

3. P.A. Hall. The Identification of High Accident Location. Presented at 11th International Study Week in Traffic Engineering and Safety, Brussels, Belgium, 1972.
4. W.F. McFarland, L.I. Griffen, J.B. Rollins, W.R. Stockton, D.T. Phillips, and C.L. Dudek. Assessment of Techniques for Cost-Effectiveness of Highway Accident Countermeasures. FHWA, Rept. FHWA-RD-79-52, Jan. 1979.
5. L.S. Lasdon. Optimization Theory for Large Systems. Macmillan, London, 1970.
6. D.W. Jorgensen. Multiple Regression Analysis of the Poisson Process. Journal of the American Statistical Association, Vol. 56, 1961.
7. D.C. Weber. Accident Rate Potential: An Application of Multiple Regression Analysis of a Poisson Process. Journal of the American Statistical Association, Vol. 66, 1971.

Publication of this paper sponsored by Committee on Methodology for Evaluating Highway Improvements.

Evaluating Need for Accident-Reduction Experiments

WILLIAM D. BERG AND CAMIL FUCHS

New traffic control devices or new applications of existing devices are frequently proposed as a means of facilitating the driving guidance and control process and thereby improving traffic safety. Before such changes can be approved at the national level, some research must be undertaken to evaluate the potential safety effectiveness of the new device. Safety effectiveness can be measured directly in terms of a reduction in accident rate or indirectly in terms of a change in an alternative measure of effectiveness. A requirement that accident data be collected before a new traffic control device standard or guideline is approved may itself be impractical and/or not cost-effective. A four-step methodology is presented for quantitatively addressing the need to undertake accident-reduction research experimentation. Statistical analysis and sampling requirements are developed first. This is followed by a determination of the minimum accident-rate reduction that would economically justify nationwide deployment of the new traffic control device treatment. The cost-effectiveness of alternative experimental designs is then evaluated. The final step is a trade-off analysis of the value of information to be derived versus the cost of obtaining the information. A case study application of the methodology is also presented.

New traffic control devices or new applications of existing devices are frequently proposed as a means of facilitating the driving guidance and control process and thereby improving traffic safety. Before such changes can be approved at the national level, some research must be undertaken to evaluate the potential safety effectiveness of the new device. In addition, the costs required to implement the new traffic control device treatment, under either an as-needed or an immediate-replacement policy, must be evaluated.

Safety effectiveness can be measured directly in terms of a reduction in accident rate or indirectly in terms of a change in an alternative measure of effectiveness. Examples of the latter include vehicle speed profiles, variance in lateral placement of vehicles within a roadway lane, driver head and/or eye movements, and various types of traffic conflicts as defined by procedures for traffic-conflicts analysis (1-3). Regardless of whether accident data or alternative measures of effectiveness are used, the principal issue is how much information is necessary to make a reasonably confident decision about potential safety cost-effectiveness.

A requirement that accident data be collected and evaluated before a new traffic control device standard or guideline is approved may itself be impractical and/or not cost-effective. If this is the case, then a decision about approval of the new traffic control device must be based on an evaluation of alternative measures of effectiveness. This would require an assumption about the true relationship between accident rate and the alternative measure. Because this is usually a qualitative judge-

ment, there can be substantial differences of opinion about potential safety effectiveness and therefore a lack of necessary support for what may actually be a very cost-effective standard or guideline.

The purpose of this paper is to present a methodology for analytically addressing these issues. The methodology was developed during research on an experimental design for evaluating the safety benefits of railroad advance-warning signs (4). The results from that case study will be used to demonstrate the application of the methodology.

EVALUATION METHODOLOGY

The evaluation methodology is designed to address the following basic questions relative to proposed research experiments of the accident-reduction potential of a new traffic control device standard or guideline:

1. What are the sampling requirements based on a treatment-control comparison?
2. What is the critical, or minimum, accident-rate reduction that the experimental design should be capable of detecting?
3. What is the cost-effectiveness of alternative experimental designs?
4. What is the value of the information to be derived from the experiment?

The evaluation methodology is therefore presented as a four-step process.

STATISTICAL ANALYSIS AND SAMPLING REQUIREMENTS

It is assumed that a treatment-control comparison is to be used in the analysis of the experimental data, although this can be supplemented with a before-and-after analysis if desired. With a treatment-control type of design, one group of sites is selected to receive or be treated with the proposed new control device application. A second group of sites would be selected as a control or base against which measured changes in accident rate at the treatment sites can be compared.

The sampling scheme is composed of two parts. First, the selected population of study sites would be divided into k homogeneous sets, each composed of n similar sites (where $n = 1, 2, 3, 4, 5,$ or larger). Then, from each of the k sets, one of the

n sites would be randomly selected to be in the treatment group. All of the other n - 1 sites would be included in the control group.

There are two advantages to this sampling scheme. First, the direct comparison between the treatment site and the respective sites in the control group for each set ensures that the comparison is performed among sites as homogeneous as possible, thereby permitting random fluctuations to be minimized. Second, the k sets of homogeneous sites can later be consolidated into a smaller number of sets, or scenarios. Although each resulting scenario would not be as homogeneous as one of the original sets, it would still be possible to determine whether the change in traffic control device had a statistically significant effect on accident experience. The various scenarios would reflect a cross-classification of the highway population in terms of appropriate design and operational characteristics.

The comparison of accident rates between treatment and control sites can be made on the basis of the overall accident rates for the two groups of sites or in terms of subsets of sites having similar characteristics. In both cases it is assumed that accidents are rare events and are therefore Poisson distributed. For the overall comparison of treatment and control sites, it can be shown that by using the normal approximation to the Poisson distribution and applying the correction for continuity, the statistic for testing the null hypothesis of no effect is

$$Z = \frac{(1/k) \sum_{j=1}^k \{ [Y_j - (1/2)] - [\bar{X}_j + [1/2(n-1)]] \}}{\left\{ \sum_{j=1}^k [Y_j + (n-1)\bar{X}_j] / (n-1) \right\}^{1/2}} \quad (1)$$

in which Y_j is the number of accidents at the treatment site in set j and \bar{X}_j is the mean number of accidents for the control sites in set j (5). The null hypothesis will be rejected at the 5 percent significance level if $Z < -1.64$; the conclusion is that the new traffic control device has a significant effect in reducing accidents.

Rather than an analysis of all the sample sites as a group, it may be of interest to examine subsets of sites, each of which exhibits similar characteristics. The subsets can be identified as scenarios, $s = 1, 2, \dots, m$, that represent combinations of the k sets of sites described previously. If it is assumed that the summations of accidents for the treatment and control sites within each scenario are Poisson distributed and if the correction for continuity is applied, it can be shown that the statistic for testing the null hypothesis to be of no effect is

$$Z_s = \frac{\{ [\bar{Y}_{.s} - (1/2n_{s1})] - [\bar{X}_{.s} + (1/2n_{s2})] \}}{\left\{ [(n_{s1} \bar{Y}_{.s} + n_{s2} \bar{X}_{.s}) / (n_{s1} + n_{s2})]^{1/2} [(1/n_{s1}) + (1/n_{s2})]^{1/2} \right\}} \quad (2)$$

in which $\bar{Y}_{.s}$ is the mean number of accidents per site for the treatment group in scenario s, $\bar{X}_{.s}$ is the mean number of accidents per site for the control group in scenario s, and n_{sg} is the number of sites in scenario s, where $g = 1$ for the treatment group and $g = 2$ for the control group (5). Under the null hypothesis, Z_s^2 has a chi-square distribution with one degree of freedom, and $\sum_{g=1}^m Z_s^2$ has a chi-square distribution with m degrees of freedom.

The null hypothesis that the accident rates for the treatment and control sites over all the scenarios are equal will be rejected if the value of $\sum_{g=1}^m Z_s^2$ exceeds the critical value in the chi-square table. Furthermore, those scenarios for which there is a statistically significant dif-

ference between the accident rates for the treatment and the control groups can be identified by examining the respective values of Z_s^2 . The sign of Z_s will indicate whether the treatment of the control group has the lower accident rate. If the sign is negative, it can be said that the new traffic control treatment provides a statistically significant reduction in the expected accident rate.

The partitioning of accident rates by treatment and control groups and by scenario also offers the opportunity to apply a two-way analysis of variance (ANOVA). This procedure as well as that for a before-and-after comparison are described elsewhere (5-8).

The sample size required for the treatment-control study depends on the following parameters:

1. The desired power of the test (β),
2. The value of the overall mean accident rate ($\bar{\lambda}$), and
3. The expected percentage of reduction in accident rate (Δ).

The power is the probability of correctly detecting a change in accident rate, if there is a change (8). For a fixed mean accident rate ($\bar{\lambda}$) and a fixed percentage of accident-rate reduction (Δ), the required sample size will vary directly with the desired power. Furthermore, there is a greater likelihood in detecting a change in accident rate when the rate of accidents is high than when accidents are a rare event. Finally, it is clear that a larger change is more likely to be detected than a smaller one.

The sample-size relationship presented below was derived by assuming that the principal statistical analysis would be the overall comparison of all treatment and control sites. A larger sample size would be required to achieve the same power for the analysis of scenarios or for before-and-after analyses. The sampling scheme is designed to create k homogeneous sets of n sites, where one site will be randomly selected as a treatment site and the remaining n - 1 sites will serve as control sites. If we assume that accidents are Poisson distributed and use the normal approximation to the Poisson distribution, it can be shown that the total number of homogeneous sets of n sites needed to test the null hypothesis that Δ equals zero at the 5 percent significance level is

$$k = [\Phi^{-1}(\beta) + 1.64]^2 [n/(n-1)] (1/\bar{\lambda}\epsilon^2) \quad (3)$$

in which $\Phi^{-1}(\beta)$ is the inverse of the standard normal distribution at point β , $\bar{\lambda}$ is the mean accident rate over all sets, and ϵ is the expected change in accident rate (Δ) expressed as a fraction (5). The value of $\Phi^{-1}(\beta)$ can be easily determined from a table of the cumulative standard normal distribution as the value of the standard variate that yields a cumulative probability of β . The total number of sets (k) of one treatment and n - 1 control sites is in effect the desired sample size.

CRITICAL ACCIDENT-RATE REDUCTION

Because highway traffic accidents at a given study site are rare events, the overall mean accident rate ($\bar{\lambda}$) can be very small. This can create the need for very large required sample sizes. It is therefore possible that no experimental design would be statistically sensitive to small changes in accident rate, be feasible in terms of site availability, and be economically practical to conduct.

The smallest relative change in accident rate

that would need to be detected can be defined as the critical accident-rate reduction. Quantitatively, this is the minimum relative reduction in accident rate that would economically justify nationwide deployment of the new traffic control device treatment. To determine the critical accident-rate reduction, an economic decision model must first be specified (9,10). By using a net-present-value (NPV) criterion, the model is expressed as

$$NPV = PVB - PVC \quad (4)$$

in which the present value of benefits (PVB) is the present dollar value of a time stream of benefits and the present value of costs (PVC) is the present dollar value of costs over the same time period. If the NPV of an investment situation is greater than zero, then that investment is considered to be economically feasible. When mutually exclusive alternatives, each with a positive NPV, are compared, that alternative with the highest NPV is preferred.

PVB is defined as the present dollar value of future accident-rate reductions attributable to the nationwide deployment of the new traffic control device treatment. PVC is defined as the present dollar value of the costs of deploying and maintaining the new treatment on a nationwide basis. In addition to a policy of immediate deployment of the new treatment, it may be appropriate to consider an as-needed replacement policy that would permit gradual implementation over a period of years.

The actual formulation of the PVB and PVC functions will depend on the nature of the proposed traffic control device treatment. In general, PVB for an immediate-deployment policy can be expressed as

$$PVB = (AAR) (\Delta) (AC) (N) (SPW_{i,n}) \quad (5)$$

where

AAR = present average annual accident rate per site,

Δ = percentage of reduction in accident rate (AAR) due to increased effectiveness of new traffic control device,

AC = average dollar cost of an accident,

N = number of sites at which treatment is to be deployed, and

$SPW_{i,n}$ = series present-worth factor for discount rate of i percent and analysis period of n years.

PVC can be expressed as

$$PVC = 2N [(\Delta C + LC + MC) + (\Delta C/m^2) (GPW_{i,m}) + (\Delta C/m) (SPW_{i,n-m}) (PW_{i,m})] \quad (6)$$

where

ΔC = dollar materials cost difference between current and proposed new traffic control device treatment,

LC = dollar labor cost for deploying each new treatment,

MC = dollar mileage cost per treatment for installation crew,

m = average life of new treatment (years),

$GPW_{i,m}$ = uniform gradient present-worth factor for discount rate of i percent over $n-m$ years, and

$PW_{i,m}$ = present-worth factor for discount rate of i percent over m years.

The smallest relative change in accident rate

that the experimental design should be capable of statistically detecting is determined by applying the economic decision model and determining the relative accident-rate reduction (Δ) that would yield an NPV of zero. Any reduction in accident rate smaller than this value would be insufficient to economically justify deployment of the proposed new traffic control treatment. Any accident-rate reduction larger than this value would mean that deployment of the new treatment could be economically justified. Thus it would be important to be able to detect accident-rate reductions as small as the critical value found at the break-even point.

EXPERIMENTAL DESIGN COST-EFFECTIVENESS

As discussed previously, the required sample size for a field evaluation of accident-reduction potential consists of k treatment sites and k ($n - 1$) control sites. The magnitude of k is given by Equation 3 and can be seen to

1. Increase with the desired power (β),
2. Decrease with the level of the overall mean accident rate (λ),
3. Decrease with the square of the change in accident rate (ϵ), and
4. Decrease slightly with the size of the sets (n).

The value selected for each of these parameters will govern both the effectiveness of the experimental design in terms of its ability to detect actual changes in accident rate and the cost of conducting the experiment and analyzing the data. Thus, several alternative experimental designs can be specified and then evaluated for their potential cost-effectiveness.

The selection of the power of the test (β) is a subjective decision. Increasing its value reduces the likelihood of not statistically detecting a change in accident rate, if in fact one occurs. From a safety standpoint, it is clearly important to make the value of this parameter as large as practical. A range of values, for example, between 50 and 90 percent, could be used to specify alternative experimental designs and sampling requirements.

The overall mean accident rate (λ) is a function of the roadway situation under study. Generally, typical accident-rate data would be available from accident records systems or safety publications. Whatever the traffic control device treatment, the annual accident rate can be expected to be relatively low. This has the effect of requiring large, and possibly very large, sample sizes. Therefore, it may be necessary to consider a multi-year rather than a one-year study period. For example, the overall mean accident rate (λ) for a three-year study is three times as large as the average annual accident rate. This would have the effect of reducing the required number of treatment and control sites by a factor of 3.

The selection of a change in accident rate (ϵ) that the experimental design should be capable of detecting can be approached in two ways. First, a subjectively established range of minimum values can be used. Alternatively, the critical accident-rate reduction based on benefit-cost considerations can be employed. In either case, the smaller the value, the larger the required sample size.

The final parameter that can be varied is the size of the treatment-control sets (n). As n increases, the required number of treatment sites is reduced but the total number of treatment plus control sites is increased. The most desirable ratio depends on the relative cost of preparing the treat-

ment sites versus the cost of data collection at all treatment and control sites.

By varying each of the above four parameters, a set of alternative sampling plans can be specified in terms of the required number of treatment sites (k) and control sites [$k(n-1)$]. The cost-effectiveness of each alternative experimental design can then be examined in terms of the cost of conducting the study, the availability of treatment and control sites, and the smallest relative change in accident rate that can be detected with an acceptable power of test. Boundary conditions may constrain the range of feasible experimental designs. For example, a required sample size may exceed the number of available sites. Alternatively, a maximum budget level may limit the number of sites that can be selected and thus the minimum relative change in accident rate that can be detected during a reasonable data-collection period. In general, the cost of conducting an accident study will increase as the relative change in accident rate to be detected is decreased, the power of the test is increased, and the data-collection period is decreased.

VALUE OF INFORMATION

The ultimate question is which, if any, of the alternative experimental designs should be undertaken. For a given study cost, the most effective design is that which has the potential of detecting the smallest relative change in accident rate with an acceptable power of test. As the study budget is increased, the effectiveness of the experimental design will generally increase, but at a diminishing rate. Therefore, the selection of an experimental design must consider that trade-off between the value of the information to be derived and the estimated cost of obtaining that information.

This trade-off can be examined in two ways. First, the smallest accident-rate reduction that is likely to be statistically detectable can be compared with the lowest rate that would economically justify the nationwide application of the new traffic control treatment. Second, the cost of undertaking the accident study can be compared with the cost of deploying the new treatment nationwide. These comparisons must then be interpreted with respect to both experimental and economic practicality.

If the sample sizes necessary to detect the critical accident-rate reduction exceed the population of available sites, then the study would clearly be impractical because no experimental design could be expected to detect all possible accident-rate reductions that would economically justify deployment of the new traffic control treatment. Similarly, if the estimated cost of conducting the most cost-effective study designs were to equal or exceed the approximate total cost of nationwide deployment of the new traffic control treatment, then the study would be economically impractical. Finally, if the estimated study cost necessary to detect the critical accident-rate reduction was simply considered to be too expensive, then the study would also be considered impractical.

If the final decision is that a practical, cost-effective experimental design is available, then the study should be initiated. However, if no such study design can be found, then three options exist. First, no further action of any type would be taken. This choice should only be favored if the proposed traffic control treatment is subjectively judged to have little merit. Second, experimental research could be conducted by using alternative safety measures of effectiveness instead of accident-rate data. Finally, the proposed new traffic

control treatment could be approved for use without further experimental research. This option should only be favored if the critical accident-rate reduction is very low and the new traffic control treatment is judged to offer positive (albeit unmeasured) safety benefits.

APPLICATION OF METHODOLOGY

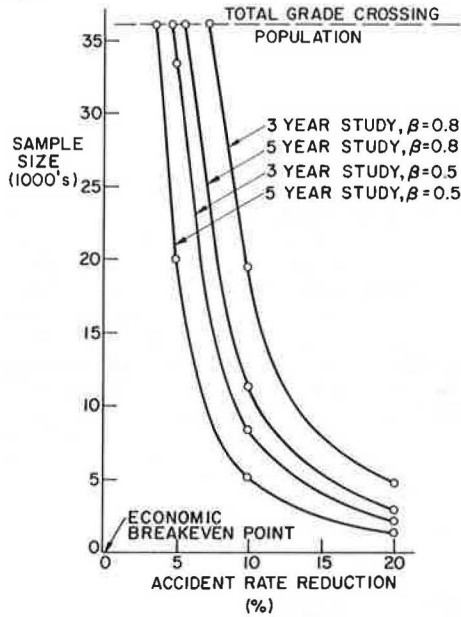
As described in detail elsewhere, the above methodology was used in a study to develop an experimental design for evaluating the safety benefits of a new railroad advance-warning sign (4,5). The sample population of interest consisted of the total nationwide population of 36 104 railroad-highway grade crossings that have reflectorized crossbucks and standard advance-warning signs. The overall mean accident rate ($\bar{\lambda}$) was 0.049 accident per crossing per year. Three-year and five-year expected accident rates were then calculated to reflect the expected mean accident rates for a three-year and a five-year study. The relative change in accident rate to be detected was varied from 5 to 20 percent, and the ratio of treatment to control sites was varied from 1:1 to 1:4. A 50 percent and an 80 percent power of test were also specified.

By using Equation 3, it was found that to be 50 percent confident of detecting an actual 5 percent reduction in accident rate over a three-year study period, it would be necessary to select almost 15 000 treatment and 15 000 control sites. The required number of treatment sites could be reduced to approximately 11 000 if desired, but the number of control sites would then have to be increased to a little more than 22 000. Thus by decreasing the ratio of treatment to control sites from 1:1 to 1:2, the total number of required sites would increase from approximately 30 000 to 33 000. Of greater significance, however, was the fact that both sample sizes nearly equaled the total population of 36 104 grade crossings. If the ratio of treatment to control crossings were to be reduced any further, the total required sample size would exceed the population size. Increasing the power of the test to 80 percent and assuming a relatively long five-year data-collection period resulted in the same finding.

The critical accident-rate reduction was then calculated and found to vary over a range of 0.01-0.13 percent, depending on the sign-deployment policy and the unit cost data used. This was equivalent to a reduction of about one grade crossing accident in five to six years over the total population of 36 104 grade crossings. These results clearly suggested that it might be both experimentally and economically impractical to attempt to determine whether the actual safety effectiveness of the new advance-warning sign would justify its deployment on a nationwide basis.

The trade-off among the smallest detectable accident-rate reduction, the required sample size, and the accident-rate reduction associated with the economic break-even point for justifying deployment of the new advance-warning sign was examined by preparing the graph shown in Figure 1. It is clear that none of the alternative experimental designs could be expected to provide the information necessary to establish whether the potential safety benefits of the new sign would exceed the total cost of nationwide deployment. When the six most cost-effective experimental design alternatives were compared with the estimated cost of nationwide deployment of the new sign, it was found that the cost of four of the alternative experimental designs would significantly exceed that total initial cost of deploying the new sign on an as-needed basis over a seven-year period. Moreover, these study costs fell within approx-

Figure 1. Sample-size requirements: total nationwide population of rural and urban grade crossings with reflectorized crossbucks and advance-warning signs.



imately 30-75 percent of the total initial cost of an immediate nationwide sign-replacement policy. It was therefore concluded that the proposed research study would be both experimentally and economically impractical and should therefore not be undertaken.

CONCLUSIONS

It is believed that the methodology described above is both generalizable and practical. It can provide a quantitative basis for decisionmaking where strong differences in personal opinion may exist regarding the need for accident data as a precondition for approving a new traffic control device treatment. Application of the methodology in these situations can assist in making rational choices, avoiding needless experimentation, and facilitating early decisions and timely realization of the benefits of meritorious new traffic control device treatments.

ACKNOWLEDGMENT

This paper is based on a study sponsored by the Office of Research of the Federal Highway Administration. The contents reflect our views, and we are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official view or policy of the U.S. Department of Transportation.

REFERENCES

1. W.A. Stimpson, W.A. Kittelson, and W.D. Berg. Methods for Field Evaluation of Roadway-Delineation Treatments. TRB, Transportation Research Record 630, 1977, pp. 25-32.
2. J.S. Koziol and P.H. Mengert. Railroad Grade Crossing Passive Signing Study. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, Final Rept. FHWA-RD-78-34, Aug. 1978.
3. W.D. Glauz and D.J. Higletz. Application of Traffic Conflicts Analysis at Intersections. NCHRP, Rept. 219, Feb. 1980.
4. W.D. Berg, C. Fuchs, and J. Coleman. Evaluating the Safety Benefits of Railroad Advance-Warning Signs. TRB, Transportation Research Record 773, 1980, pp. 1-6.
5. W.D. Berg. Experimental Design for Evaluating the Safety Benefits of Railroad Advance Warning Signs. FHWA, Rept. RHWA-RD-79-78, April 1979.
6. F. Mosteller, and J.W. Tukey. Data Analysis and Regression. Addison-Wesley, New York, 1977.
7. F.I. Anscombe and J.W. Tukey. Examination and Analysis of Residuals. Technometrics, Vol. 5, 1963, pp. 141-160.
8. K.A. Brownlee. Statistical Theory and Methodology in Science and Engineering. Wiley, New York, 1960.
9. R. Winfrey. Economic Analysis for Highways. International Textbook Company, Scranton, PA, 1969.
10. A Manual on User Benefit Analysis of Highway and Bus-Transit Improvements. American Association of State Highway and Transportation Officials, Washington, DC, 1977.

Publication of this paper sponsored by Committee on Methodology for Evaluating Highway Improvements.