Evaluation of Wide Edgelines on Two-Lane Rural Roads

BENJAMIN H. COTTRELL, JR

The effect of 8-in.-wide edgelines on the incidence of run-off-the-road (ROR) and related accidents was evaluated. The treatment locations consisted of three two-lane rural road sections totaling 60.7 miles. A before-and-after design with a comparison group and a check for comparability was used to analyze data. Five years of accident data, covering the 3 years before wide edgeline installation and the 2 years after installation, were used. It was concluded that there is no evidence to indicate that wide edgelines significantly affected the incidence of ROR and related accidents for any individual treatment location or for the locations combined. The related accidents include ROR accidents that involved driving under the influence of alcohol or drugs, ROR accidents on curves, ROR accidents during darkness, and opposite-direction accidents.

There are a high number of run-off-the-road (ROR), drunken driving, and night accidents in rural areas. In 1985, there were 19,385 ROR accidents in rural areas in Virginia (*1*). Of this total, 268 (1.4 percent) were fatal accidents, 9,434 (48.6 percent) were injury accidents, and 9,683 (50.0 percent) were property damage accidents. ROR accidents accounted for 29.1 percent of all rural accidents, 40.7 percent of the fatal accidents (the largest percentage for any type of accident), and 35.6 percent of the injury accidents in rural areas. Individuals driving under the influence of alcohol or drugs (DUI) were involved in 9,878 (14.8 percent) of all rural accidents. Accidents involving DUI accounted for 34.4 percent of fatal accidents, 20.1 percent of injury accidents, and 11.0 percent of property damage accidents in rural areas. There were 22,570 accidents during darkness, which constituted 33.9 percent of all accidents in rural areas.

Edgelines are used to delineate the right edge of the roadway to provide guidance to motorists. The standard edgeline width is 4 in. The edgeline is one element in a pavement marking system that provides warning and guidance information to the driver without diverting attention from the roadway (*2*). Reflectorized pavement markings are the most common form of delineation at night, when reduced visibility creates a greater need for guidance information. Edgelines 8-in. wide may reduce the probability of a driver running off the road and increase the probability that a driver will position his vehicle close to the centerline. However, since it is possible that wide edgelines will influence the lateral position of the vehicle in this way, the probability of centerline encroachment may increase as well.

Virginia Transportation Research Council, P.O. Box 3817 University Station, Charlottesville, Va. 22903.

OBJECTIVES AND SCOPE

The objectives of this research were to evaluate the effect of wide edgelines on the incidence of ROR, DUI, and other related types of accidents, as well as on the lateral placement and speed of vehicles. The scope was limited to two-lane rural roads. Primary routes were selected because accident data are more detailed and more readily available for these than for secondary routes.

The subject of this paper is the incidence of accidents. The report that documented the evaluation of lateral placement and speed may be summarized as follows (*3*):

- There were no statistically significant differences between the 4- and 8-in.-wide edgelines in lateral placement, lateral placement variance, encroachments by automobiles and trucks, mean speed, and speed variance.
- The mean lateral placement was significantly lower for the 8-in.-wide edgeline. The difference was small, however, and of no practical significance.
- Lateral placement and speed were not practically affected by a change from a 4-in. to an 8-in.-wide edgeline.

STUDY DESIGN

Experimental Plan

After testing several procedures for evaluating highway safety improvements, a before-after design with a comparison group and a check for comparability was selected. A detailed description of this procedure is given by Griffin (*4*). The procedure he described is condensed and discussed later in this section. The before-after design with a comparison group and a check for comparability provides some relief from two fallacies. By using a comparison group, the influence of extraneous factors is at least partially controlled; therefore there is some relief from the post hoc ergo propter hoc (after the fact, therefore because of the fact) fallacy. By using multiple before and after readings (e.g., each year represents a reading), some relief is obtained from the regression to the mean fallacy (*4*). Consequently, this evaluation design is more rigorous and more valid than a simple before-after design and a before-after design with a comparison group.

The comparability is determined by the difference in the rate of change in the frequency of accidents at the treatment and comparison locations during the before and after periods (Figure 1). The rates of change in accident frequencies are expressed as natural logarithms. When the rates of change in accident frequencies of the treatment and comparison groups

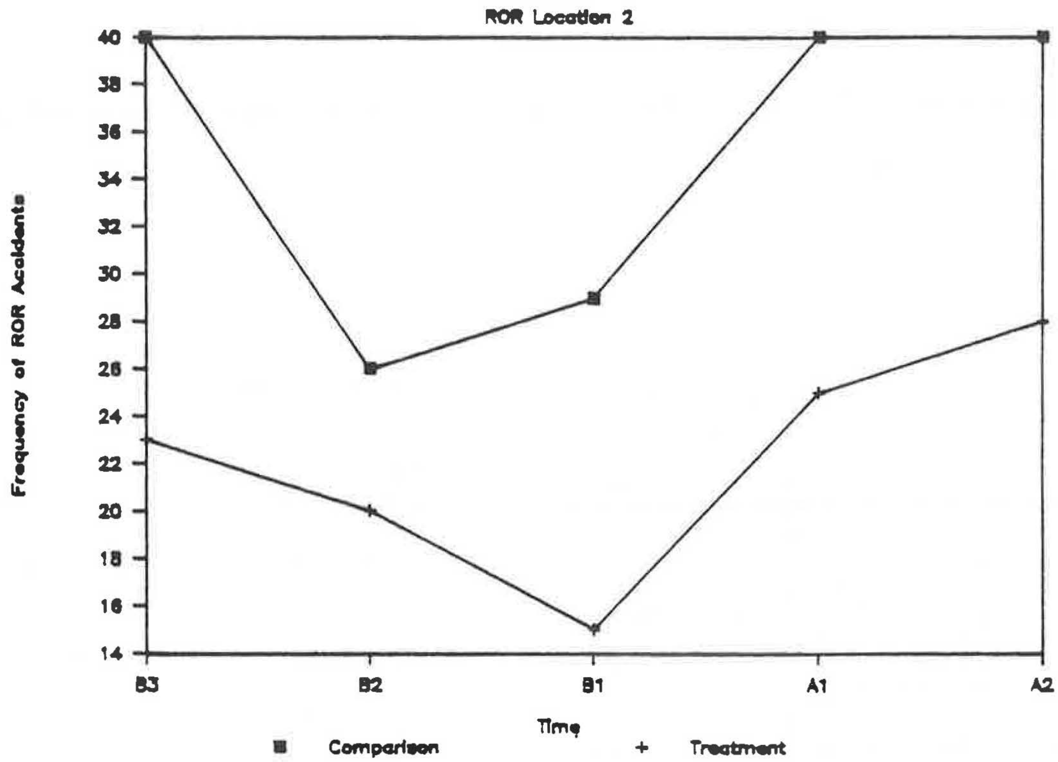

FIGURE 1 Frequency graph.

are equivalent, the slopes of the natural log (ln) frequency over time are the same, and therefore they are parallel (Figure 2). The procedure involves two steps:

Step 1: Check for Comparability. If the slopes on the treatment and comparison functions of ln frequency versus time deviate by more than chance expectation during the before and after periods, then the comparison group is not comparable to the treatment group, and further analysis is not appropriate. If the slopes do not deviate, there is no reason to doubt the comparability of the comparison group (*4*).

Step 2: Effect of the Treatment. In the second step, the treatment and comparison groups are collapsed across the before and after periods. If the slopes on the treatment and comparison functions do not deviate by more than chance expectation from before to after, then there is no evidence that the treatment imposed affected the incidence of accidents. If the slopes do deviate, then the treatment is said to have produced an effect. If the slope on the treatment is more negative (or less positive) than the slope in the comparison function, the treatment is beneficial. If the slope on the treatment function is less negative (or more positive) than the slope on the comparison function, the treatment is harmful (*4*).

Statistical Equations

The calculations used to answer the questions are based on the likelihood ratio chi-square (G^2) test. A $2 \times n$ contingency table, where n = total number of years of data, is developed. The overall goodness-of-fit test, G^2 total, is equal to the sum of G^2 Comparability and G^2 Treatment. In other words, the contingency table is partitioned into two parts:

- G^2 Comparability for the goodness of fit within the before and after periods for homogeneity of the treatment and comparison group, and
- G^2 Treatment for the goodness of fit from the before and after periods for the association of the treatment and comparison groups (*4, 5*).

The critical G^2 values that are compared with G^2 Comparability and G^2 Treatment are based on a 0.05 level of significance and are 7.81 and 3.84, respectively.

The formula for the likelihood ratio chi-square (G^2) test is (*4*):

$$G^2 = -2 \sum_i \sum_j X_{ij} \ln \frac{\hat{m}_{ij}}{X_{ij}} \qquad (1)$$

where X_{ij} = observed accident frequency in cell ij row (i) and column (j);

$$\hat{m}_{ij} = \frac{X_{i+} X_{+j}}{X_{++}}$$

for G^2 Before when $i = 1, 2, 3$, and $j = 1, 2$; for G^2 After and G^2 Treatment when $i = 1, 2$ and $j = 1, 2$

$X_{i+} = \sum_i X_{ij}$ (sum of row i)

$X_{+i} = \sum_j X_{ij}$ (sum of column j)

$X_{++} = \sum_i \sum_j X_{ij}$ (sum of the partitioned contingency table being tested)

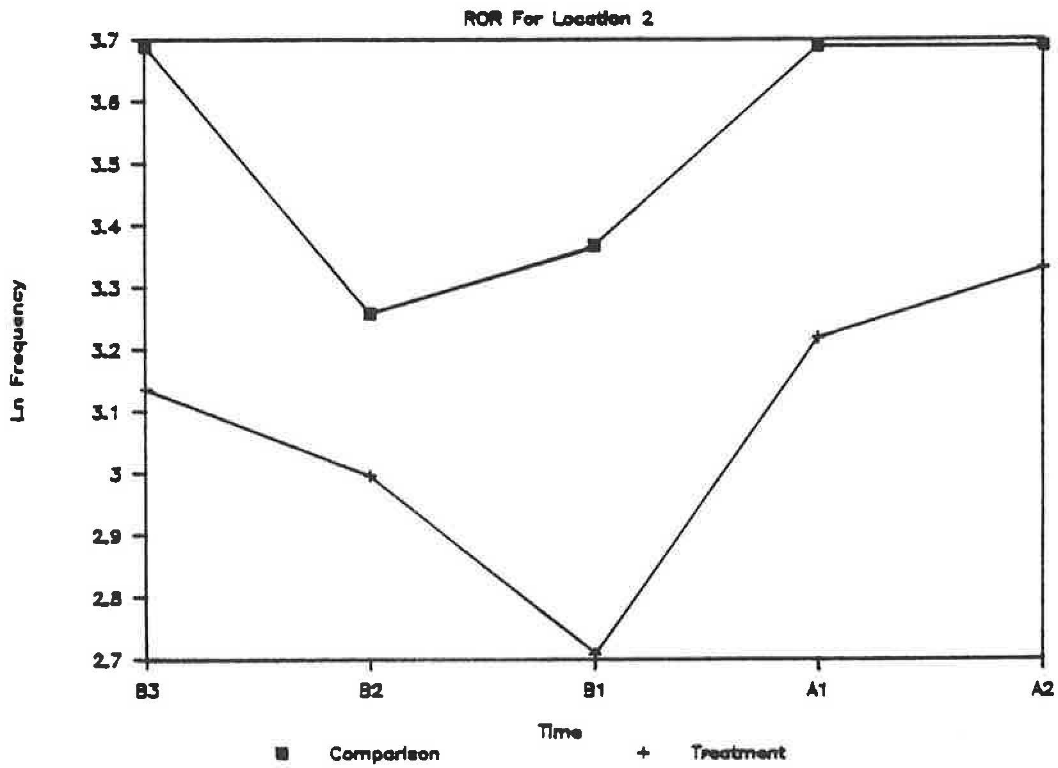

FIGURE 2 Ln frequency graph.

An alternative for calculating G^2 Treatment uses the same 2×2 table as in step 2, in which the treatment and comparison groups are collapsed across the before and after periods. The following equations are used:

	Comparison	*Treatment*
B3–B1	X_{11}	X_{12}
A1–A2	X_{21}	X_{22}

$$\tau = \frac{X_{11} X_{22}}{X_{12} X_{21}} \qquad (2)$$

where τ is the cross-products ratio: $(\tau - 1) \times 100$ equals the apparent percentage change in accidents attributable to the treatment. Equation 3 is used to determine if the apparent treatment effect is significant:

$$Z = \frac{\ln \tau}{(1/X_{11} + 1/X_{12} + 1/X_{21} + 1/X_{22})^{1/2}} \qquad (3)$$

For $\alpha = 0.05$ and a two-tailed test, the confidence interval lies between −1.96 and 1.96.

The advantage to using this alternative is that the apparent change in accidents attributable to the treatment is obtained. Both methods of calculating G^2 Treatment were used in the analysis.

A limitation should be noted in using this study design. To avoid dividing by 0, which results in an undefined G^2 value, each cell in the 2×5 contingency table must be greater than 0. Note that frequencies are used in contingency tables instead of rates. Moreover, exposure was a factor in the selection of the comparison groups.

Combining Treatment Sections

So that the effects of all three treatment sections can be examined together, the logarithms of the odds ratios are combined by using a technique commonly called Gart's procedure (*5–7*). Gart's procedure combines 2×2 contingency tables with the natural logarithm of the odds (or the maximum likelihood) ratio as the measure of association. The log odds ratio for each location is weighted on the basis of the accident frequency. Figure 3 displays the worksheet used for the procedure, along with the equations used. The chi-square statistic for testing the homogeneity of the odds ratio, X^2 homogeneity with 2 degrees of freedom, indicates the existence of insignificant differences among the three odd ratios. An acceptable X^2 homogeneity indicates no significant difference. The chi-square statistic for testing the significance of the mean log odds ratio, X^2 association with 1 degree of freedom, indicates the existence of insignificant differences between the comparison and treatment groups. The chi-square total is equal to the sum of X^2 homogeneity and X^2 association.

There are benefits to combining the three locations. By increasing the amount of data available for testing, the statistical power is increased. In other words, combining the locations improves the opportunity to identify a treatment effect if one is present.

Treatment Locations

Three sections of roadway—17.2-mi, 19.1-mi, and 24.4-mi long—served as the treatment locations. Wide (8-in.) edgelines were painted at these sites during spring and summer 1984. The wide edgelines were repainted approximately one year later. The actual edgeline width varied from 7.0 to 10.0 in. The study

Accident Type:

Study Number	Comparison Before a	Comparison After b	Treatment Before c	Treatment After d	Estimated Weights W_i	Relative Risk L_i	Product $W_i L_i$	$W_i (L_i)^2$
1								
2								
3								
Total					ΣW_i		$\Sigma W_i L_i$	$\Sigma W_i (L_i)^2$

$W_i = \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)^{-1}$ - estimated weight

$L_i = \ln \left(\frac{a}{b} \cdot \frac{d}{c}\right)$ - natural logarithm of the odds ratio

$\chi^2_{total} = \Sigma W_i (L_i)^2$

$\chi^2_{homogeneity} = \Sigma W_i (L_i)^2 - \frac{(\Sigma W_i L_i)^2}{\Sigma W_i}$

$\chi^2_{association} = \chi^2_{total} - \chi^2_{homogeneity}$

Critical χ^2

df	α=0.05
2	5.99
1	3.84

FIGURE 3 Worksheet for combining the three locations.

sections were in four districts, so four different paint crews were used. On the basis of 12 sample site studies in the interim report for lateral placement and speed changes, the average edgeline width for each treatment location was:

- 17.2-mi section: 7.6 in.
- 19.1-mi section: 7.4 in.
- 24.4-mi section: 9.3 in.

Comparison Locations

Several measures were used as a guide in selecting locations for comparison with the three treatment locations. The primary objective was to identify locations that were similar to the treatment locations for the following characteristics: two-lane rural roads, overall roadway geometrics, average daily traffic, total accident frequencies, run-off-the-road accident frequencies, and alcohol- and drug-related accident frequencies. Also, no changes that would influence the frequency of accidents were planned for the road sections.

The key to the appropriateness of a comparison location is the check for comparability. If the results of the check were that the treatment location was not comparable with the comparison location, the alternative comparison location would be the treatment location with all other accident types. The use of all other accident types on the treatment location as a comparison location is generally acceptable. The alternative comparison locations eliminate extraneous factors such as exposure, roadway geometrics, alignment, and weather because the alternative comparison and treatment road sections are the same. Information on the treatment and comparison locations is presented in Table 1.

Measures of Effectiveness

This evaluation focuses on the effectiveness of the wide edgelines in reducing accidents, especially ROR, DUI, and other related types of accidents. ROR accidents were the primary type of accident evaluated. Also, ROR accidents involving four other factors in addition to DUI were selected for a detailed analysis. ROR accidents at curves were considered because horizontal alignment is a factor in ROR accidents. Because edgelines are important in delineating the roadway during darkness, ROR accidents during darkness were selected as a measure. ROR accidents in inclement weather were selected as a measure because inclement weather is an extraneous factor that

TABLE 1 DESCRIPTION OF THE TREATMENT AND COMPARISON LOCATIONS

Location	1985 ADT (vehicles)	1985 Daily Distance Traveled (vehicle-mi)	Roadway Width (ft)	Roadway Description
1 T: Route 20, Buckingham	2,275	43,340	20.0	Straight road
C: Route 40, Pittsylvania and Halifax	2,180	54,085	20.0	Straight road
2 T: Route 20, Albemarle	3,685	66,090	20.5	Winding road
C: Route 8, Floyd and Montgomery	3,670	71,385	22.0	Winding road
3 T: Route 501, Bedford and Rockbridge	2,580	48,710	19.4	Winding and straight roadway
C: Route 20, Albemarle and Orange	2,700	40,095	20.1	Winding and straight roadway

NOTE: T: = treatment location; C: = comparison location. The alternative comparison location was the treatment location with all other accidents.

may contribute to ROR accidents. Because of concern about drivers encroaching on the centerline because of wide edgelines, opposite-direction accidents were evaluated. In all, six measures of effectiveness were used.

Data

Accident data were obtained from the Virginia Department of Transportation's computerized traffic accident-reporting system. Three years of before-data with 4-in.-wide edgelines and 2 years of after-data with 8-in.-wide edgelines were used. The accident data were based on accident reports completed by the state or local police officer who responded to the accident. The presence of a curve, darkness, or inclement weather was determined by the police officer. Similarly, DUI was noted as a contributing factor on the basis of tests administered by the police officer or when DUI was suspected because of the situation, evidence, or testimony of witnesses.

ACCIDENT DATA ANALYSIS

The analysis results for each measure of performance will be described for each treatment section and for all sections combined. Although two levels of significance, 0.05 and 0.10, were examined, only 0.05 is displayed in the analysis tables. Unless otherwise stated, the conclusions on the effect of the treatment are the same for both levels.

Run-off-the-Road Accidents

The analysis data for ROR accidents are presented in Table 2. In the check for comparability, treatment Location 1 was not comparable to its original comparison location. Therefore the alternative comparison location of all non-ROR accidents on the treatment location was used and found to be comparable for all treatment locations. On the basis of $\alpha = 0.05$, there was no evidence that the wide edgelines significantly affected the incidence of ROR accidents for any of the three treatment locations individually or combined. However, for a level of significance of 0.10, Location 1 shows a significant decrease in ROR accidents. The apparent percentage reduction is 55 percent. The low accident frequency in the A1 period probably accounts for the significant decrease in accidents because the A2 period accident frequency is the highest of the 5-year period.

ROR Accidents Involving DUI

The analysis data are presented in Table 3. Because there were 0 values in the original comparison location for Treatment Location 1, the alternative comparison location of all other accidents on the treatment locations was used and found to be comparable for all treatment locations at $\alpha = 0.05$. There was no evidence that the wide edgelines significantly affected the incidence of accidents involving both ROR and DUI on all treatment locations. On the basis of X^2 homogeneity and X^2 association, the combined locations are acceptable and there is no indication of a significant effect.

TABLE 2 ROR ACCIDENTS

(a) Analysis for Each Location[a]

	Accident Frequency					
	Location 1		Location 2		Location 3	
Year	C	T	C	T	C	T
B1	1	4	40	23	20	21
B2	10	10	26	20	18	16
B3	5	10	29	15	23	16
A1	13	6	40	25	14	12
A2	12	11	40	28	25	20

	G^2 Values for Locations				
Source	1	2	3	*df*	Critical X^2 $\alpha = 0.05$
Comparability					
Before	2.02	0.93	0.84	2	5.99
After	1.15	0.10	0.02	1	3.84
Treatment	3.14	0.11	0.04	1	3.84
Total	6.31	1.14	0.90	4	
Apparent Change (%)	−55	9	−6		

(b) Combining Locations With Gart's Procedure[b]

Source	X^2	*df*	Critical X^2 $\alpha = 0.05$
Homogeneity	2.91	2	5.99
Association	0.31	1	3.84
Total	3.22	3	

NOTE: C = Comparison, T = Treatment. The comparison group is all other accidents at the treatment location. Apparent change values are given only if locations are comparable.

[a]Conclusion for each location: Comparability—Acceptable for each; Treatment—No significant effect for each.

[b]Conclusion for combined locations: Homogeneity—Acceptable; Association—No significant effect.

ROR Accidents on Curves

All three treatment locations were comparable with the alternative comparison locations. There was no evidence that the wide edgelines significantly affected the incidence of ROR accidents on curves for the treatment locations (Table 4). When the locations are combined, the X^2 homogeneity was acceptable, and X^2 association indicated no significant effect.

ROR Accidents During Darkness

The analysis data are presented in Table 5. All three pairs of treatment and original comparison locations were comparable. Furthermore, for all three locations, there was no evidence to suggest that the wide edgelines significantly affected the incidence of ROR accidents during darkness. The apparent percentage increase for Location 2 of 122 percent is high but ineffective statistically.

X^2 homogeneity is acceptable, and X^2 association indicates that there is no significant effect for the combined locations. When the alternative comparison groups are used, the treatment and alternative comparison groups for Locations 1 and 2 are comparable. There are no significant effects for the two sites individually nor combined.

TABLE 3 ROR ACCIDENTS INVOLVING DUI

(a) Analysis for Each Location[a]

	Accident Frequency					
	Location 1		Location 2		Location 3	
Year	C	T	C	T	C	T
B1	2	3	60	3	34	7
B2	18	2	36	10	30	4
B3	11	4	43	1	38	1
A1	16	3	60	5	22	4
A2	19	4	64	4	42	3

	G^2 Values for Locations				Critical X^2
Source	1	2	3	*df*	α = 0.05
Comparability					
Before	5.52	11.80	5.31	2	5.99
After	0.02	0.17	1.35	1	3.84
Treatment	0.44	0.55	0.02	1	3.84
Total	5.99	12.52	6.69	4	
Apparent Change (%)	−31	–	−7		

(b) Combining Locations With Gart's Procedure[b]

Source	X^2	*df*	Critical X^2 α = 0.05
Homogeneity	0.25	2	5.99
Association	0.20	1	3.84
Total	0.45	3	

NOTE: C = Comparison, T = Treatment. The comparison group is all other accidents at the treatment location. Apparent change values are given only if locations are comparable.
[a]Conclusion for each location: Comparability—Acceptable for each; Treatment—No significant effect for each.
[b]Conclusion for combined locations: Homogeneity—Acceptable; Association—No significant effect.

TABLE 4 ROR ACCIDENTS ON CURVES

(a) Analysis for Each Location[a]

	Accident Frequency					
	Location 1		Location 2		Location 3	
Year	C	T	C	T	C	T
B1	4	1	46	17	25	16
B2	17	3	33	13	23	11
B3	11	4	36	8	29	10
A1	15	4	50	15	19	7
A2	19	4	49	19	29	16

	G^2 Values for Locations				Critical X^2
Source	1	2	3	*df*	α = 0.05
Comparability					
Before	0.72	1.55	1.64	2	5.99
After	0.09	0.41	0.57	1	3.84
Treatment	0.01	0.02	0.00	1	3.84
Total	0.82	1.99	2.21	4	
Apparent Change (%)	−6	4	0		

(b) Combining Locations With Gart's Procedure[b]

Source	X^2	*df*	Critical X^2 α = 0.05
Homogeneity	0.02	2	5.99
Association	0.01	1	3.84
Total	0.03	3	

NOTE: C = Comparison, T = Treatment. The comparison group is all other accidents at the treatment location. Apparent change values are given only if locations are comparable.
[a]Conclusion for each location: Comparability—Acceptable for each; Treatment—No significant effect for each.
[b]Conclusion for combined locations: Homogeneity—Acceptable; Association—No significant effect.

ROR and Weather

Because there were 0 values for each treatment and comparison location in the contingency table, it was not possible to analyze ROR and weather. The low frequency of ROR accidents in inclement weather demonstrates that weather is not a substantial influence in ROR accidents. Consequently, it was concluded that there was an insufficient number of ROR accidents in inclement weather to determine a statistical effect.

Opposite Direction

The analysis data are presented in Table 6. Because Treatment Location 1 had three 0 values in the contingency table, it was not possible to analyze this location. Because the original comparison location for Treatment Location 2 had a 0 in the table, the alternate comparison location of all nonopposite-direction accidents was used. Treatment Locations 2 and 3 were comparable with their alternative comparison locations. There was no evidence that wide edgelines affected the incidence of opposite-direction accidents. Similarly, the X^2 homogeneity and X^2 association were acceptable and showed no evidence of a significant effect.

Results from a Before-After Design With a Comparison Group

As noted previously, the before-after design with a comparison group and check for comparability has more statistical power and is more statistically valid then the traditional before-after design with a comparison group. These differences can be illustrated by reviewing the results of this study against the more familiar before-after design with a comparison group. The evaluation procedure used accident rates in accidents per million vehicle-mi and the Poisson distribution for testing (*8*). The B2 and B3 years were the before period. The results, presented in Table 7, are mixed, inconsistent, and inconclusive. Again, two advantages of the before-after design with a comparison group and check for comparability are that the comparability of the comparison group is tested and the test locations can be combined and evaluated.

Summary

On the basis of the analysis of the six measures of effectiveness, there is no evidence to indicate that wide edgelines significantly affected the incidence of ROR accidents and related accident types. This is also true when the level of significance was increased to 0.10 for a lower level of confidence.

TABLE 5 ROR ACCIDENTS DURING DARKNESS

(a) Analysis for Each Location[a]

	Accident Frequency					
	Location 1		Location 2		Location 3	
Year	C	T	C	T	C	T
B1	2	2	6	8	5	15
B2	1	5	4	12	9	8
B3	1	5	5	8	9	5
A1	3	5	4	13	13	8
A2	2	7	3	16	10	13

Source	G^2 Values for Locations 1	2	3	*df*	Critical X^2 α = 0.05
Comparability					
Before	1.64	1.18	5.96	2	5.99
After	0.48	0.34	1.50	1	3.84
Treatment	0.08	2.37	0.49	1	3.84
Total	2.19	3.90	7.95	4	
Apparent Change (%)	−20	122	−25		

(b) Combining Locations With Gart's Procedure[b]

Source	X^2	*df*	Critical X^2 α = 0.05
Homogeneity	2.28	2	5.99
Association	0.58	1	3.84
Total	2.86	3	

NOTE: C = Comparison, T = Treatment. The comparison group is all other accidents at the treatment location. Apparent change values are given only if locations are comparable.
[a]Conclusion for each location: Comparability—Acceptable for each; Treatment—No significant effect for each.
[b]Conclusion for combined locations: Homogeneity—Acceptable; Association—No significant effect.

TABLE 6 OPPOSITE DIRECTION ACCIDENTS

(a) Analysis for Each Location[a]

	Accident Frequency					
	Location 1		Location 2		Location 3	
Year	C	T	C	T	C	T
B1	5	0	56	7	36	5
B2	20	0	41	5	29	5
B3	15	0	37	7	34	5
A1	16	3	57	8	20	6
A2	22	1	65	3	41	4

Source	G^2 Values for Locations 1	2	3	*df*	Critical X^2 α = 0.05
Comparability					
Before	–	0.67	0.11	2	5.99
After	–	2.82	2.64	1	3.84
Treatment	–	1.32	0.03	1	3.84
Total	–	4.80	2.77	4	
Apparent Change (%)	–	−36	8		

(b) Combining Locations With Gart's Procedure[b]

Source	X^2	*df*	Critical X^2 α = 0.05
Homogeneity	0.79	2	5.99
Association	0.51	1	3.84
Total	1.30	3	

NOTE: C = Comparison, T = Treatment. The comparison group is all other accidents at the treatment location. Apparent change values are given only if locations are comparable.
[a]Conclusion for each location: Comparability—Acceptable for each; Treatment—No significant effect for each.
[b]Conclusion for combined locations: Homogeneity—Acceptable; Association—No significant effect.

TABLE 7 STUDY RESULTS FROM THE BEFORE-AFTER DESIGN WITH A COMPARISON GROUP AND POISSON TEST

Accident Type	Location 1, Route 20, Buckingham County	Location 2, Route 20, Albemarle County	Location 3, Route 501, Bedford and Rockbridge Counties
ROR	48.9% decrease at 90% CL	84.4% increase at 99% CL	23.3% decrease at 90% CL
ROR and DUI	Not performed because of division by 0	Not significant	66.7% increase at 95% CL
ROR and curve	66.7% increase at 95% CL	21.5% decrease at 90% CL	Not significant
ROR and darkness	52.7% decrease at 99% CL	90.8% increase at 99% CL	51.1% increase at 95% CL
Opposite direction	Not performed because of division by 0	53.0% increase at 99% CL	99.6% increase at 99% CL

NOTE: CL = confidence level.

Moreover, these findings concur with the results of an evaluation of wide edgelines in New Mexico where, by using a before-after design with a comparison group, 100 mi of wide (8-in.) edgelines were compared with 353 mi of a comparison group with the standard 4-inch edgelines (*9*).

CONCLUSION

The before-and-after design with a comparison group and a check for comparability was used, along with Gart's procedure (for combining the accident data from the three study locations and the respective comparison groups), to analyze the data. There was no evidence that wide edgelines significantly affected the ROR accident frequency or the frequency of related accident types for each study location nor for the combined locations at 0.05 level of confidence. The accident types included ROR accidents, ROR involving DUI, ROR on curves, ROR during darkness, ROR and weather, and opposite direction. The findings are based on 5 years of accident data, 3 years before wide edgeline installation and 2 years after installation.

ACKNOWLEDGMENTS

The assistance of Lindsay I. Griffin III in the application of the experimental plan is acknowledged and very much appreciated. Griffin's work described the before-after design with a comparison group and a check for comparability. The research reported here was financed with Highway Planning and Research funds administered by the Federal Highway Administration.

REFERENCES

1. *Virginia Traffic Crash Facts.* Virginia Department of Motor Vehicles, Richmond, 1985.
2. *Manual on Uniform Traffic Control Devices for Streets and Highways.* FHWA, U.S. Department of Transportation, 1978.
3. B. H. Cottrell, Jr. *Evaluation of Wide Edgelines on Two-Lane Rural Roads.* Interim Report. Virginia Highway and Transportation Research Council, Charlottesville, 1985.
4. L. I. Griffin III. Three Procedures for Evaluating Highway Safety Improvement Programs. Presented at the American Society of Civil Engineers Annual Convention, New Orleans, La., 1982.
5. J. L. Fleiss. *Statistical Methods for Rates and Proportions,* 2nd ed. John Wiley and Sons, New York, 1981.
6. J. J. Gart. On the Combination of Relative Risks. *Biometrics,* Vol. 18, 1962, pp. 601–610.
7. P. R. Sheehe. Combination of Log Relative Risks in Retrospective Studies of Disease. *American Journal of Public Health,* Vol. 56, 1966, pp. 1745–1750.
8. *Evaluation of Highway Safety Projects: Procedural Guide.* FHWA, U.S. Department of Transportation, 1979.
9. J. W. Hall. *Evaluation of Wide Edgelines.* University of New Mexico, Bureau of Engineering Research, Albuquerque, 1986.

DISCUSSION

JOHN A. DEACON
University of Kentucky, Lexington, Ky. 40506-0046.

Because wide edgelines have been found to be cost-effective alternatives to standard edgelines if they further reduce crashes by no more than about 1 percent (*1*), considerable effort is under way in the United States to assess their possible safety effect. The Virginia study is one of the first of these current initiatives to be reported. The purpose of this discussion is to offer additional perspective on this study—and others like it that seek to document small safety benefits by using crash data—and to urge caution in the interpretation of its findings.

In analyzing crash data, the highway safety researcher is first interested in learning the effect of the treatment on crash frequency or crash rate. The two most significant types of crashes that are likely to be influenced by wide edgelines are run-off-road (ROR) and opposite-direction (OD) crashes. The hypothesis is that ROR crashes might be reduced as a result of the enhanced conspicuity of the wider edgelines but that OD crashes might be increased as drivers steer toward more central positions on the roadway. In Table 8, the frequencies of these combined crash types as observed in Virginia are tabulated, and by using a procedure common to before-after studies with control sections, the likely reduction in crashes in the 2-year after period are estimated. When the three treatment sections were considered together, there were about 17 fewer ROR and OD crashes in the 2-year period following application of 8-in. edgelines than would otherwise be expected. This translates to a 7 percent reduction on the basis of the total number of crashes observed in the after period and a 13.6 percent reduction on the basis of the observed number of ROR and OD crashes. Although these benefits appear small in magnitude, they greatly exceed the levels necessary for cost-effective application (*1*) and hence may be of considerable practical significance.

Unfortunately for safety researchers, crash frequency is a highly variable quantity, and simple analyses such as those just mentioned must be supplemented by more sophisticated techniques in an attempt to assure that the observed effect is not simply due to chance occurrence. These extended analyses attempt to minimize the risk of erroneous conclusion. One error that the safety researcher wants to avoid is the conclusion that a treatment is effective in reducing crashes when, in fact, it is not. This is termed an error of the first kind (or Type I error), and the probability of committing this error is called the level of significance. Statistical testing procedures can be designed to keep the risk of committing a Type I error to a small level. The Virginia study tested two levels of significance, 0.05 and 0.10, which are indicative of the range commonly employed by highway researchers. For neither of these two levels of significance was the crash effect of the wide edgelines in Virginia found to be statistically significant. While this certainly might mean that the differential effect of wide edgelines was nil, it also might mean that the sample size was such that large variability in the crash data was allowed to mask a small treatment effect.

In any event, the safety researcher also wants to minimize risk of committing another kind of error, a Type II error or error of the second kind. A Type II error results when a treatment is concluded to be ineffective when, in fact, it is effective. A large risk of committing a Type II error is expected when the treated

TABLE 8 SUMMARY OF EXPECTED EFFECTS OF 8-IN. EDGELINES ON RUN-OFF-ROAD AND OPPOSITE-DIRECTION CRASHES

	Crash Frequency					
	Observed				Expected After Period	
	Before Period		After Period			
Location	All Other[a]	ROR + OD	All Other[a]	ROR + OD	ROR + OD	Reduction
1	16	24	21	21	31.50	10.50
2	76	77	69	64	69.91	5.91
3	46	68	29	42	42.87	0.87

NOTE: Total crashes observed in after period: 246
Total ROR + OD crashes observed in after period: 127
Total reduction in ROR + OD crashes expected: 17.28
Reduction as a percentage of ROR + OD crashes: 13.61
Reduction as a percentage of crashes of all types: 7.02

[a] At each location in the Virginia study, crash frequencies used for control were the frequencies of all other crash types at that same location. In this table, all crash types except ROR and OD are included in this category.

mileage and safety effect are small and when the variability in crash frequency is large. Unfortunately, although it appears that such conditions characterized the Virginia experiment, no assessment was made of the Type II risk, and apparently no attempt was made to maintain the risk at an acceptable level by selecting a sample of adequate size.

In an attempt to illustrate the sample size problem, a hypothetical analysis was undertaken of a before-after crash study with matched or paired treatment and control sites. This is illustrative of the kind of study conducted in Virginia, except that testing for comparability of the matched sites was unnecessary. In the absence of real data, the following assumptions were made:

- Extent of crash data: Two years before and two years after treatment,
- Length of each site: Five miles,
- Traffic volume: 2,000 vehicles per day,
- Crash rate: Five crashes per million vehicle miles,
- Mean crash frequency at each site: 36.5 crashes in 2 years (from above),
- Variance in crashes: 25 percent of the mean frequency or 9.1 crashes in 2 years,
- Correlation coefficients: 0.50 between crashes before and after time of treatment and 0.25 between before-after crash decrement at treatment site and that at the matched control site, and
- Size of treatment effect to be detected: 1 percent of the untreated mean crash frequency to reflect the approximate size necessary for cost effective treatment.

For the above experiment, the number (n) of site pairs required in the sample to maintain acceptable levels of risk is (2):

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{d^2} \quad (4)$$

in which z is the normal variate, α is the level of significance, β is the probability of committing a Type II error, and d is given as follows:

$$d = \frac{(m_T - m_C)}{\sigma} \quad (5)$$

in which $(m_T - m_C)$ is the value of the average difference in crash frequency that is to be detected and σ is the standard deviation of the relative change in crash frequency of a site pair resulting from treatment. Given the above assumptions, the value of d can be shown to be 0.09904 and

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{0.009809} \quad (6)$$

The typical crash investigation of a feature such as wide edgelines might involve 10 to 30 test pairs (100 to 300 mi of roadway in the context of the example given here). The level of significance is commonly 0.05. Given the above assumptions, and by using Equation 6, the probability of not detecting a real 1 percent crash effect (i.e., the probability of committing a Type II error) is very large, of the order of 0.87 to 0.91. If the level of significance is relaxed to 0.10 and if β is set to 0.40 (still a large risk), the required number of test pairs is about 240, corresponding to about 2,400 mi of highway. Seldom is crash data of the type required for such an analysis available for more than 2,000 mi of roadway, and even when it is, the risk of error remains large if the treatment effect is small.

Although it is important to avoid direct comparisons between the Virginia experiment and this hypothetical example, the example does illustrate that large mileages may be required for crash studies in which observation of small treatment effects is important. It further illustrates that interpretation of the Virginia findings is incomplete without an assessment of the risks of committing a Type II error. In summary, although the Virginia study has produced useful new data, it has not conclusively established the safety effects of wide edgelines. Because wide edgelines offer promise as a cost-effective accident countermeasure, a great deal could be lost if continued experimentation is prematurely abandoned.

REFERENCES

1. W. E. Hughes, et al. A Cost-Effectiveness Evaluation of Alternative Edgeline Widths for Two-Lane Rural Roads. Report FHWA/RD-S3, Bellomo-McGee Inc., Vienna, Va., Aug. 1983.
2. M. G. Natrella. *Experimental Statistics*. Handbook 91, National Bureau of Standards, Washington, D.C., Aug. 1963.

DISCUSSION

Olga J. Pendleton
Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843.

Although this paper is commendable in both the application and use of the statistical before-after methodology, the description of the methodology, including mathematical notations, is taken directly from Griffin's work. Although Griffin is acknowledged, the author should ensure that Griffin receives adequate recognition for his work, which is literally duplicated in parts of this paper. As an alternative, the author might have wished to eliminate the portions of text that come from Griffin's article and simply refer the reader to that article for detail.

Given that the results of the traditional before-after design using the less powerful Poisson test were, in the author's words, "mixed, inconsistent, and inconclusive," Table 7 should have been omitted. Inclusion of these results will naturally lead to misuse and misinterpretation.

The conclusion of this paper is too strong and is not supported by the sparse and limited data presented. All that can be said is that at these locations, no significant safety effect due to edgelining could be found. To make inference to all edgelined sections on the basis of these three nonrandomly selected locations is not valid.

Finally, the author has made the all-too-common error of interpreting partitioned chi-squares when the overall chi-square is not significant. This is equivalent to the analysis of variance analog of interpreting model parameters when the overall F-statistic is not significant. Specifically, for Locations 2 and 3 of Table 2, since the overall (total) chi-square is not significant, it is incorrect to attach any meaning to the chi-square statistics for homogeneity or treatment.

AUTHOR'S CLOSURE

I would like to thank John A. Deacon and Olga J. Pendleton for commenting on this paper. Their comments and interest in my work are greatly appreciated. The paper has been improved by their contributions.

Deacon's point is that the results of the paper should be viewed with caution because the sample size was too small to detect a small treatment effect, such as a 1 percent decrease in accident frequency. Before this research was initiated, I carefully reviewed the reference cited by Deacon. This work describes several scenarios of a sample size range of values for α, β, and the size of the treatment effect to be detected. From this, it was apparent that 2,000 mi or more of roadway was necessary to obtain "acceptable" values of α, β, and the size of the treatment effect to be detected. Limited resources did not permit me to pursue such a large-scale effort.

It is noted that the value of d in Deacon's hypothetical analysis should be 0.12, yielding $\beta = 0.94$ for $n = 10$ to 30. A 1 percent treatment effect is small enough to be quite difficult to detect. If study data for ROR accidents are used, for $\alpha = 0.10$, $n = 3$, and size of treatment effect to be detected = 10 percent; β is equal to 0.70. Admittedly, this β value is high. Note that even for $n = 240$ with $\alpha = 0.05$, treatment effect to be detected = 0.01 and $\beta = 0.40$, as stated by Deacon, "the risk of error remains large if the treatment effect is small." Because of the sample size, no inferences are made.

Given the limitation on the sample size, a special effort was made to improve the statistical method of the research by using (a) a statistically rigorous and valid study design and (b) Gart's procedure to combine the three study locations, thereby increasing the power of the test. Moreover, study locations with fairly high ROR accident frequencies were selected to permit the study of a higher number of ROR accidents in lieu of additional miles of roadway.

In this research, the study locations were combined by using Gart's procedure, including weighting the locations on the basis of accident frequency. Higher weight is given to locations with higher accident frequency. When weighting is used to combine the data for the three study locations in Table 8, the following values are obtained:

	Recommended	*Deacon's*
Total reduction in ROR + OD crashes expected	5.02	17.28
Reduction as a percentage of ROR + OD crashes	3.95	13.61
Reduction as a percentage of all types	2.04	7.02

These values are substantially lower than those in Table 8 because the location with the largest reduction had the lowest accident frequency. The recommended values are more appropriate, better reflect the actual changes, and are more reliable.

Three changes were made in the paper in response to Pendleton's discussion. I have emphasized the contributions of Lindsay I. Griffin III through additional discussion in the paper and the acknowledgments. Griffin's review of the paper and his contributions to this closure are also greatly appreciated. The methodology was discussed because (a) most readers are probably unfamiliar with the methodology and (b) the methodology has not been cited in journals or reports widely distributed to transportation professionals.

The discussion on Table 7 was revised to emphasize that by including Table 7, the advantages of the before-after design with a comparison group and check for comparability are demonstrated.

The conclusion in the preprint was incorrectly stated because it was inconsistent with analysis findings and the limited sample size. The revised conclusion is that there was no evidence that wide edgelines significantly affected the ROR accident frequency or related accident types for each study location or for the combined locations at a 0.05 level of significance.

The opinions, findings, and conclusions expressed in this paper are those of the authors and not necessarily those of the sponsoring agencies.

Publication of this paper sponsored by Committee on Traffic Control Devices.