11. P.A. Gimotty and T. Chirachavala. Injury Models for the Prediction of Severe Accidents. In Proceedings of the 26th Conference of the American Association of Automotive Medicine, 1982, pp. 63-76.
12. Y.M.M. Bishop, S.E. Fienberg, and P.W. Holland. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, Mass., 1975.
13. R.R. Hocking and R.N. Leslie. Selection of the Best Subset in Regression Analysis. Technometrics, Vol. 9, 1967, pp. 531-540.

Problem of Identifying Hazardous Locations Using Accident Data

EZRA HAUER and BHAGWANT N. PERSAUD

ABSTRACT

Most agencies with responsibility for extensive road systems use some variant of the rate-and-number method to identify hazardous locations or blackspots. Sites so identified are later examined in detail to diagnose deficiencies and to suggest remedial measures. In this paper the degree to which the rate-and-number method is successful in identifying the unidentified, and what proportion of the sites that are subjected to detailed examination are not deviant at all, is examined. The first part of the paper is devoted to the analysis and development of the mathematical machinery. In the second part the use of the analytical results is illustrated by application to two data sets--one dealing with highway ramps in Ontario and the other with California drivers. The main result of this research is the facility to examine the performance of various identification procedures on the basis of measures of performance that are easy to understand. Such an examination should lead to a realistic assessment of what can be attained when identification for treatment is made on the basis of past accident history.

In most agencies with jurisdiction over extensive road systems it is common practice to try and rectify so-called accident blackspots. Ordinarily a two-stage process is used. In the first stage the past accident history of all sites is reviewed to select a limited number of apparently dangerous locations for further examination. In the second stage the selected sites are studied in more detail, often in the field, in order to devise cost-effective remedial projects for some of the sites.

The two-stage process is required because detailed examination of all sites is impractical. It is hoped that the first stage of the process will act as a sieve. A good sieve is one that allows through all sites that do not require remedial action and retains all sites that do require detailed study. Conversely, an inefficient sieve is one that retains a large number of sites that do not need close scrutiny and allows most blackspots to pass through its holes and thus escape identification. The purpose of this paper is to examine the quality and performance of a commonly used sieve.

Most sieves in current use are a variant of the rate-and-number method. Sites that register an unusually high number of accidents during a specified period of time or an unusually high accident rate (accidents per vehicle kilometer) are selected for inspection. Accidents are often weighted according to their severity. The rationale for the rate-and-number method appears to be left unspecified in the literature. However, a plausible line of reasoning for its raison d'être goes as follows:

> If the accident history of a site is found to deviate from the norm for its class, there surely is some reason for it. If so, a responsible agency and its professionals should examine the cause for this deviation and, if a cost-effective remedy can be found, should remove the cause of substandard performance.

It should be evident that a sieve that screens sites on the basis of number of accidents, accidents per vehicle kilometer, or accident severity is aimed only at establishing deviancy. It is not an indication of how easy or how difficult remedial treatment might be.

Some causes of substandard performance are random and fleeting in nature and essentially unrelated to the physical characteristics of the site. (Consider, for example, a local snow squall or rainstorm that causes several accidents to occur within a few minutes.) Other causes for deviation from the norm are more permanent in nature (sharp curves, polished pavement, narrow bridge, and so forth). These are the causal factors that are subject to remedial action. Accordingly, the object of the exercise is to identify sites for which the deviation from the safety norm is attributable to some permanent properties of the site.

It is well-known that the actual number of accidents occurring on a site fluctuates from year to year. It is only the average number of accidents in the long run that can be linked to the permanent properties of the site. This gives rise to the fundamental difficulty facing the screening process.

> Researchers wish to identify those sites for which, say, the "average number of accidents in the long run" deviates from the norm. However, in the identification process, researchers are restricted to the use of accident histories that are subject to pronounced random fluctuation.

This inescapable difficulty affects the quality of all sieves. When the number of accidents occurring on a site in the last 2 or 3 years is higher than the average in the long run for that site, the site will be caught by the sieve and subjected to detailed inspection, possibly unnecessarily. Conversely, sites with permanent properties such that their average in the long run is considerably higher than the norm will often escape detection because of a random down-fluctuation.

Accordingly, several questions are raised: How good are the accident-history-based sieves for blackspot identification? Do they capture most of the truly deviant sites? How many normal sites are lumped with the deviant ones that are labeled blackspots? How many deviant sites escape detection?

These questions translate into the following figures of merit by which the quality of the sieve should be measured:

1. The number of sites selected for closer examination (this is a measure of the effort required at the later stage when site-specific deficiencies are identified, remedies are designed, and economics are examined);
2. The number of truly deviant sites among those selected for closer inspection (these are often called correct positives);
3. The number of sites that are not deviant yet have been captured by the sieve and selected for closer inspection (these are the false positives); and
4. The number of truly deviant sites that are not identified as requiring attention (these are called the false negatives).

The central issues are easy to visualize with the aid of a Venn diagram. Let the box in Figure 1 symbolize the collection of all sites. The set of all deviant sites is delimited by curve 1. Thus all nondeviant sites are outside curve 1. The set of all sites selected for closer inspection is enclosed by curve 2. An ideal sieve would be one for which curves 1 and 2 coincide. However, no real sieve or screening process is ideal. Therefore, curves 1 and 2 delineate three distinct sets. Set A contains sites that are deviant but are not selected for closer inspection. These are the false negatives. Set B contains deviant sites that are selected for inspection. These are the correct positives. Set C contains sites that are not deviant but are selected for inspection. These are the false positives.

The union of sets A and B is the collection of all deviant sites in the population. Curve 1 corresponds to a specific definition of what site is considered deviant. A more stringent definition of deviancy would be associated with a smaller ellipse. This will result in fewer deviant sites captured by the sieve (smaller set B) (i.e., the rarer the hunted animal, the more difficult is its capture).

The union of sets B and C is the collection of all sites selected for inspection. Curve 2 corresponds to a specific criterion by which sites are selected for further examination. A less-stringent selection criterion corresponds to the large circle. This results in a larger number of sites that require close inspection and also a larger number of deviant sites captured by the sieve.

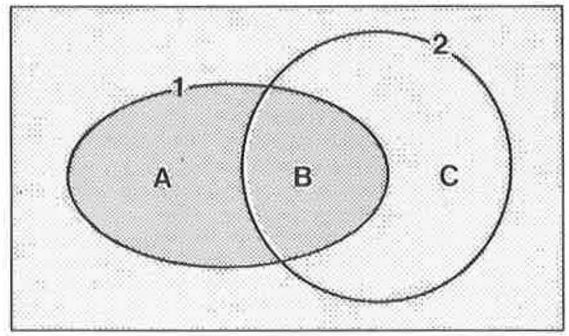

FIGURE 1 False negatives $\{A\}$, correct positives $\{B\}$, and false positives $\{C\}$.

In general, the more stringent the criterion of deviancy, the more difficult it is to identify deviant sites. The more stringent the selection criterion, the smaller the number of deviant sites captured by the sieve.

In present practice a site is considered to be a blackspot if its accident record deviates k standard deviations from the norm. The value of k is linked to statistical level of significance, and the practice in this case is borrowed from industrial quality control. What value to use for k is largely a matter of custom, with no apparent rationale. This is why it appears sensible to examine whether it is possible to discard what is arbitrary and use instead measures of performance that have clear meaning.

In this paper the focus is on blackspots that occur on a road system. This is why researchers speak of sites, road sections, ramps, intersections, and so forth. It is worth noting that identical issues arise when trying to identify deviant drivers, and that the results of analysis apply equally in both cases. To underscore this point, one example will deal with the population of drivers instead of the population of road sections.

MEASURES OF EFFICIENCY FOR A SIMPLE SIEVE

The simplest case is usually the easiest to analyze. Once the categories of thought and lines of argument for the simple case are established, the examination of more complex sieves can be undertaken.

A mathematical notation was not introduced in the first section because the central issues could be explained without burdening the reader with symbols. However, the main content of this section is analysis, and it would be inefficient to postpone the use of a precise notation any longer. Therefore, let λ be the expected (average in the long run) number of recorded accidents prevailing at a site during a specified period of time, and let x be the number of accidents actually recorded for that site and period of time.

In this section the performance of a sieve is examined, the aim of which is to identify sites for which λ is larger than some limiting value λ^*. This is done by selecting for inspection sites for which x is not less than some limiting value x^*.

For a specific site, λ is never known. What is known is x. Therefore, the question is: What can be said about the λ of a site if its x is known? The answer is best stated in terms of a conditional probability distribution. The corresponding symbol has to be added to the notational arsenal. Thus let $F(\lambda|x)$ be the probability that the expected number of accidents at a site was less than or equal to λ when the number of accidents actually recorded was x.

To provide the reader with a sense of direction, it is best to first show that $F(\lambda|x)$ is the kingpin on which everything hinges. Indeed, when $F(\lambda|x)$ is known, the performance of a sieve can be described with ease and precision. How to estimate $F(\lambda|x)$ will be described later.

The information on which analysis is based is the knowledge of x for each site. Let n(x) be the number of sites (N) that had x accidents, x = 0, 1, 2, ...

1. When sites for which $x \geq x^*$ are selected for inspection, the number of sites (S) to be inspected is

$$S(x^*) = \sum_{x^*}^{\infty} n(x) \quad (1)$$

This corresponds to the number of sites in the union of sets B and C in Figure 1.

2. When sites for which $\lambda > \lambda^*$ are considered deviant, the expected number of deviant sites (D) in the population is

$$D(\lambda^*) = \sum_{0}^{\infty} n(x)\,[1 - F(\lambda^*|x)] \quad (2)$$

This corresponds to the expected number of sites in the union of sets A and B in Figure 1.

3. With x* as the selection criterion and λ^* as the criterion for deviancy, the expected number of false positives (FP) is

$$FP(x^*, \lambda^*) = \sum_{x^*}^{\infty} n(x)\, F(\lambda^*|x) \quad (3)$$

This is the expected number of sites in set C of Figure 1.

4. Because S(x*) corresponds to the union of B and C, whereas FP(x*, λ*) corresponds to set C alone, it follows that the expected number of correct positives (CP) is

$$CP(x^*, \lambda^*) = S(x^*) - FP(x^*, \lambda^*) \quad (4)$$

This corresponds to the number of sites in set B.

5. Because D(λ*) corresponds to the union of A and B, whereas CP(x*, λ*) corresponds to set B alone, the expected number of false negatives (FN) is

$$FN(x^*, \lambda^*) = D(\lambda^*) - S(x^*) + FP(x^*, \lambda^*) \quad (5)$$

This corresponds to the number of sites in set A.

It follows that knowledge of n(x) and $F(\lambda|x)$ will enable researchers to find all figures of merit that describe the performance of the screening process. Because n(x) is obtained from the raw data, it remains to find $F(\lambda|x)$. This is the subject of the next section.

ESTIMATION OF SIEVE EFFICIENCY

Each site of a population of sites has associated with it an unknown value λ. Regarding λ as a continuous random variable within the population of sites, let $g(\lambda)$ denote its probability density function. Furthermore, let $P(x|\lambda)$ denote the probability of recording x accidents on a site where their expected number is λ. According to Bayes' theorem,

$$f(\lambda|x) \propto P(x|\lambda)g(\lambda) \quad (6)$$

Integration of $f(\lambda|x)$ yields $F(\lambda|x)$. The coefficient of proportionality is selected to make $\int_0^{\infty} F(\lambda|x)d\lambda = 1$.

It is assumed, as is common, that accident occurrence obeys the Poisson probability law. Thus

$$P(x|\lambda) = \lambda^x e^{-\lambda}/x! \quad (7)$$

The only missing link in Equation 6 is $g(\lambda)$. The clues for the estimation of $g(\lambda)$ are hidden in the numbers n(x). Because the number of accidents recorded on a site is a reflection of λ for that site (see Equation 7), the number of sites with x accidents [n(x)] must be a reflection of the distribution of λ among all sites. This is captured by the following relationship:

$$\text{Expected proportion of sites with x accidents} = E\left\{n(x)/[\Sigma n(x)]\right\}$$
$$= \int_0^{\infty} P(x|\lambda)g(\lambda)d\lambda \quad (8)$$

The problem here is to extract the function $g(\lambda)$ from Equation 8. It is a well-researched problem [see, for example, Maritz (1)]. In consequence, it is possible to make use of results obtained by others. One specific case that appears to be of practical interest when $g(\lambda)$ is a gamma probability density function will be described in detail. This assumption is common in actuarial literature [see, for example, Buhlmann (2) or Freifelder (3)] and is used to describe the distribution of expected claim frequencies for a population of insureds. The results to follow were obtained and used by Jarrett et al. (4) when estimating the magnitude of the regression to the mean in before-and-after comparisons.

When $g(\lambda)$ is a gamma probability density function and Equation 7 holds, the probability that a site selected at random has x accidents is given by the negative binomial probability law. Therefore, the parameters of $g(\lambda)$ can be estimated easily from the sample mean and sample variance of x as follows.

1. Calculate sample mean and variance (unless indicated otherwise, summation is over all values of x):

$$\bar{x} = \Sigma\, x\, n(x)/\Sigma n(x) \quad (9)$$

$$s^2 = [\Sigma (x - \bar{x})^2 n(x)]/\Sigma n(x) \quad (10)$$

2. Estimate parameters α and β and write $g(\lambda)$:

$$\hat{\alpha} = \bar{x}/(s^2 - \bar{x}) \quad (11)$$

$$\hat{\beta} = \bar{x}^2/(s^2 - \bar{x}) \quad (12)$$

With this,

$$g(\lambda) = \alpha^{\beta}\lambda^{\beta-1}e^{-\alpha\lambda}/\Gamma(\beta), \quad \text{when } \lambda \geq 0 \quad (13)$$

By using the results of Equations 7, 8, and 13,

$$f(\lambda|x) \propto \lambda^{x+\beta-1}e^{-\lambda(1+\alpha)} \quad (14)$$

which is also a gamma probability density function with

$$E\{\lambda|x\} = (x + \beta)/(1 + \alpha) \quad (15)$$

$$VAR\{\lambda|x\} = (x + \beta)/(1 + \alpha)^2 \quad (16)$$

It follows that $f(\lambda^*|x)$ is a gamma probability distribution function and estimates of its parameters α and β are known. $F(\lambda^*|x)$ may be found by using numerical integration on

$$\int_0^{\lambda^*} \lambda^{x+\beta-1} e^{-\lambda(1+\alpha)} d\lambda / \int_0^{\infty} \lambda^{x+\beta-1} e^{-\lambda(1+\alpha)} d\lambda \quad (17)$$

What remains to be done is to apply these results to some actual cases.

TWO ILLUSTRATIVE EXAMPLES

The theory developed so far suffices to describe the performance of a simple screening process. The numerical examples in the following sections will serve to show what may have been obscured by convoluted mathematical arguments.

Illustrative Example 1: Ontario Highway Ramps

The second column of Table 1 lists the number of Ontario highway ramps that, in 1978, had x = 0, 1, 2,...,14 accidents. The third column lists what should be expected if, indeed, the distribution of λ is as has been assumed in Equation 13. It appears that there is satisfactory support for making this assumption.

TABLE 1 Accidents on Ontario Highway Ramps in 1978

No. of Accidents (x)	No. of Ramps with x Accidents [n(x)]	No. of Ramps Expected by Negative Binomial Model
0	2,254	2,278
1	286	249
2	95	98
3	48	48
4	21	26
5	7	15
6	8	9
7	6	5
8	5	3
9	3	2
10	0	1
11	1	1
12	0	-
13	1	-
14	1	-

By using Equations 9-12, it is shown that $\bar{x}$ = 0.3414, s^2 = 1.0677, $\hat{\alpha}$ = 0.47, and $\hat{\beta}$ = 0.16. Therefore, by using Equation 13,

$$g(\lambda) = 0.152\, \lambda^{-0.84} e^{-0.47\lambda} \quad (18)$$

This is an estimate of how λ was distributed in the population of Ontario highway ramps in 1978 and is of considerable interest by itself.

In Figure 2 the probability distribution function (PDF) of λ, based on Equation 17, is shown. It appears that 10 percent of the ramps (276 ramps) have $\lambda > 1$ accidents per year, 5 percent of the ramps have $\lambda > 1.8$ accidents per year, and so forth. It is the ramps with relatively high values of λ that may demand closer examination and that the screening procedure should identify.

The probability distribution of λ in the subpopulation of sites that had x accidents can also be

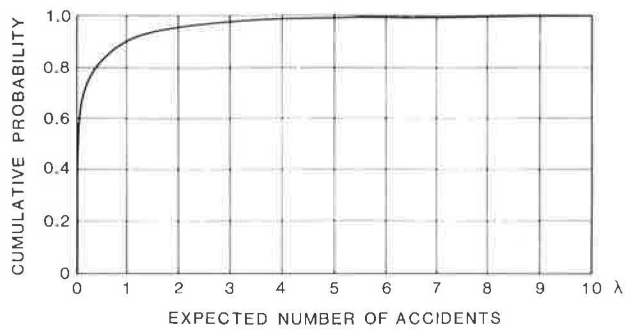

FIGURE 2 PDF of λ in the population of Ontario highway ramps (1978).

shown. This is accomplished by making use of Equations 14-16. In Figure 3, $F(\lambda = 1|x)$ for x = 1, 2, 3, 5, and 10 are shown.

Suppose that on the basis of Figure 2 researchers wish to identify those ramps for which $\lambda > 1$. There are some 276 such ramps. By using the terminology established earlier, $\lambda^* = 1$. This is shown by the vertical line in Figure 3.

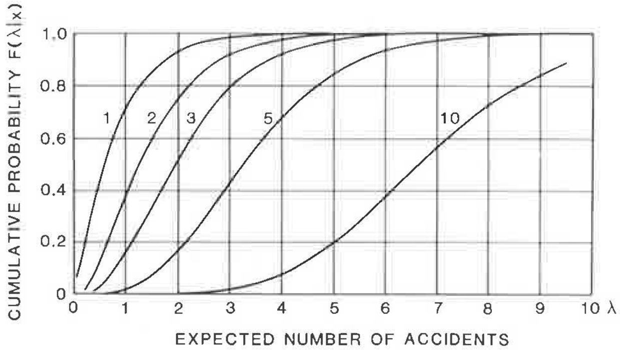

FIGURE 3 PDF of λ on ramps with 1, 2, 3, 5, and 10 accidents in 1978.

Consider now the 95 ramps that had two accidents (Table 1). From Figure 3, some 39 percent of those are expected to have $\lambda < 1$. Thus in this group of ramps it should be expected that 95 x 0.39 = 37 false positives and 95 - 37 = 58 correct positives. Similarly, the 48 ramps with three accidents each are expected to contain 48 x 0.16 = 8 false positives and 40 correct positives. Proceeding in this fashion, the data in Table 2 can be generated.

Columns 1 and 2 of Table 2 are the raw data copied directly from Table 1. The cumulation from below of the entries in column 2 yields S(x*) in column 3. Thus if sites with three or more accidents are selected for detailed scrutiny, 101 ramps have to be inspected.

The mathematical machinery assembled in the previous section and, in particular, Equations 14-16 facilitate the calculation of $F(\lambda^* = 1|x)$ in column 4. Because the computation is tedious, a FORTRAN computer code has been written for that purpose.

The products of entries in columns 2 and 4 are estimates of the number of false positives to be expected in the group of ramps that had x reported accidents, as explained earlier. Column 6 is the cumulation from below of the entries in column 5 and are therefore the estimates of the number of false posi-

TABLE 2 Measures of Performance for Ontario Highway Ramps with $\lambda^* = 1$

1	2	3	4	5	6	7	8
Recorded Number of Accidents	Number of Ramps With x Recorded Accidents	Number of Ramps With x* or More Recorded Accidents	Proportion of Sites With $\lambda<1$	Expected Number of False Positives	Cumulative Expected Number of False Positives	Expected Number of Correct Positives (3)-(6)	Expected Number of False Negatives CP(0,λ) - CP(x*,λ*)
x	n(x)	S(x*)	F(1|x)	(2)x(4)	FP(x*,1)	CP(x*,1)	FN(x*,1)
0	2254	2736	0.980	2209	2460	276	0
1	286	482	0.718	205	251	231	45
2	95	196	0.386	37	46	150	126
3	48	101	0.158	8	9	92	184
4	21	53	0.052	1	1	52	224
5	7	32	0.017	-	-	32	244
6	8	25	0.004	-	-	25	251
7	6	17	0.001	-	-	17	259
8	5	11	-	-	-	11	265
9	3	6	-	-	-	6	270
10	0	3	-	-	-	3	273
11	1	3	-	-	-	3	273
12	0	2	-	-	-	2	274
13	1	2	-	-	-	2	274
14	1	1	-	-	-	1	275

tives in the selected ramps. Thus if x* = 3, then in the 101 selected ramps it should be expected that there will be 9 ramps for which $\lambda < 1$. This makes the number of correct positives equal to 92, which is the entry in column 7.

The topmost entry in column 7 is the number of correct positives in the entire set of 276 ramps. It is, therefore, the expected number of deviant sites in the population: $D(\lambda^*) = 276$. The number of false negatives (those ramps not captured by the sieve) is calculated by subtracting from 276 the entry in column 7. Thus if x* = 3 in the group of 101 ramps selected for inspection, there are 92 ramps that have $\lambda > 1$, which leaves the remaining 276 - 92 = 184 deviant ramps undetected in the population.

For ease of visual representation, the main results from Table 2 are shown in Figure 4. Thus 84 percent of the deviant sites can be captured with x* = 1. But this means that more nondeviant than deviant sites are selected for close inspection and the inspection effort is large. With x* = 2, the inspection effort and the number of false positives are reduced. However, almost half of the deviant ramps remain undetected. This illustrates the main tradeoffs and also describes the power and limitations of this screening process. With a small x*, the majority of deviant sites can be identified at the cost of having to examine a large number of them in the field. Included in the selected sites will be many that are not deviant, and their inspection may be a waste of time. With a large x*, the number of sites to be inspected can be reduced and it can also be ensured that almost all inspected sites are deviant. In this case, however, many deviant sites will not be selected for inspection.

Thus if λ^* is given as a criterion of deviancy, the analyst can trade the cost of field inspection against the penalty of leaving a deviant site undetected.

The last issue to explore is the effect of deciding on what is to be considered deviant by the choice λ^*.

There are, on average, 0.34 accident per ramp. Setting $\lambda^* = 1$, as in Table 2, defines as deviant ramps for which the expected number of accidents is about 3 times the population average. Had $\lambda^* = 1.5$ been chosen (Table 3), the number of deviant ramps is, of course, much smaller (176). Because the object of the search is now rarer, it is more difficult to capture. Thus, although for $\lambda^* = 1$, x = 3, [100(48 - 8)/48] = 84 percent of the 48 sites with x = 3 were correct positives, for $\lambda^* = 1.5$ the

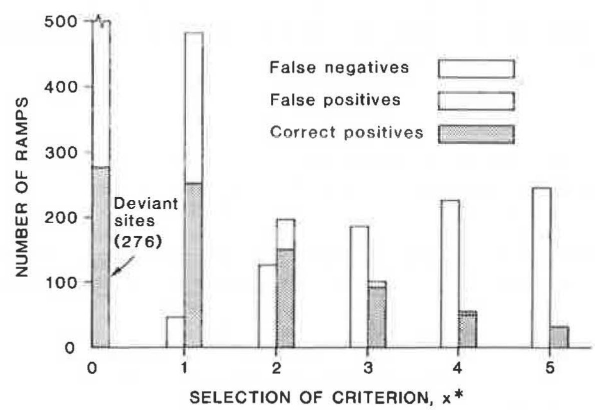

FIGURE 4 Measures of performance for $\lambda^* = 1$.

TABLE 3 Measures of Performance for Ontario Highway Ramps with $\lambda^* = 1.5$

Recorded Number of Accidents	Number of Ramps With x Recorded Accidents	Number of Ramps With x* or More Recorded Accidents	Proportion of Ramps with $\lambda<1$	Expected Number of False Positives	Cumulative Expected Number of False Positives	Expected Number of Correct Positives (3)-(6)	Expected Number of False Negatives CP(0,λ)-CP(x*,λ*)
x	n(x)	S(x*)	F(1.5|x)	(2)x(4)	FP(x*,1.5)	CP(x*,1.5)	FN(x*,1.5)
0	2254	2736	0.993	2238	2560	176	0
1	286	482	0.859	246	322	160	16
2	95	196	0.604	57	76	120	56
3	48	101	0.343	16	19	82	94
4	21	53	0.160	3	3	50	126
5	7	32	0.070	-	-	32	144
6	8	25	0.024	-	-	25	151
7	6	17	0.007	-	-	17	159
8	5	11	0.002	-	-	11	165
9	3	6	0.001	-	-	6	170
10	0	3	-	-	-	3	173
11	1	3	-	-	-	3	173
12	0	2	-	-	-	2	174
13	1	2	-	-	-	2	174
14	1	1	-	-	-	1	175

same 84 percent yield is reached only for a larger x = 4.

The variation in the measures of performance of this screening process in dependence on λ^* is shown in Figure 5.

Illustrative Example 2: California Drivers

The records of 86,726 California drivers have been examined, and the number of reported accidents during 1961 have been noted (5). The number of drivers with 0, 1, 2, or 3 accidents is given in column 2 of Table 4. Column 3 gives the number of drivers in each category if the distribution of λ is as in Equation 13. This assumption is well supported. From this,

$\bar{x} = 0.08839$, $s = 0.0939$, $\hat{\alpha} = 16.092$, $\hat{\beta} = 1.422$, and $g(\lambda) = 58.7\,\lambda^{0.4224}\,e^{-16.1\lambda}$

On this basis, the data in Tables 5 and 6 are constructed, as in the previous numerical example. In Table 5, $\lambda^* = 0.25$, which is about 3 times the average number of accidents per driver. The difficulties of identifying deviant drivers are obvious. With $x^* = 3$, only half of those identified are deviant, yet the overwhelming majority of the 3,425

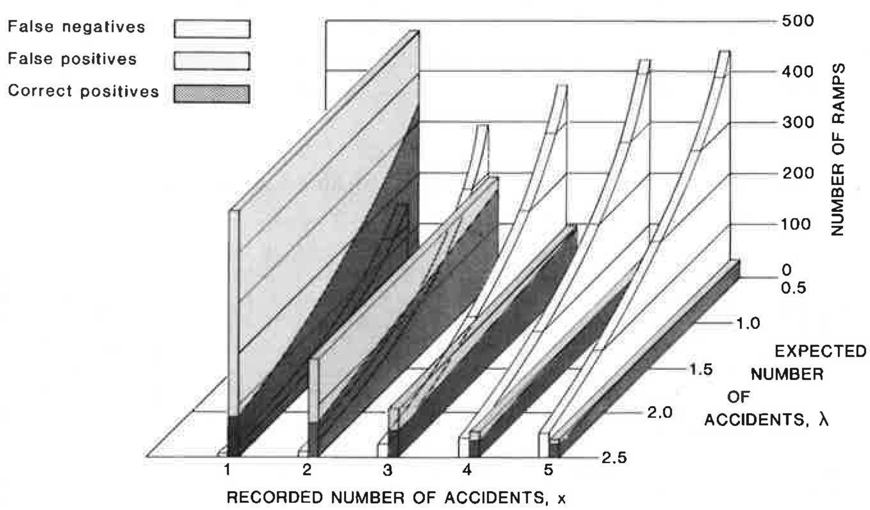

FIGURE 5 Measures of performance as a function of λ^*.

TABLE 4 Accidents to California Drivers in 1961

No. of Accidents (x)	No. of Drivers [n(x)]	No. of Drivers Predicted by Negative Binomial Model
0	79,595	79,598
1	6,638	6,624
2	451	469
3	42	31

drivers remain unidentified. It does not help much to select drivers with $x^* = 1$ because a large majority are false positives (6,205). The performance of the sieve is even worse when a more severe criterion of deviancy is considered ($\lambda^* = 0.50$ in Table 6).

SUMMARY, DISCUSSION, AND FUTURE RESEARCH

Normally, a two-stage process is used for the identification of blackspots. In the first stage a limited number of apparently dangerous locations are selected from all sites on the basis of their accident history. The sites so selected are examined in more detail in the second stage.

The data in this paper deal with the first stage of the blackspot identification process, which is likened to a sieve. A good sieve retains most sites that require detailed examination and allows through most sites that need not be examined any further.

Accordingly, a concept of sieve efficiency is proposed in which the number of sites to be inspected and the expected numbers of correct positives, false positives, and false negatives serve as measures of performance.

This concept is converted into a procedure for a special but common case, and it is applied to two illustrative examples. One deals with the population of Ontario highway ramps, and the other deals with California drivers.

In both cases the objective of the screening process is to identify units for which the expected number of accidents exceeds a given norm. What can and cannot be achieved is illustrated. Because the measures of performance are explicit, rationality in decision making and design are facilitated.

The screening process used in practice is more complex than what has been analyzed. In particular, the accident rate (accidents per vehicle kilometer), which is the most important selection criterion, is not used here. Thus the theory and computational process need to be extended so as to be applicable to the realistic blackspot identification procedures. This extension appears to be straightforward. The corresponding research work is under way.

The procedure relies on the assumption that λ obeys the gamma distribution. This may not be a good assumption in some cases. Accordingly, it is necessary to develop numerical methods to free the procedure from reliance on this assumption.

Therefore, the quality-control approach to blackspot identification does not give the analyst clues about how well or how poorly his sieve is working. In contrast, the approach suggested in this paper provides measures of performance that describe the efficiency of the sieve in intuitively clear terms.

ACKNOWLEDGMENT

The support of research by Transport Canada is gratefully appreciated.

REFERENCES

1. J.S. Maritz. Empirical Bayes Methods. Methuen and Company, Ltd., London, England, 1970.

TABLE 5 Measures of Performance for California Drivers with $\lambda^* = 0.25$ Accidents per Year

Recorded Number of Accidents	Number of Drivers With x Recorded Accidents	Number of Drivers With x* or More Recorded Accidents	Proportion Of Drivers with λ<1	Expected Number of False Positives	Cumulative Expected Number of False Positives	Expected Number of Correct Positives (3)-(6)	Expected Number of False Negatives CP(0,λ) - CP(x*,λ*)
x	n(x)	S(x*)	F(0.25 x)	(2) x (4)	FP(x*,0.25)	CP(x*,0.25)	FN(x*,0.25)
0	79595	86726	0.969	77096	83301	3425	0
1	6638	7131	0.882	5854	6205	926	2499
2	451	493	0.729	329	351	142	3284
3	42	42	0.537	23	23	19	3406

TABLE 6 Measures of Performance for California Drivers with $\lambda^* = 0.50$ Accidents per Year

Recorded Number of Accidents	Number of Drivers With x Recorded Accidents	Number of Drivers With x* or More Recorded Accidents	Proportions of Drivers With λ<1	Expected Number of False Positives	Cumulative Expected Number of False Positives	Expected Number of Correct Positives (3)-(6)	Expected Number of False Negatives CP(0,λ) - CP(x*,λ*)
x	n(x)	S(x*)	F(0.5|x)	(2) x (4)	FP(x*, 0.5)	CP(x*,0.5)	FN(x*,0.5)
0	79595	86726	0.999	79550	86648	79	0
1	6638	7131	0.996	6613	7097	34	45
2	451	493	0.985	444	484	9	70
3	42	42	0.956	40	40	2	77

2. H. Bühlmann. Mathematical Methods in Risk Theory. Springer Verlag, Berlin, 1970.
3. L.R. Freifelder. A Decision Theoretic Approach to Insurance Ratemaking. Irwin, Homewood, Ill., 1976.
4. D.F. Jarrett, C. Abbess, and C.C. Wright. Bayesian Methods Applied to Road Accident Blackspot Studies: Some Recent Progress. In Proc., Seminar on Short-Term and Areawide Evaluation of Safety Measures, Institute for Road Safety Research, SWOV, Amsterdam, Netherlands, 1982, pp. 69-74.
5. R.C. Peck, R.S. McBride, and R.S. Coppin. The Distribution and Prediction of Driver Accident Frequencies. Accident Analysis and Prevention, Vol. 2, No. 4, 1971, pp. 243-299.

Comparison of Two Methods for Debiasing Before-and-After Accident Studies

BHAGWANT N. PERSAUD and EZRA HAUER

ABSTRACT

When corrective treatment is applied to road sections, intersections, drivers, or vehicles that had a poor accident record in the past, the safety effect of the treatment is properly estimated by comparing the number of accidents in a post-treatment period with the number of accidents that would have occurred in this period without the treatment. Earlier papers have shown that simple before-and-after comparisons are consistently biased; that is, treatments appear to be more effective than they really are. Accordingly, two methods--a nonparametric method and a Bayesian method--have been separately proposed for purging this bias. The nature of the bias and the two debiasing methods are reviewed. In the main body of the paper several data sets are used to compare the performance of the methods. In most cases the Bayesian method was found to yield better estimates.

Before-and-after accident comparisons are a common method for assessing the safety effect of a treatment applied to road sections, intersections, drivers, and so forth. Conclusive evidence exists to show that when treatment is administered to systems with a poor safety record, simple before-and-after comparisons are biased (1). The bias is caused by the erroneous assumption that the number of accidents on a system in the period before treatment is an unbiased estimate of what should be expected to occur on the system during an equivalent after period had treatment not been applied. Systems with above-average accident numbers or rates in one period must be expected to show a decrease in a subsequent period even without treatment, and vice versa. This phenomenon, identified as regression-to-the mean, was demonstrated to be significant and can, in simple before-and-after comparisons, make safety treatments appear to be more effective than they really are.

To illustrate, Table 1, taken from Hauer (2), presents accident data for 20,762 1-km road sections in Ontario. Sections were grouped according to the number of accidents in 1 year. As shown by the data in the table, 12,859 sections had no accidents in that year; 4,457 had one accident, and so forth. Column 3 shows that, for each group, the average number of accidents recorded in the subsequent year revealed a reduction in the number of accidents in the second year for each group of sections with accidents in the first year. These reductions are balanced by the 12,859 sections that had no accidents in the first year but experienced an increase to 0.404 accident per section in the second year.

This is the essence of the regression-to-the-mean. When a random down-fluctuation occurs, as for the group with no accidents, an upward return to the mean for that group should be expected; when a random up-fluctuation occurs as it does for all the other groups, a downward return to the group mean should be expected.

Although there has been an increasing awareness of the phenomenon, its effect has often been dismissed because it will rarely be statistically sig-

TABLE 1 Regression-to-the-Mean: Ontario Data *(2)*

No. of Sections in Group	No. of Accidents for Avg Section in Group: First Year	Second Year	Change (%)
12,859	0	0.404	–[a]
4,457	1	0.832	-16.8
1,884	2	1.301	-35.0
791	3	1.841	-38.6
374	4	2.361	-41.0
160	5	3.206	-35.9
95	6	3.695	-38.4
62	7	4.968	-29.0
33	8	4.818	-39.8
14	9	6.930	-23.0
33	≥ 10[b]	10.390	-22.0

[a]Increase. [b]Average = 13.33.