USING BEFORE-AND-AFTER DATA TO ESTIMATE THE EFFECTIVENESS OF ACCIDENT COUNTERMEASURES IMPLEMENTED AT SEVERAL TREATMENT SITES
by

Lindsay I. Griffin, III

December 1989

Safety Division
Texas Transportation Institute
The Texas A\&M University System
College Station, TX 77843

## Introduction

Each year in the United States, thousands of safety projects are implemented throughout the nation's highway system in the hope of reducing traffic accidents. Although most of these projects are never evaluated to determine if (or to what degree) they are reducing accidents, others are evaluated. For those projects that are evaluated, the evaluation design most often selected to assess project effectiveness is the before-and-after design.

The before-and-after design is a popular method for evaluating safety projects for several reasons. First of all, this design requires the collecting of relatively few data - just counts of accidents before and after treatment. Secondly, the statistical tests employed with this design [e.g. tests based upon the Poisson, chi-square (Michaels, 1966), or standard normal distribution (Gerlough and Barnes, 1971)] are fairly simple to use. Third, and perhaps most importantly, this design has an intuitive appeal - if accidents decrease by 30 percent following treatment, that 30 percent reduction in accidents is assumed to have resulted from the treatment imposed. ${ }^{1}$

When the before-and-after design is used to estimate the overall effect of treatment at several different sites, some evaluators choose to compare "the sum of the before accidents" to "the sum of the after accidents" at the treatment sites and to carry out their analysis "as if" they were assessing one large project. This practice is statistically suspect - and a squandering of information.

In this article an alternative method of using before-and-after data to estimate the overall (average) accident reduction effectiveness of remedial projects implemented at several different treatment sites will be demonstrated. The data used in this demonstration are taken from the state of New York [Larsen (1986)] and are reproduced in Table 1 (along with several columns of statistics that will be used shortly).

Each of the 20 remedial projects in Table 1 was implemented (with geometric

[^0]design exceptions) at a different site as part of the state's non-freeway, 3 R program. For each of these 20 projects, six years of accident history were available: three years before and three years after treatment. ${ }^{2}$

Table 1: Three Years of "Before" and Three Years of "After" Accident Data (and Assorted Statistics) for Twenty, Non-Freeway, 3R Projects in New York

| Site | Accidents |  | L | $L_{\text {se- }}$ | W | WL | $\dot{w}^{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Before | After |  |  |  |  |  |
| 1 | 10 | 14 | 0.336 | 0.414 | 5.833 | 1.960 | 0.659 |
| 2 | 20 | 11 | -0.598 | 0.375 | 7.097 | -4.244 | 2.538 |
| 3 | 15 | 7 | -0.762 | 0.458 | 4.773 | -3.637 | 2.771 |
| 4 | 28 | 29 | 0.035 | 0.265 | 14.246 | 0.499 | 0.017 |
| 5 | 40 | 18 | -0.799 | 0.284 | 12.414 | -9.919 | 7.925 |
| 6 | 5 | 4 | -0.223 | 0.671 | 2.222 | -0.496 | 0.111 |
| 7 | 40 | 36 | -0.105 | 0.230 | 18.947 | -1.989 | 0.209 |
| 8 | 10 | 28 | 1.030 | 0.368 | 7.368 | 7.589 | 7.817 |
| 9 | 57 | 45 | -0.236 | 0.199 | 25.147 | -5.935 | 1.401 |
| 10 | 82 | 51 | -0.475 | 0.178 | 31.444 | -14.936 | 7.095 |
| 11 | 2 | 8 | 1.386 | 0.791 | 1.600 | 2.218 | 3.074 |
| 12 | 26 | 29 | 0.109 | 0.270 | 13.709 | 1.494 | 0.163 |
| 13 | 16 | 3 | -1.674 | 0.629 | 2.526 | -4.229 | 7.079 |
| 14 | 14 | 17 | 0.194 | 0.361 | 7.677 | 1.489 | 0.289 |
| 15 | 29 | 17 | -0.534 | 0.305 | 10.717 | -5.723 | 3.056 |
| 16 | 27 | 17 | -0.463 | 0.310 | 10.432 | -4.830 | 2.236 |
| 17 | 17 | 21 | 0.211 | 0.326 | 9.395 | 1.982 | 0.418 |
| 18 | 18 | 16 | -0.118 | 0.344 | 8.471 | -1.000 | 0.118 |
| 19 | 33 | 24 | -0.318 | 0.268 | 13.895 | -4.419 | 1.405 |
| 20 | 5 | 5 | 0.000 | 0.632 | 2.500 | 0.000 | 0.000 |
|  | 494 | 400 |  |  | 210.413 | -44.126 | 48.381 |

Larsen estimates that accidents were reduced by 19 percent at the 20 treatment sites shown in Table 1.

$$
\begin{aligned}
E & =[(0 b-\text { ex }) \div \text { ex] } 100 \\
& =[(400-494) \div 494] 100 \\
& =-19.0
\end{aligned}
$$

${ }^{2}$ One treatment site shown in Larsen (1986) sustained no accidents during the before or after period. That site was omitted from Table 1.

Where

$$
\begin{aligned}
E= & \text { percent change in accidents, with }(-) \text { indicating } \\
& \text { reductions and }(+) \text { indicating increases, } \\
o b= & \text { observed accidents in the after period, and } \\
\text { ex }= & \frac{\text { expected accidents in the after period, as reflected }}{\text { in the number of accidents in the before period }}
\end{aligned}
$$

Larsen's overall estimate of a 19 percent reduction in accidents is, in actuality, the weighted average of the accident reductions at each of 20 sites $\left(E_{i}\right)$, with weights defined in proportion to the number of before accidents at each site ( $b_{i}$ ).

$$
\begin{align*}
E & =\left[\Sigma b_{i} E_{i} \div \Sigma b_{i}\right] 100  \tag{2}\\
& =[(o b-e x) \div e x] 100
\end{align*}
$$

Where, $\quad E_{i}=$ percent change in accidents at the $i^{\text {th }}$ site, and $b_{i}=$ before accidents at the $i^{\text {th }}$ site

## An Alternative Method of Analysis

In this section the data in Table 1 will be reanalyzed using a different method than the one employed by Larsen. In the first part of this section, a procedure for evaluating the effectiveness of a project (e.g., lane widening, shoulder widening, etc.) implemented at one site is explained. In the second part of this section, the procedure is extended to permit evaluation of a group of projects, i.e., projects implemented at several sites.

## Estimating the Effect of <br> Treatment at One Site

Let us begin our alternative analysis of Larsen's data by defining a test statistic for each of the 20 sites in Table l. For a given site, let $U$ equal the ratio of accidents in the after period (A) to accidents in the before period (B). Taking the second site in Table 1 as an example, we calculate:

$$
\begin{align*}
U & =A \div B  \tag{3}\\
& =11 / 20 \\
& =0.550
\end{align*}
$$

And, the percent change in accidents at this site (following treatment) may be expressed as a function of $U$ :

$$
\begin{align*}
E & =(U-1) 100  \tag{4}\\
& =(0.550-1) 100 \\
& =-45.0
\end{align*}
$$

(a 45.0 percent reduction)

Now, the question might reasonably be asked: Is this apparent 45.0 percent reduction in accidents in the present example a bona fide effect, or, might a change in accidents of this magnitude have occurred, say, 5 or more times in a hundred by chance? To answer this question, let us reflect for a moment upon $U$, and its natural logarithm (L).

Under the null hypothesis that the treatment imposed is ineffective, accidents would not be expected to change from before to after. That is to say, we would expect $U$ to equal 1 , plus or minus random error, if the imposed treatment were ineffective. Nevertheless, the sampling distribution for $U$ can range from 0 to $+\infty$. Values of $U$ between 0 and 1 represent an apparent decrease in accidents from before to after; values between 1 and $+\infty$ represent an apparent increase in accidents from before to after.

Because the sampling distribution for $U$ is highly skewed with a mean of 1 and a range from 0 to $+\infty, U$ is often transformed by taking its natural logarithm. This expression for $L$ is referred to as a logit.

$$
\begin{align*}
L & =\ln U  \tag{5}\\
& =-0.598
\end{align*}
$$

The sampling distribution for $L$ is symmetric and asymptotically normal with a mean of 0 and a range from - $\infty$ to $+\infty$. Furthermore, the standard error for $L$ $\left(L_{\text {se }}\right)$ may be approximated as:

$$
\begin{align*}
L_{\text {se }} & =(1 / A+1 / B)^{0.5}  \tag{6}\\
& =0.375
\end{align*}
$$

(for $A=11 ; B=20$ )

Eqs 5 and 6 may then be used to calculate a standard normal ( $Z$ ) score to determine if $L$ is significantly different from 0 , which is equivalent to determining if $U$ is significantly different from 1 . And, if $U$ is significantly different from 1, the null hypothesis is rejected in favor of the alternative that the treatment imposed had an effect on accidents, i.e., that the 45.0 percent reduction in accidents observed in the current example is a real effect.

$$
\begin{aligned}
Z & =L \div L_{\text {se }} \\
& =\ln (A / B) \div(1 / A+1 / B)^{0.5} \\
& =\ln (11 / 20) \div(1 / 11+1 / 20)^{0.5} \\
& =-1.59
\end{aligned}
$$

Since the calculated $Z$ does not exceed $\pm 1.96, L$ is not significantly different from zero, $U$ is not significantly different from 1 , and the apparent 45.0 percent reduction in accidents is not significantly different from zero (at $\alpha=0.05$ )

To place a 95 percent confidence interval around the estimated 45.0 percent reduction in accidents at the second site in Table 1 , proceed as follows: ${ }^{3}$

$$
\begin{align*}
L_{1-u} & =L \pm 1.96 L_{\text {se }}  \tag{8}\\
& =-0.598 \pm 1.96(0.375) \\
& =-1.333 \text { to } 0.137 \\
U_{1-u} & =e^{L_{1}} \text { to } e^{L_{u}}  \tag{9}\\
& =e^{-1.333} \text { to } e^{0.137} \\
& =0.264 \text { to } 1.147
\end{align*}
$$

${ }^{3}$ The subscripts 1 and $u$ represent the lower and upper limits of the 95 percent confidence interval.

$$
\begin{aligned}
E_{1-u} & =\left(U_{1}-1\right) 100 \text { to }\left(U_{u}-1\right) 100 \\
& =(0.264-1) 100 \text { to }(1.147-1) 100 \\
& =-73.6 \text { to }+14.7
\end{aligned}
$$

Note that when $L_{1-u}$ brackets zero (or $U_{1-u}$ brackets 1 ), $E_{1-u}$ will bracket zero and the calculated effect of treatment will not be significant. In the present example, the 95 percent confidence interval around the estimated treatment effect (i.e., around an estimated 45.0 percent reduction in accidents) ranges from a 73.6 percent reduction in accidents to a 14.7 percent increase.

Figure 1 further explains Eqs 3 through 10. In the top half of this figure, two functions are shown. The upper, diagonal function passes through the origin and has a slope of 1 . This function represents the null hypothesis that accidents will not change from before to after. The lower function has a slope of ( 0.550 ) and represents the alternative hypothesis. It passes through the origin and one data point (20,11). Apparent treatment effect (E) is the vertical distance between the two functions, relative to the distance between the diagonal and the horizontal axis. In this case, the vertical distance between the two functions is 45.0 percent of the distance between the diagonal and the horizontal axis.

In the lower half of Figure 1, a 95 percent confidence interval (depicted as a shaded area) has been placed around the alternative hypothesis. $U_{1}(0.264)$ represents the lower bound of the 95 percent confidence interval; $U_{u}$ (1.147) represents the upper bound of the 95 percent confidence interval. Since the diagonal (i.e., the null hypothesis) is contained within the shaded area, the apparent 45.0 percent reduction in accidents is not significant. If the diagonal had lain outside the shaded area, the calculated treatment effect would have been significant (at $\alpha=0.05$ ) .

## Estimating the Effect of Treatment at Several Sites

To determine the overall effect of treatments implemented at several sites, we will proceed in a manner somewhat similar to Larsen. Namely, we will calculate the weighted average logit ( $\bar{L}$ ) for the several treatment sites being evaluated and assess the statistical significance of the calculated $[$. The weighted average


logit will then be transformed into an overall estimate of the percent change in accidents following treatment ( $\bar{E}$ ), and a 95 percent confidence interval will be placed around $\bar{E}$. Data from Table 1 will be used to demonstrate the equations that follow.

To begin we will calculate a weight (w) for each logit (L). The calculated weights will be the reciprocals of the squares of the standard errors of the logits. (See, for example, Netter, Wasserman and Kutner, 1989, Chapter 16.) ${ }^{4}$

$$
\begin{align*}
& w=1 \div L_{s e}^{2}  \tag{11}\\
& w=1 \div(1 / A+1 / B)
\end{align*}
$$

For the second site in Table 1, the appropriate weight is 7.097 .

$$
\begin{aligned}
W & =1 \div(1 / 11+1 / 20) \\
& =7.097
\end{aligned}
$$

The weighted average logit ( $\bar{L}$ ) is simply:

$$
\begin{align*}
\bar{I} & =\Sigma w L \div \Sigma w  \tag{12}\\
& =-44.126 \div 210.413 \\
& =-0.210
\end{align*}
$$

And, the standard error for $[$ is:

$$
\begin{align*}
\Sigma_{\text {se }} & =1 \div(\Sigma w)^{0.5}  \tag{13}\\
& =1 \div 210.413^{0.5} \\
& =0.069
\end{align*}
$$

[^1]To determine if the overall, weighted average logit is significantly different from zero (i.e., to determine if the overall treatment effect is significant), a standard normal (Z) test is employed once again.

$$
\begin{align*}
Z & =\left[\div \bar{L}_{\text {se }}\right.  \tag{14}\\
& =-0.210 \div 0.069 \\
& =-3.04
\end{align*}
$$

The weighted average logit ( $\bar{L}$ ) may be transformed to $\bar{U}$ by taking its antilogarithm:

$$
\begin{align*}
\bar{U} & =e^{\bar{L}}  \tag{15}\\
& =e^{-0.210} \\
& =0.811
\end{align*}
$$

The estimated overall percent change in accidents ( $\bar{E}$ ) attributable to the treatments imposed is:

$$
\begin{align*}
\bar{E} & =(\bar{U}-1) 100  \tag{16}\\
& =(0.811-1) 100
\end{align*}
$$

$$
=-18.9 \quad \text { (an } 18.9 \text { percent reduction })
$$

The 95 percent confidence interval around $\bar{E}$ is derived from the following three equations:

$$
\begin{align*}
\bar{L}_{1-u} & =\bar{L} \pm 1.96 \bar{L}_{\text {se }}  \tag{17}\\
& =-0.210 \pm 1.96(0.069) \\
& =-0.345 \text { to }-0.075 \\
\bar{U}_{1-u} & =e^{\bar{I}_{1}} \text { to } e^{\bar{L}_{u}}  \tag{18}\\
& =e^{-0.345} \text { to } e^{-0.075} \\
& =0.708 \text { to } 0.928
\end{align*}
$$

$$
\begin{aligned}
\bar{E}_{1-u} & =\left(\bar{U}_{1}-1\right) 100 \text { to }\left(\bar{U}_{u}-1\right) 100 \\
& =(0.708-1) 100 \text { to }(0.928-1) 100 \\
& =-29.2 \text { to }-7.2
\end{aligned}
$$

In summary, the overall reduction in accidents at the 20 sites shown in Table 1 (from before to after) is estimated to be 18.9 percent, which is significant at $\alpha=0.05$ (and very similar to Larsen's estimate of 19.0 percent). The 95 percent confidence interval around this estimate ranges from a 7.2 percent reduction in accidents to a 29.2 percent reduction in accidents.

Figure 2 may help to further explain Eqs 11 through 19. In the upper half of Figure 2, the 20 before-after data points from Table 1 are shown. The diagonal function passing through the origin represents the null hypothesis that accidents do not change from before to after. The lower function passing through the origin and having a slope of $\bar{U}(0.811)$ represents the alternative hypothesis, that accidents do change from before to after, based upon 20 data points. The vertical distance between the two functions, relative to the vertical distance between the diagonal and the horizontal axis, represents the percent change in accidents from before to after. In the present case, the vertical distance between the two functions is 18.9 percent of the vertical distance between the diagonal and the horizontal axis.

In the lower half of Figure 2, the 95 percent confidence interval around the alternative hypothesis is shown as a shaded area. The upper boundary for this interval ( $\bar{U}_{u}$ ) passes through the origin and has a slope of 0.928 . The lower boundary $\left(\mathrm{U}_{1}\right)$ passes through the origin and has a slope of 0.708 . Since the shaded area does not contain the diagonal (i.e., the null hypothesis), the estimated 18.9 percent reduction in accidents $(\vec{E})$ is significant at $\alpha=0.05$.

Having determined that, overall, the 20 treatment sites demonstrated a significant 18.9 percent reduction in accidents from before to after, it is important to ask the question: Were the changes in accidents homogeneous (i.e., consistent) across all treatment sites? Or, were accident changes (from before to after) at the different treatment sites heterogeneous (i.e., inconsistent)?

If the imposed treatments were equally effective (i.e., homogeneous) across all sites, the data points in the upper half of Figure 2 would fall along the



FIGURE Z: BEFOFE-AFTER ANALYSIS OF, TREATMENTS IMFLEMENTED AT ZO SITES
lower function (i.e., along the alternative hypothesis that passes through the origin with a slope of 0.811). To the extent that the data points are scattered around this function, individual estimates of treatment effect are not homogeneous, not consistent. The question now becomes: Is it reasonable to assume that the scattering of the data points around the alternative hypothesis represents random error about a consistent treatment effect? Or, is it more reasonable to assume that inconsistent treatment effects (i.e., "apples and oranges") are being averaged together? If inconsistent treatment effects are being averaged together, the overall weighted average is at least partially dependent upon which remedial projects are included in the analysis (i.e., upon the relative numbers of apples and oranges included in the analysis).

Fortunately, a simple test statistic referred to as chi-square ( $\mathrm{x}^{2}$ ) homogeneity is available for determining the consistency of several estimates of treatment effect, as outlined in Table 2. ${ }^{5,6}$

Table 2: Calculation of $x^{2}$ Treatment, $x^{2}$ Homogeneity and $x^{2}$ Total

| Source | $\mathrm{x}^{2}$ | Degrees of Freedom |
| :---: | :---: | :---: |
| Treatment | $L^{2}(\Sigma w)$ | 1 |
| Homogeneity | $\Sigma w(L-\bar{L})^{2}$ | $\mathrm{N}-1$ |
| Total | $\Sigma w L^{2}$ | $N$ |

$x^{2}$ treatment is a measure of the significance of the departure of $\bar{U}(0.811)$ from 1 , or the significance of the departure of $[(-0.210)$ from zero. By squaring the $Z(-3.04)$ associated with $[$, we get a value that is algebraically
${ }^{5}$ For more detail on chi-square homogeneity see Woolf (1955) or Fleiss (1973, Chapter 10). See also Berkson's minimum logit chi-square which is equivalent (Bishop, Fienberg and Holland, 1975, Chapter 10).
${ }^{6}$ The program used to perform the analyses in this article was written in SAS (the Statistical Analysis System ${ }^{\mathrm{tm}}$ ). A copy of the program is provided as Appendix A. The output for the current data set is provided as Appendix B.
equivalent to $x^{2}$ treatment: $\left[^{2}(\Sigma \omega) .^{7}\right.$ There is one degree of freedom associated with $x^{2}$ treatment.
$x^{2}$ total is equal to the sum of the of the $Z^{2 \prime}$ s (i.e., $x^{2 \prime} s$ ) at each of the treatment sites. ${ }^{8}$ The number of degrees of freedom associated with $x^{2}$ total is equal to the number of treatment sites. If there are $N$ treatment sites, there are $N$ degrees of freedom associated with $x^{2}$ total.
$x^{2}$ homogeneity is the difference between $x^{2}$ total and $x^{2}$ treatment, which is algebraically equivalent to $\Sigma w(L-\bar{L})^{2}$. In effect, $x^{2}$ homogeneity is a measure of "goodness of fit" - a measure of the significance of the scatter of the data around the overall estimate of treatment effect. When individual estimates of treatment effect (i.e., $L, U$ or $E$ ) at each of $N$ sites are equal, $x^{2}$ homogeneity will equal zero. When calculated treatment effects differ by nothing more than random error from site to site, $x^{2}$ will be relatively small (i.e., within chance expectation), and there will be no reason to reject the notion that the individual estimates of treatment effect are homogeneous
${ }^{7}$ The significance of the weighted average logit is assessed with Z :

$$
\begin{align*}
Z & =\bar{L} \div \bar{L}_{\text {se }}  \tag{fromEq14}\\
Z^{2} & =\bar{L}^{2} \div \bar{L}_{\text {se }}{ }^{2} \\
& =\bar{L}^{2} \div(1 / \Sigma \mathrm{W})  \tag{seeEq13}\\
& =\bar{L}^{2}(\Sigma \mathrm{~W})
\end{align*}
$$

(Recall that chi-square with one degree of freedom is equivalent to $z^{2}$.)
${ }^{8}$ The significance of a logit at a given treatment site is assessed with Z :

$$
\begin{align*}
Z & =L \div L_{s e}  \tag{fromEq7}\\
Z^{2} & =L^{2} \div L_{s e}{ }^{2} \\
& =L^{2} \div(1 / w)  \tag{seeEq11}\\
& =W L^{2}
\end{align*}
$$

The sum of the individual $Z^{2 \prime} s$ is, therefore, $\Sigma \mathrm{wL}^{2}$, as shown in Table 2.
(consistent). When calculated treatment effects differ substantially from site to site, $x^{2}$ homogeneity will be relatively large (i.e., larger than we would have expected by chance), and we will conclude that the overall estimate of treatment effect is an average derived by combining heterogeneous (inconsistent) effects (i.e., an average based upon apples and oranges). There are N - 1 degrees of freedom associated with $x^{2}$ homogeneity.

Table 3 is a chi-square analysis of the data in Table 1 . The chi-square homogeneity shown in this table is large: $x^{2}=39.127$ with 19 degrees of freedom. The probability of getting a $x^{2}$ of 39.127 (or larger) with 19 degrees of freedom, if the effects of treatment at all 20 sites are really homogeneous (equal), is small: 0.0024. Therefore, we will conclude that the effects of treatment at the 20 sites shown in Table 1 were heterogeneous, inconsistent from site to site. This finding of heterogeneity suggests that other factors - factors not accounted for in the present analysis - may explain why some sites benefitted more than others from treatment.

Table 3: Summary Chi-Square Analysis of Larsen's Data (Table 1)

| Source | $\mathrm{x}^{2}$ | df | Probability |
| :---: | :---: | :---: | :---: |
| Treatment | 9.254 | 1 | 0.0024 |
| Homogeneity | 39.127 | 19 | 0.0043 |
| Total | $\overline{48.381}$ | 20 | 0.0004 |

This finding of heterogeneity also suggests that the 18.9 percent reduction in accidents (overall) results, at least partially, from the particular admixture of projects in Table 1 . Were this analysis to be repeated with a different admixture of projects, a different overall treatment effect might very well be computed.

## Discussion

When accident data are recorded at a site before (B) and after (A) treatment, the difference between $B$ and $A$ may result from any one of three factors (or combinations of factors):

1. Treatment Effect
2. Random Error
3. Non-Random Error (e.g., confounding factors, selection bias, etc.)

In the analyses performed in this article, we assume that all of the differences between B and A result from (1) treatment effect and/or (2) random error. Then, if we find random error an unlikely explanation for the difference between $B$ and $A$, we assume that the observed difference between $B$ and $A$ results from treatment (Eqs 7 and 14). But, what about non-random error?

What if the "after" data in our analysis were collected during an economic recession or an oil embargo?

What if the state in which our data were collected raised its accident reporting threshold from $\$ 250$ to $\$ 1,000$ just as we began collecting "after" data?

What if the years during which we collected "after" data were unusually cold and icy, or rainy?

What if average daily traffic (ADT) was increasing rapidly at our treatment sites between the "before" and "after" period?

What if the sites selected for treatment were specifically chosen because they demonstrated unusually high numbers of accidents during the "before" period?

If economic recessions, oil embargoes, accident reporting thresholds, weather conditions and ADT affect accident probability - and if any of these factors occur unevenly in the "before" and "after" periods of the evaluation -non-random error is introduced into our analysis. If individual treatment sites are selected for high numbers of accidents during the "before" period, accidents at that those sites will (other things being equal) regress back toward the mean during the "after" period, and, once again, non-random error will be introduced
into our analysis. (See, for example, Griffin et al., 1975).
When non-random error is present but not accounted for in our analysis, estimates of treatment effect may be seriously in error. Treatments may appear to be significant when they are not; treatments that do not appear to be significant may, in fact, be significant.

Again, the statistical procedure presented in this article assumes that non-random errors (confounding variables, selection bias, etc.) are not at play in our analysis. If this assumption is correct, the procedure presented above is valid. But, to the extent that this assumption is incorrect, the procedure presented above is invalid.

## References

Bishop, Y.M.M., S.E. Fienberg and P.W. Holland. Discrete Multivariate Analysis: Theory and Practice. Cambridge, MA: The MIT Press, 1975.

Campbell, D.T. and J.S. Stanley. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally and Company, 1972.

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

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

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.

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, June 1975.

Larsen, G.E. Non-Freeway Programwide 3R Accident Analysis: New York. Pp 22-30 in J.F. Carney (ed.) Effectiveness of Highway Safety Improvements. New York: American Society of Civil Engineers, 1986.

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

Neter, J., W. Wasserman and M.H. Kutner. Applied Linear Regression Models. Boston: Irwin, 1989.

Woolf, B. On Estimating the Relationship between Blood Group and Disease, Annals of Human Genetics, 19, 251-253, 1955.
L.I. Griffin, III

APPENDIX A

```
data ba;
options nodate nonumber linesize = 80 pagesize = 52;
input b a;
Labe1 b = 'Before'
        a = 'After'
        Lse = 'Lse'
        w = 'w'
        WL = 'WL'
        wL2 = 'wL**2';
if b = 0 then b = b + .5;
if a = 0 then a = a + .5;
E = ((a/b)-1)*100;
E = round (E,.1);
L = log(a/b);
Lse = sqrt((l/a)+(l/b)); Lse = round (Lse,.001);
L = round (L, .001);
Z = L/Lse
Z = round (Z, .01);
w = 1/((1/a)+(1/b)); w = round (w,.001);
wL = w*L;
wL = round (wL,. .001);
wL2 = wL*L;
wL2 = round (wL2, .001);
n + l;
sumw + w;
sumwL + wL;
sumwL2 + wL2;
cards;
10 14
2011
157
```

- 

.
-
3324
55
run;
proc print noobs label uniform;
var b a E L Lse Z w wL wL2;
sum b a w WL wL2;
run;
data newba;
set ba end = last;
if last ${ }^{\wedge}=1$ then delete;
keep $n$ sumw sumwL sumwL2;
$M=$ sumwL/sumw;
Mse $=1 /($ sqrt(sumw $)) ;$

```
MZ = M/Mse;
MU = exp(M);
ME = (MU-1)*100;
M1 = M - 1.96*Mse;
M2 = M + 1.96*Mse;
MU1 = exp(M1);
MU2 = exp(M2);
ME1 = (MU1-1)*100;
ME2 = (MU2-1)*100;
ct = (M**2)*sumw;
ch = sumwL2 - ct;
chdf= n-1;
pr_treat = 1 - probchi(ct,1);
pr_homo = 1 - probchi(ch,chdf);
pr_tot = 1 - probchi(sumwL2,n);
file print;
put page_;
put T------------------------------------------------------------------';
put ' SUMMARY ANALYSIS';
put '-------------------------------------------------------------------
put 'The weighted average logit' @ 60 M 8.3;
put 'The antilogarithm of the weighted average logit' @ 60 MU 8.3;
put 'The apparent overall percentage change in accidents' @ 60 ME 8.3;
put 'The Z associated with this change in accidents' @ 60 MZ 8.3;
put '-------------------------------------------------------------------
put
put
95 PERCENT CONFIDENCE INTERVALS';
put 'Weighted average logit: Upper limit' @ 60 M2 8.3;
put 'Weighted average logit: Lower limit' @ 60 Ml 8.3;
put ' ';
put 'Antilogarithm of weighted average logit: Upper limit' @ 60 MU2 8.3;
put 'Antilogarithm of weighted average logit: Lower limit' @ 60 MU1 8.3;
put ' ';
put 'Percent change in accidents: Upper limit' © 60 ME2 8.3;
put 'Percent change in accidents: Lower limit' @ 60 MEl 8.3;
put ',------------------------------------------------------------------------
put ' CHI-SQUARE ANALYSIS';
put '--------------------------------------------------------------------';
put 'Source Chi-Square df Probability';
put '-------------------------------
put 'Treatment 1' @ 21 ct 7.3 @ 43 pr_treat 7.4;
put 'Homogeneity' @ 21 ch 7.3 @ 35 chdf 2.0 @ 43 pr_homo 7.4;
put '---------- ----------------------';
put 'Total' @ 21 sumwL2 7.3 @ 35 n 2.0 @ 43 pr_tot 7.4;
put '-----------------------------------------------------------------
put
-----------------------------------------------------------------------------
run;
```


## L.I. Griffin, III

## APPENDIX B

SAS

| Before | After | E | $L$ | Lse | $Z$ | $w$ | $W L$ | $W L * * 2$ |
| ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| 10 | 14 | 40.0 | 0.336 | 0.414 | 0.81 | 5.833 | 1.960 | 0.659 |
| 20 | 11 | -45.0 | -0.598 | 0.375 | -1.59 | 7.097 | -4.244 | 2.538 |
| 15 | 7 | -53.3 | -0.762 | 0.458 | -1.66 | 4.773 | -3.637 | 2.771 |
| 28 | 29 | 3.6 | 0.035 | 0.265 | 0.13 | 14.246 | 0.499 | 0.017 |
| 40 | 18 | -55.0 | -0.799 | 0.284 | -2.81 | 12.414 | -9.919 | 7.925 |
| 5 | 4 | -20.0 | -0.223 | 0.671 | -0.33 | 2.222 | -0.496 | 0.111 |
| 40 | 36 | -10.0 | -0.105 | 0.230 | -0.46 | 18.947 | -1.989 | 0.209 |
| 10 | 28 | 180.0 | 1.030 | 0.368 | 2.80 | 7.368 | 7.589 | 7.817 |
| 57 | 45 | -21.1 | -0.236 | 0.199 | -1.19 | 25.147 | -5.935 | 1.401 |
| 82 | 51 | -37.8 | -0.475 | 0.178 | -2.67 | 31.444 | -14.936 | 7.095 |
| 2 | 8 | 300.0 | 1.386 | 0.791 | 1.75 | 1.600 | 2.218 | 3.074 |
| 26 | 29 | 11.5 | 0.109 | 0.270 | 0.40 | 13.709 | 1.494 | 0.163 |
| 16 | 3 | -81.3 | -1.674 | 0.629 | -2.66 | 2.526 | -4.229 | 7.079 |
| 14 | 17 | 21.4 | 0.194 | 0.361 | 0.54 | 7.677 | 1.489 | 0.289 |
| 29 | 17 | -41.4 | -0.534 | 0.305 | -1.75 | 10.717 | -5.723 | 3.056 |
| 27 | 17 | -37.0 | -0.463 | 0.310 | -1.49 | 10.432 | -4.830 | 2.236 |
| 17 | 21 | 23.5 | 0.211 | 0.326 | 0.65 | 9.395 | 1.982 | 0.418 |
| 18 | 16 | -11.1 | -0.118 | 0.344 | -0.34 | 8.471 | -1.000 | 0.118 |
| 33 | 24 | -27.3 | -0.318 | 0.268 | -1.19 | 13.895 | -4.419 | 1.405 |
| 5 | 5 | 0.0 | 0.000 | 0.632 | 0.00 | 2.500 | 0.000 | 0.000 |
| $=====$ | $=====$ |  |  |  |  | $=======$ | $======$ | $=====$ |
| 494 | 400 |  |  |  |  | 210.413 | -44.126 | 48.381 |


| SUMMARY ANALYSIS |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| The weighted | ge logit |  |  | -0.210 |
| The antiloga | of the weig | ave | e logit | 0.811 |
| The apparent | 11 percenta | hang | n accidents | -18.918 |
| The $Z$ associ | ith this ch | in | idents | -3.042 |
| 95 PERCENT CONFIDENCE INTERVALS |  |  |  |  |
| Weighted average logit: Upper limit |  |  |  | -0.075 |
| Weighted average logit: Lower limit |  |  |  | -0.345 |
| Antilogarithm of weighted average logit: Upper limit Antilogarithm of weighted average logit: Lower limit |  |  |  | 0.928 |
|  |  |  |  | 0.708 |
| Percent change in accidents: Upper limit <br> Percent change in accidents: Lower limit |  |  |  | -7.188 |
|  |  |  |  | -29.166 |
| CHI-SQUARE ANALYSIS |  |  |  |  |
| Source Chi-Square df Probability |  |  |  |  |
| Treatment Homogeneity | 9.254 | 1 | 0.0024 |  |
|  | 39.127 | 19 | 0.0043 |  |
| Total | 48.381 | 20 | 0.0004 |  |


[^0]:    ${ }^{1}$ Unfortunately, analyses based upon the before-and-after design can be, and often are, invalid for a variety of reasons that will not be fully addressed in this article. See, however, Campbell and Stanley (1966); Griffin, Powers and Mullen (1975); Griffin (1981).

[^1]:    ${ }^{4}$ If a given site sustains many before and after accidents, the logit for that site is estimated with relatively little error (i.e., $L_{\text {se }}$ is small) and, therefore, that site should (and does) have a relatively large weight (w) by Eq 11. Conversely, if a given site sustains few before and after accidents, the logit for that site is estimated with less precision (i.e., $L_{\text {s }}$ is large) and the weight associated with that site should be (and is) relatively small.

