# New Directions for Learning About the Safety Effect of Measures

EZRA HAUER and JANE LOVELL

ABSTRACT

Much of what is known about the safety effect of various measures must be extracted from implementations in real life rather than from experiments that are staged to meet the dicta of rigorous scientific experimental design. The tools for extracting usable knowledge from data must be tailored to suit this reality. Methods of estimation that appear well suited for this task are reported here. First it is shown that what is commonly done is incorrect; it is incorrect to compare the count of "before" accidents with the count of "after" accidents and from this to draw conclusions about the safety effect of a measure. A simple method is provided for the correct analysis of "before" and "after" data. Next the likelihood function is introduced; it serves a dual purpose: First, it allows the assessment of the accuracy with which the safety effect is known. Second, it is a coherent formal device by which results from diverse studies can be accumulated. The ability to accumulate empirical evidence from many small studies is the key to progress in research on safety. The test of the advocated methods is in application; in this case, the examination of the effect on intersection safety of a change from two-way to multiway stop control. Details are given in two companion papers appearing elsewhere in this Record.

What is known about the safety effect of some treatment or measure is based mostly on data derived from instances of implementation. The implementation of a real measure is usually fashioned by the circumstances of the real world and only seldom by the requirements of scientific experimental design. For measures that are in the orbit of highway or traffic engineering, most data come in the form of before and after accident counts, perhaps supplemented by the corresponding changes in exposure. This is why much of the traditional knowledge about the effect of such measures is based on before-and-after studies.

The before-and-after study is almost always too small to be statistically conclusive. It is also vulnerable to a variety of threats to the validity of the inferences that the data permit. This is why the purist will often refuse to consider evidence based on uncontrolled studies of this kind. However, it helps little to bemoan the fact that the before-and-after experimental design is subject to threats and that its results are not statistically significant. The challenge is to devise methods that minimize such threats and that allow this ubiquitous source of information to be used constructively.

In this paper the authors report on an approach that appears well suited for the task of extracting useful information from uncontrolled before-and-after studies and that facilitates the accumulation of empirical evidence from diverse studies, each of which, standing alone, may be inconclusive.

Even though the focus of this paper is methodological, the authors rely more on common sense and intuition than on mathematical formalism; in this way they hope to convince a wide readership that it is unsound and therefore unprofessional to draw conclusions about the safety effect of a measure from a simple comparison of accidents before to accidents after (even when changes in exposure and the secular time-trend are taken into account). This is discussed next. In the subsequent section the correct methods of analysis are discussed. Later the concept of the likelihood function is introduced and its attraction and use are explained. In the final section of the paper accomplishments are reviewed and some of the problems yet unresolved are discussed.

Transport Safety Studies Group, Department of Civil Engineering, University of Toronto, Ontario M5S 1A4, Canada.

## WHY IT IS INCORRECT TO COMPARE THE COUNT OF "BEFORE" ACCIDENTS WITH THE COUNT OF "AFTER" ACCIDENTS

A typical before-and-after study follows a simple pattern: at some time a measure (treatment) that affects safety is implemented on a few entities. Entities may be intersections or drivers, cities, road sections, or vehicles. The count of accidents on these entities before treatment is compared with the record of accident occurrence after treatment. On the basis of such a comparison, inferences are made about the effect of the measure or treatment.

The mind is so accustomed to this kind of comparison that the logic behind it is seldom examined; a crucial assumption that turns out to be incorrect is overlooked. It is not incorrect because of some theoretical niceties but because it is contradicted by mountains of empirical evidence. To recognize the faulty assumption it must be spelled out.

To learn about the effect of the treatment, what would have happened during the after period had the treatment not been implemented is compared with what actually has happened during the after period with the treatment in place.

This simple logical construct is behind all experimental designs, no matter how sophisticated or how simple. It can never be known "what would have happened ...." To avoid this difficulty the tendency is to assume that what has happened during the period before treatment implementation is a good indication of what would have happened during the after period had the treatment not been implemented.

This is the assumption the authors claim to be contrary to empirical fact. Only one piece of empirical evidence is given here; however, the authors have examined literally dozens of data sets and every one of those sets corroborates the conclusion that the aforementioned assumption is incorrect. The reader is invited to furnish his or her own evidence. All that is needed is at least 2 years of accident data about several hundred entities that remained largely unchanged. Such data are easy to find. When the data are examined, as in the following example, the conclusion is inescapable.

Consider the entries in Table 1. The table is based on the count of accidents occurring during the years 1974 and 1975 at 1,142 intersections in San Francisco. All intersections in this population had stop signs on the two approaches carrying the lesser flows and remained virtually unchanged in these 2 years. Column 1 gives the number of intersections [n(x)] on which the count of accidents in 1974 was x = 0, 1, 2... as shown in Column 2. Column 3 gives the average of the count of accidents [M(x)] for the same n(x) intersections during 1975.

TABLE 1 Accident Count at 1,142 Intersections, 1974-1975

| 1<br>Number of Intersections<br>[n(x)] | 2<br>Number of Accidents per Intersection in 1974<br>[x] | 3<br>Average Number of Accidents per Intersection in 1975<br>[M(x)] |
|---|---|---|
| 553 | 0 | 0.54 |
| 296 | 1 | 0.97 |
| 144 | 2 | 1.53 |
| 65 | 3 | 1.97 |
| 31 | 4 | 2.10 |
| 21 | 5 | 3.24 |
| 9 | 6 | 5.67 |
| 13 | 7 | 4.69 |
| 5 | 8 | 3.80 |
| 2 | 9 | 6.50 |

Note: Two intersections had 13 accidents, one had 16.

Were the assumption correct, it should be observed that if an intersection registered, for example, x = 3 accidents in 1974 and if it remained largely unchanged, it should record, on the average, three accidents in 1975. However, inspection of Table 1 reveals that intersections that registered three accidents in 1974, registered 1.97 accidents on the average in 1975. Similar discrepancies between the entries of Columns 2 and 3 exist for all values of x (except for x = 1, which will turn out to be the rule-confirming exception). These discrepancies cannot be reasonably attributed to chance; nor are they likely to reflect a sudden, large, and peculiarly systematic change between these 2 years. (The total number of accidents at these intersections was 1,253 in 1974 and 1,216 in 1975). It must be concluded therefore, that in this case the 1974 count of accidents is not a good indication of the average count in 1975 for any value of x (except for x = 1). Therefore, the accident count "before" is a systematically bad guess of what would have happened after.

Tables 2 and 3 give similar information for the same 1,142 intersections during the pairs of years 1975-1976 and 1976-1977. The preceding conclusion remains unchanged. It follows that there was nothing unique or peculiar about the years 1974-1975 (Table 1); what happened "before" did not prove to be a good indication of what happened after in 1975-1976 and 1976-1977 either. It is worth noting that there is a pronounced similarity between the corresponding entries of the third columns in the three tables. This regularity will be explored in the section: How to Analyze Before-and-After Data.

TABLE 2 Accident Counts at 1,142 Intersections, 1975-1976

| 1<br>Number of Intersections<br>[n(x)] | 2<br>Number of Accidents per Intersection in 1975<br>[x] | 3<br>Average Number of Accidents per Intersection in 1976<br>[M(x)] |
|---|---|---|
| 559 | 0 | 0.55 |
| 286 | 1 | 0.98 |
| 144 | 2 | 1.41 |
| 73 | 3 | 1.82 |
| 35 | 4 | 1.97 |
| 18 | 5 | 2.50 |
| 11 | 6 | 3.91 |
| 9 | 7 | 4.22 |
| 3 | 8 | 2.00 |
| 1 | 9 | 3.00 |
| 2 | 10 | 2.50 |
| 1 | 11 | 5.00 |

TABLE 3 Accident Counts at 1,142 Intersections, 1976-1977

| 1<br>Number of Intersections<br>[n(x)] | 2<br>Number of Accidents per Intersection in 1976<br>[x] | 3<br>Average Number of Accidents per Intersection in 1977<br>[M(x)] |
|---|---|---|
| 562 | 0 | 0.53 |
| 287 | 1 | 0.94 |
| 155 | 2 | 1.37 |
| 74 | 3 | 1.72 |
| 33 | 4 | 2.61 |
| 13 | 5 | 3.00 |
| 11 | 6 | 2.64 |
| 4 | 7 | 2.25 |
| 1 | 8 | 1.00 |
| 2 | 9 | 3.50 |

The results in Tables 1, 2, and 3 are not an exception or aberration. They are used here merely to illustrate a general phenomenon found in many other data sets. Based on diverse and ample empirical evidence it can be concluded that the assumption (what has happened during the period before treatment implementation is a good indication of what "would have happened during the after period had the treatment not been implemented") is contrary to empirical fact and is therefore wrong.

Because the assumption on which the simple before-and-after comparison is based is incorrect, so must be conclusions drawn from such comparisons. To illustrate, consider a site (similar to those in Table 1) that recorded, for example, three accidents before treatment and one accident during a corresponding period after treatment. The incorrect comparison is between three and one. It is clear from Tables 1, 2, and 3 that sites that record three accidents in the before period, when left untreated, record approximately two accidents in the after period. Therefore, the correct comparison is between two and one accidents. To be sure, no conclusions will be drawn from a few accidents recorded at one site. Accident counts from many sites will usually be added and the sums compared. However, if every term in the addition is incorrect, so will be the sum. The errors in the sum of before accidents will cancel only if treatment is administered at random and is implemented at very many sites. In practice, the number of treated sites is limited and professionals do not usually treat sites at random.

It is concluded, therefore, that it is incorrect to compare the count of before accidents to the count of after accidents as is common practice in before-and-after studies. A valid comparison requires that there be a way to estimate "what would have happened had the treatment not been implemented," which is in accord with empirical fact. How to obtain such estimates is described next.

## HOW TO ANALYZE BEFORE-AND-AFTER DATA

The task is to obtain a good estimator to replace that which is in common use but is shown to be faulty. One wishes to estimate the number of accidents expected to be recorded during the after period had the treatment not been implemented if, during the before period, the entity recorded x accidents. The symbol $\varepsilon(x)$ will be used to denote an estimator. It turns out that there are several candidate estimators of which two [$\varepsilon_1(x)$, $\varepsilon_2(x)$] are recommended for use in practice.

One obvious option is to use $\varepsilon(x) = M(x)$ (see Column 3 in Tables 1-3). The symbol M(x) stands for "average after-period count of accidents on those entities that recorded x accidents in the before period and were left without treatment." The use of $\varepsilon(x) = M(x)$ amounts to stating: "It is expected that had the treated entity, which in the before period recorded x accidents, been left untreated, it would have recorded during the after period, on the average, what has in fact materialized on similar entities that were left untreated." In essence, the entities with x before accidents, which were left untreated, are regarded as a control group.

The trouble with $\varepsilon(x) = M(x)$ is that, to have an accurate estimate, a sufficient number of "similar entities" have to be found that during the before period had the same number of accidents as the treated entity but that were left without treatment. This is often difficult to do. Ordinarily, it is the entities with many accidents that are treated, and there are not many such entities to begin with. Once some entities have been treated, only a few remain for the calculation of M(x). This difficulty is easy to see in Tables 1-3. In the lower reaches of these tables (where few intersections are used to calculate the average) the values of M(x) fluctuate widely. Furthermore, were some of these intersections treated, even fewer could be used for the calculation of the M(x), making it even less reliable. This renders the estimator M(x) of little use in practice.

Another estimator that can be justified on theoretical grounds (1) is $\varepsilon(x) = (x+1) \cdot n(x+1)/n(x)$. The advantage of this estimator is that only data about accidents occurring during the before period are needed. However, because in the lower reaches of Tables 1-3, the n(x) are small, the ratio n(x+1)/n(x) is subject to vagaries of chance that are similar to those that plague M(x). Thus, the problem is how to smooth out the random fluctuations that plague both estimators. Two sensible ways to obtain smooth estimates are discussed in the following paragraph.

First, a continuous function $\varepsilon_1(x)$ can be fit to the points $(x+1) \cdot n(x+1)/n(x)$. Thus, using the data in Columns 1 and 2 of Table 1, values of $(x+1) \cdot n(x+1)/n(x)$ for x = 0, 1, ..., 8 were calculated. These are the ordinates of the points in Figure 1. The bars around each point designate one standard deviation. In this case, a straight line appears to be the sensible choice of a function to fit the data points. The ordinate of the fitted function at x = 0, 1, 2... is the estimate $\varepsilon_1(x)$. When fitting a smooth function to the data points two technical issues must be given attention. First, the data points have different standard deviations. When fitting a curve, each point is to be weighed in inverse proportion to its variance, which is estimated by $[(x+1) \cdot n(x+1)/n(x)]^2 \cdot [1/n(x+1) + 1/n(x)]$. Second, for reasons of logical consistency, one would like to ensure that $\Sigma[\varepsilon_1(x) \cdot n(x)] = \Sigma[x \cdot n(x)]$ when the summation is over all values x.

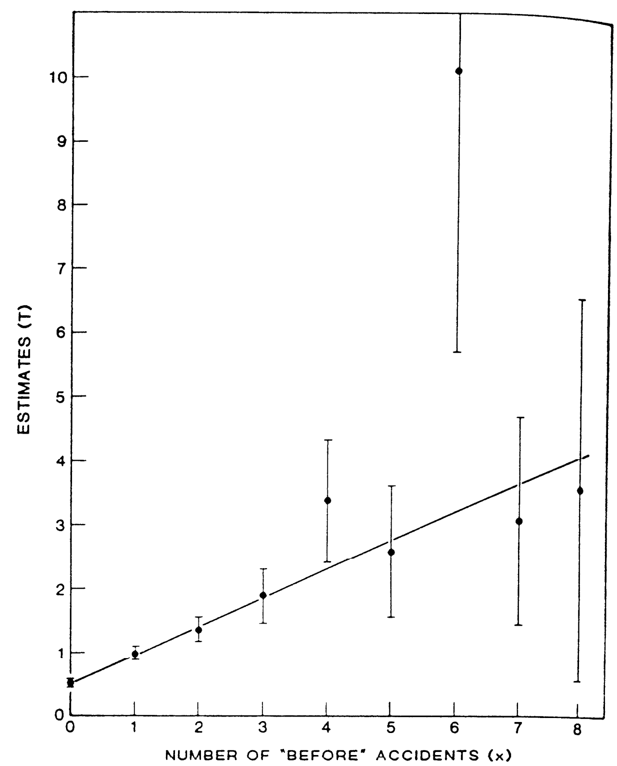


FIGURE 1 Least-squares fit of a linear function $\varepsilon_1(x)$ to points $(x+1) \cdot n(x+1)/n(x)$ based on data in Table 1.

Except for these guidelines to curve fitting, the approach is perfectly general and requires no assumptions. It consists of two basic steps: (a) values of data points $(x+1) \cdot n(x+1)/n(x)$ for x = 0, 1, 2... are calculated and plotted, and (b) a legitimate function $\varepsilon_1(x)$ is selected and fitted to the data points.

If a linear fit to the data points appears sensible, the task of curve fitting may be replaced by a much simpler and more transparent estimator, $\varepsilon_2(x)$. First, the sample mean and sample variance are calculated by using

$$\bar{x} = \Sigma[x \cdot n(x)]/\Sigma n(x) \quad (1)$$

$$s^2 = \Sigma[(\bar{x}-x)^2 \cdot n(x)]/\Sigma n(x) \quad (2)$$

Then, by using $\bar{x}$ and $s^2$

$$\varepsilon_2(x) = x + (\bar{x}/s^2) \cdot (\bar{x}-x) \quad (3)$$

Equation 3 is not magic nor does it contain "fudge factors." It is a rigorous result obtained by deduction, and it holds under broad conditions described elsewhere (2). Its main appeal is simplicity in use and clarity in interpretation.

The two terms of the sum in Equation 3 have recognizable meaning. The first term is the count of before accidents; the second term is a correction for regression-to-the-mean. The larger the difference between the count of before accidents (x) and its mean in the population of similar entities ($\bar{x}$), the larger is the correction required. It is positive when $x < \bar{x}$ (see first line in Table 1), negligible when $x \cong \bar{x}$ (see second line in Table 1) and negative when $x > \bar{x}$ (see lines below line 2 in Table 1).

The role of the sample mean-to-variance ratio ($\bar{x}/s^2$) is also interesting to examine. If it was known that all entities in the population had the same expected number of accidents (and if accident occurrence obeys the Poisson probability law), the ratio would approach 1. Under such conditions Equation 3 instructs that the expected number of accidents for a specific entity be estimated by $\bar{x}$ (not by x !). On reflection, this is as should be. If, on the other hand, the entities in the population are very different in terms of their expected number of accidents, $s^2 >> \bar{x}$, the correction will be small. In this case, $\xi_2(x)$ is very close to x. This is also as should be.

Before summarizing, it is of interest to examine in Table 4 the performance of the candidate estimators on the basis of the data in Table 1.

TABLE 4 Comparison of Estimates

| 1 [n(x)] | 2 [x] | 3 M(x) | 4 (x+1)n(x+1)/n(x) | 5 $\xi_1(x)$ Curve Fit | 6 $\xi_2(x)$ Eqn. 3 |
|---|---|---|---|---|---|
| 553 | 0 | 0.54 | 0.54 | 0.53 | 0.44 |
| 296 | 1 | 0.97 | 0.97 | 0.98 | 1.04 |
| 144 | 2 | 1.53 | 1.35 | 1.43 | 1.64 |
| 65 | 3 | 1.97 | 1.91 | 1.88 | 2.24 |
| 31 | 4 | 2.10 | 3.39 | 2.32 | 2.84 |
| 21 | 5 | 3.24 | 2.57 | 2.77 | 3.44 |
| 9 | 6 | 5.67 | 10.11 | 3.22 | 4.04 |
| 13 | 7 | 4.69 | 3.08 | 3.67 | 4.64 |
| 5 | 8 | 3.80 | 3.60 | 4.11 | 5.25 |
| 2 | 9 | 6.50 | n.a. | 4.56 | 5.85 |

A few points deserve mention. First, having established earlier that it is incorrect to use the raw number of before accidents in before-and-after comparisons, it was necessary to show what should be used instead. Several candidate methods of estimation have been presented. No matter which estimator is used, all correct estimates differ from the raw number of before accidents, which have been shown to be systematically biased.

Second, Column 3 in Table 4 is M(x) and therefore indicates what actually happened during the after period for entities that were left untreated. Where M(x) is a reliable average it could be used as a yardstick against which to judge the performance of $\xi_1(x)$ and $\xi_2(x)$. These estimates are observed to approximate the entries in Column 3 with varying degrees of success. The agreement is good in the upper part of Table 4 where the entries of Column 3 are quite accurate. Not much can be made of the discrepancies in the lower part of the table because here the entries of Column 3 (being averages over only a few intersections) are unreliable.

Third, in the authors' view, either $\xi_1(x)$ or $\xi_2(x)$ should be used because in the domain of interest they smooth out some of the fluctuations due to randomness. When the plot of points indicates a nonlinear fit, use least-squares curve fitting to find $\xi_1(x)$; otherwise use Equation 3 to obtain $\xi_2(x)$.

It remains to be demonstrated how these results are to be used in the context of a before-and-after study. This is done by example. Assume that 49 intersections (similar to those used to construct Table 1) were converted from two-way to four-way stop control. (Data and estimates are summarized in Table 5.)

Row 2 in Table 5 gives the number of intersections, which, during the before period had the number of accidents listed in Row 1. Row 3 gives the number of accidents for each group of intersections. Thus, for example, the 8 intersections, which during the before period recorded 2 accidents each, had together 16 accidents. During the before period there were 172 accidents at the 49 sites. Row 4 gives the number of accidents during the corresponding after period, which totaled 50. The authors argued that it is wrong to compare 172 to 50. Row 5 gives an estimate (here $\xi_2(x)$ from Table 4) of the number of accidents that should be expected had the intersections not been converted to four-way stop control. In this illustration, the changes in exposure and the secular trend in accidents is disregarded. Complete details are given in the paper "The Safety Effect of Conversion to All-Way Stop Control" elsewhere in this Record. In Row 6 the number of accidents in each group of intersections that should be expected had they remained unconverted has been calculated. Thus the seven intersections, which during the before period recorded no accidents, should be expected to record 7 x 0.44 = 3.1 accidents during the after period had they remained with two-way stop control. The sum of these expected accidents is 124.8. The effectiveness of the conversion should be judged by comparing what "would have happened without treatment" (124.8) with what actually transpired (50 accidents). In this numerical example it is estimated that the effect of conversion to all-way stop control was to reduce the expected number of accidents by 60 percent [= 100 x (124.8-50)/124.8].

## THE LIKELIHOOD FUNCTION AND ITS USE

Point estimates of the type mentioned at the end of the previous section (a 60 percent reduction in expected accidents) are often the main figure of merit when it comes to practical decisions. However, for sound decisions it is necessary to have, in addition to the point estimate, a good idea about the uncertainty surrounding it. Unfortunately, real-life studies are almost without fail small in the sense

TABLE 5 Before-and-After Comparison

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | No. of before accidents per intersection | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Row sums |
| 2 | No. of such intersections in sample | 7 | 6 | 8 | 7 | 4 | 6 | 4 | 3 | 0 | 2 | 2 | 49 |
| 3 | No. of before accidents (1x2) | 0 | 6 | 16 | 21 | 16 | 30 | 24 | 21 | – | 18 | 20 | 172 |
| 4 | No. of after accidents | 2 | 3 | 7 | 5 | 4 | 12 | 6 | 5 | – | 2 | 4 | 50 |
| 5 | Estimate $\xi_2(x)$ from Table 4 | 0.44 | 1.04 | 1.64 | 2.24 | 2.84 | 3.44 | 4.04 | 4.64 | | 5.85 | 6.45 | |
| 6 | No. expected w/o treatment (5x2) | 3.1 | 6.2 | 13.1 | 15.7 | 11.4 | 20.6 | 16.2 | 13.9 | – | 11.7 | 12.9 | 124.8 |

that the estimates of safety effect derived from them are inaccurate. Also, taken singly, such estimates are only a frail guide for sound decisions. Thus, to make progress it is necessary to combine the information contained in many small studies in order for reliable knowledge to gradually emerge. The likelihood function is proposed for both purposes: (a) to characterize the accuracy with which the safety effect of a measure is known, and (b) to accumulate information obtained from diverse studies. It is best to postpone giving the reasons for this choice until after the use of the likelihood function in this context is explained.

To illustrate the use and interpretation of the likelihood function, another numerical example is introduced. It is concerned with the safety effect on right-angle accidents of converting 10 rural intersections in Michigan to all-way stop control (4). The data for the likelihood function are given in Table 6.

**TABLE 6 Data for the Likelihood Function for Michigan Right-Angle Accidents (4)**

| Site | $\alpha_i$ | $\beta_i$ | $(\varepsilon'/\varepsilon)_i$ | $x_i$ Accidents Before | $x'_i$ Accidents After | $B_i$ Years Before | $A_i$ Years After |
|---|---|---|---|---|---|---|---|
| 1 | 1.5603 | 0.1434 | 1.2237 | 14 | 6 | 3 | 3 |
| 2 | 1.6187 | 0.1457 | 1.0657 | 16 | 3 | 3 | 3 |
| 3 | 1.5603 | 0.1434 | 1.0189 | 18 | 9 | 3 | 3 |
| 4 | 1.5603 | 0.1434 | 1.0549 | 28 | 7 | 3 | 3 |
| 5 | 1.5603 | 0.1434 | 1.1387 | 15 | 3 | 3 | 3 |
| 6 | 1.4733 | 0.1339 | 1.0643 | 28 | 1 | 3 | 3 |
| 7 | 1.7044 | 0.1592 | 0.9976 | 4 | 0 | 2 | 2 |
| 8 | 1.5603 | 0.1434 | 0.8642 | 1 | 3 | 3 | 3 |
| 9 | 1.7044 | 0.1597 | 0.9659 | 6 | 2 | 2 | 2 |
| 10 | 1.4733 | 0.1339 | 1.0069 | 6 | 2 | 3 | 3 |

In Table 6, $\hat{\alpha}_i$, $\hat{\beta}_i$ are estimates of parameters given by

$$\hat{\alpha}_i = \bar{x}_i/(s_i^2 - \bar{x}_i)$$

$$\hat{\beta}_i = \bar{x}_i^2/(s_i^2 - \bar{x}_i) \qquad (4)$$

where $\bar{x}_i$ and $s_i^2$ are the sample mean and variance of the number of before accidents in the population of entities of which the treated entity i is a part. $(\varepsilon'/\varepsilon)_i$ is the ratio of exposures of the after to the before period for entity i. $x_i$ is the number of accidents (of a certain type) occurring on entity i (i = 1, 2, ..., n) during a before period that for this entity is $B_i$ years long. $x'_i$ is the number of accidents (of the same type) occurring on entity i during an after period which, for this entity, is $A_i$ years long.

The likelihood function for this case can be written as

$$L(\theta) = \prod_{i=1}^{n} \theta^{x'_i} [B_i + \alpha_i + (\varepsilon'/\varepsilon)_i A_i \theta]^{-(x_i+\beta+x'_i)} \qquad (5)$$

The variable θ serves here as the index of safety effect. If a measure reduces the expected number of accidents to, for example, 90 percent of its previous value, θ = 0.90. If it causes an increase of 5 percent, θ = 1.05. (The detailed derivation of Equation 5 may be found in study by Hauer et al. (4).

Using the entries in Table 6, the likelihood function (Equation 5) takes on the form

$$\begin{aligned} L(\theta) = \; & \theta^{6} [4.5603 + (3 \times 1.2237\theta)]^{-20.14} \\ & \theta^{3} [4.6187 + (3 \times 1.0657\theta)]^{-19.15} \\ & \theta^{9} [4.5603 + (3 \times 1.0189\theta)]^{-27.14} \\ & \theta^{7} [4.5603 + (3 \times 1.0549\theta)]^{-35.14} \\ & \theta^{3} [4.5603 + (3 \times 1.1387\theta)]^{-18.14} \\ & \theta^{1} [4.4733 + (3 \times 1.0643\theta)]^{-29.13} \\ & \theta^{0} [3.7044 + (2 \times 0.9976\theta)]^{-4.16} \\ & \theta^{3} [4.5603 + (3 \times 0.8642\theta)]^{-4.14} \\ & \theta^{2} [3.7044 + (2 \times 0.9659\theta)]^{-8.16} \\ & \theta^{2} [4.4733 + (3 \times 1.0069\theta)]^{-8.13} \end{aligned} \qquad (6)$$

Each line in Equation 6 corresponds to one of the 10 sites and thus to one row of Table 6.

With the stage set the meaning and use of the likelihood function can be discussed. The likelihood function has two important properties: (a) it preserves, in a condensed form, the entire information content of the data, and (b) it makes the merging of information contained in separate data sets simple. Thus, for example, the first line in Equation 6 captures all that can be learned (about the safety effect on right-angle accidents of conversion from two- to four-way stop control) from what has been observed at Site 1 alone. The corresponding likelihood function is shown by curve A in Figure 2.

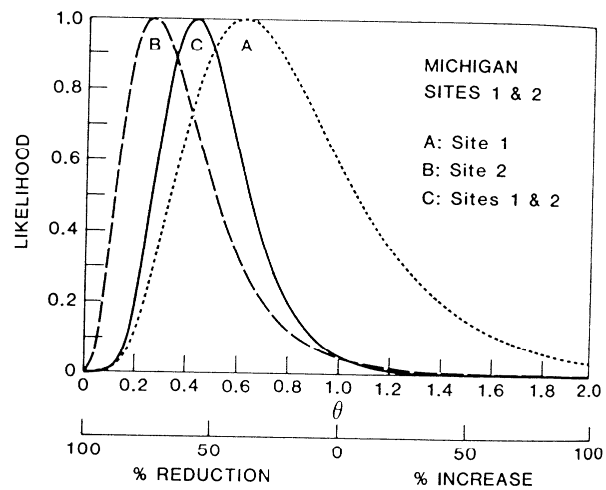


**FIGURE 2 Likelihood functions and their combination.**

The ordinate for a certain value of θ is proportional to the probability of recording 14 right-angle accidents during a 3-year before period and 6 right-angle accidents during a 3-year after period if that index of safety effect (θ) actually prevailed. The larger this probability, the more likely is the value of θ said to be. Values of θ for which the likelihood is small compared to its largest value (scaled to be equal to 1) are deemed unlikely.

The information contained in a reduction in number of accidents from 14 to 6 [when α, β, and (ε'/ε) are as in line 1 of Table 6] is meager. This is reflected in Figure 2 by the fact that curve A is quite flat near its peak, and a wide range of θs has likelihoods that are not much lower than 1. Although it is meager, whatever information the 14- to 6-accident reduction contains is now preserved. In a similar manner the second line in Equation 6 preserves all the information that can be extracted from the accident history of Site 2. The likelihood function for Site 2 is shown in Figure 2 by curve B.

How can the results from Sites 1 and 2 be combined? As indicated by Equation 5 (or 6), the ordinate of the joint likelihood function is proportional to the product of the two component ordinates. For computational convenience, the sum of the logarithms is used. The joint likelihood function for Sites 1 and 2 is shown by curve C in Figure 2. It represents all that can be learned from the data of Sites 1 and 2 taken together.

To complete the illustration, imagine one study encompassing Sites 1 to 4 and a later study encompassing Sites 5 to 10. The likelihood function for the first study is shown in Figure 3 by curve A. When, at some later time, data from Sites 5 to 10 become available (curve B), the two data sets can be combined to yield the joint likelihood function for all 10 sites (curve C).

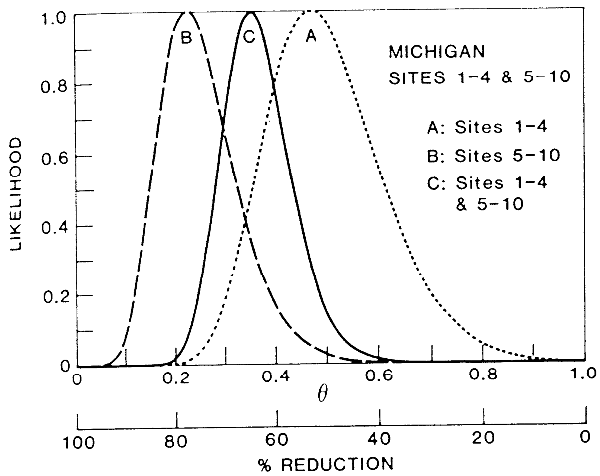


FIGURE 3 Likelihood functions for right-angle accidents at 10 Michigan intersections.

The reasons for choosing the likelihood function to represent, preserve, and accumulate information about the safety effect of a measure are now clear. The likelihood function (a) identifies the most likely value of θ and represents the uncertainty surrounding it in an intuitively clear fashion; (b) preserves in condensed form all that can be extracted from a data set; (c) represents a structured process for the accumulation of information and learning from experience. At any point in time it represents the current state of knowledge. When new data become available, the corresponding likelihood function is used to revise the existing data and to create a new (current) state of knowledge; and (d) facilitates the use of formal decision analysis and is an essential ingredient for making coherent decisions.

With all its merits, routine use of the likelihood function to combine information extracted from diverse data sets is not free of difficulties. The central question (presently unresolved) can be explained with reference to Figure 3. Is there some real difference between the group of Sites 1-4 (curve A) and the group of sites 5-10 (curve B), which is the reason why the same treatment (conversion to four-way stop control) may affect the safety of both groups differently?

If there is such a difference, the two likelihood functions should not be fused into curve C. Rather, an attempt should be made to describe the difference. Thus when conversion to four-way stop control for another intersection is contemplated, one will be in a position to assess whether curve A or curve B applies. If the treatment effect varies randomly from site to site, curves A and B should be fused into curve C. In this case, curve C properly represents the uncertainty surrounding the estimate of the average safety effect of the treatment. It is the role of further research to shed light on this important and difficult question.

## SUMMARY AND DISCUSSION

The authors have attempted to devise a methodology that facilitates the extraction of useful information from real-world instances of treatment implementation. Such instances are the predominant source of information about the safety effect of highway and traffic engineering measures. Therefore, the methodology devised here appears particularly suited for the creation of substantial knowledge in this field.

A simple comparison of before-and-after accident counts is shown to be incorrect even if corrections for exposure and secular trend are applied. Inasmuch as most of the traffic and highway engineering traditional knowledge about the effect of safety measures is based on such simple (and incorrect) before-and-after comparisons, a wholesale revision of this body of knowledge is in order.

Two smoothed estimates, $[\varepsilon_1(x)$ and $\varepsilon_2(x)]$, are recommended for use. To calculate their values, some additional data are required. The needed additional data are the count of before accidents on all similar entities.

What constitutes a similar entity? The answer that the analyst gives to this question influences the estimate and therefore introduces into the analysis an element of the arbitrary. Procedures that allow the analyst some freedom of choice tend to be viewed with suspicion. Two arguments can be raised in defense.

In practice the determination of what constitutes a sensible choice of the population of similar entities does not appear unduly difficult. The choice is seen to be severely circumscribed by what data can be obtained and by the interpretation of what can be described as a homogeneous population of entities. However, that in practice the choice is narrow, is only a weak defense against the charge that scientific methods should be devoid of the arbitrary. A stronger defense is that all known methods for the statistical interpretation of data require a similar measure of the arbitrary. Thus, for example, were it at all possible to match a control group (of entities left untreated) to the treated entities, a judgment would have to be made as to what entities are to be considered similar for the purposes of matching. This is precisely the judgment required to delineate a population of similar entities. If one is accepted as scientifically defensible, so must be the other.

The recommended procedure for before-and-after comparisons is an improvement on two counts. First, it is asymptotically unbiased and automatically eliminates regression-to-the-mean effects. Second, the accuracy of estimation is enhanced. However, it suffers from an ugly asymmetry. Although variance-reducing methods are devised for the utilization of before data, the use of after data remains primitive. Future research in this direction might lead to further improvements in estimation accuracy.

An attempt has been made to erect a methodological basis for extracting information from data and for the accumulation of such data. The likelihood function appears to be well suited for this purpose. The formal logic is sound, the application is straightforward, and the interpretation is relatively free from obfuscation.

In the methodological domain an intriguing conceptual question remains to be explored: When should likelihood functions derived from different data sets be combined? It has its practical translation: When are results obtained in city A applicable to city B? The commonly voiced contention: "but our conditions are different," which exerts a paralyzing effect on rational safety management, stems from the same source.

It has been shown that the chosen methodological framework worked well when the safety effect of conversion to all-way stop control was examined (Lovell and Hauer, and Persaud elsewhere in this Record). It is not surprising that a storehouse of empirical information on this issue exists. After all, there is more than half a century of application and use to rely on. It proved relatively simple to assemble the additional data that were required to do the proper analysis. The net result of this effort is a defensible current estimate of the safety effect of this measure.

Much of the traditional knowledge about the safety effect of highway and traffic engineering measures is based on simple before-and-after comparisons, and estimates based on these comparisons are now known to be incorrect. Furthermore, a method for analyzing before-and-after data has been presented in this paper, and experience indicates that past data can be used to set the record straight. It follows that a concerted effort to do so appears appropriate.

ACKNOWLEDGMENT

The work discussed in this paper was performed as part of a study sponsored jointly by Transport Canada and the U.S. Department of Transportation. The authors are grateful to both sponsors for their support and guidance.

REFERENCES

1. H. Robbins. An Empirical Bayes Approach to Statistics. Proc., Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California Press, Berkeley, 1956, pp. 157-163.
2. E. Hauer. On the Estimation of the Expected Number of Accidents. Accident Analysis and Prevention, Vol. 18, No. 1, 1986.
3. E. Hauer, J. Lovell, and B.N. Persaud. New Directions for Learning About Safety Effectiveness. FHWA/RD-86/015. FHWA, U.S. Department of Transportation, Jan. 1986.
4. E. Hauer, J. Lovell, and B.N. Persaud. The Safety Effect of Conversion from Two-Way to Four-Way Stop Control in Michigan. Publication 84-17. Department of Civil Engineering, University of Toronto, Canada, 1984.

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