installed and no other significant changes were made in roadway geometrics or operations data after August 1, 1987, were selected as reference (comparison) sites. Videodiscs of these sections were reviewed to ensure that roadway characteristics such as markings, curvature, etc. were comparable. Since these comparison group sites were not matched or "yoked" to the treatment sites, but rather were selected to represent the population of sites from which the treatment sites were selected, they can also be considered as reference sites.

Daytime accidents at the same sites were selected as the comparison condition for the treatment condition of nighttime accidents. *Day* was defined as 1 hour after sunrise to 1 hour before sunset, and *night* was defined as 1 hour after sunset to 1 hour before sunrise. Sunrise and sunset hours were selected based on a chart provided by the Federal Highway Administration (FHWA) that contained daily readings for sunrise and sunset for Lansing, Michigan. Locations not close to Lansing were adjusted according to the estimated time differences.

Table 4 represents the accident counts for the comparison group illustration. Nighttime accidents at the treated sites were compared to nighttime accidents at the reference (comparison) group sites.

Table 4. Total before/after accidents for raised pavement marker example.

Time	Reference (Comparison)	Treatment	Total
Before	656	286	942
After	645	275	920
Total	1301	561	1862

The expected values for the observed values of table 4 are shown in table 5.

Table 5. Expected values for before/after accidents for raised pavement marker with comparison group.

Time	Reference (Comparison)	Treatment
Before	658.19	238.81
After	642.81	277.19

The treatment effect as estimated by the odds ratio of equation 5 is:

$$O.R. = \frac{656/645}{286/275} = 0.98 \qquad \textbf{(8)}$$

The treatment effect is then 100(0.98-1.0) = -2.2 percent, i.e., a 2.2 percent reduction in nighttime accidents is estimated after installing the markers.

To determine if this reduction is statistically significant, the likelihood ratio chi-square of equation 6 is computed from the observed and expected values of tables 4 and 5. This value is -0.05, which, when compared to the critical value of 3.84, is not statistically significant at the 5 percent level of significance.

Because raised pavement markers are thought to be more effective at night, accidents at the treated sites can be partitioned into day/night categories and the daytime accidents can be used as a comparison condition. Table 6 lists the accident counts for the treatment and comparison conditions at the sites where the markers were installed. (Although 2 years of before data were available and a test for comparability can be done, the 2-year totals for before-and-after accidents are used both in tables 4 and 6.)

Table 6. Raised pavement marker data at treated sites.

Time	Treatment - Day	Treatment - Night	Total
Before	281	286	567
After	289	275	564
Total	570	561	1131

The odds ratio is:

$$\frac{281/289}{286/275} = 0.935 \qquad (9)$$

The treatment effect is:

$$(0.935 - 1) * 100 = -6.5\ percent \qquad (10)$$

The likelihood ratio chi-square is 0.32, which is less than 3.84. Thus, although a 6.5 percent reduction in nighttime accidents as compared to daytime accidents at the raised pavement marker sites was evidenced, this is not a statistically significant reduction, and raised pavement markers do not appear to reduce nighttime accidents relative to daytime accidents.

1.c. Before/after treatment with comparison groups and multiple measurements in time.

Description and numerical computations

The assumption that the comparison group is valid can be tested if more than 1 year of data is available before treatment. This design is called a before/after design with test for comparability. Typically, the test for comparability consists of a chi-square analysis performed on the table of counts for the treatment and comparison group for the before period only. For example, suppose 3 years of accident counts are available before treatment for both the treated and comparison groups (conditions). The table for the before data is represented by table 7, where B1, B2, and B3 represent the 3 before years.

Table 7. Generic before table for 3 years of before data.

Time	Comparison	Treatment
B1	A	D
B2	B	E
B3	C	F

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

Expected values for the cells in table 7 would be computed as for the usual (Pearson) chi-square analysis as shown in equation 11. For the first cell of the table, B1, for the comparison group, the expected value is:

$$e_{11} = \frac{(A + D)(A + B + C)}{(A + B + C + D + E + F)} \qquad (11)$$

The classical analysis would actually use the likelihood ratio chi-square, rather than the Pearson chi-square, because of the additive nature of that statistic. The formula for the likelihood ratio chi-square is the same as equation 6. The degrees of freedom are computed as the product of the number of rows (years) minus 1 and the number of columns (always 2) minus 1. For the above example with 3 years of data, this would be 2 (3 minus 1 times 2 minus 1). The degrees of freedom will obviously vary depending upon the number of before years available for the comparison.

Numerical example

Again using the Michigan raised pavement marker study as an example, the before data for the treatment and comparison conditions (nighttime and daytime accidents) are listed in table 8 for each of the 2 before years:

Table 8. Accidents for test for comparability analysis.

Time	Treatment - Day	Treatment - Night	Total
Before year 1	149	167	316
Before year 2	132	119	251
Total	281	286	567

The expected frequencies for the data in table 8 are displayed in table 9.

Table 9. Expected frequencies for comparison condition analysis.

Time	Treatment - Day	Treatment - Night
Before year 1	156.6	159.4
Before year 2	124.4	126.6

The chi-square test statistic value is 1.66. With 2 before years, the degree of freedom is still 1 as in the before/after comparison (2 rows minus 1 times 2 columns minus 1). Since 1.66 is less than 3.84 (the chi-square value at the 0.05 level of significance), the comparison condition of daytime accidents would be considered a legitimate comparison group for this analysis and the treatment effect can be tested for these data.

1.d. Before/after with comparison groups or conditions with replication.

Description and numerical computations

Another type of yoked comparison group design that is very strong is one where the comparison site and the treatment site are part of the same site. An example of this would be the evaluation of a new resurfacing treatment where the new treatment is applied to the southbound lane and the standard treatment is applied to the northbound lane of a 0.16-km (0.10-mi) area. The comparison site, assuming that the daily traffic is approximately the same on both lanes, is as similar to the treatment site as possible. In these designs, and in matched pair designs, it is possible to compute a treatment effect at each site. Multiple sites, therefore, can be considered replicates. When combining the separate treatment effects, rather than averaging these effects and giving each site equal weight, each treatment effect can be weighted based on the amount of "information" at each site. This weighted average is the average of each site's log odds ratio multiplied by the reciprocal of the variances. This weight is called w and is defined as:

$$w = ((1/A) + (1/B) + (1/C) + (1/D)) \tag{12}$$

A graphical procedure to better appreciate and interpret these analyses is called a nomograph. In this graph, the abscissa (x-axis) represents "expected accidents at the treatment locations during the after period" adjusted on the change in accidents from before to after at the paired comparison locations. The ordinate (y-axis) represents "observed accidents at the treatment locations during the after period." The diagonal line in this figure, with a slope of 1.0, represents "no treatment effect." If all of the data points lay on the diagonal, the observed after treatment would equal the expected after treatment. To the extent that the data do not fall on the solid line, but are instead scattered around the line in some heterogeneous fashion, the question arises: is the observed scatter in the data so great that some other factor or factors may be interacting with the imposed treatment — or is the scatter within a chance level of expectation? If the scatter is not within the chance level of expectation, our estimate of the overall effectiveness may be misleading. That is to say, the effectiveness of the treatment may vary from location to location for reasons (factors) that are not yet known.

To determine whether the scatter can be explained by chance, we will carry out a likelihood ratio chi-square (G^2) test of the following form:

$$G^2 = \sum w_i[\ln(O.R.)_i - \ln(O.R.)_{mean}]^2 \tag{13}$$

where $(O.R.)_i$ is the odds ratio defined in equation 5, and w_i is the weight for the i^{th} replicate site.

The $\ln(O.R.)_{mean}$ is:

$$\ln(O.R.)_{mean} = \sum_{i=1}^{n} w_i(O.R.)_i \div \sum_{i=1}^{n} w_i \tag{14}$$

There are n minus 1 degrees of freedom associated with this likelihood ratio chi-square, where n equals the number of treatment-comparison locations.

To obtain the simple overall (mean) odds ratio for the n treatment-comparison sites, we exponentiate the mean log odds ratio:

$$OR_{mean} = e^{\ln(O.R.)_{mean}} \tag{15}$$

The overall weighted average treatment effect over the n sites is then:

$$E_{overall} = (OR_{mean} - 1.0)100 \tag{16}$$

The G^2 statistic compared to the chi-square value with n minus 1 degrees of freedom tells us whether or not this average treatment effect is significant.

Numerical example

Another example will be used to illustrate this methodology. Assume that the purpose of this study is to evaluate the "wet-weather, accident-reduction effectiveness" of a new resurfacing technique. A 16-km (10-mi) section of four-lane, divided highway is resurfaced in the northbound lanes using the standard resurfacing technique. A new resurfacing technique is used on the same 16 km (10 mi) of southbound lanes. We then compare the wet-weather accident frequencies in the northbound lanes to the wet-weather accident frequencies in the southbound lanes for a period of 4 years — 2 years before and 2 years after the treatments were imposed. The following data were collected:

Table 10. Hypothetical wet-weather accident data for a 16-km (10-mi) highway segment that was resurfaced by "traditional" and "new" resurfacing techniques.

Wet-Weather Accidents	Comparison Condition: Traditional Resurfacing	Treatment Condition: New Resurfacing
2 Years Before	(A) 104	(C) 125
2 Years After	(B) 153	(D) 134

The odds ratio is:

$$OR = \frac{(104/153)}{(125/134)} = 0.729 \tag{17}$$

The treatment effect is:

$$E = (O.R. - 1.0)100 = -27.1 \qquad \textbf{(18)}$$

As before, the first question that we should ask ourselves is this: Is this apparent 27.1 percent reduction in accidents a bona fide effect, or might this apparent 27.1 percent reduction have been expected by chance as often, say, as 5 times in 100? To answer this question, we will use a different, but equivalent, test statistic as the likelihood ratio chi-square — the Z-test. This test relies upon the fact that the sampling distribution for the natural logarithm of O.R. is symmetric and asymptotically normal, with a standard error (SE) that may be approximated (for relatively large samples) as:

$$SE = (1/A + 1/B + 1/C + 1/D)^{0.5} \qquad \textbf{(19)}$$

For this example, the SE is:

$$SE = (1/104 + 1/153 + 1/125 + 1/134)^{0.5} = 0.178 \qquad \textbf{(20)}$$

Now we can perform a standard Z-test to determine if our apparent 27.1 percent reduction is more logically attributed to the treatment imposed, or to chance fluctuation in the data. The formula for this test is:

$$Z = \frac{\ln(O.R.)}{SE} \qquad \textbf{(21)}$$

The absolute value is compared to the normal Z-value of 1.96 at the 5 percent level of significance. (The reason that this test is equivalent to the likelihood ratio chi-square is that Z^2 is approximately equal to the likelihood ratio chi-square and the critical value of $1.96^2 = 3.84$, the critical chi-square value for one degree of freedom at the 5 percent level. This alternative test is presented in this report since it is used nearly as frequently as the likelihood ratio chi-square.)

Returning to the example, the value of Z is -1.78, which in absolute value does not exceed 1.96. Therefore, at the standard alpha level (i.e., α=0.05), the calculated Z shown above is not significant, i.e., it does not lie beyond plus or minus 1.96. Thus, we cannot conclude that the treatment being evaluated (a new resurfacing technique) reduces wet-weather accidents.

But what if we had carried out our evaluation at three locations rather than just one? What if we had collected wet-weather accident data at three different treatment-comparison locations, as shown in the next table? Considering these data, we estimate that at locations 1, 2, and 3, accidents were apparently reduced by 27.1, 18.9, and 18.3 percent, respectively. If these estimates are not too discrepant, we would like to combine them to come up with some overall estimate of the effectiveness of the new resurfacing technique. By combining data from multiple treatment-comparison locations, we gain two benefits:

- By increasing the number of treatment-comparison locations and thereby increasing sample sizes (i.e., accidents), we increase statistical power.
- By carrying out our evaluation at multiple treatment-comparison locations, we better ensure that any calculated "significant" reduction in accidents is more likely to result from the treatment being evaluated, rather than from some unique characteristic of a particular treatment-comparison location — or the interaction of the treatment imposed with the particular characteristics of a given treatment-comparison location.

Table 11. Hypothetical wet-weather accident data for a 16-km (10-mi) highway segment by location.

Location	Wet-Weather Accidents	Comparison Condition: Traditional Resurfacing	Treatment Condition: New Resurfacing
1	2 Years Before	(A) 104	(C) 125
	2 Years After	(B) 153	(D) 134
2	2 Years Before	(A)161	(C) 213
	2 Years After	(B) 193	(D) 207
3	2 Years Before	(A) 101	(C) 109
	2 Years After	(B) 127	(D) 112

The best overall estimate (i.e., the best linear, unbiased estimate) of treatment effectiveness at multiple treatment-comparison locations is derived from the weighted average of the individual log odds ratios at each location, with the weights defined as the reciprocals of the variances (i.e., the squared standard errors) of the individual treatment-comparison locations.

The weights used in equation 12 for the three treatment-comparison locations shown in the previous table are:

$$\begin{aligned} w_1 &= 31.632 \\ w_2 &= 47.805 \\ w_3 &= 27.872 \end{aligned} \quad \textbf{(22)}$$

The log odds ratios for locations 1, 2, and 3 are -0.317, -0.210, and -0.202, respectively. Each of these log odds ratios are weighted by the small w's shown above and the average (mean) calculated.

$$\begin{aligned} \ln(O.R.)_{mean} &= [31.632(-0.317) + 47.805(-0.210) + 27.872(-0.202)] + 107.309 \\ &= -0.239 \end{aligned} \quad \textbf{(23)}$$

We exponentiate the mean log odds ratio to obtain the simple overall (mean) odds ratio for the three treatment-comparison locations:

$$O.R_{mean} = e^{-0239} = 0.787 \qquad \textbf{(24)}$$

The average treatment effect is then:

$$E_{overall} = (0.787 - 1.0)100 = -21.3 \qquad \textbf{(25)}$$

To better appreciate what this analysis is showing us, consider the graph in figure 1. In this graph, the abscissa (x-axis) represents "expected accidents at the treatment locations during the after period" adjusted by the change in accidents from before to after at the paired comparison locations. The ordinate (y-axis) represents "observed accidents at the treatment locations during the after period." The diagonal line in this figure, with a slope of 1.0, represents "no treatment effect." Had the three data points (the three treatment-comparison locations) lain on the diagonal, "observed wet-weather accidents after treatment" would equal "expected wet-weather accidents after treatment." But, the best fit to the three data points is approximated by a line with a slope of 0.729, the average (mean) odds ratio ($O.R._{mean}$) for the three treatment-comparison locations.

Note that the "best fit" line in figure 1 lies 21.3 percent below the diagonal and represents the estimated overall effect of treatment $E_{(overall)}$. Had the same estimate of treatment effectiveness been calculated for all three treatment-comparison locations, all three data points would have fallen on the solid line. The estimates would have been homogeneous. To the extent that the data do not fall on the solid line, but are instead scattered around the line in some heterogeneous fashion, it behooves us to ask: Is the observed scatter in the data so great that some other factor or factors may be interacting with the imposed treatment — or is the scatter within a chance level of expectation? If the scatter is not within the chance level of expectation, our estimate of the overall effectiveness may be misleading. That is to say, the effectiveness of the treatment (a new resurfacing technique) may vary from location to location for reasons (factors) that are not yet known.

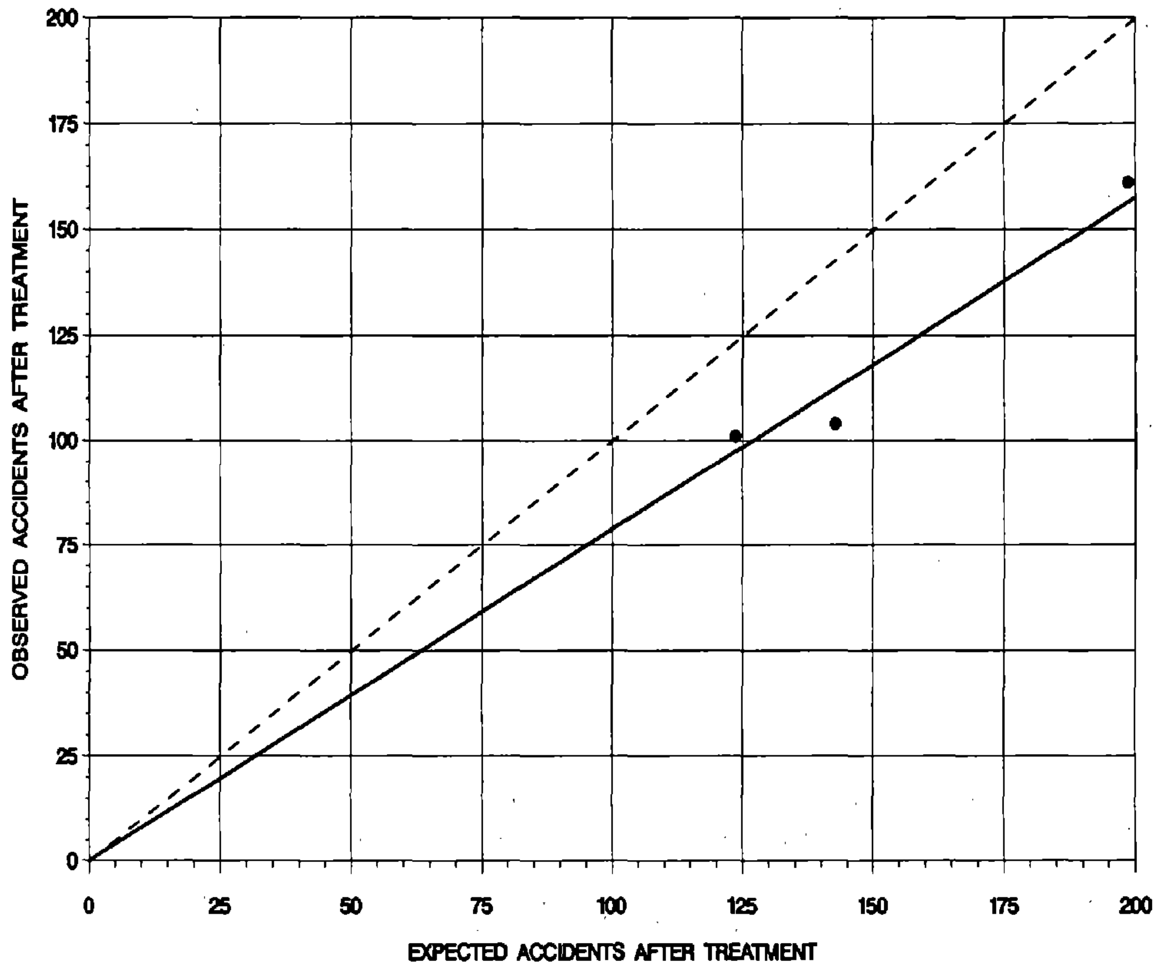

Figure 1. Graphical depiction of the multiple before-and-after design with yoked comparisons.

To determine whether the scatter can be explained by chance, we will carry out a likelihood ratio chi-square test of equation 13. For these data, this value is 0.270. There are two (i.e., 3 sites minus 1) degrees of freedom associated with this likelihood ratio chi-square. Because this chi-square is *not* significant ($G^2_{(2)} = 0.270$; p-value = 0.874), we will conclude that the data from the three locations are reasonably homogeneous and may be combined to yield an overall estimate of treatment effect. Had the chi-square homogeneity been large, i.e., had the probability that the data were homogeneous been small, we should be reluctant to combine these data into one overall estimate of treatment effect.

CRITIQUE

The before/after methodology without a comparison group (1.a.) is the weakest of all before/after analyses and should be avoided. It is only valid if there is absolute certainty that there are no before/after differences that could confound a treatment effect, such as changes in traffic volume, weather, etc. Otherwise, every attempt should be made to, in some way, control for the effect of any changes over time. If there is no reason to suspect a regression-to-the-mean effect, the comparison group/condition methodologies (1.b.) are fairly sound and established. There may be cases when the treatment group truly represents an unbiased and nearly random sample. For example, if the intended treatment is changing posted speed limits and the sites selected for the change were not selected based on their accident histories, but for policy or other reasons, regression to the mean may be a non-issue. A recent FHWA study applied the regression-to-the-mean adjustment to several data sets and found little, if any, differences between the classical before/after comparison group methods and the empirical Bayes method.[10]

A critical assumption of the before/after with comparison group/condition methodology is that the groups are truly comparable. The methodology (1.c.) to test for this is straightforward and allows the opportunity to test for this assumption. When multiple years of before data are unavailable, the assumption cannot be tested. Thus, it is strongly encouraged that whenever possible, such data be attained and the test for comparability be performed. More years of data increase the degrees of freedom of the test, and as many reliable years as possible should be used. However, it is essential that the before period not extend past what would be considered "comparable." That is, if data quality or other factors would make 5 years of before data questionable, then it would be better to shorten the before period.

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

The main disadvantage of the weighted site-averaging methodology (1.d.) is the same disadvantage seen in the simple before-and-after design with a yoked comparison group. Namely, this design is subject to regression toward the mean bias. If treatment sites are specifically chosen on the basis of egregiously high accident counts during the before period, this design should not be used.

A second disadvantage of this design results from the fact that there is no guarantee that the comparison group or condition is, in fact, comparable to the treatment group or comparison being evaluated. In the absence of random assignment, the "comparability" of the comparison group or condition against which the treatment group or condition is being evaluated is always subject to interpretation — and debate.

If regression toward the mean is not seen as a problem, then this design is fairly simple and straightforward. As previously stated, it offers two advantages over the simple before-and-after design with a yoked comparison. By collecting data at multiple locations, statistical power is increased and, thus, other things being equal, our ability to detect a real treatment effect when one exists is enhanced. Secondly, by carrying out our evaluation at multiple locations, we tend to rule out the likelihood that any calculated effectiveness of the treatment imposed is the result of some unique set of factors or conditions at a given location. By going to multiple, yoked treatment comparison locations, we are, in effect, running multiple replications of our evaluation, all within the course of one analysis.

These methodologies are generally used when exposure data are unavailable. However, they may be superior even when exposure data are available. When exposure data are available, accident rates are calculated and used as the primary measure of safety. Inherent in the rate calculation is the assumption that accident frequency and exposure (VMT, AADT, etc.) are linearly (constantly) related. Some studies have found this assumption to be questionable.[(10,11)] If this is indeed the case, rate-based analyses could be erroneous since the very calculation of rate assumes a linear relationship between these variables. In this case, it is possible that methodologies that successfully adjust for exposure are actually preferable. This relationship has not been fully explored. In addition, it would be interesting to compare results of analyses on the same data with and without using the exposure method.

CHAPTER 3: CASE-CONTROL STUDIES

INTRODUCTION, HISTORICAL SUMMARY, AND DESCRIPTION OF THE METHODOLOGY

Let us begin this discussion of case-control studies with a definition:[1]

> The *case-control* study, also commonly called a *retrospective study*, follows a paradigm that proceeds from effect to cause. In a case-control study, individuals with a particular condition or disease (the *cases*) are selected for comparison with a series of individuals in whom the condition or disease is absent (the *controls*). Cases and controls are compared with respect to existing or past attributes or exposures thought to be relevant to the development of the condition or disease under study.

To put this definition in perspective, and to change the discussion from diseases to a more familiar highway safety topic, let us consider four basic methodologies that could (theoretically) be used to evaluate the effectiveness of safety belts in reducing driver fatalities:

- True experimental evaluation.
- Cohort evaluation (current or prospective).
- Cohort evaluation (historic or retrospective).
- Case-control evaluation.

True experimental evaluation: By the tenets of a true experimental evaluation of the effectiveness of safety belts in reducing driver fatalities in traffic accidents, we would be obliged to ensure (on a random basis) that some drivers always used safety belts, while others never used safety belts. Although this procedure is relatively simple in the laboratory — half the rats (the treatment group) are randomly selected and administered a drug, and the other half (the control group) are administered a placebo — there is no realistic way to accomplish this object in a real-world setting with human subjects. We can neither ethically nor practically dictate which drivers will always use safety belts and which will forego the use of safety belts. Thus, true experimental methodology is not a viable means of evaluating the effectiveness of safety belts in reducing driver fatalities in crashes.

Cohort evaluation (current or prospective): Suppose that we could, theoretically, identify two groups of drivers: one group (the treatment group) is somehow known to always use their safety belts; the other group (the comparison group) is known to never use their safety belts. For purposes of this example, further assume that both groups contain 1,000 drivers. Note that we did not assign these drivers to their respective groups, we simply "discovered" these two separate and mutually exclusive groups.

We could follow these two groups into the future and record the traffic accidents in which the members of both groups were involved and tally the numbers of fatally injured drivers in both groups. The differences in fatalities for the two groups would reflect the effectiveness of safety belts

[1]Schlesselman, J.J. *Case-Control Studies*. Oxford University Press, p. 14, 1982.

in reducing fatalities *if* both groups were comparable except for their use of safety belts. If, for example, the safety belt users were healthier, drove fewer miles in a more massive vehicle, and drove more "responsibly" than the non-users, then the difference in recorded fatalities for the treatment and comparison groups would yield a biased estimate of the fatality reduction benefits of safety belts. To get around this bias, some sort of ad hoc weighting of the data might be undertaken to render the two groups more comparable and, thereby, to adjust the calculated benefits of safety belts to reflect the lack of comparability in the two groups.

It might also be noted that there are only 2,000 drivers involved in this evaluation, half in the treatment group and half in the comparison group. Because traffic accidents that result in driver fatalities are (thankfully) rare events, as a practical matter we would have to follow these two groups for a long time to have sufficient data (i.e., driver fatalities) to draw reasonable estimates of the benefits of safety belts.

Cohort evaluation (historical or retrospective): For purposes of evaluating the effectiveness of safety belts in reducing driver fatalities, this design overcomes some of the problems encountered in the last design. The historical or retrospective cohort design starts with an "accident data set" that includes belted and unbelted drivers who have or have not been fatally injured in traffic accidents. Rather than selecting two groups of drivers and tabulating their fatalities in future accidents (as we did in the last design), we start with the accident data. We divide those data into two groups: Drivers using safety belts in their accidents (the treatment group) and drivers not using safety belts in their accidents (the comparison group). We then proceed to compare the proportions of fatally injured drivers in both groups. The difference in these two proportions is representative of the effectiveness of safety belts in reducing driver fatalities *if* the two groups are comparable except for their use of safety belts. Again, if the two groups are not comparable, any simplistic assessment of the benefits of safety belts based on the difference in the proportions of fatally injured drivers in the two groups may be biased and misleading. Ad hoc adjustment is sometimes used in these circumstances to parcel out the effects of these differences and to reduce the bias in the estimate of treatment effectiveness.

In summary, there are three advantages of this design over the previous design for purposes of evaluating safety belt effectiveness. First, it is easier to determine whether or not a given driver in a given accident was or was not using a safety belt than to ensure that in all future accidents one group of drivers will always be belted and the other will not. Second, by starting with historical accident data, we can conduct our analyses now — we do not have to delay our analyses until some future date. Finally, by starting from a basic accident data set, rather than a set of drivers, we can much more easily collect larger numbers of accident-involved drivers who were or were not fatally injured while using or not using safety belts.

Case-control evaluation: In this evaluation methodology, we start with an outcome (fatally injured drivers) and work backwards. To add specificity to this example, assume that we collect data on 100 fatally injured drivers (the cases, i.e., the treatment group) — some of whom were belted and some of whom were not belted at the time of their accident. We then collect a *matched* sample of, say, 500 drivers operating in traffic (the controls, i.e., the comparison group). These matched drivers are selected for comparability to the fatally injured drivers in terms of other factors or variables that are thought to be influential in determining the likelihood that an accident-involved driver would be fatally injured in the collision, e.g., age and sex of the driver, type (mass) of the vehicle the driver operates, etc. Then we compare the proportions of belted drivers in the case and control groups. The differences in these two proportions are a measure of the effectiveness of safety belts in reducing driver fatalities.

To the extent that the matching procedures used in case-control methodology are effective in ensuring comparability between the treatment and comparison groups, this design may offer a reasonable means of evaluating safety belt effectiveness. It should be noted, however, that any comparability that is achieved through this design results from matching the controls to the cases (fatally injured drivers). And the cases — fatally injured drivers — are a fairly atypical group of drivers.

The great strength of this design is the efficiency with which the data of interest (driver fatalities) can be collected. Experimental designs and cohort designs are forced to collect large quantities of data (driver data, accident data) to ensure that relatively few driver fatalities are included in the analysis. The case-control design, by contrast, selects only the outcome variable of interest (driver fatalities).

To better appreciate the similarities and differences between these four evaluation designs, consider the following summary table based on the safety belt example that has been used for the purpose of discussion:

Table 12. Comparison of case-control methodologies.

Evaluation Design	Assignment to Treatment and Control/Comparison Groups	Elements or Members of the Treatment Group	Elements or Members of the Control/Comparison Groups
Experimental	Random Samples	Randomly Selected Drivers Who Are Required to Always Use Safety Belts	Randomly Selected Drivers Who Are Not Allowed to Use Safety Belts
Cohort: Current or Proactive	Opportunistic or Naturalistic Samples, Often With *Ad Hoc* Adjustments of the Data	Drivers Who Are Known to Always Use Safety Belts	Drivers Who Are Known to Never Use Safety Belts
Cohort: Historic or Retroactive	Opportunistic or Naturalistic Samples, Often With *Ad Hoc* Adjustments of the Data	Drivers Who Were Using a Safety Belt When Involved in an Accident	Drivers Who Were Not Using a Safety Belt When Involved in an Accident
Case-Control	Samples Based on Outcomes (Cases) and Matches to Outcomes (Controls)	Fatally Injured Drivers	Matched, Cross Section of Drivers

In the highway safety field, traditional case-control studies have been used to assess the effects of alcohol on pedestrian fatalities; the effects of roadway geometry on fatal, ran-off-road accidents; and the effects of truck defects on truck accidents.[2] More recently, Leonard Evans has been popularizing the use of a methodology that he refers to as the "Double Pair Comparison" method. In fact, this method is a very clever application of the statistical procedures developed by Woolf (1955) (discussed in a previous section) in a case-control context.[3]

NUMERICAL EXAMPLE

To see how Evans' method works, consider the data in the table on the next page that were taken from Evans' 1988 article. In this study, Evans sought to determine the effectiveness of restraints in reducing fatal injuries to right-rear-seat occupants. The data were selected from the Fatal Accident Reporting System (FARS) for the calendar years 1975-1985 and included approximately 10,265 accident-involved passenger cars. Each accident-involved car contained a right-rear passenger (R) and a driver (D), or a right-front passenger (F), or a left-rear passenger (L). In addition, one or more fatalities were recorded in each car in the data set.

In table 13, the comparison (C) vehicle occupants [i.e., the driver (D), right-front passenger (F), and left-rear passenger (L)] and the right-rear occupant (R) are either restrained ($_R$) or unrestrained ($_U$), as indicated by the subscripts. The comparison occupants (C) and the right-rear passengers (R) were either fatally injured (^{F}) or were not fatally injured (^{N}), as indicated by the superscripts.

Considering the fatalities in the first two rows in this table, we see that 6,264 unrestrained drivers and 4,876 unrestrained right-rear passengers were killed while riding in 9,678 accident-involved passenger cars. In addition, we see that 48 restrained drivers and 31 restrained right-rear passengers were killed while riding in 70 accident-involved passenger cars. Using the unrestrained drivers as a comparison group, Evans sets up the following odds ratio (U):

$$U = \frac{31/48}{4876/6264} = 0.83 \qquad (26)$$

---

[2]Haddon, W., Jr.; P. Valien; J.R. McCarroll; and C.J. Umberger. "A Controlled Investigation of the Characteristics of Adult Pedestrians Fatally Injured by Motor Vehicles in Manhattan," *Journal of Chronic Diseases*, Vol. 14, pp. 655-678, 1961. (Reprinted in Haddon, W., Jr.; E.A. Suchman; and D. Klein. *Accident Research: Methods and Approaches*, New York: Harper and Row Publishers, 1964.)

Jones, I. and H.S. Stein. "Defective Equipment and Tractor-Trailer Crash Involvement," *Accident Analysis and Prevention*, Vol. 21(5), pp. 469-481, 1989.

Wright, P.H. and L.S. Robertson. "Studies of Roadside Hazards for Projecting Fatal Crash Sites," *Transportation Research Record*, 609, Transportation Research Board, Washington, DC, pp. 1-8, 1976.

[3]Evans, L. "Rear Seat Restraint System Effectiveness in Preventing Fatalities," *Accident Analysis and Prevention*, Vol. 20, pp. 129-136, 1988.

Table 13. Seatbelt effectiveness data.

	Accident-Involved Passenger Cars					Fatalities	
Seating Configuration	C^FR^N	C^FR^F	C^NR^F	C^NR^N	Total	Comparison	Right Rear
1 D_UR_U	4802	1462	3414		9678	6264	4876
2 D_UR_R	39	9	22		70	48	31
3 D_RR_U	116	44	184		344	160	228
4 D_RR_R	63	30	80		173	93	110
5 F_UR_U	4245	1747	2643		8635	5992	4390
6 F_UR_R	38	8	20		66	46	28
7 F_RR_U	86	44	121		251	130	165
8 F_RR_R	59	28	71		158	87	99
9 L_UR_U	1697	870	1666		4233	2557	2536
10 L_UR_R	17	6	6		29	23	12
11 L_RR_U	14	2	14		30	16	16
12 L_RR_R	29	11	31		71	40	42

This odds ratio (U) indicates that fatalities to right-rear passengers are 17 percent less likely when they are restrained, when controlling for the likelihood of fatality to unrestrained drivers riding in the same vehicles.

By similar logic, Evans compares rows 3 and 4, 5 and 6, 7 and 8, 9 and 10, and 11 and 12. For each of these pairings, he calculates an odds ratio. He then takes the weighted average log odds ratio associated with these six pairings as proposed by Woolf (1955) to come up with an overall estimate of the effectiveness of restraints in reducing fatalities to right-rear passengers.

It should be noted that in this analysis, Evans is reusing his data. That is to say, the fatally injured right-rear passengers in the first two rows are that same fatalities that appear in rows 3 through 12. No adjustment is made in Evans' calculations to allow for this reuse of data.

CHAPTER 4: INDUCED-EXPOSURE METHODOLOGY

INTRODUCTION AND HISTORICAL SUMMARY OF THE METHODOLOGY

Induced-exposure methodologies are additional methods of bypassing the use of measured exposure. Brown defines *induced exposure* as an attempt to "induce" what the exposure is from the accident data. In these methodologies, the fundamental concept is that the likelihood of a non-responsible combination of being involved in an accident is proportional to meeting that combination on the road. Basically, this can be stated in two assumptions:

- All drivers involved in accidents can be divided into two groups — responsible and non-responsible.

- The number of not responsible drivers in any group is proportional to that group's exposure.

Thorpe is credited as the first to attempt to apply the method of induced exposure to the field of highway safety. Thorpe uses single-vehicle accident counts as surrogates for at-fault driving exposure in multiple-vehicle accidents. His methodology is based on five basic assumptions:

1. Single-vehicle accidents are caused by attributes of the driver-vehicle combination of interest.

2. Most collisions are caused by the first two vehicles, so only the first two vehicles in multiple-vehicle accidents are analyzed.

3. In each accident, the responsible and not-responsible drivers are identified.

4. The relative likelihood of a responsible driver being involved in a multiple-vehicle accident is the same as the likelihood of a driver in a single-vehicle accident.

5. The likelihood of any particular driver-vehicle combination being innocently involved in an accident is the same as the likelihood of meeting that combination anywhere on the road.

Thorpe computes a "relative accident likelihood (R.A.L.)" for each group partitioning and applies this to Australian accident data for 1961-1962 for personal injury accidents only. When partitioning by age, he does not find the traditional U-shaped curve relating accident frequency and age. Given what we know now, this was probably due to interacting factors. Thorpe then computed these R.A.L.'s by driver experience and age and presented them in a two-way table. He concluded that as age and experience increased, the relative risk of accidents decreased. To check the validity of his assumption based on single-vehicle accidents, Thorpe computed the ratio of single vehicles in an age group to the proportion of licensed drivers in that age group; this he labeled "the involvement index (I.I.)." When these indices were compared to the R.A.L.'s, there was a "striking" similarity in their quantities, although no statistical tests of equality were run. Additional analyses were performed on vehicle types and drinking drivers.

Carr later modified Thorpe's measure and called it a relative risk measure. The relative risk is simply the ratio of two frequencies — the frequency of accidents that have the characteristic over the frequency of those that do not. For example, in examining the probability of being at fault in an accident by driver age, the relative risk would be computed for each age category and it would be defined as the number of accidents where the driver was at fault over the number of accidents where the driver was not at fault. Such a measure, Carr claims, has the advantage of providing a measure of exposure that is directly related to the risk of having accidents and controls for all potential confounding environmental factors. More recently, empirical Bayes methods have been introduced into the induced-exposure methodology.

DESCRIPTION AND NUMERICAL COMPUTATIONS

Induced exposure in highway safety applications seems to often focus on at-fault and innocent victim categorization of drivers, although there is no reason that some other category could not be defined and the same procedure applied. Whatever the category, the ratio (which takes on various terms in the literature) is computed:

$$I_i = \frac{f_i}{nf_i} \tag{27}$$

where I_i is the induced-exposure index for group i, f_i is the number of at-fault drivers, and nf_i is the number of not-at-fault drivers. Group i can be any categorization, but the most prevalent one in the literature is age grouping. These categorizations are generally driver- or vehicle-related characteristics for which it is difficult to obtain, or even define, the appropriate exposure. For example, in studying the problem of whether elderly drivers have more difficulty with left-turn lanes, what is the proper exposure? Is it the number of elderly drivers that pass through each intersection each day? If so, how do we estimate this quantity? Questions such as these gave rise to the methodology of induced exposure. If we compute the induced-exposure index for elderly drivers and all other drivers, these quantities should be equal if elderly drivers are not having any more of a problem with left-turn lanes than anyone else. Inherent in this calculation is the assumption that the number of not-at-fault drivers in each category will represent the same age distributions as the at-fault drivers. In other words, there should be just as many elderly drivers who are victims in accidents as there are those who are at fault.

Once the induced-exposure index is calculated, it can be used in various ways. Again, one of the most common uses is to compute it for various age groups and to compare these ratios to see if certain ages of drivers are more at fault than others. When only two groups are of interest, an odds ratio can be computed to statistically compare the groups, similar to the odds ratio in section 1. Suppose the induced-exposure example is divided into two groups — "young" and "old" as defined by some age criteria, and "at fault" or "not." Table 14 reflects a generic table of this type.

Table 14. Generic induced-exposure example.

		Innocent Victim	
		Young	Old
At-Fault Driver	Young	A	C
	Old	B	D

The "effect" (young versus old at-fault) is computed as the odds ratio of this table:

$$O.R. = \frac{A/B}{C/D} \qquad (28)$$

where A/B are the odds of a young versus an old at-fault driver being involved with a young victim and C/D are the same odds for the old victims. Any differences will cancel out, assuming that the same differences occurred in both groups. If the equality exposure assumptions for the two age groups hold, this is a valid method. The statistical test applied to this estimate is the same Z-test described in section 1.d. If this estimate is not significant, then the assumption of equal exposure for the two age groups (random victim selection) holds.

The second step in this process is to test if the accident frequencies of the two groups (young versus old) are equal. This is done by essentially computing an odds ratio on the marginal totals of the table and again using the Z-test. Consider, now, the generic table corresponding to table 14:

Table 15. Table of marginal totals.

	Young	Old
At-Fault Drivers	A+C (x_i)	B+D (x_j)
Victims	A+B (y_i)	C+D (y_j)

This example is taken from a paper written by Davis.[9] In this paper, Davis' notation uses the x's and y's in table 16. The O.R. to estimate the differences in young and old at-fault driver accident frequencies is then:

$$O.R. = \frac{(A+C)(C+D)}{(A+B)(B+D)} = \frac{x_i\ y_j}{x_j\ y_i} \qquad (29)$$

Rearranging these numbers as ratios reveals the motivation for this computation:

$$O.R. = \frac{(A+C)/(A+B)}{(C+D)/(B+D)} \qquad (30)$$

which is the ratio of young at-fault drivers to victims over the ratio of old at-fault drivers to victims or the odds of young drivers being at fault to the odds of older drivers being at fault. If this is significant when tested using the log odds Z-test of equation 21, then there is a difference and the direction of the difference is determined by the sign of the odds ratio minus 1. A negative sign would indicate that the odds of the older drivers being at fault are greater than the odds of the younger drivers being at fault.

NUMERICAL EXAMPLE

The data used in this example were originally presented by Lyles et al. and then later used by Davis.[9,14] The study investigated the incidence of accidents in nighttime interstate accidents by driver sex in Michigan in 1988. The data are presented in table 16.

Table 16. Nighttime interstate accident frequencies for induced-exposure example.

At-Fault Driver		Innocent Victim		
		Male	Female	Total At-Fault
	Male	2232	894	3126
	Female	605	256	861
	Total Victims	2837	1150	3987

In this example, male takes the place of "younger" driver in the previous discussion and female takes the place of "older" driver. Of course, the ordering is arbitrary, but it will have an effect on the interpretation of the results.

The first step is to check for the assumption of random victim selection. The odds ratio of equation 28 is:

$$O.R. = \frac{2232/605}{894/256} = 1.056 \qquad \textbf{(31)}$$

The log odds Z-test of equation 31 is:

$$Z = \frac{0.055}{0.036} = 0.65 \qquad \textbf{(32)}$$

Comparing this to the normal value at the 5 percent level — 1.96 — we conclude that the O.R. is not significant, indicating that the assumption of random victim selection is valid. In other words, it does appear that the distribution of gender among the victims is the same as the distribution among the at-fault drivers. Testing if there is a difference in the accident frequencies by gender, we compute the odds ratio on the following table:

Table 17. Table of marginal totals.

Group	Male	Female
At-Fault Drivers	3126	861
Victims	2837	1150

The odds ratio for this table is 1.471 and the Z-test is 7.43, which is greater than 1.96. Therefore, there is a statistically significant difference in the at-fault accident frequencies by gender. Since the O.R. minus 1 is positive, the male at-fault accident frequency is significantly greater than for the females in nighttime interstate accidents.

CRITIQUE

The primary criticism of the induced-exposure methodology is in the selection of the appropriate "exposure" surrogate group. The validity of the at-fault/victim partitioning may be dependent on the hypothesis being tested. For example, if the hypothesis is to compare accident frequencies by age, is it true that the "victims" in accidents exist in the same proportional age groupings as the at-fault drivers? Suppose the area of interest is a retirement community, so that the resident population consists of older persons. The "younger" drivers are expected to be fewer in number. If most of those few drivers are at fault in accidents due to a more reckless driving behavior, the number of victims will obviously be few, whereas the number of older drivers will be in greater quantity, both as victims and as at-fault drivers. The test for random victim selection supposedly would discover this. But if it does, what then? The hypothesis cannot be tested, yet the researcher will not discover this until after the data have been collected. This may be a very costly way to find out that no conclusions can be reached. So the user of this methodology should expend great care in thinking through the assumptions and should be very sure that there is no confounding among these partitionings. Given the difficulty of measuring exposure for the applications for which induced exposure is used, however, the methodology has its merits.

CHAPTER 5: SUMMARY AND CONCLUSIONS

Because of the difficulty in defining exposure for certain highway safety applications, there have been numerous attempts to perform accident analyses without using a direct measure of exposure. Indirect methods are generally used when exposure data are either unavailable, unreliable, or undefinable. This report presented the most common methods used in before/after treatment evaluation studies. Each methodology was defined and worked examples were presented. A critique of each methodology, including the pros and cons, was included. The readers of this report should have been able to determine, from this presentation, which methods would be most suitable for their own particular application and to follow the step-by-step computational methods for applying the methodology to their own research.

The methodologies contained in this report focused primarily on before/after safety evaluation studies. Specifically, these methods are the before/after analyses, with and without comparison or control groups and test for comparability; before/after with replications that allow averaging of multiple site effects; case-control studies; and induced-exposure methodologies. Each methodology contained a discussion, including an introduction and historical application of the methodology, a description, numerical computations and examples, and a critique of the methodology.

Each methodology has its pros and cons. The before/after methodology without a comparison group (1.a.) is the weakest of all before/after analyses and should be avoided. It is only valid if there is absolute certainty that there are no before/after differences that could confound a treatment effect, such as changes in traffic volume, weather, etc. Otherwise, every attempt should be made, in some way, to control for the effect of any changes over time. If there is no reason to suspect a regression-to-the-mean effect, the comparison group/condition methodologies (1.b.) are fairly sound and established. There may be cases when the treatment group truly represents an unbiased and nearly random sample. For example, if the intended treatment is changing posted speed limits and the sites selected for the change were not selected based on their accident histories, but for policy or other reasons, regression to the mean may be a non-issue. A recent FHWA study applied the regression-to-the-mean adjustment to several data sets and found little, if any, differences between the classical before/after comparison group methods and the empirical Bayes method.[10]

A critical assumption of the before/after with comparison group/condition methodology is that the groups are truly comparable. The methodology (1.c.) to test for this is straightforward and allows the opportunity to test for this assumption. When multiple years of before data are unavailable, the assumption cannot be tested. Thus, it is strongly encouraged that whenever possible, such data be attained and the test for comparability be performed. More years of data increase the degrees of freedom of the test and as many reliable years as possible should be used. However, it is essential that the before period not extend past what would be considered "comparable." That is, if data quality or other factors would make 5 years of before data questionable, then it would be better to shorten the before period.

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

The main disadvantage of the weighted site-averaging methodology (1.d.) is the same disadvantage seen in the simple before-and-after design with a yoked comparison group. Namely, this design is subject to regression-toward-the-mean bias. If treatment sites are specifically chosen on the basis of egregiously high accident counts during the before period, this design should not be used.

A second disadvantage of this design results from the fact that there is no guarantee that the comparison group or condition is, in fact, comparable to the treatment group or comparison being evaluated. In the absence of random assignment, the "comparability" of the comparison group or condition against which the treatment group or condition is being evaluated is always subject to interpretation — and debate.

If regression toward the mean is not seen as a problem, then this design is fairly simple and straightforward. As previously stated, it offers two advantages over the simple before-and-after design with a yoked comparison. By collecting data at multiple locations, statistical power is increased, and thus, other things being equal, our ability to detect a real treatment effect when one exists is enhanced. Secondly, by carrying out our evaluation at multiple locations, we tend to rule out the likelihood that any calculated effectiveness of the treatment imposed is the result of some unique set of factors or conditions at a given location. By going to multiple, yoked treatment comparison locations, we are, in effect, running multiple replications of our evaluation — all within the course of one analysis.

These methodologies are generally used when exposure data are unavailable. However, they may be superior even when exposure data are available. When exposure data are available, accident rates are calculated and used as the primary measure of safety. Inherent in the rate calculation is the assumption that accident frequency and exposure (VMT, AADT, etc.) are linearly (constantly) related. Some studies have found this assumption to be questionable.[10,11] If this is indeed the case, rate-based analyses could be erroneous, since the very calculation of rate assumes a linear relationship between these variables. In this case, it is possible that methodologies that successfully adjust for exposure are actually preferable. This relationship has not been fully explored. In addition, it would be interesting to compare results of analyses on the same data with and without using the exposure method.

Four commonly used case-control study experimental designs were presented. The true experimental evaluation would be obliged to ensure (on a random basis), for example, that some drivers always used safety belts, while others never used safety belts in a seatbelt evaluation. Although this procedure is relatively simple in a laboratory, there is no realistic way to accomplish this objective in a real-world setting with human subjects. In general, true experimental methodology is not a viable means of performing highway safety evaluations. The cohort evaluation (current or prospective) is likewise impractical as it would require following observations (highway sites, drivers, etc.) for long periods of time. Cohort historical or retrospective evaluations are the most common and practical of designs, but they have their flaws and limitations. However, there are three advantages of this design over the previous design for purposes of evaluating safety belt effectiveness, for example. First, it is easier to determine whether or not a given driver in a given accident was or was not using a safety belt than it is to ensure that in all future accidents one group of drivers will always be belted and the other will not. Second, by starting with historical accident data, we can conduct our analyses now — we do not have to delay our analyses until some future date. Finally, by starting from a basic accident data set, rather than a set of drivers, we can much more easily collect larger numbers of accident-involved drivers who were or were not fatally injured while using or not using safety belts.

In the case-control evaluation, we start with an outcome (fatally injured drivers) and work backwards. To the extent that the matching procedures used in case-control methodology are effective in ensuring comparability between the treatment and comparison groups, this design may offer a reasonable means of evaluating safety belt effectiveness. It should be noted, however, that any comparability that is achieved through this design results from matching the controls to the cases (fatally injured drivers). And the cases — fatally injured drivers — are a fairly atypical group of drivers.

The great strength of this design is the efficiency with which the data of interest (driver fatalities) can be collected. Experimental designs and cohort designs are forced to collect large quantities of data (driver data, accident data) to ensure that relatively few driver fatalities are included in the analysis. The case-control design, by contrast, selects only the outcome variable of interest (driver fatalities).

The primary criticism of the induced-exposure methodology is in the selection of the appropriate "exposure" surrogate group. The validity of the at-fault/victim partitioning may be dependent on the hypothesis being tested. For example, if the hypothesis is to compare accident frequencies by age, is it true that the "victims" in accidents exist in the same proportional age groupings as the at-fault drivers? Suppose the area of interest is a retirement community, so that the resident population consists of older persons. The "younger" drivers are expected to be fewer in number. If most of those few drivers are at fault in accidents due to more reckless driving behavior, the number of victims will obviously be few, whereas the number of older drivers will be in greater quantity, both as victims and as at-fault drivers. The test for random victim selection supposedly would discover this. But if it does, what then? The hypothesis cannot be tested, yet the researcher will not discover this until after the data have been collected. This may be a very costly way to find out that no conclusions can be reached. So the user of this methodology should practice great care in thinking through the assumptions and should be very sure that there is no confounding among these partitionings. Given the difficulty of measuring exposure for the applications for which induced exposure is used, however, the methodology has its merits.

In conclusion, each design has its merits, strengths, weaknesses, and assumptions. The best design to use will ultimately depend on the application, validity of assumptions, and objective inferences of the study. This report has attempted to summarize those most commonly used designs and methodologies in highway safety treatment evaluations. By carefully assessing the applicability of each to one's individual research study, validating the necessary assumptions and weighing the weaknesses and pitfalls presented in the critiques for each methodology, it is hoped that future highway researchers may better plan, conduct, and analyze such studies to the betterment of future highway safety decisions.

REFERENCES

1. Michaels, R.M. "Two Simple Techniques for Determining the Significance of Accident-Reducing Measures," *Traffic Engineering*, pp. 45-46, September 1966.

2. Gerlough, D.L. and F.C. Barnes. *The Poisson and Other Probability Distributions in Highway Traffic*, Saugatuck, Connecticut: Eno Foundation, 1971.

3. Tanner, J.C. "A Problem in the Combination of Accident Frequencies," *Biometrica*, 45, pp. 331-341.

4. Woolf, B. "On Estimating the Relationship Between Blood Group and Disease," *Annals of Human Genetics*, 19, pp. 251-253, 1955.

5. Griffin, L.I. *Three Procedures for Evaluating Highway Safety Improvement Programs*, Report Number TARE 51, Texas Transportation Institute, Texas A&M University, College Station, TX, October 1982.

6. Griffin, L.I. *A Systematic Framework for Analyzing Categorical, Before-and-After Data*, Texas Transportation Institute, Texas A&M University, College Station, TX, April 1989a.

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

8. Arnold, S.F. and C.E. Antle. "An Empirical Bayes Solution for the Problem Considered by Williford and Murdock," *Accident Analysis and Prevention*, 20, No. 4, pp. 229-301, 1988.

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

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

11. Christiansen, C.L. and C.N. Morris. *Hierarchical Poisson Regression Modeling*, Report #HCP-1994-2, Harvard Medical School, Cambridge, MA, 1994.

12. Griffin, L.I. *A Criticism of Evans' Double Pair Comparison Method*, Texas Transportation Institute, Texas A&M University, College Station, TX, in-house report, December 1989c.

13. Pendleton, O.J. *Evaluation of Accident Analysis Methodologies*, Publication No. FHWA-RD-96-039, Federal Highway Administration, Washington, DC, report in press, December 1995.

14. Lyles, R.; N. Stamatiadis; and D. Lightizer. "Quasi-Induced Exposure Revisited," *Accident Analysis and Prevention*, Vol. 23, pp. 275-285, 1991.

APPENDIX A: LITERATURE REVIEW OF INDIRECT-EXPOSURE HIGHWAY SAFETY EVALUATION STUDIES

This literature review consists of brief reviews of major publications on methodologies that have been used to evaluate highway safety studies, but which do not directly use exposure methods (such as rate-based methodologies) in the statistical analysis. These methodologies are referred to as *indirect* methods that attempt to account or adjust for exposure through the experimental design of the study. The most prominent methodologies are the before-and-after methods with comparison groups. More recently, the methodology of *induced* exposure has been proposed. Since this methodology is more recent and has not been explored in depth, the literature review will focus heavily on these methods. The more traditional "non-induced" methodologies, which span a very long period of time and have been used in numerous studies and applications, will be reviewed more or less historically. For this reason, this literature review will be divided into two sections — non-induced-exposure methodologies and induced-exposure methodologies. The non-induced methodology sources will be listed chronologically as the methods were introduced into the literature. However, due to the large numbers of induced-exposure sources, these will be listed alphabetically by first authors to facilitate referencing these sources at a later date.

NON-INDUCED-EXPOSURE METHODOLOGIES

When we conduct a before-and-after analysis of the effectiveness of an accident countermeasure, it is not uncommon to ask: "Was exposure comparable in the before-and-after periods?" If not, we would like to be able to adjust our data to take into account any change in exposure that may have transpired between the before-and-after periods.

Typically, when we think of exposure, we think of measures such as "millions of vehicle-miles traveled," "number of vehicles entering the intersection," or even "number of months in the before-and-after periods." Each of these exposure measures is (or is assumed to be) a surrogate for the "opportunity for an accident to occur." If exposure (i.e., accident opportunity) is increasing from the before to the after period (independent of the treatment imposed), then any simplistic analysis of treatment effect may *underestimate* the benefits of treatment. Conversely, if exposure is decreasing from the before to the after period, any simplistic analysis of treatment effect may *overestimate* the benefits of treatment.

As an alternative to the use of exposure measures such as those cited above, some authors have chosen instead to use "comparison accidents" for the purposes of calibrating the effectiveness of accident countermeasures. These comparison accidents are typically accidents recorded at an adjacent site, before and after the implementation of an accident countermeasure at a treatment site. Just as we might use increases or decreases in exposure to adjust our analysis of the effectiveness of an imposed treatment, so too, increases or decreases in accidents at the comparison site might be used to adjust our estimates of treatment effectiveness.

Several articles will be briefly reviewed (and others referenced) to show the development of procedures that have been advanced for evaluating accident countermeasures through the use of comparison accidents. Several of these articles come from the epidemiological literature. That is to say, these articles were not originally written to assist highway safety evaluators. Nevertheless, the logic and the arithmetic underlying these methodologies can be easily transformed for present purposes.

Evans, L. "Rear Seat Restraint System Effectiveness in Preventing Fatalities," *Accident Analysis and Prevention*, Vol. 20, pp. 129-136, 1988a.

In 1986, Evans developed an evaluation methodology that he referred to as the "Double Pair Comparison" method. In fact, this method is nothing other than a very clever application of Woolf's methodology. See Griffin (1989c).

Other examples of the application of the double-pair comparison method include Evans (1988b), and Evans and Frick (1986 and 1988).

To see how Evans' method works, consider the following data taken from his 1988a article. In this study, Evans sought to determine the effectiveness of restraints in reducing fatal injuries to right-rear-seat occupants. The data were selected from the Fatal Accident Reporting System (FARS) for the calendar years 1975-1985 and included approximately 10,265 accident-involved passenger cars. Each accident-involved car contained a right-rear passenger (R) and a driver (D), or right-front passenger (F), or left-rear passenger (L). In addition, one or more fatalities were recorded in each car in the data set.

In the table shown below, the comparison (C) vehicle occupants [i.e., the driver (D), right-front passengers (F), and left-rear passengers (L)] and the right-rear occupants (R) are either restrained ($_R$) or unrestrained ($_U$), as indicated by the subscripts. The comparison occupants (C) and the right-rear passengers (R) were either fatally injured (^{F}) or not fatally injured (^{N}), as indicated by the superscripts.

	Accident-Involved Passenger Cars					Fatalities	
Seating Configuration	$C^F R^N$	$C^F R^F$	$C^N R^F$	$C^N R^N$	Total	Comparison	Right Rear
1 $D_U R_U$	4802	1462	3414		9678	6264	4876
2 $D_U R_R$	39	9	22		70	48	31
3 $D_R R_U$	116	44	184		344	160	228
4 $D_R R_R$	63	30	80		173	93	110
5 $F_U R_U$	4245	1747	2643		8635	5992	4390
6 $F_U R_R$	38	8	20		66	46	28
7 $F_R R_U$	86	44	121		251	130	165
8 $F_R R_R$	59	28	71		158	87	99
9 $L_U R_U$	1697	870	1666		4233	2557	2536
10 $L_U R_R$	17	6	6		29	23	12
11 $L_R R_U$	14	2	14		30	16	16
12 $L_R R_R$	29	11	31		71	40	42

Considering the fatalities in the first two rows in this table, we see that 6,264 unrestrained drivers and 4,876 unrestrained right-rear passengers were killed while riding in 9,678 accident-involved passenger cars. In addition, we see that 48 restrained drivers and 31 restrained right-rear passengers were killed while riding in 70 accident-involved passenger cars. Using the unrestrained drivers as a comparison group, Evans sets up the following odds ratio (U):

$$U = 31/48 \div 4876/6264 = 0.83$$

This odds ratio (U) indicates that fatalities to right-rear passengers are 17 percent less likely when they are restrained, when controlling for the likelihood of fatality to unrestrained drivers riding in the same vehicles.

By similar logic, Evans compares rows 3 and 4, 5 and 6, 7 and 8, 9 and 10, and 11 and 12. For each of these pairings, he calculates an odds ratio. He then takes the weighted average log odds ratio associated with these six pairings as proposed by Woolf (1955) to come up with an overall estimate of the effectiveness of restraints in reducing fatalities to right-rear passengers.

It should be noted that in this analysis, Evans is reusing his data. That is to say, the fatally injured right-rear passengers in the first two rows are those same fatalities that appear in rows 3 through 12. No adjustment is made in Evans' calculations to allow for this reuse of data. See Griffin (1989c) and Kahane (1987).

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

The standard method by which traffic engineers often seek to assess the effectiveness of accident countermeasures is the simple before-and-after design. The difference in the number of accidents recorded during the before-and-after period is assumed to be a measure of the effect of treatment.

Two statistical tests are offered to determine whether the difference in before-and-after accidents is within chance variation. If the difference is within chance variation, the treatment is termed *not significant*.

The first of Michaels' two tests is based on the Poisson distribution. In this test, the number of accidents recorded in the before period is defined as the expected number of accidents in the after period. Then, a simple one-sample Poisson test is used to compare the expected and observed accidents in the after period. This test, which Michaels refers to as a *liberal test*, assumes that the number of accidents in the before period is known without error. Given this very generous assumption, this test is, indeed, liberal. [Note: This test is not often seen in before-and-after evaluations conducted in this country, although it may be more popular in France (see, for example, Garwood and Newby (1970), pp. 42-43).]

The second of Michaels' two tests is based on the chi-square distribution. In this test, Michaels assumes that accidents would be equally distributed across the before-and-after periods (plus or minus chance fluctuation) if the imposed treatment was totally ineffective. Although Michaels does not show the derivation of this chi-square equation, the derivation has been shown elsewhere (Griffin, 1989a). Michaels' chi-square is nothing other than McNemar's chi-square (McNemar, 1947):

$$\chi^2 = (A\text{-}B)^2/(A+B)$$

where A and B represent accidents recorded after and before treatment, respectively.

Supplemental Comments: Other authors have offered alternative tests of significance for the before-and-after design. Gerlough and Barnes (1971), for example, transform Michaels' (McNemar's) chi-square into a Z-test by taking the square root of both sides of the equation shown above. With a correction for continuity, this equation becomes:

$$Z = (A\text{-}B\text{-}1)/\sqrt{(A+B)}$$

Griffin (1989a,b) defines a Z-test for the before-and-after design based on the logit (i.e., the log odds):

$$Z = \log(A/B)/\sqrt{(1/A+1/B)}$$

It can be shown that for relatively large sample sizes, the Z-test based on the square root of McNemar's chi-square and the Z-test based on the logit are comparable. There are, however, some advantages to defining the before-and-after data in terms of a log odds. It will not be demonstrated here, but by use of the odds ratio (and its natural logarithm), it is possible to not only assess the statistical significance of the treatment imposed, but to estimate the percent reduction (or increase) in accidents (from before to after) and to define a confidence interval about that estimate.

For additional comments on the before-and-after design, see Griffin (1982; 1990a,b); Griffin, Powers, and Mullen (1975); Joksch (1968); and Laughland, Haefner, Hall, and Clough (1975).

Tanner, J.C. "A Problem in the Combination of Accident Frequencies," *Biometrica*, Vol. 45, pp. 331-341, 1958.

This article adds two extensions to the simple before-and-after design that has just been discussed:

- It allows for the use of a comparison group to adjust the accidents recorded at the treatment site before and after the treatment was imposed.

- It recognizes that multiple treatment and comparison sites might be employed in a single evaluation of a given treatment.

Imagine that 100 accidents were recorded at a given section of highway in the last 3 years (1989-1991). Then, the highway section was treated and in the next 3 years only 80 accidents were recorded (1992-1994). There was an apparent 20 percent reduction in accidents following treatment.

Now imagine that on highway segments adjacent to the treatment section, 200 accidents were recorded during the before period (1989-1991) and 180 were recorded during the after period (1992-1994), i.e., there was an apparent 10 percent reduction in accidents from before to after at the comparison section.

Tanner derives a modified version of McNemar's chi-square to test the significance of the reduction in accidents at the treatment section that takes into account the reduction in accidents at the comparison section:

$$\chi^2 = (A\text{-}Bc)^2/(A+Bc)$$

where A and B still equal accidents after and before treatment, respectively, at the treatment site; and c (the "comparison ratio") equals the ratio of after accidents to before accidents at the comparison section, i.e., c = 200/180 = 1.1111.

It should be noted that c is assumed to be known without error — a very generous assumption.

In this article, Tanner also considers the situation in which multiple treatment and comparison groups are included in the evaluation. He offers: (1) a test to determine if individual treatment-comparison pairings are yielding consistent estimates of treatment effectiveness and (2) a procedure for combining the information from several treatment-comparison pairings. The specifics of Tanner's technique for combining information from several treatment-comparison pairings will not be commented upon here because the arithmetic is fairly complex and, more important, the technique has been found to be less useful than other, simpler techniques that will be commented upon shortly, e.g., Woolf's technique.

Woolf, B. "On Estimating the Relationship Between Blood Group and Disease," *Annals of Human Genetics*, Vol. 19, pp. 251-253, 1955.

Although the methodology developed in this article was addressed to medical and epidemiological problems, this article has proved to be most useful for assessing the effectiveness of accident countermeasures. This methodology is based on the odds ratio and its natural logarithm.

Woolf's log odds ratio (expressed in accident countermeasure terms) is simply:

$$L = \log(A/B \div a/b)$$

where A and B are accidents at the treatment site after and before treatment, respectively; and a and b are accidents at the comparison site after and before treatment, respectively.

The sampling distribution of L is asymptotically normal with a standard error (for relatively large samples) of:

$$L_{se} = \sqrt{[(1/A) + (1/B) + (1/a) + (1/b)]}$$

A test of significance based on the standard normal (Z) distribution can be offered as the ratio of the last two equations:

$$Z = L/L_{se}$$

Note that the comparison ratio (a/b) in Woolf is not assumed to be known without error.

This last equation can be used to test the significance of treatment imposed at a single treatment-comparison pairing, but what about the situation in which there are multiple treatment-comparison pairings? Consider the data in the following table:

	First Treatment-Comparison Pairing		Second Treatment-Comparison Pairing		Aggregated Data	
Time	Comparison	Treatment	Comparison	Treatment	Comparison	Treatment
Before	87	20	24	50	111	70
After	87	18	48	90	135	108
"Effect"	apparent 10% *reduction*		apparent 10% *reduction*		apparent 26.8% *increase*	
Z	-0.29		-0.34		+1.19	

In this example, the data from two treatment-comparison pairings are summed and the effectiveness of the summed data is calculated. Here we see that two treatment-comparison pairings, both of which show an apparent 10 percent reduction in accidents associated with treatment, when added together show an apparent 26.8 percent increase in accidents associated with the imposition of treatment. This anomalous outcome is often referred to as Simpson's Paradox (Simpson, 1951). See also Goddard (1991).

Rather than summing the data from individual treatment-comparison pairings and then calculating one overall log odds ratio, Woolf recommends calculating the log odds ratio of each treatment-comparison pairing and then taking the weighted average of several pairings to calculate an overall log odds ratio. The weights associated with the individual log odds ratios are the reciprocals of their variance. Thus, for the first treatment-comparison in the table shown above, the standard error about the estimated log odds ratio is:

$$L_{se} = \sqrt{[(1/18) + (1/20) + (1/87) + (1/87)]}$$

And, the weight (w) assigned to this log odds ratio is:

$$w = 1/[(1/18) + (1/20) + (1/87) + (1/87)]$$

Woolf goes on to provide a test of the "homogeneity" (i.e., the similarity) of the individual log odds ratios. Woolf reasons that if the individual log odds ratios differ by more than chance fluctuation, then some other factor or factors are at play. In the presence of these other factors, the weighted average log odds ratio may be misleading. That is to say, some other admixture of treatment-comparison pairings might yield a very different weighted average log odds ratio.

Supplemental Comments: Woolf's procedure is discussed in a number of texts, e.g., Breslow and Day (1980), Fleiss (1973), Kahn and Sempos (1989), and Selvin (1991). An application of Woolf's procedure to assess the effectiveness of accident countermeasures can be found in Griffin (1990b).

For a good discussion of the use — and limitations — of comparison groups in evaluating the effectiveness of accident countermeasures, see Hauer (1991a,b). For a discussion of a procedure for assessing the "comparability" of comparison groups, see Griffin (1982), and Griffin and Hatfield (1982).

Alternative procedures for combining information from two-by-two tables are to be found in Cochran (1954), and Mantel and Haenszel (1959).

NON-INDUCED-EXPOSURE METHODOLOGY REFERENCES

Breslow, N.E. and N.E. Day. *Statistical Methods in Cancer Research, Volume 1: The Analysis of Case-Control Studies*, World Health Organization, Lyon, France: International Agency for Research on Cancer, 1980.

Cochran, W.G. "Some Methods of Strengthening the Common χ^2 Tests," *Biometrics*, Vol. 10, pp. 417-451, 1954.

Evans, L. "Double Pair Comparison — A New Method to Determine How Occupant Characteristics Affect Fatality Risk in Traffic Crashes," *Accident Analysis and Prevention*, Vol. 20, pp. 129-136, 1986.

Evans, L. "Rear Seat Restraint System Effectiveness in Preventing Fatalities," *Accident Analysis and Prevention*, Vol. 20, pp. 129-136, 1988a.

Evans, L. "Risk of Fatality From Physical Trauma Versus Sex and Age," *The Journal of Trauma*, Vol. 28(3), pp. 368-378, 1988b.

Evans, L. and M.C. Frick. "Safety Belt Effectiveness in Preventing Driver Fatalities Versus a Number of Vehicular, Accident, Roadway, and Environmental Factors," *Journal of Safety Research*, Vol. 17, pp. 143-154, 1986.

Evans, L. and M.C. Frick. "Helmet Effectiveness in Preventing Motorcycle Driver and Passenger Fatalities," *Accident Analysis and Prevention*, Vol. 20, pp. 447-458, 1988.

Fleiss, J.L. *Statistical Methods for Rates and Proportions*. New York: John Wiley and Sons, 1973.

Garwood, F. and R.F. Newby. "Use of Chi-Squared Distributions for Comparing Accident Frequencies," in *Road Research — Symposium on the Use of Statistical Methods in the Analysis of Road Accidents*. Paris: Organisation for Economic Co-operation and Development, pp. 37-43, 1970.

Gerlough, D.L. and F.C. Barnes. *The Poisson and Other Probability Distributions in Highway Traffic*, Saugatuck, Connecticut: Eno Foundation, 1971.

Goddard, M.J. "Constructing Some Categorical Anomalies," *The American Statistician*, Vol. 45(2), pp. 129-134, 1991.

Griffin, L.I. *Three Procedures for Evaluating Highway Safety Improvement Programs*, Report Number TARE 51. Texas Transportation Institute, Texas A&M University System, College Station, TX, October 1982.

Griffin, L.I. *A Systematic Framework for Analyzing Categorical, Before-and-After Data*. Texas Transportation Institute, Texas A&M University System, College Station, TX, April 1989a.

Griffin, L.I. *Using Before-and-After Data to Estimate the Effectiveness of Accident Countermeasures Implemented at Several Treatment Sites*. Texas Transportation Institute, Texas A&M University System, College Station, TX, December 1989b.

Griffin, L.I. *A Criticism of Evans' Double Pair Comparison Method*, TTI-sponsored report, December 1989c.

Griffin, L.I. *Engineers' Guide to Program and Product Evaluation*, Publication No. FHWA-SA-93-028, Federal Highway Administration, Washington, DC, April 1990a.

Griffin, L.I. *Using the Before-and-After Design with Yoked Comparisons to Estimate the Effectiveness of Accident Countermeasures Implemented at Multiple Treatment Locations*, TTI-sponsored report, May 1990b.

Griffin, L.I. and N.J. Hatfield. "Evaluation of a Selective Traffic Enforcement Program (STEP) in Tyler, Texas," *Traffic Safety Evaluation Research Review*, Vol. 1(4), pp. 13-24, 1982.

Griffin, L.I.; B. Powers; and C. Mullen. *Impediments to the Evaluation of Highway Safety Programs*. Highway Safety Research Center, University of North Carolina, Chapel Hill, NC, 1975.

Hauer, E. "Should Stop Yield? Matters of Method: Safety," *ITE Journal*, Vol. 61(9), Institute of Traffic Engineers, September 1991a.

Hauer, E. "Comparison Groups in Road Safety Studies: An Analysis," *Accident Analysis and Prevention*, Vol. 23(6), 1991b.

Joksch, H.C. "Comments on *Significance Tests for Accident Reductions Based on Classical Statistics and Economic Consequences*," *Transportation Science*, Vol. 2(2), May 1968.

Kahane, C.J. *Fatality and Injury Reducing Effectiveness of Lap Belts for Back Seat Occupants* (870486), Detroit, MI: Society of Automotive Engineers, pp. 45-51, 1987.

Kahn, H.A. and C.T. Sempos. *Statistical Methods in Epidemiology*. Oxford: Oxford University Press, 1989.

Laughland, J.C.; L.E. Haefner; J.W. Hall; and D.R. Clough. *Methods for Evaluating Highway Safety Improvements (NCHRP 162)*, Transportation Research Board, Washington, DC, 1975.

Mantel, N. and W. Haenszel. "Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease," *Journal of the National Cancer Institute*, Vol. 22, pp. 719-748, 1959.

McNemar, Q. "Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages," *Psychometrika*, Vol. 12, pp. 153-157, 1947.

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

Selvin, S. *Statistical Analysis of Epidemiological Data*. Oxford: Oxford University Press, 1991.

Simpson, E. "The Interpretation of Interaction in Contingency Tables," *Journal of the Royal Statistical Society*, Ser. B, Vol. 13, pp. 238-241, 1951.

Tanner, J.C. "A Problem in the Combination of Accident Frequencies," *Biometrica*, Vol. 45, pp. 331-341, 1958.

Woolf, B. "On Estimating the Relationship Between Blood Group and Disease," *Annals of Human Genetics*, Vol. 19, pp. 251-253, 1955.

INDUCED-EXPOSURE METHODOLOGIES

Brown, I.D. "Exposure and Experience Are a Confounded Nuisance in Research on Driver Behavior," *Accident Analysis and Prevention,* Vol. 14(5), pp. 345-352.

The models posed by Brown and others are divided into two parts — one representing the driver's capabilities and the other representing traffic demands. These models imply that the two functions vary independently and do not consider interacting effects.

Brown states that crude measures of exposure, such as vehicle-miles traveled (VMT), provide inadequate methods for controlling accident data if the objective is to assess individual differences in liability. The example he gives to show this is trying to relate driver speed to accidents. Because of individuals' differing abilities to drive safely at different speeds, two drivers may have identical "exposures" as defined by VMT, yet one has a greater risk of having an accident than the other. He terms this *self-induced exposure risk.*

Brown criticizes Thorpe and Haight's method of "induced exposure," which makes the assumption that in single-vehicle accidents, the driver's exposure is equal to that of the guilty driver in two-vehicle collisions. Brown claims that this is a very questionable assumption due to a variety of personal characteristics, such as aggression, extroversion, and anxiety. These characteristics will tend to produce greater self-induced driver risk exposure when other drivers are present. Brown also emphasizes the need to validate these assumptions with data.

Basically, this is a non-mathematical discussion of some theories of human factors with regard to accident risk and it does not really provide any suggestions as to how to analyze data when induced exposure is a necessary factor.

Carr, B.R. "A Statistical Analysis of Rural Ontario Traffic Accidents Using Induced Exposure Data," *Accident Analysis and Prevention*, Vol. 1, pp. 343-358, 1969.

Carr presents a method of induced exposure using a relative risk (RR) function. The advantages of this method are:

- Provides a measure of exposure that is directly related to the risk of having accidents.
- Exposure is provided by categories.
- It is quick, easy, and inexpensive to apply.
- Takes advantage of the great quantities of mass accident data already collected.
- Controls for all environmental conditions.
- Permits further analysis within and between exposure populations.

Data used in this study were based on 1966 and 1967 Ontario police-reported accidents. The variables available for analysis were driver age, sex, and driving experience; vehicle condition; driver residence; and driver condition. A modification of Thorpe's method was used. Whereas Thorpe relied on single-vehicle collisions as a surrogate measure for the exposure of the "at-fault driver" in two-vehicle collisions, Carr assesses culpability in each two-vehicle accident from the accident reports. The basic requirements of Carr's method are:

- Responsibility for any single-vehicle accident lies sole with the driver-vehicle combination involved.
- Only two-vehicle collisions were included.
- Drivers in two-vehicle collisions can be assigned culpability.
- Frequency of involvement of any driver-vehicle combination as the non-responsible combination is representative of the exposure for that driver-vehicle combination.

The last assumption is the core of the induced-exposure methodology. The relative risk index is defined as:

$$RR_i = \frac{\text{frequency of at-fault driver accidents for category i}}{\text{frequency of innocent victims for category i}}$$

Carr computed this index for each age category and by single- and multiple-vehicle collision. In the multiple-vehicle collisions, a plot of this index by age group takes on the typical U-shape, i.e., the relative risk index is highest at the young and old age groups. For single-vehicle collisions, the function is monotonically decreasing with age. When this same process was used for driving experience, the functions were both monotonically decreasing, but they crossed at around 10 years of experience. With 10 or more years of driving experience, the relative risk for single-vehicle collisions was less than that for multiple-vehicle collisions, where for less than 10 years, the relative risk was higher for single-vehicle collisions. When the data were further partitioned and only drivers with 5+ years of experience were analyzed, the relationship of relative risk and age changed from the sharp U-shape for multiple-vehicle accidents to a more backward J-shape, i.e., the relative risk increased more sharply for the older age groups than for the younger age groups. The relative risk for males and females with age was similar and monotonically decreasing for single-vehicle accidents. Carr then proceeded with building regression models and doing a factor analysis. The regression model is highly suspect, however, as the dependent variable was defined as either "2" for non-responsible or "3" for responsible — a very strange coding indeed. A factor analysis was performed on 20 dichotomous variables and is also somewhat suspect.

Cerrilli, Ezio C. "Driver Exposure: The Indirect Approach for Obtaining Relative Measures." *Accident Analysis and Prevention,* Vol. 5, pp. 147-156, 1973.

This study used the method of indirect exposure to study the relationship between driver age and accident involvement. The two basic assumptions required of this approach are:

- All drivers involved in two-vehicle accidents can be divided into two groups — responsible and not responsible.
- The number of not-responsible drivers in any group is proportional to that group's exposure.

Exposure is defined as a measure of movement. At the time of this study (1972), the only sources of data for estimating the amount of travel had been tax revenues for fuel consumption. Cerrilli recognized that even where sample traffic counts were being taken on roadway systems, neither of these estimates of "exposure" was suitable for conducting safety studies on driver characteristics, such as driver age.

Thorpe is credited as the originator of the induced-exposure concept in 1964. The fundamental concept is that the likelihood of a not-responsible combination being involved in an accident is proportional to meeting that combination on the road. Carr and Haight followed with mathematical models that were slightly different, but the basic concept of induced exposure was the same. Cerrilli stated three reasons why induced exposure is such a popular concept:

1. It is based on already available accident data.
2. Finer exposure measurements are possible because relative exposure measures for any driver-vehicle class can be derived from any environmental condition.
3. No other method is available for handling exposure for driver-vehicle safety studies.

For this study, 1969 accident data from 24 States were subdivided into 14 age groups. The actual number of drivers in each age group for each State was also tabulated to obtain a percentage distribution by driver group (the author now lists 25, rather than 24, States). Only multiple-vehicle accidents were used in this study and the accident data files contained information on culpability (was the driver responsible?). At this point, I am a bit confused as Texas is listed as one of the States and yet Texas does not have a category for assigning culpability to drivers in accidents.

The author then computes three indices from these data:

1. Relative exposure index (R.E.I.), which is the ratio of the percentage of innocent victims to the percentage of drivers in a particular age group.

2. Liability index (L.I.), which is the percentage of responsible (at-fault) drivers over the percentage of licensed drivers in each age category.

3. Hazard index (H.I.), which is the percentage ratio of the L.I. to the R.E.I. In this case, the percentage of drivers in each category cancels out and this index then provides an induced-exposure method for relating driver responsibility by age group without the direct measure of exposure by age group.

In this study, the author admits that it was tempting to continue to partition these indices by other variables, but he recognized the loss of power in continuing to subdivide the data. At this point, the paper disappointingly ends. No conclusions based on statistical validity are presented. The R.E.I., which should be a measure for validating the assumption that the innocent victims are indeed proportional to the actual drivers in each age group, varies dramatically. If this index is significantly different from 1, this would imply that the basic assumption on which induced exposure is based is invalid. The author makes no mention of this and ignores what would appear to be the most critical factor in the study.

Chapman, Roger. "The Concept of Exposure," *Accident Analysis and Prevention,* Vol. 5, pp. 95-110, 1973.

Chapman claims to explore the concept of exposure from its general to its particular use and this work serves as a literature review of exposure up to 1972. For purposes of this review, we will only report on the induced-exposure portion of this paper, although it covers a much broader range of exposure data.

Chapman defines the concept of exposure as an attempt to take account of the amount of opportunity for accidents that the driver or traffic system experiences. Brown defines *induced exposure* as an attempt to "induce" what the exposure is from the accident data. Thorpe (1964) is credited with first proposing this and Carr modified this in 1969. Thorpe uses the relative frequency of single-vehicle accidents as a surrogate for the guilty driver in multiple-vehicle accidents based on five assumptions. In Carr's data, the driver culpability was actually available and no assumption as to guilt had to be made. The *Research on Road Safety* in 1963 computed the ratio of involvement in all accidents to the involvement in fixed-vehicle collisions for a number of age groups. A comparison of Thorpe's and Carr's methods is available here as the report also computes exposure for these age groups using the "at-fault" or innocent victim criterion. The two methods were found to be similar. Haight then presented an idea, based on Thorpe's method, for making "exposure correcting" transformations. Koornstra presented an even more general approach, using a multivariate analysis of categorical data.

In the remaining portions of the paper not dealing with induced exposure, Chapman discusses exposure as opportunity and exposure for specific locations by types of accidents, including a discussion of what is the appropriate exposure measure for intersections.

Davis, Gary A. and Y. Goa. "Statistical Methods to Support Induced Exposure Analyses of Traffic Accident Data," *Transportation Research Record,* 1401, 1990.

This study presents an induced-exposure-type methodology using at-fault versus innocent driver as the induced-exposure factor. A definition of *induced exposure* is given as "the opportunity to be involved in the event (accident)." Accident *rate* is defined as "the tendency to have accidents." Accident counts are used to make inferences about accident rates, which require knowledge of *exposure*.

The induced-exposure model used in this study requires that in two-vehicle accidents, one driver is identified as the "at-fault" driver and the other is identified as the "innocent victim." The innocent victim is assumed to be "selected at random" by the at-fault driver from a selected pool of available drivers. The probability that the innocent victim is a member of a given subgroup is assumed to be directly proportional to that subgroup's *exposure* at the accident site. In this way, the same measure of exposure reflects both the "at-fault" driver and the "innocent victim." Thus, in the formation of the odds ratio to assess the relationship of a specific factor, such as gender or age, the "exposure" in the numerator and denominator cancel out.

Davis and Gao present two analyses using this methodology, one using gender and daytime and nighttime accidents, and the other using age and two highway locations.

In the first example, originally presented by Lyles et al., male and female drivers in accidents in Michigan in 1988 were assessed as to the at-fault driver and innocent victim. Daytime non-rush-hour accidents and nighttime accidents were tabulated in two-by-two tables, with at-fault being the row variable and innocent victim being the column variable, each partitioned by male/female. In both cases (daytime and nighttime), no association was found between the innocent victim and the at-fault driver. This provides validation of the assumption of random victim selection that allows pursuit of the next step — computing a log-rate ratio for comparing accident "rates" (the term *rate* here may be inappropriately used at this point in the paper because these are really ratios of proportions) between male and female drivers. In both cases, daytime and nighttime, males were found to have significantly higher accident "rates" (approximately 40 percent higher).

In the second example, Minnesota accidents for two Minnesota highways were grouped by age (25-55 and older than 55) and at-fault or innocent victim. These data represent accidents occurring at the signalized intersections along two Minnesota trunk lines (MNTH 47 and MNTH 65) during 1988-1989. The random selection assumption was valid for both trunk lines; there was no significant difference in driver age accident "rates" at MNTH 47, but at MNTH 65, there was some indication that older drivers had more accidents.

The study then provided another methodology — an empirical Bayes method for identifying high hazard locations. This was motivated by the results of the previous study, indicating that older drivers had more accidents at MNTH 65 since this corridor might be a candidate for safety improvements targeted at older drivers. There were 29 signalized intersections along this corridor and they could have differing risk potentials to older drivers.

Haight, F. "Induced Exposure," *Accident Analysis and Prevention*, Vol. 5, pp. 111-126, 1973.

Haight noted that the concept of *exposure to accidents* was derived from the epidemiological concept of *exposure to disease*. Haight referred to a symposium held at the University of California Institute of Transportation and Traffic Engineering in the early 1950's where a list of more than 100 potential exposure factors was given, ranging from total vehicle weight tax payable, to the density of bars and taverns (no specific references to any publications from this symposium are given).

Haight termed exposure definitions, such as the ones developed at this symposium, *direct* definitions. *Indirect* definitions are quantities other than traffic parameters, such as registered drivers, population, fuel tax, etc. *Induced* exposure is categorized as a subset of indirect exposures. Again, Thorpe is credited with first defining *induced exposure* that is based solely on accident experience. Haight defines *quasi-induced* exposure as the rest of the literature defines *induced exposure*, i.e., exposure based on two factors — accident experience and culpability — the "innocent victim" definition. Haight's paper deals with the *fully* induced definition and deals only marginally with the *quasi-induced* and other peripheral concepts.

Fully induced exposures are further subdivided into the categories of "narrow" sense and "wide" sense. The distinction between these definitions is not at all clear to me. Haight uses the analogy of a car hitting a tree. In the "narrow" definition, this accident could not be due to exposure, but to the same factor influencing the guilty party in a two-car collision. In the "wide" sense, this accident would not differ from a two-car collision except that the "not guilty" object was moving and was not fixed. Induced exposure in the "narrow" sense refers to exposure to automobile collisions only, and assumes that the proportion of single-vehicle collisions in a given category (say age) are equal to the proportion of guilty drivers in two-vehicle collisions for that same category. Induced exposure in the wide sense applies to all types of hazards divided into four categories. For two-car collisions, there are *internal* and *external* factors; the internal factor is proneness to an accident and the external factor is exposure to the accident. For single-vehicle accidents, the categories are "fixed" or "moving."

Haight then proceeds to develop weighting factors based on a normalized measure of exposure, e.g., for each age category, exposure would be computed as the proportion of vehicle-miles traveled (VMT) for that age group relative to total VMT. The claim is that Haight's weighting scheme is slightly different from the one proposed by Thorpe. Basically, these methodologies for accounting for exposure do so by computing weighting factors, as opposed to what Haight refers to as *quasi-induced* methods that assume that the exposures cancel out in the ratios of innocent victim and guilty driver proportions.

Haight makes the point that these induced-exposure models need validation based on real data. All of the models proposed are purely theoretical and no one really knows if they are realistic. This paper does not give any numerical examples to illustrate how to apply these theoretical models to actual data.

Koornstra, M.J. "A Model for Estimation of Collective Exposure and Proneness From Accident Data," *Accident Analysis and Prevention*, Vol. 5, pp. 157-173, 1973.

This study was inspired by Haight's attempt to weaken Thorpe's axioms and his attempt to test the fit of the model on the basis of more equations than unknowns. This is an extremely mathematical paper that attempts to define abstract probability models to act as surrogates for exposure data. It is based entirely on assumptions. Note that examples are given using real data and there is no indication that any of the numerous assumptions that are made regarding these probabilities holds.

Stamatiadis, Nikoforos; William C. Taylor; and Francis X. McKelvy. "Accidents of Elderly Drivers and Intersection Traffic Control Devices," *Journal of Advanced Transportation*, Vol. 24, No. 2, pp. 99-112, 1990.

This research examined the relationship between accidents of elderly drivers and intersection traffic control devices. The exposure to the accident (induced exposure) is defined as the "probability of being the driver not cited for the accident occurrence." The motivation for applying this methodology to the elderly driver problem was that elderly drivers are increasing in the population. In examining the driving population in two time periods — 1965-1975 and 1965-1985 — in both periods, elderly drivers showed the fastest growth.

The research of Carr, Haight, and Thorpe were cited as previous studies that used the induced-exposure methodology. The induced-exposure methodology is based on the assumption that the accident exposure by any age group of drivers is proportional to the innocent victim involvements in multiple-vehicle accidents by that age group of drivers. This definition was originally attributed to Taylor.

Induced exposure was measured using a ratio called the *R.A.I.R.* — the percentage of accidents in which the driver was at fault to the proportion of drivers not at fault from the same age group. This distribution of the R.A.I.R. follows a U-shaped distribution, as documented in previous studies by McKelvy. The R.A.I.R. was greater than 1.0 for drivers younger than age 25 and older than age 60. The in-between age groups had R.A.I.R.'s of less than 1.0.

The data used in this study were based on multiple-vehicle intersection accidents on the Michigan trunk line system between 1983-1985. The total number of accidents during this time was 148,134. The data were divided into three age groups: younger than 25, 25-59, and older than 59. The R.A.I.R.'s were 1.19, 0.86, and 1.22, respectively.

The authors defined a *one-way* analysis of these data as a "partitioning of the R.A.I.R.'s by variables thought to affect elderly drivers differently than younger drivers." A *two-way* analysis was defined as "partitioning the R.A.I.R.'s by two factors (examining the interactions of variables with respect to R.A.I.R.'s)." The conclusions of these analyses were that:

- Elderly female drivers had higher R.A.I.R.'s in snowy weather.
- Female drivers at night had higher R.A.I.R.'s.
- Female drivers on multiple-lane highways had higher R.A.I.R.'s.
- Elderly drivers at night in rural areas had higher R.A.I.R.'s.
- Elderly drivers at signalized intersections on multiple-lane two-way roads had higher R.A.I.R.'s.

- In general, elderly drivers do not exhibit different accident patterns between signalized and unsignalized intersections under similar conditions (this conclusion was not supported by numbers or specifics such as what is meant by "similar conditions"). The only exception to this was "elderly female drivers at signalized intersections under snowy conditions." (This would appear to be a *three-way* partitioning — although the authors never mention this.)

The report continues in this fashion with various other hypotheses and conclusions. However, no numbers or statistical analyses and tests of hypotheses are ever given. No definition of *at-fault* or *how at-fault* was provided. Because of this, the conclusions can only be taken as the opinions of the authors, rather than statistically validated results. The primary contribution of this study appears to be in the definition of and use of a relatively unknown index of induced exposure.

Thorpe, J.D. "Calculating Relative Involvement Rates in Accidents Without Determining Exposure," *Australian Road Research*, Vol. 2, pp. 25-36, 1964.

Thorpe's paper is being credited as the first attempt to apply the method of induced exposure to the field of highway safety. It is written in a very strange style — each point made in the paper is numbered from 1 to 43. Basically, Thorpe uses single-vehicle accident counts as surrogates for at-fault driving exposure in multiple-vehicle accidents. It is based on five assumptions:

1. Single-vehicle accidents are caused by attributes of the driver-vehicle combination of interest.

2. Most collisions are caused by the first two vehicles, so only the first two vehicles in multiple-vehicle crashes are analyzed.

3. In each accident, a responsible and a non-responsible driver can be identified.

4. The relative likelihood of a responsible driver being involved in a multiple-vehicle accident is the same as the likelihood of a driver being in a single-vehicle accident.

5. The likelihood of any particular driver-vehicle combination being innocently involved in an accident is the same as the likelihood of meeting that combination anywhere on the road.

Thorpe computes a relative accident likelihood (R.A.L.) for each group, using a partitioning similar to that of Carr. He then applies this to Australian accident data for 1961-1962 for personal injury accidents only. When computing these over age, he did not find the traditional U-shaped curve, but rather a monotonically decreasing function of the likelihood over age. Given what we now know, this was probably due to other interacting factors that were not taken into account. Thorpe then computed these R.A.L.'s by driver experience and age, and he presented them in a two-way table. He concluded that as age and experience increase, the relative risk of accidents decreases.

To check the validity of his assumption based on single-vehicle accidents, Thorpe computed the ratio of single-vehicle accidents in an age group to the proportion of licensed drivers in that age group; this he labeled the involvement index (I.I.). When these indices were compared to the R.A.L.'s, there was a striking similarity in their quantities, although no statistical tests of equality were run. Additional analyses were performed on vehicle types and drinking drivers. Rather than draw conclusions regarding the relative risks by vehicle type, Thorpe merely compared R.A.L.'s and I.I.'s to show their similarities. The relative risk of drinking drivers was higher than that for non-drinking drivers, although no statistical tests were conducted.

In an appendix, Thorpe derives confidence limits for R.A.L.'s, but he does not use them for the inferences he makes throughout the paper.

INDUCED-EXPOSURE METHODOLOGY REFERENCES

Carlson, W.L. *Induced Exposure Revisited*, HIT Lab Report, University of Michigan, Ann Arbor, 1969.

Cooper, P. "Differences in Accident Characteristics Among Elderly Drivers and Between Elderly and Middle-Aged Drivers," Accident Analysis and Prevention, Vol. 22, pp. 499-508, 1990.

Garber, N. and R. Srinivasan. "Risk Assessment for Elderly Drivers at Intersections," *Transportation Research Record,* 1325, TRB, Washington, DC, 1991.

Haight, F.A. "Accident Proneness, The History of an Idea," *Automobilisimo & Automobilisimo Industriale*, Vol. 12, pp. 534-546, 1964.

Haight, F.A. "A Crude Framework for Bypassing Exposure," *Journal of Safety Research,* Vol. 2, pp. 26-29, 1970.

Haight, F.A. *Indirect Methods for Measuring Exposure Factors as Related to the Incidence of Motor Vehicle Traffic Accidents*, Report No. DOT-HS-041-1-049. University Park, PA: Pennsylvania State University, September 1971.

Hall, W.K. *An Empirical Analysis of Accident Data Using Induced Exposure*, HIT Lab Report, University of Michigan, Ann Arbor, 1970.

Hodge, G.A. and A.J. Richardson. "The Role of Accident Exposure in Transport Safety Evaluations I: Intersection and Link Exposure," *Journal of Advanced Transportation,* Vol. 19(2), 1984.

Hodge, G.A. and A.J. Richardson. "The Role of Accident Exposure in Transport System Safety Evaluations II: Group Exposure and Induced Exposure," *Journal of Advanced Transportation*, Vol. 19, pp. 201-213, 1985.

Koornstra, M.J. "Empirical Results on the Exposure-Proneness Model," *Accident Analysis and Prevention,* Vol. 5, pp. 175-189.

Lyles, R.W.; N. Stamatiadis; and D.R. Lighthizer. "Quasi-Induced Exposure Revisited," *Accident Analysis and Prevention*, Vol. 21, pp. 275-285, 1991.

Maleck, T. and H. Hummer. "Driver Age and Highway Safety," *Transportation Research Record,* 1059, TRB, Washington, DC, pp. 6-12, 1987.

McKelvy, F.X.; T.L. Maleck; N. Stamatiadis; and D. Hardy. *Relationship Between Driver Age and Highway Accidents, Phase Two.* College of Engineering, Michigan State University, 1987.

McKelvy, F.X.; T.L. Maleck; N. Stamatiadis; and D. Hardy. "Highway Accidents and the Older Driver," *Transportation Research Record*, 1172, TRB, pp. 47-57, 1988.

McKelvy, F.X. and N. Stamatiadis. *Accidents and the Older Driver at Signalized and Non-Signalized Intersections*. College of Engineering, Michigan State University, TRB Preprint, 1988.

Mengert, P. *Analysis and Testing of Koornstra-Type Induced Exposure Methods*, Report No. FR-45-U-FAA-86-01, Cambridge, MA: U.S. DOT, TSC, 1985.

Roine, Matti and Risto Kulmala. *Induced Exposure and Random Sample Models for Road Accidents*, Technical Research Centre of Finland, Transport Research, Espoo, Finland.

Stanton, H. *Inventory of Exposure Data*. Report CR 18, Office of Road Safety, DOT, Australia, 1981.

Taylor, W.C. and M.W. Delong. *Validation of the "Innocent Victim" Concept*. College of Engineering, Michigan State University, 1986.

Thorpe, J.T. "Calculating Relative Involvement Rates in Accidents Without Determining Exposure." *Traffic Safety Research Review*, March 3-8, 1967.

Van der Molen, H. "Child Pedestrian's Exposure, Accidents and Behaviour," *Accident Analysis and Prevention*, Vol. 13(4), pp. 193-224, 1981.

Waller, P.F.; D.W. Reinfurt; J.L. Freeman; and P.B. Imrey. "Method for Measurement in Exposure to Automobile Accidents," Presented at the 101st Annual Meeting of the American Public Health Association, 1974.

Wigan, M.R. "Bicycle Ownership, Use and Exposure in Melbourne 1978-79," *Australian Road Research Board*, A.I.R., pp. 380-383, 1982.

Wolfe, A.C. "The Concept of Exposure to the Risk of a Road Traffic Accident and an Overview of Exposure Data Collection Methods," *Accident Analysis and Prevention*, Vol. 14(5), pp. 397-405.