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.

nificant, and it is not often likely to lead to serious results (according to 1965 data from the Road Research Laboratory). Column 4 in Table 1 converts these regressions to percent changes, which show that, contrary to this opinion, the phenomenon is consistent, real, and nothing short of dramatic. Hauer (2) showed that not only road sections are subject to this phenomenon. In fact, any element of the transport system for which events occur randomly will be subject to regression-to-the-mean.

In summary, the number of accidents on a system in the before period does not, on average, remain the same in an equivalent after period. When safety treatment is applied, an estimate of the number of accidents that would have occurred in a subsequent period without the treatment needs to be made. In the next section, procedures for doing so are reviewed.

REVIEW OF METHODS FOR DEBIASING BEFORE-AND-AFTER COMPARISONS

Establishment of control groups, where possible, is perhaps the best method for obtaining estimates of what the number of after period accidents would have been without treatment. When doing so is not practical, two analytic methods are available. Details of these methods are given elsewhere (1-4).

Method 1

The nonparametric (NP) method (1,2) is simple to apply yet is based on intricate statistical reasoning. To estimate the number of accidents α_k expected to occur during an equivalent after period on a system that had k accidents in the before period, the following factors need to be known: N_k = number of systems with k accidents in the population of similar systems, and N_{k+1} = number of systems with (k + 1) accidents in the population of similar systems. Then,

$$\alpha_k = [(k+1)\,N_{k+1}]/N_k \tag{1}$$

The simple formula relies on the sole assumption that accidents on any system are Poisson distributed. Unlike the alternative method, no assumptions are made about the underlying distribution of accidents in the population of systems. To illustrate the use of Equation 1, α_3 was estimated for the Ontario data in Table 1. Here N_3 = 791 and N_4 = 374 from column 2. Thus, based on first-year data, the estimate of the number of accidents in the second year on a section that in the first year had three accidents is given by

$$\alpha_3 = (4 \times 374)/791 = 1.891.$$

This compares to 1.841 actually observed in that year.

If there is interest in estimating total accidents for cumulative groups with k or more accidents, it is not necessary to apply the nonparametric method individually for each accident group. It can be shown [see Hauer (2) for proof] that

$$A_k = N_{k+1}^{(+)} \tag{2}$$

where A_k is the estimated total number of accidents on systems that in the before period had k or more accidents, and $N_{k+1}^{(+)}$ is the total number of accidents on those systems that in the before period had (k + 1) or more accidents.

Method 2

The empirical Bayesian (EB) method (3,4) is just as simple to apply as method 1, but it is based on stronger assumptions and requires accident data for the entire population of systems. As before, it is assumed that the number of accidents for a system obeys the Poisson with a mean characteristic of the system. Furthermore, it is assumed that the distribution of these means in a population of systems can be approximated by a gamma distribution. With these two assumptions, the number of systems of a population with k accidents must obey the negative binomial distribution except for a rare situation discussed later in this section.

The expected number of accidents α_k' in the after period on a system that had k accidents in the before period is given by

$$\alpha_k' = [(k+1)\,N_{k+1}']/N_k' \tag{3}$$

N_k' is the number of systems expected by the negative binomial distribution to have k accidents. (Note the similarity between Equation 3 and Equation 1 for the nonparametric method, and recall that in Equation 1, N_k was the actual number of systems with k accidents.)

To employ this method, the before period accident data are used to get the sample mean (m) number of accidents and sample variance (s^2) for the population of systems. From these, estimates of the parameters b, c of the gamma distribution can be obtained, as follows:

$$b = m^2/(s^2 - m) \qquad m < s^2 \tag{4}$$

$$c = m/(s^2 - m) \qquad m < s^2 \tag{5}$$

As shown by Jarrett et al. (3) and by Abbess et al. (4) Equation 3 then reduces to

$$\alpha_k' = (b+k)/(c+1) \qquad m < s^2 \tag{6}$$

It should be noted that if the negative binomial distribution were to fit all of observed frequencies perfectly, then the two methods would give identical estimates.

For the rare situations when the sample mean is not less than the sample variance ($m \geq s^2$), Equations 4, 5, and 6 do not apply. Instead, the distribution of means in the population of systems approximates the limiting form of the gamma distribution, where each system has the same expected number of accidents. Therefore, instead of Equation 6,

$$\alpha_k' = m \qquad \text{for } m \geqslant s^2 \tag{7}$$

To illustrate the more common case, suppose again that k = 3 for the Ontario data. The sample mean of the number of accidents in the first year is m = 0.707 and the sample variance is s^2 = 1.6491. From Equations 4 and 5 the following estimates are obtained: $\hat{b}$ = 0.5345, and $\hat{c}$ = 0.7540. Therefore,

$$\alpha_3' = 3.5345/1.7844 = 2.015.$$

This also compares favorably with the observed 1.841 (Table 1). A systematic comparison of the performance of both methods is the subject matter of the next section.

In using the EB method to estimate total accidents A_k' for cumulative groups, the equivalent expression to Equation 2 is

$$A_k' = N_{k+1}'^{(+)} \tag{8}$$

where $N_{k+1}^{'(+)}$ is the total number of systems expected by the negative binomial distribution to have (k + 1) or more accidents. Recall that to get α_k in Equation 6, it was not necessary to get the $N_k^{'}$'s, so it may not always be convenient to apply this shortcut with the Bayesian method.

COMPARISON OF THE TWO METHODS

Given the differences between the two methods, it is of interest to compare estimates obtained by each method against what was actually recorded to see if there are circumstances in which one or the other should be preferred.

Data

Eleven primary data sets were used in this comparison. Some of the data sets contain several years of accident history, so it was possible to effectively increase the number of comparisons by varying the before and after periods. In addition, one of the driver accident data sets was disaggregated into five age groups. Thus the comparisons were done for a total of 42 data sets that involved a variety of systems (driver accidents, driver violations, road sections, intersections, and roundabouts), and that covered a variety of countries (the United States, Canada, Sweden, Israel, and the United Kingdom) and a variety of before period lengths. A total of 293 comparisons were obtained. These data sets are identified in Table 2.

Analysis and Results

To illustrate the nature of the performance comparisons, the data in Table 3 present the results for the Ontario data set. Columns 1, 2, and 6 merely repeat the data in Table 1. As shown in the first line, by using the negative binomial distribution, it would be estimated that 13,222 sections (column 3) are expected to have 0 accidents (column 1) compared with the 12,859 sections (column 2) actually counted. The nonparametric method estimates that one such section chosen at random would average 0.35 accident (column 4) during the second year, whereas by the Bayesian method the estimate is 0.31 accident (column 5). These estimates are compared to the 0.40 accident per section actually recorded in the second year. In Figure 1, 0.35 on the ordinate plotted against 0.40 on the abscissa is point A, which is designated by an empty circle; 0.31 plotted against 0.40 is point B shown by a full circle. Thus data in Table 3 yield 10 pairs of circles.

Similar tables for all of the data sets produced the data for Figures 1 and 2, where estimates from each of the two methods are plotted against what was recorded. For clarity and for reasons discussed later, the driver and the road data sets are plotted separately. In these figures the empty circles represent the nonparametric estimates, whereas the full circles plot the Bayesian estimates. Some observations follow.

For both driver and road systems, the full circles tend to hug the diagonal somewhat closer than the empty circles. Thus it is concluded that the Bayesian method is likely to give somewhat better estimates.

For the drivers (Figure 2), the nonparametric method consistently overestimates the number of accidents (or violations) per driver from about 0.2 accident per year on. In an earlier paper (1), it was speculated that this is a reflection of maturation and possibly the effect of accidents or convic-

TABLE 2 Comparison of Parametric and Nonparametric Bayesian Estimates

N - indicates the non-parametric method is better; P - indicates the parametric method is better

	χ^2		Mean		Variance		No. of Accidents										
DATASET DESCRIPTION	Before	After	Before	After	Before	After	0	1	2	3	4	5	6	7	8	9	10
N. Carolina Driver Accidents (Years 1,2 Before)	156	304	.122	.130	.143	.151	N	N	P	P	P	P	P				
N. Carolina Driver Accidents (Years 3,4 Before)	304	156	.130	.122	.151	.143	NP	N	P	P	P	P	P				
Ontario Road Sections	100	-	.707	-	1.649	-	N	N	N	N	NP	P	P	P	P	N	P
Sweden Road Junctions "Reported"	38	-	.833	-	1.946	-	N	P	N	N	P	P	P				
Sweden Road Junctions "Personal Injury"	2	-	.197	-	.393	-	P	P	P	N							
Driver Violations-North Carolina (Years 1,2 Before) (3,4 After)	550	664	.225	.252	.367	.401	N	N	N	P	P	P	P				
Driver Violations-North Carolina (Years 3,4 Before) (1,2 After)	664	550	.252	.225	.401	.367	N	N	N	P	P	P	P	P	P		
U.K. Roundabouts	5	-	3.911	-	17.245	-	N	P	P	P	P	P	P	P	P		
New Mexico Run Off Road (80,81 Before)(82 After)	55	67	.773	.383	1.747	.659	N	N	N	P	P	P	P	P	N	P	P
New Mexico Run Off Road (82 Before) (80,81, After)	67	55	.383	.773	.659	1.747	N	N	N	P	P	P	P	P	P		
New Mexico Fixed Object (80,81 Before)(82 After)	28	9	.263	.129	.459	.184	N	N	P	P	P	P	P	P	P		
New Mexico Fixed Object (82 Before)(80,81 After)	9	28	.129	.263	.184	.459	N	N	N	N	N	P	P				
North Carolina Driver Accidents (Yrs.1,2,3 Before)(Yr. 4 After)	473	74	.187	.064	.231	.070	N	N	P	P	P	P	N				
North Carolina Driver Accidents (Year 4 Before)(Year 1,2,3 Aft.)	74	473	.064	.187	.070	.231	NP	N	P	P	P	P					

TABLE 2 (continued)

N - indicates the non-parametric method is better; P - indicates the parametric method is better

DATASET DESCRIPTION	X^2 Before	X^2 After	Mean Before	Mean After	Variance Before	Variance After	No. of Accidents 0	1	2	3	4	5	6	7	8	9	10
North Carolina Driver Accidents (Year 1 Before)(Yrs.2,3,4 After)	21	1038	.061	.191	.066	.235	NP	N	P	P	P	P					
North Carolina Driver Accidents (Yrs.2,3,4 Before)(Yr.1 After)	1038	21	.191	.061	.235	.066	P	N	NP	P	P	P	P				
North Carolina, 22-25 year olds (Yrs. 1,2 Before)(Yrs.3,4 After)	3	26	.160	.170	.191	.203	NP	NP	N	N	P	P	P				
North Carolina, 22-25 year olds (Yrs.3,4 Before)(Yrs. 1,2 After)	26	3	.170	.160	.203	.191	N	N	N	P	P	P	P				
North Carolina, 26-39 year olds (Yrs.1,2 Before)(Yrs.3,4 After)	160	89	.127	.134	.151	.158	N	N	P	P	P	P	P				
North Carolina, 26-39 year olds (Yrs.3,4 Before)(Yrs.1,2 After)	89	160	.134	.127	.158	.151	P	N	P	P	P	P	N	P			
North Carolina, 40-59 year olds (Yrs.1,2 Before)(Yrs.3,4 After)	68	91	.107	.114	.122	.129	N	N	P	N	N	P	P	P	P		
North Carolina 40-59 year olds (Yrs.3,4 Before)(Yrs.1,2 After)	91	68	.114	.107	.129	.122	P	N	P	P	P	P	P	P			
North Carolina 60+years (Yrs.1,2 Before)(Yrs.3,4 After)	17	115	.111	.114	.125	.131	N	N	P	N	P	P	P	P			
North Carolina 60+years (Yrs.3,4 Before)(Yrs.1,2 After)	115	17	.114	.111	.131	.125	N	N	P	P	P	P	P	P			
North Carolina 21 year olds (year 1 Before)(Yrs.2,4 After)	1	1	.106	.098	.118	.108	NP	N	P	P	N						
North Carolina 21 year olds (Year 2 Before)(Year 1 After)	1	1	.098	.106	.108	.118	NP	P	N	P	N						
Israeli Road Sections (Yrs.2,3,4 Before)(Yrs.5,6,7Aft)	3	16	1.685	1.909	4.379	4.095	N	P	P	P	P	P	P				
Israeli Road Sections (Yrs.5,6,7 Bef.)(Yrs.2,3,4 Aft.)	17	3	1.909	1.685	4.095	4.379	N	P	N	P	P	P	P				

N - indicates the non-parametric method is better; P - indicates the parametric method is better

DATASET DESCRIPTION	X^2 "Before"	X^2 "After"	Mean Before	Mean After	Variance 'Before'	Variance 'After'	No. of Accidents 0	1	2	3	4	5	6	7	8	9	10
Israeli Road Sections (Year 1 Before)(Year 2 After)	1	3	.577	.589	.972	.943	P	P	N	P							
Israeli Road Sections (Year 2 Before)(Year 1 After)	3	1	.589	.577	.943	.972	P	P	P	N							
Israeli Road Sections (Year 6 Before)(Year 7 After)	1	2	.661	.705	1.01	1.03	N	N	P	N							
Israeli Road Sections (Year 7 Before)(Year 6 After)	2	1	.705	.661	1.03	1.01	N	P	N	P							
Israeli Road Sections (Yrs.1,2,3 Before)(Yrs.4,5 After)	7	9	1.710	1.799	3.675	3.88	N	P	P	N	P	N	P	P			
Israeli Road Sections (Yrs.4,5,6 Before)(Yrs.1,2,3Aft)	9	7	1.799	1.710	3.88	3.675	N	P	N	P	N	P	N				
Israeli Road Sections (yrs.1,2 Bef.)(Yrs.3,4 After)	4	2	1.124	1.169	2.245	2.30	P	P	P	N	P	P					
Israeli Road Sections (Yrs.3,4 Bef.)(Yrs.1,2 After)	2	4	1.169	1.124	2.30	2.245	N	P	N	N	P	N					
Israeli Road Sections (Yrs.3,4 Before)(Yrs.5,6 Aft.)	2	8	1.127	1.220	2.295	2.233	N	P	P	N	P	P					
Israeli Road Sections (Yrs.5,6 Bef.)(Yrs.3,4 After)	8	2	1.220	1.127	2.233	2.295	N	N	P	P	P	P					
Westminister Blacksites	60	-	3.223	-	19.29	-	not reported	N	P	P	P	P	P	P	P		
California Driver Accidents (72,73 Before)(74 After)	9	2	.133	.048	.149	.051	P	N	P	P	P	P					
California Driver Accidents (74 Before)(72,73 After)	2	9	.048	.133	.051	.149	NP	P	N	P							
Philadelphia Intersections (68 Before)(69 After)	3	4	.759	.797	.970	.893	N	P	N	P	P						
Philadelphia Intersections (69 Before)(68 Before)	4	3	.797	.759	.893	.970	P	P	P	P	P						

Note: N indicates that the nonparametric method is better, and P indicates that the parametric method is better. The North Carolina driver data sets are from Stewart and Campbell *(5)*, the United Kingdom roundabout data sets are from Helliar-Symons *(6)*, and the Sweden intersection data sets are from Brude and Larsson *(7)*.

TABLE 3 Application of Estimating Methods to Ontario Road Sections Data

No. of Accidents (k)	No. of Sections with k Accidents	No. of Sections Estimated by Negative Binomial Distribution to Have k Accidents	Nonparametric Estimate	Bayesian Estimate	Recorded After
0	12,859	13,222	0.35	0.31	0.40
1	4,457	4,029	0.84	0.87	0.83
2	1,884	1,762	1.26	1.44	1.30
3	791	850	1.89	1.99	1.84
4	374	428	2.14	2.59	2.36
5	160	221	3.56	3.16	3.20
6	95	116	4.57	3.73	3.69
7	62	62	4.26	4.30	4.96
8	33	33	3.82	4.87	4.81
9	14	8	5.71	5.44	6.93

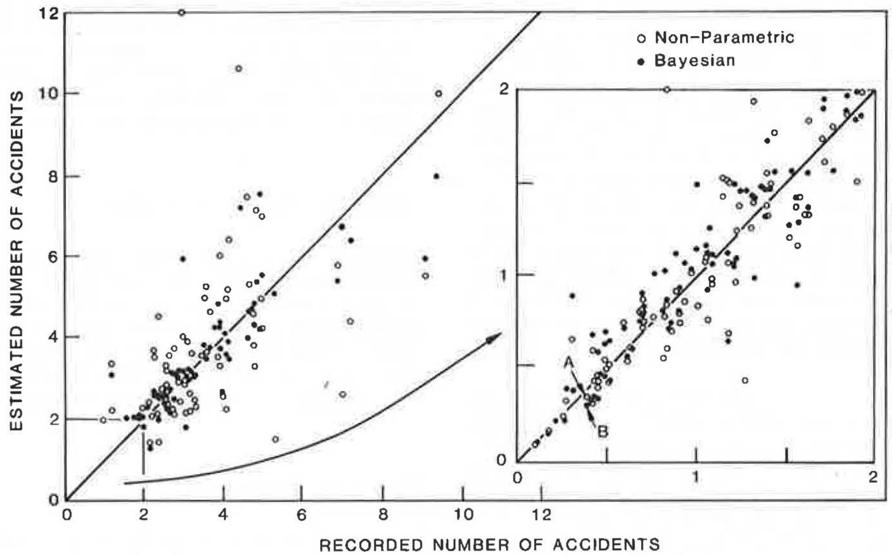

FIGURE 1 Application of debiasing methods to road data sets.

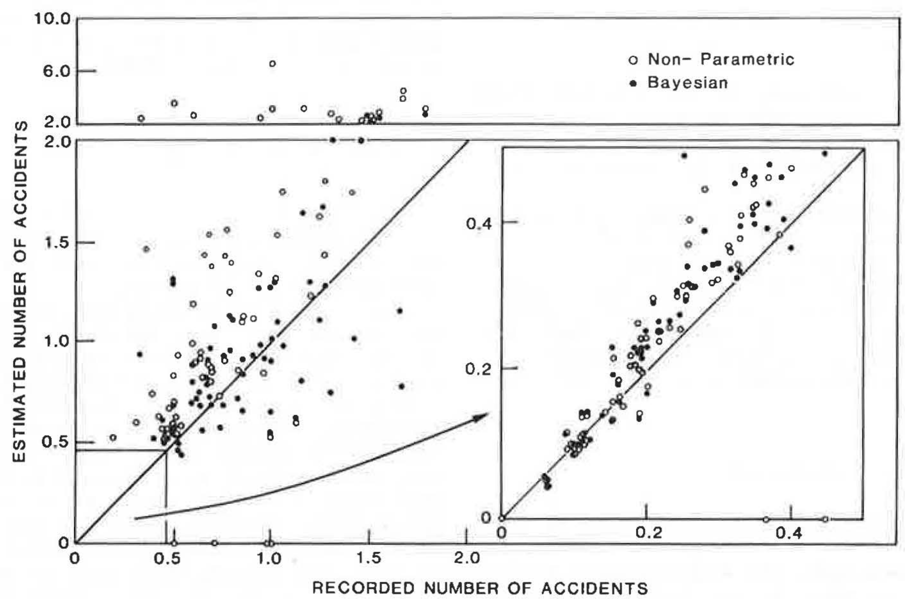

FIGURE 2 Application of debiasing methods to driver data sets.

tions on the subsequent driving record. This explanation now appears to be incorrect, for, when the order of comparison is reversed (call a later year before and an earlier year after), the same results are obtained. As indicated by the squares in Figure 2, for the Bayesian method this problem is not as severe. The overestimates are not as large or as consistent.

The data in Table 2 present the results in a different form along with some additional information to assist in the discussion that follows. To illustrate, the first line in Table 2 gives information for the entire North Carolina driver population, with the first 2 years of data representing the before period and the second 2 years representing the after period. In fitting a negative binomial distribution to the before period frequencies, a chi-square value of 156 was calculated (column 2), whereas for the after period a value of 304 (column 3) was obtained. The average driver had 0.122 accidents (column 4) in the before period and 0.130 accidents (column 5) in the after period. The sample variances associated with these two means were 0.143 and 0.151 (columns 6 and 7), respectively. For this data set, for k = 0 or 1 accident (column 8), the nonparametric estimate was closer to what was recorded, whereas for other values of k the Bayesian method was closer. For the second entry, the second 2 years of data were used for the before period with the first 2 years as the after period, and so on. Note that when the lengths of the before and after periods differ, so do the orders of magnitude of the mean numbers of accidents (e.g., the New Mexico data).

The data in Table 2 confirm that, on the whole, the Bayesian method gives better results and, in addition, reveals something that is not immediately apparent in Figures 1 and 2. For systems with k = 0 or 1, the nonparametric method performs, in most cases, at least as well as the Bayesian method. This finding has important implications, as the effect of treatments on systems with k = 0 or 1 accident is often of interest.

Discussion of Results

Effect of Type of System

From the data sets examined, it is apparent that there is a need to distinguish between road systems and driver systems. Why this is so remains an interesting research question that is currently being investigated.

Disaggregation of the data sets does not appear to have any influence on the performance of the methods. In a real application, a treatment program may be aimed at a fairly narrow group (e.g., young drivers, signalized intersections, head-on collisions). Consequently, it is important that the population be defined to include only similar systems. This issue of population definition is being researched further.

Effect of Number of Accidents (k)

For any value of k larger than 1, the number of road (nondriver) systems tends to be relatively small, so it is not surprising that the nonparametric method does not perform as well as the Bayesian method. This finding lends empirical confirmation to statements made by other researchers (1-4) about the influence of random variations in observed frequencies when the number of systems is small. By smoothing these frequencies, the Bayesian method provides more reliable estimates for the smaller groups of systems. For a more general discussion of this issue, see Maritz (8).

Effect of Number of Systems in a Group

For road systems at least, it is expected that the size of a group with k accidents would be a more direct index of the relative performance of the methods than the value of k. However, from the examination of the data, it appears that statements about the relative performance of the two methods based on group size are not clear-cut. The best that can be said about the methods is that the nonparametric method is at least as good as the Bayesian method when the number of road systems with k accidents is larger than 200. If this was made into a rule, however, there would be many exceptions. For drivers, although it appears reasonable that the size of the group must play a role in the performance of the methods, the overestimation problem prevents this issue from being examined.

Effect of Chi-Square Values

Analysis of a wide range of data sets with diverse chi-square values (see Table 2) suggests that, contrary to intuition, chi-square values for the before period data do not appear to be a good index of the performance of either of the methods. Even when the after period data also have small chi-square values, a reliable estimate is not guaranteed.

Effect of Parameters

For the Bayesian method, the sample parameters appear to be more relevant than chi-square values in determining the performance with respect to road systems. Once the sample means and sample variances for the before period data are close to these values for the after period, the Bayesian method tends to give more reliable estimates for road systems. The same conclusion cannot be made for the nonparametric method or for driver systems.

SUMMARY AND CONCLUSIONS

In this paper the regression-to-the-mean phenomenon was reviewed along with two analytic methods for purging the resulting bias from the results of before-and-after comparisons.

The focus of the paper was on an empirical comparison of the two methods: the nonparametric method where only observed accident frequencies are used to estimate the expected number of future accidents, and the Bayesian method where an assumed underlying statistical distribution smooths these frequencies before using them in estimations. The comparison, based on a large number and variety of data sets, indicated that, in general, the Bayesian method gives somewhat better estimates and should be used in assessing the safety effect of a treatment. However, for systems with zero or one accident, the nonparametric method gives slightly better results and might be preferred if the future expected number of accidents on these systems is of interest. The nonparametric method is also preferred if, in revising estimates for previous studies, accident data are only available for high-accident locations.

Discussion

Olga J. Pendleton*

In the paper by Persaud and Hauer, the authors attempt to show that, by way of example, the Bayesian method for estimating accidents in the after period is better than the nonparametric method. Whereas the data sets to which this comparison was applied appear to support this claim, it should be noted that (a) the authors do not apply any statistical methods in making the comparison and appeal only to graphical and numerical descriptive measures to support their claim, and (b) an example is not a proof.

Addressing the first comment (a), this paper would be greatly enhanced if the authors applied relatively simple statistics in making the comparison between methods. For example, along with the plots depicting the relationships of the two methods that compare actual and estimated values, statistics such as the correlation coefficient and the mean squared error of deviation from the line representing equality could be reported for the two methods and equality could be statistically tested. It also appears that there is a region of accident frequency where the comparison of these methods may yield different results (e.g., <0.5 and $\geq$0.5 in Figure 2). The nonparametric technique might even be better at <0.2. Another statistical test that could be made is a simple t-test on the differences of the methods or for a more nonparametric approach, a χ^2 test of observed versus expected for each method. These statistics would be easy to apply and lend more credence to the authors' claims.

The second comment (b) is motivation for future research. Either a rigorous mathematical proof that compares the power of the two techniques or a simulation study would be interesting. In light of the difficulty of this task, this paper did a suitable job of attempting to answer this question in a less rigorous but informative and interesting manner.

*Operations Research and Analysis Program, Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843

Authors' Closure

We thank Pendleton for her interest in our paper and for highlighting an apparent shortcoming. We think that the shortcoming is not so much in the omission of statistical tests as in the absence of any rationalization of this omission.

After reviewing our results, we chose not to perform any statistical tests as we felt that the conclusion that "the Bayesian method is likely to give somewhat better estimates" was ably supported by the plots. We did not seek a stronger conclusion because, as Pendleton notes, an example is not proof and, equally important, because a stronger conclusion would have detracted from our findings with regard to other issues. The effects of the type of system and the number of before accidents are issues well in keeping with our original intention "to see if there are circumstances in which one or the other method should be preferred."

In summary, although we agree in principle with Pendleton's proposals, we believe that they are outside the scope of this paper.

REFERENCES

1. E. Hauer and B. Persaud. Common Bias in Before-and-After Comparisons and Its Elimination. In Transportation Research Record 905, TRB, National Research Council, Washington, D.C., 1982, pp. 164-174.
2. E. Hauer. Bias-by-Selection: Overestimation of the Effectiveness of Safety Countermeasures Caused by the Process of Selection for Treatment. Accident Analysis and Prevention, Vol. 12, No. 2, June 1980.
3. 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 of Safety Research, SWOV, Amsterdam, Netherlands, 1982.
4. C. Abbess, D. Jarrett, and C.C. Wright. Accidents at Blackspots: Estimating the Effectiveness of Remedial Treatment, with Special Reference to the "Regression-to-the-Mean Effect." Traffic Engineering and Control, Vol. 22, No. 10, 1981, pp. 535-542.
5. J.R. Stewart and B.J. Campbell. The Statistical Association Between Past and Future Accidents and Violations. Highway Safety Research Center, University of North Carolina, Chapel Hill, Dec. 1972.
6. R.D. Helliar-Symons. Yellow Bar Experimental Carriageway Markings: Accident Study. TRRL Report 1010. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
7. U. Brude and J. Larsson. The Regression-to-the-Mean Effect--Some Empirical Examples Concerning Accidents at Road Junctions. In Proc., Seminar on Short-Term and Areawide Evaluation of Safety Measures, Institute of Safety Research, SWOV, Amsterdam, Netherlands, 1982.
8. J.S. Maritz. Empirical Bayes Methods. Methuen and Company, Ltd., London, England, 1970.