Common Bias in Before-and-After Accident Comparisons and Its Elimination

EZRA HAUER AND BHAGWANT PERSAUD

When treatment is applied to road sections, intersections, drivers, or vehicles that have had a poor accident record in the past, a simple before-and-after comparison of accidents will usually make useless treatments appear effective and overestimate the effect of useful treatments. Because much of what we know about the effect of various safety countermeasures comes from such studies, it is important to demonstrate that this bias can be very large, to show that it can be relatively easily purged from before-and-after comparisons, and to examine whether the method works. These are the three central aims of this paper. To render the statements credible, several data sets are used. The data come from Canada (Ontario), Sweden, the United Kingdom, Israel, and the United States (North Carolina and California) and relate to road sections, intersections, traffic circles, driver violations, and driver accidents. These data sets are used to demonstrate the magnitude of the bias, to illustrate the technique for its elimination, and to examine the success of this debiasing procedure.

It is common practice to apply treatment to those elements of the transport system (intersections, drivers, vehicles) that have had a poor accident record. This makes good sense. The effect of such treatments is often assessed by comparing the accident histories before and after treatment. The cumulative experience gained from several such before-and-after comparisons tends to become the current lore about the effectiveness of the treatment.

We will argue below that when systems are selected for treatment because of their poor accident experience, the simple before-and-after comparison leads to consistently biased conclusions; it makes the treatment appear to be more effective than it actually is. In the long run, the cumulative result of consistent overestimation of treatment effectiveness can lead to misallocation of resources.

The simple before-and-after comparison is and will continue to be an important source of information about the effect of corrective treatments. Therefore, it is important that the existence and size of the bias be recognized and that it be eliminated from the results of before-and-after comparisons.

Accordingly, this paper is divided into three principal parts. The first will show that the bias by selection is not a figment of the theorists' imagination. On the contrary, it is an all-pervasive and empirically substantiated phenomenon. We will also show that the size of this bias is often comparable with what one might expect the result of an effective corrective treatment to be. The second part of the paper will present procedures for the removal of the bias from results of before-and-after studies. In the third part we will examine how these procedures perform when applied to real data.

REGRESSION TO THE MEAN

More than a century ago, Sir Francis Galton (1) observed that the offspring of tall parents were, on the average, shorter than their progenitors. This phenomenon became known as the regression to the mean. As this phrase suggests, when a random deviation from the mean occurs (upward or downward), one should expect the next "trait" to be a return (regression) to the mean. One can show other examples of the same phenomenon that are closer to our common experience. We list below the scores from a professional golf tournament. A large number of golf players are let loose on the course. At the end of the first two days of play, those with the lowest scores are said to have "made the cut" and are allowed to proceed to the final two days of the tournament. (In golf, the score is a count of strokes needed to place a small white ball into a fixed number of holes in the ground. Thus, the smaller the score, the better one played.) The scores for the first and subsequent pairs of days of the 50 golfers who have made the cut are listed below. Thus, one golfer who scored 134 during the first two days scored 149 during the next two days; the two golfers who scored 138 on the first two days had an average score of 144.5 during the next two days, etc.

No. of Golfers in Group	Total Score for Avg Golfer in Group: Days 1 and 2	Days 3 and 4
1	134	149.0
2	138	144.5
1	139	143.0
3	140	144.3
7	141	143.4
5	142	145.0
4	143	148.7
10	144	146.1
9	145	147.0
8	146	146.9

It is quite apparent that those who got in a few lucky shots on days 1 and 2 returned to their average game on days 3 and 4.

What are the salient features common to the two examples that link golf and heredity and how are these relevant to before-and-after comparisons in safety?

In both cases, people were selected for subsequent monitoring on the basis of some trait or score that was found to be higher (or lower) than the average score for the population from which the selection is made. When later the same trait or score was observed again, there was a tendency for it to be closer to the population average.

This does not mean that selection procedures that identify groups on the basis of high (or low) scores will not identify groups whose future scores will be higher (or lower) than the average scores of the population from which they are selected. In fact, such selection procedures will often identify groups whose future scores will continue to be higher (or lower) than average. The crucial issue is the fact that the averages of the original selection scores will almost invariably be higher (or lower) than the subsequent scores of the selected groups.

The parallel in safety is simple. Intersections, drivers, vehicles, etc., are often selected for treatment or improvement if they have an unusually high number of accidents or a high accident rate. Due to the same regression-to-the-mean phenomenon, one should expect these systems to have less accidents subsequently even if nothing is done to them. Yet in a simple before-and-after study, such an observed reduction is normally interpreted as an indication that the countermeasure has been effective.

Skeptics might doubt that what holds for golf scores or stature applies also to the occurrence of

Table 1. Regression to the mean: Ontario data.

No. of Sections in Group	No. of Accidents for Avg Section in Group: First Year	No. of Accidents for Avg Section in Group: Second Year	Change (%)
12 859	0	0.404	Increase
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[a]	10.39	-22.0

[a]Avg = 13.33.

accidents. To convince them that the phenomenon is real and important, we will use accident data taken from several countries and that describe a variety of circumstances.

System of Road Sections: Ontario, Canada

The King's Highways (excluding freeways) are divided into 20 762 sections 1 km in length. As shown in Table 1, 12 859 of these had no nonintersection accidents in the first year of a two-year period, 4457 had one such accident, etc. The average number of accidents per kilometer during the first year was 0.707 and during the second year, 0.746.

On many of the 12 859 road sections that had no accidents in the first year (first line in Table 1), some accidents occurred during the second year. That this should be so is in accord with intuition. It will facilitate the understanding of the regression-to-the-mean phenomenon if we attempt to make the basis of this intuitive understanding more explicit.

Every road section that is open to traffic will register an accident from time to time. Thus, the long-term mean number of accidents per year for any road section, however small, is always larger than zero. A year in which no accidents occurred on a road section is therefore an event that by chance is below the long-term mean value. During the next year there could again be zero accidents. However, there could also be 1, 2, 3, ... accidents. The average number of accidents in the long term is therefore larger than zero.

This is the essence of the regression to the mean. When a random down fluctuation occurs, one should expect a return to the mean; when a random up fluctuation occurs, one should also expect a return to the mean.

Inspection of Table 1 reveals that this is in fact what happens. Road sections with no accidents in the first year have, on the average, 0.404 accident in the second year. Thus, 0.404 is an estimate of the long-term mean number of accidents for this group of road sections. Similarly, the 4457 road sections that all recorded one accident in the first year regress to the mean and have, on the average, 0.832 accident in the second year. Road sections with two accidents in the first year regress to 1.301 accidents in the second, etc. Naturally, road sections that had no accidents in the first year have a lower long-term mean (0.404) than road sections with, say, eight accidents in the first year (4.818).

The last column of Table 1 shows that the phenomenon is consistent, real, and nothing short of dramatic. To be sure, there are several confounding factors at play. Treatment in one form or the other may have occurred, road and environmental conditions may have been better during the second year, law enforcement may have improved. However, none of these potentially confounding factors can begin to explain a substantial part of the observed convergence on the mean.

Imagine now what would have happened had the Ministry hired a shaman to pray for accident reduction on road sections that had in the first year seven or more accidents. The apparent effectiveness of this "treatment" is, from the figures in Table 1, about 30 percent.

In summary, road sections that in one period had more than the average number of accidents will usually have less accidents in the next period even if no corrective treatment is applied. This is the essence of our concern. If one is to have a realistic assessment of the effect of some corrective treatment on the basis of a before-and-after comparison, the bias due to the regression to the mean has to be accounted for.

It is natural to ask at this point whether this bias can be substantially reduced (or even rendered negligible) if systems are selected for treatment on the basis of accident data that have been accumulated over a long period of time. The hope is that such data closely represent the mean rather than a random up or down fluctuation. The empirical evidence can be examined on the basis of the following data set.

System of Road Sections: Israel

A seven-year history of nonintersection accidents was obtained for each of 828 km of rural roads in Israel. This data set enabled us to examine the relationship between the magnitude of the regression to the mean and the duration of accident history. To do so, we regard the last year (year 7) as the "after" period; years 6, 5+6, 4+5+6, etc., are considered "before" periods, which are 1, 2, 3, ... years in duration.

Table 2 summarizes the empirical evidence. Several interesting observations follow.

First, the regression to the mean for road sections does not recognize national boundaries. It is as evident for Israeli as for Canadian road sections. Thus, prayers for accident reduction on road sections that in one year had three or more accidents (Table 2, bottom) might seem to bring about a 55 percent reduction in the number of accidents.

Second, as the duration of the before period increases (column 1 of Table 2), the relative size of the regression to the mean diminishes (column 5). This is as anticipated. Note, however, that even with a before history of six years, the size of the phenomenon is far from negligible.

At the peril of boring the converted and belaboring unnecessarily a point that is already clear, we will present further empirical evidence. The intent is to demonstrate by real cases that what holds for human stature, golf scores, and road sections also holds for traffic circles, intersections, driver accidents, and traffic law violations. We hope that the breadth of geographical coverage will show universality and that the variety of circumstances will underscore the all-pervasive nature of this phenomenon.

The first two accident-related data sets dealt with road sections. The next two illustrations focus on intersections.

Approaches to Traffic Circles: United Kingdom

Transverse yellow bar markings with gradually dimin-

Table 2. Effect of duration of before period: Israeli data.

Duration of Before Period (years)	No. of Road Sections	No. of Accidents per Year Before	After	Change (%)
Road Sections with One or More Accidents per Year				
1	337	530	317	-40
2	258	393	277	-30
3	231	321	250	-28
4	191	292	230	-21
5	178	272	224	-18
6	170	258	222	-14
Road Sections with Two or More Accidents per Year				
1	126	319	164	-49
2	64	159	91	-43
3	47	111	78	-30
4	45	107	75	-30
5	37	92	70	-24
6	28	72	55	-24
Road Sections with Three or More Accidents per Year				
1	45	157	71	-55
2	16	53	31	-41
3	5	17	15	-12
4	4	15	10	-33
5	7	25	21	-16
6	7	24	19	-21

Table 3. Regression to the mean: U.K. roundabout data.

No. of Accidents per Site in Year 1	No. of Sites in Group	Aggregate Accidents for Each Group Year 1	Year 2	Change (%)
10+	5	74	54	-37
9	2	18	10	-44
8	7	56	49	-13
7	2	14	6	-57
6	9	54	48	-15
5	9	45	35	-22
4	7	28	29	+4
3	11	33	44	+33
2	15	30	46	+53
1	9	9	13	+44
0	6	0	7	Increase
	82	361	341	

ishing spacing are known to reduce speed. Results of a study on the effect of such markings on accidents at approaches to traffic circles (roundabouts) have been published (2). These will be used later to show that when before-and-after comparisons are cleansed of the bias by selection, proper conclusions can be reached even when it is not practical to include corresponding control sites in the study. At present, however, we will use the data only to show that had the researchers not taken proper precautions, the regression to the mean would have destroyed the validity of their results.

For the 82 roundabout approaches included in the study, the number of injury accidents during each of two years prior to treatment is shown in Table 3.

Since there was no treatment or any other important change between the two years, the observed effect is largely due to the regression to the mean. The mean annual number of accidents per approach is 4.5. Sites that had more than four accidents in year 1 tend to experience a reduction in year 2; approaches with less than the mean in year 1 had an increase.

The danger is obvious. Had the researchers decided to treat with yellow bar markings approaches

Table 4. Regression to the mean: reported and injury accidents at intersections in Sweden.

No. of Intersections in Group	No. of Accidents per Intersection During Before Period	Avg No. of Accidents per Intersection During Equivalent After Period	Change (%)
All Reported Accidents			
1500	0 (0)	0.37	Increase
657	1 (0.85)	0.77	(-9)
244	2 (1.70)	1.04	(-39)
101	3 (2.56)	1.79	(-30)
53	4 (3.41)	1.94	(-43)
39	5 (4.26	2.69	(-37)
17	6 (4.11)	3.05	(-40)
26	8.92 (7.60)[a]	5.46	(-28)
Injury Accidents			
2039	0 (0)	0.19	Increase
441	1 (0.85)	0.42	(-51)
119	2 (1.70)	0.71	(-59)
24	3 (2.56)	1.33	(-48)
14	4.143 (3.53)[b]	1.50	(-57)

Note: Figures in parentheses are exposure adjusted.
[a]These figures are for sites with seven or more accidents in 1972-1975.
[b]These figures are for sites with four or more accidents during 1972-1975.

that have many accidents (more than four per year), they would have observed a reduction even if the treatment was useless. With a useful treatment, the observed reduction would be an overestimate of treatment effectiveness.

The consistency with which the phenomenon appears in such a limited data set is quite remarkable. A more extensive study is examined below.

System of Intersections: Sweden

This illustration is based on 2637 unsignalized rural road junctions that were unaltered during the period 1972-1978 (3). The four-year stretch 1972-1975 is regarded as the before period and the three years 1976-1978 as the after period. The top part of Table 4 is based on all police-reported accidents, whereas the bottom half is based only on personal-injury accidents.

In Table 4, the number of before-period accidents has been adjusted to reflect differences in length of the before and after periods and in exposure. Unlike the road sections examined earlier, only intersections that remained physically unaltered are included in this data set. Therefore, the change in the number of accidents between the before and after periods is attributable almost exclusively to the regression to the mean.

It is interesting to note that the change in injury accidents is larger than that for all reported accidents. The exact reason will become clear later. However, the main element of the explanation is already evident. It is the difference in the mean towards which the number of accidents regresses. Since there are less injury accidents than all reported accidents, the size of the regression to the mean for injury accidents is larger than that for all reported accidents.

The four data sets examined so far dealt with accidents occurring on elements of the road system. The regression-to-the-mean effect that has been shown to exist in these illustrations must be taken into account when safety benefits of corrective treatments to road sections and intersections are estimated.

Corrective treatment is often aimed at the road user (rather than at the road system). Thus, for example, those convicted repeatedly for impaired driving may be sent for a week-long course, those who accumulated a certain number of demerit points may receive a warning letter, etc. As before, when the selection for treatment is on the basis of a bad record, the regression to the mean may be at work. Accordingly, one should expect an improvement in the record even when no treatment is administered. That this is in fact the case will be demonstrated in the following examples.

System of Drivers: United States (North Carolina)

In an investigation about the predictability of accidents and violations on the basis of past performance, the records of some 2.5 million drivers in North Carolina were examined (4).

In the tabulation below, drivers are placed in seven groups according to the number of traffic law violations during the first two-year period. For each group, the average number of violations in the subsequent two-year period is also given. The existence and size of the regression-to-the-mean effect are evident from the examination of the last column. (The total number of drivers in the group included 674 drivers with seven or more violations in the first period.)

No. of Drivers in Group	No. of Violations for Avg Driver in Group: First Two-Year Period	Second Two-Year Period	Change (%)
2 096 935	0	0.191	Increase
298 645	1	0.457	-54.3
73 216	2	0.763	-61.9
21 907	3	1.024	-65.9
7 224	4	1.266	-68.4
2 597	5	1.459	-70.8
1 042	6	1.500	-75.0
2 502 240	0.225	0.252	+12.0

As in earlier illustrations, the regression to the mean is not the only cause of the noted change. First, the average number of violations per driver has changed somewhat (from 0.225 to 0.252). Second, many violations are linked to accidents the occurrence of which may have a substantive effect on one's driving. Third, the level of enforcement, driver maturation, possible corrective treatments, etc., all may exert some influence. Nevertheless, as will be demonstrated later, the main reason for the drop from, say, 6 to 1.5 is the fact that the group includes drivers who, on the average, have 1.5 violations per year and who due to the laws of chance happened to record six in the first two years. Similarly, the estimated long-term mean number of violations in two years for the group of drivers who during the first two-year period had no violations is 0.191. The regression is always to the long-term mean specific to the selected group.

In the table below, the same North Carolina drivers are placed into seven groups according to the number of accidents they had in the first two-year period. For each group, the average number of accidents in the subsequent two-year period is listed. The Change column gives an indication of the regression to the mean. Some of this change may be due to reduced exposure and change in driving style, which is linked to the accident trauma experienced in the first period. How much of the change is attributable to such causes and what part is due to regression to the mean will be explored later. (The total number of drivers included 10 with seven or more accidents in the first period.)

No. of Drivers in Group	No. of Accidents for Avg Driver in Group: First Two-Year Period	Second Two-Year Period	Change (%)
2 234 577	0	0.117	Increase
235 080	1	0.216	-78.4
27 919	2	0.348	-82.6
3 953	3	0.499	-83.4
584	4	0.703	-82.4
99	5	0.848	-83.0
18	6	0.944	-84.3
2 502 240	0.122	0.130	+6.1

These tabulations lead to the following observations:

First, drivers with one or more violations or accidents in the first period are seen to have, on the average, 60-80 percent less violations or accidents in the second period, in spite of the increase in the number of violations and accidents from the first to the second period. Note that the magnitude of the change is much larger for drivers than that for the elements of the road system.

Second, as observed earlier in the case of the intersections in Sweden, the smaller the mean towards which observations regress, the larger the change.

System of Drivers: United States (California)

This last illustration endeavors to reinforce two points made earlier: (a) regression to the mean is not bound by geography--accident records of drivers in California show the same effect as the records of North Carolina drivers; (b) the magnitude of the effect remains large even when the before period is three years long.

In the table below, some 93 000 randomly selected drivers who have driven in California for at least six years have been placed into four groups according to the number of accidents they had during the three-year period 1969-1971. (Data were obtained from R. Peck, Department of Motor Vehicles, State of California, March 1982.) For each group, the average number of accidents in the subsequent three-year period (1972-1974) is listed. The Change column gives an indication of the regression to the mean. (The total number of accidents for 1969-1971 was not reported for the last group of drivers.)

No. of Drivers in Group	No. of Accidents for Avg Driver in Group: 1969-1971	1972-1974	Change (%)
79 327	0	0.152	Increase
11 897	1	0.238	-76.2
1 525	2	0.374	-81.3
250	3+	0.548	--

Entries in the previous two tables are remarkably similar in spite of the difference in geography, driver population, and time frame. One is again led to conclude that the phenomenon is universal. Even though three years of data are used for the before period in the California study, the effect remains very large.

Having shown the existence of the regression to the mean in a diverse set of circumstances, we will regard it as empirically substantiated. From here, we take it as "common ground" that if a system (road section, intersection, driver, vehicle, etc.) is observed to have a higher-than-mean number of accidents in one period, it should be expected to have less accidents in the next period if no corrective treatment is applied to it.

It follows that if corrective treatment is applied to systems that had a poor accident record, it is impossible to estimate the safety benefits of the treatment from its accident record after treatment without first accounting for the effect of the regression to the mean. How to do so is explained next.

OBTAINING UNBIASED ESTIMATES IN UNCONTROLLED BEFORE-AND-AFTER COMPARISONS

To assess the safety effect that some corrective treatment had on a system to which it has been applied, one compares the safety performance of the system after the treatment against what one would expect the safety performance of the system to have been during the same period of time had the treatment not been applied. This is the logical premise on which all attempts to estimate the safety effect of countermeasures are founded, no matter how simple or sophisticated the associated experimental design.

In light of this logical premise, it is important to note that a simple before-and-after comparison is equivalent to stating, "I assume that the safety performance before treatment is a good estimate of what the safety performance would have been during the after period had treatment not been applied."

The first part of this paper has been devoted to the demonstration that this assumption runs counter to abundant empirical evidence and that when systems are selected for corrective treatment because of their poor safety record, one must expect them to have an improved accident record in the subsequent period if no treatment is applied. The central task of this section is to provide a credible answer to the question, What accident record should one expect during the after period if no treatment is administered?

To cast the problem in more precise terms, let b denote the number of accidents occurring on a system during some period before treatment and α denote the number of accidents expected to occur on the same system during an after period of the same duration had treatment not been applied. Our task is to estimate α.

When α is estimated, what evidence should count? A general answer to this simple question is not easy to provide. However, when the system is selected for treatment out of a larger group of candidates, there are at least two pieces of information that must affect the estimate:

1. The safety performance of the selected system during the before period (b) and
2. The safety performance during the before period of all other candidate systems.

The relevance of the first piece of information is self-evident. The second piece of information is relevant because the safety record of the selected system was compared with the safety records of other candidates when the selection was made.

The estimation of α turns out to be astonishingly simple.

Estimation Rule 1

The estimate of the number of accidents expected to occur in the after period on systems that during the before period had k accidents is the number of accidents occurring on systems that had (k + 1) accidents in the before period.

This rule is best illuminated by a simple example. Consider the data in Table 1 and pose the following question: How many accidents should one expect to occur during the second period on the 62 Ontario road sections that in the first period had seven accidents? [We know, of course, from Table 1 that the actual number of accidents on these road sections during the second period was 308 = (62 x 4.968). But for the moment we imagine that we are just at the end of the first period and have to provide an estimate.]

Following estimation rule 1, we find α = 8 x 33 = 264 accidents. It is tempting to immediately compare this estimate with what actually transpired. Such temptation should be resisted until a systematic juxtaposition of estimates and the actual number of accidents is carried out later in this section. Suffice to note that were one to assume (as is normal practice) that road sections with seven accidents in the before period are likely to have the same number of accidents (seven) in the after period, a gross overestimate would be obtained.

Estimation rule 1 can be recast into an alternative (equivalent) form.

Estimation Rule 2

The estimate of the number of accidents expected to occur in the after period on a system that during the before period had k accidents is the number of accidents occurring during the before period on systems that had (k + 1) accidents divided by the number of systems on which k accidents occurred during the before period.

To continue the previous illustrative example, an Ontario road section that during the first period had seven accidents should be expected to have (8 x 33)/62 = 4.3 accidents during the second period.

It should be noted that when the two periods differ in duration or in exposure, a correction should be applied to the estimate of α in the customary manner.

In some cases it is convenient to use a third variant of estimation rule 1.

Estimation Rule 3

The estimate of the number of accidents expected to occur during the after period on systems that during the before period had k or more accidents is the number of accidents occurring on systems that had (k + 1) or more accidents during the before period.

To return to the data in Table 1, Ontario road sections that during the first period had seven or more accidents should be expected to have during the second period 830 = (8 x 33 + 9 x 14 + 13.33 x 33) accidents. As can be easily ascertained, the actual number of accidents on these road sections was 907.

The above estimation rules have been obtained in an entirely rigorous way, as shown in the next section. The proof rests on the universally accepted but empirically unproven assumption that accident occurrence on each system obeys the Poisson probability law. Thus, while each road section or driver has its own characteristic mean (number of accidents per unit time), the probability of a specific realization is specified by the Poisson distribution.

Since the proof of the pudding is in the eating, the last section is a patient juxtaposition of unbiased estimates and the empirical evidence contained in the data sets already introduced in the first part of this paper.

Determination of Estimation Rules

Out of n systems, those that during time period 1 recorded k or more accidents are chosen. We wish to find the number of accidents one should expect to occur on the chosen systems during period 2 of the same duration. Let B(k) be the expected number of accidents on the chosen systems in period 1 and A(k) be the expected number of accidents on the chosen systems in period 2.

Let m_i, i = 1,2,...,n, be the expected number of accidents on candidate system i during periods 1 and 2 and let p_{ij} denote the probability that system i will have j accidents. With this notation,

$$P\,(\text{system i is chosen}) = \sum_{j=k}^{\infty} p_{ij} \tag{1}$$

and the contribution of system i to B(k) is

$$\sum_{j=k}^{\infty} jp_{ij}$$

If we sum over all systems,

$$B(k) = \sum_{i=1}^{n} \sum_{j=k}^{\infty} jp_{ij} \tag{2}$$

If we assume that the number of accidents on a system obeys the Poisson probability law,

$$p_{ij} = \exp(-m_i)m_i^j/j!$$

for system i and

$$jp_{ij} = m_i p_{ij-1} \tag{3}$$

If we substitute Equation 3 into Equation 2,

$$B(k) = \sum_{i=1}^{n} m_i \sum_{j=k}^{\infty} p_{ij-1}$$

$$= \sum_{i=1}^{n} m_i \sum_{j=k-1}^{\infty} p_{ij} \tag{4}$$

On the other hand, the expected number of accidents on system i during period 2 is m_i. By using Equation 1,

$$A(k) = \sum_{i=1}^{n} m_i \sum_{j=k}^{\infty} p_{ij} \tag{5}$$

If we compare Equations 4 and 5,

$$A(k) = B(k + 1) \tag{6}$$

Equation 6 leads directly to estimation rule 3.

Note that to obtain Equation 6, one assumes only that accident occurrence is governed by the Poisson probability law. The mean number of accidents (m_i) differs from system to system.

By using the result in Equation 6, we can write

$$A(k) - A(k + 1) = B(k + 1) - B(k + 2) \tag{7}$$

The expression on the left-hand side is the expected value of the difference (number of accidents in period 2 on systems that in period 1 had k or more accidents minus number of accidents in period 2 on systems that in period 1 had k + 1 or more accidents). This is, of course, the expected number of accidents in period 2 on systems that in period 1 had k accidents and will be denoted by $\alpha(k)$.

Similarly, the expression on the right-hand side in Equation 7 is the expected number of accidents occurring during period 1 on systems with k + 1 accidents and will be denoted by $\beta(k + 1)$. Thus,

$$\alpha(k) = \beta(k + 1) \tag{8}$$

Equation 8 leads directly to estimation rule 1.

EMPIRICAL EXAMINATION OF WHETHER ESTIMATION RULES YIELD UNBIASED ESTIMATES

The main content of this section is an examination of the estimated number of after-period accidents (or violations) compared with the number of after-period accidents (violations) actually recorded. The hope is that they are in good agreement.

Were one to plot estimates along one axis of a rectangular coordinate grid and the recorded number of accidents on the other axis, an ideal agreement would place all such points on the diagonal. There are at least three reasons why such an ideal relationship should not be expected.

First, as is evident from reading the estimation rules, what plays the role of an estimate is a (Poisson-distributed) random variable. As such it is, of course, subject to considerable variability. Second, the recorded number of accidents to which the estimate is compared is similarly subject to large random variation. Thus, even if the expected location of the point was on the diagonal, the random variation in its ordinate and abscissa will cause it to deviate from it. Third, as has been pointed out several times earlier, the before period can never be regarded as equivalent to the after period even if precautions are taken to consider only untreated systems and correction for exposure is applied. This source of uncertainty will cause even the (unknown) expected location of the points to be off the diagonal. Thus, a realistic assessment of what constitutes good agreement is that the points are fairly close to the diagonal and a regression line through the points has a slope close to unity.

Since the estimation rules are essentially equivalent, it would be redundant to apply all three to each data set. We begin by applying all three rules to the Ontario data and discuss the application in some detail. This will allow us to deal with all other data sets more summarily. Finally, results of a controlled evaluation are compared with what is obtainable from an uncontrolled study from which the bias is eliminated.

Application to Ontario Road Sections

The numbers in Table 5 summarize the application of all three estimation rules to the data on Ontario road sections that were introduced earlier and described in Table 1. The left part of Table 5 deals with the application of rules 1 and 2, the right part with estimation rule 3.

Column 2 gives the number of road sections that during the before period had k = 0,1,2,... accidents. These values are copied from Table 1. Column 3 gives the total number of accidents occurring on these sections during the before period. Thus, for example, on the 374 sections that during the before period had 4 accidents, the total number of accidents was 374 x 4 = 1496.

By estimation rule 1, one should expect that during the after period, the total number of accidents on road sections that during the before period had 3 (= 4 - 1) accidents will be 1496. This is the entry in row 4 of column 5. Note that the estimate in column 5 is simply the entry in column 3 lifted by one row. The number of accidents actually recorded during the after period is given in column 7. It is the comparison of the entries in columns 5 and 7 by which the performance of estimation rule 1 is to be judged. Thus, while on the basis of the accidents in the before period one would have estimated that 1496 accidents will occur on the 791 road sections, the number of accidents actually recorded was 1456. Were one to follow the common practice and assume

Table 5. Application of estimation rules to Ontario data.

	Estimation Rules 1 and 2[a]								Estimation Rule 3[b]			
		Total No. of Accidents								Aggregate No. of Accidents at Identified Sites		
				After Period (second year)							One Year After	
		Before Period (first year)		Estimated		Recorded						
No. of Accidents (k)	No. of Sections with Exactly k Accidents	Rule 1	Rule 2	Rule 1	Rule 2	Rule 1	Rule 2		No. of Sections with k or More Accidents During Before Period	One Year Before	Estimated	Recorded
0	12 859	0	0	4457	0.354	5199	0.404		20 762	14 648	14 648	15 467
1	4 457	4457	1	3768	0.845	3706	0.832		7 903	14 648	10 191	10 268
2	1 884	3768	2	2373	1.260	2452	1.301		3 446	10 191	6 503	6 562
3	791	2373	3	1496	1.891	1456	1.841		1 562	6 503	4 130	4 110
4	374	1496	4	800	2.139	883	2.361		771	4 130	2 634	2 654
5	160	800	5	570	3.563	513	3.206		397	2 634	1 834	1 771
6	95	570	6	434	4.568	351	3.695		237	1 834	1 264	1 258
7	62	434	7	264	4.258	308	4.968		142	1 264	830	907
8	33	264	8	126	3.818	159	4.818		80	830	566	599
9	14	126	9	80	5.714	97	6.930		47	566	440	440
10	8	80	10			74			33	440	360	343
≥11	25	360				269			25	360		

[a] For sites with exactly k accidents before. [b] For sites with k or more accidents before.

that what happened before is an indication of what should be expected after, one would expect 2373 accidents to occur on these road sections, an obvious overestimate. Furthermore, were some treatment applied to such sections, one might erroneously conclude that the apparent reduction from 2373 to 1456 can be attributed to the treatment when in fact it is but an artifact of the regression to the mean.

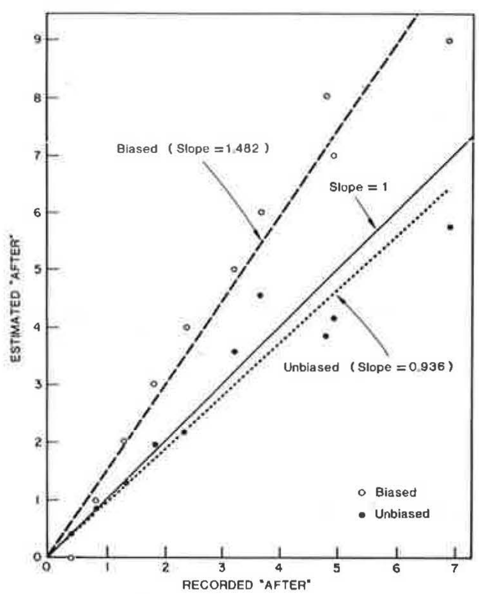
Columns 4, 6, and 8 give the entries relating to the application of estimation rule 2. Thus, column 4 gives the number of accidents during the before period for an average road section in the group. Entries in column 6 are the estimates of the number of accidents expected to occur on such a road section during the after period. To apply estimation rule 2, one has to divide the entry in column 5 by the entry in column 2. For example, a section for which k = 3 should be expected to have, during the after period, 1496/791 = 1.891 accidents. The performance of estimation rule 2 is to be judged by the comparison of columns 6 and 8. Note that entries of column 8 are taken from Table 1.

It is easy to visualize the performance of the estimation rules with the aid of the graphical representation in Figure 1. Each pair of numbers from columns 6 and 8 is represented by a solid circle. Perfect agreement between the unbiased estimate and the recorded number of accidents would place the solid circle on the diagonal. It is easy to see that the solid circles do hug the diagonal.

The performance of the biased estimates is represented by open circles based on the entries of columns 4 and 8. The vertical distance between the open circle and the diagonal indicates the magnitude of the bias due to the regression to the mean. A line through the origin has been fitted to each set of points and its slope is indicated in the figure.

Estimation rules 1 and 2 apply to sections with k accidents; rule 3 applies to systems with k or more accidents. Accordingly, one has to first put the data from columns 2 and 3 into the cumulative form shown in columns 9 and 10. Starting from the bottom of column 2, 25 + 8 = 33 is the number of road sections with 10 or more accidents during the before period as listed in column 9. Similarly, 360 + 80 = 440 is the number of accidents occurring on these road sections during the before period and listed in column 10. By estimation rule 3, one should expect that in the after period the number of accidents on road sections that during the before period had, say, nine or more accidents, will be the same as the

Figure 1. Application of rule 2 to Ontario data.

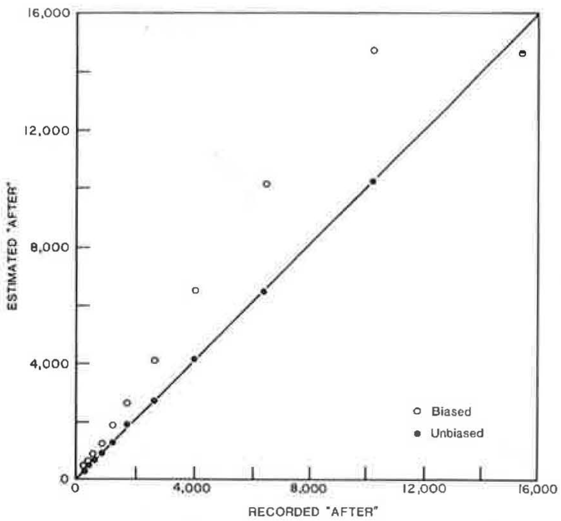


number of before-period accidents on road sections with 10 or more accidents, i.e., 440. This is the entry in column 11. As before, to obtain estimates, raise the entries in column 10 by one row. The performance of estimation rule 3 is judged on the basis of the comparison of the estimate in column 11 and the recorded number of after-period accidents in column 12. The entries of column 12 are obtained by the cumulation from below of the numbers in column 7. The performance of rule 3 in the case of Ontario road sections is shown in Figure 2.

Application to Other Data Sets

Estimation rules 2 and 3 have been applied to all data sets introduced earlier. We forego here the presentation of detailed comparisons as in Table 5 and describe the performance of biased and unbiased estimates in graphical form only. Figures 3-12 relate the number of accidents recorded during the after period to what one would expect on the basis of the number of accidents occurring during the before period. Solid circles depict the correspondence obtained by using the unbiased estimates and open circles the correspondence of the biased estimates. The slope of the best-fitting line to the solid circles and to the open circles is shown on each rule-2 graph. Figures 10 and 12 are plotted on

Figure 2. Application of rule 3 to Ontario data.



a log-log scale because of the range of numbers.

Some important observations can be made on the basis of these figures:

1. As is indicated by the slopes of the best-fitting lines, estimates obtained by using the recommended rules appear to be largely free of bias. Thus, in spite of the unavoidable limitation of comparing the safety of a system in two different periods of time, the theoretical considerations that led to the formulation of estimation rules seem to be supported by empirical data.

2. When estimates are based on a small number of accidents, they are unreliable, as should be expected. To illustrate, consider, for example, Figure 5, which is based on data in Table 3. Since each point is based only on a few accidents, the difference between the biased and unbiased estimates is entirely obscured by the random variations. However, when accidents are accumulated as for the application of estimation rule 3 (Figure 6), the same data clearly show the superiority of unbiased estimation.

Driver violations and accidents are overestimated by the unbiased method as well. This is not surprising due to the maturation of the driver population. Drivers who have had accidents or violations during the before period might be expected to have fewer accidents or violations subsequently, not only because they are more careful, but possibly because

Figure 3. Application of rule 2 to Israeli data.

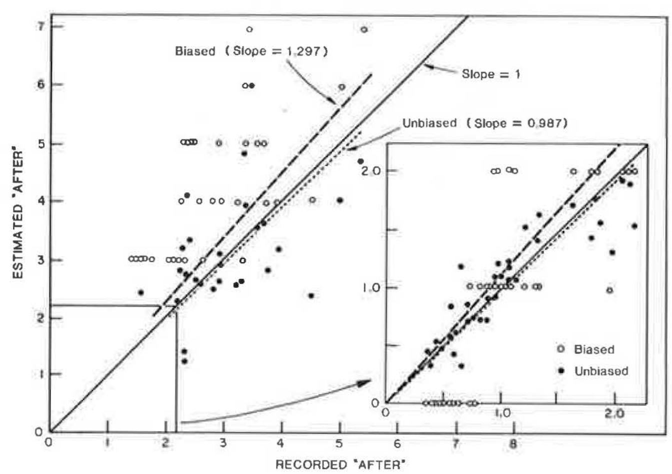
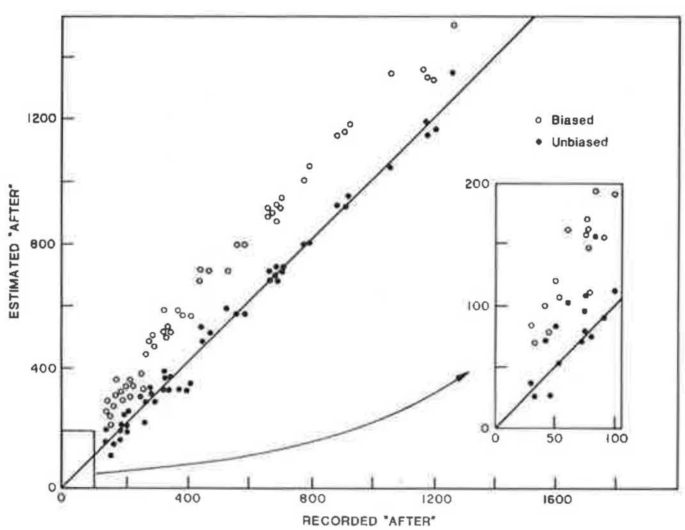


Figure 4. Application of rule 3 to Israeli data.



they tend to drive less due to hospitalization, license revocation, and similar reasons.

Comparing Results of Controlled Study to Study Without Control

The best way to avoid problems due to regression to the mean is to match to each system treated another system with the same accident experience that is left untreated and draw conclusions from their performance during the after period. However, when this is not practical, the suggested estimation rules should be applied.

The study on the safety effect of yellow bar

Figure 5. Application of rule 2 to U.K. roundabout data.

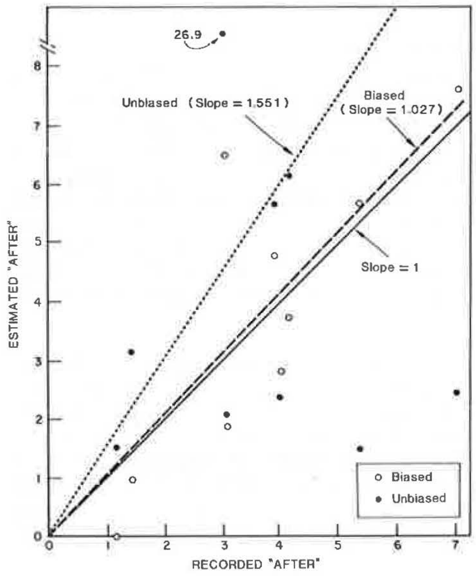


Figure 6. Application of rule 3 to U.K. roundabout data.

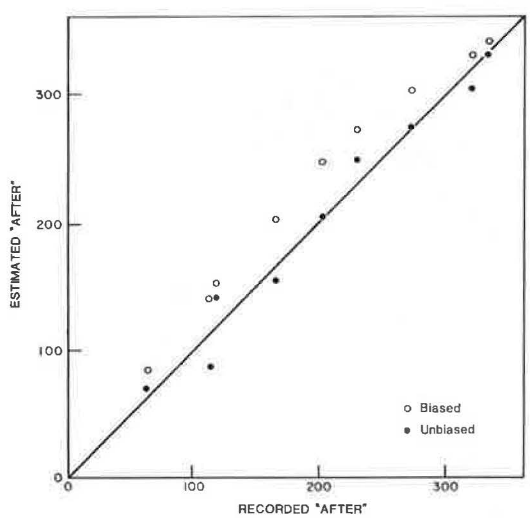


Figure 7. Application of rule 2 to Swedish intersection data.

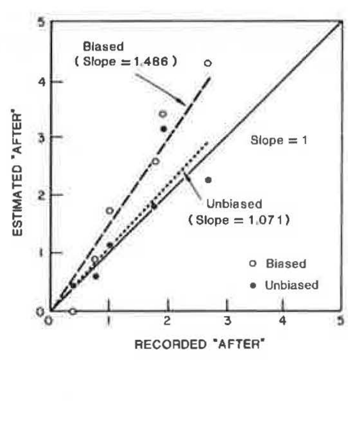


Figure 8. Application of rule 3 to Swedish intersection data.

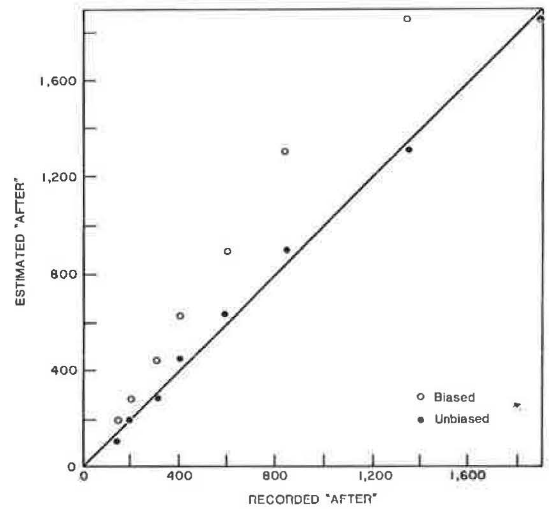


Figure 9. Application of rule 2 to North Carolina driver violations.

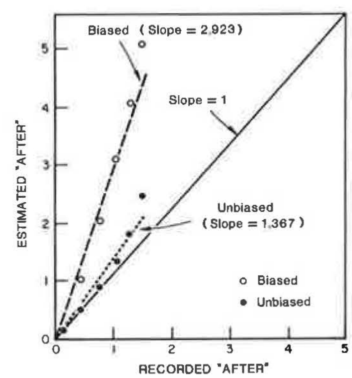


Figure 10. Application of rule 3 to North Carolina driver violations.

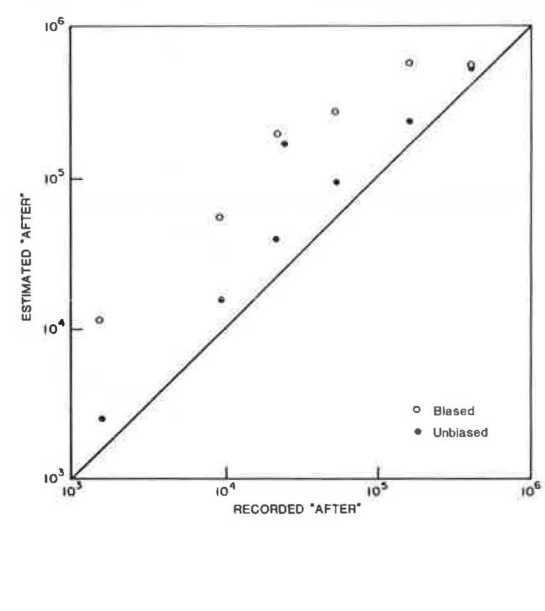


Figure 11. Application of rule 2 to driver-accident data.

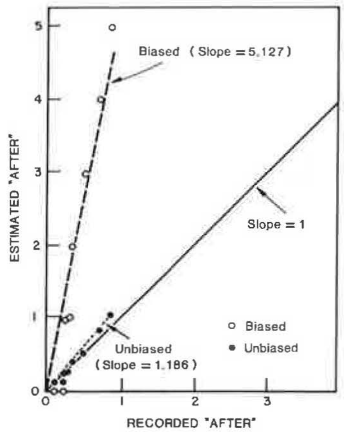
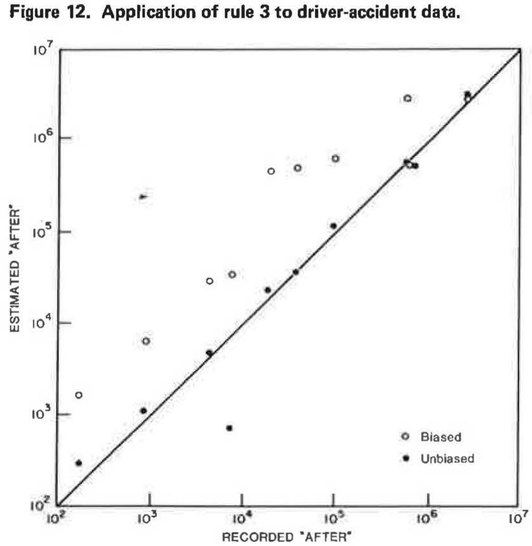


Figure 12. Application of rule 3 to driver-accident data.



markings was of the controlled type, and therefore the reported 57 percent reduction in susceptible accidents is free from bias. It is interesting to examine what the estimate of effectiveness obtained by the use of estimation rule 3 would be had the researchers not enjoyed the luxury of being able to identify appropriate control sites.

By using only data about treated sites and applying estimation rule 3, we obtain the results shown below:

No. of Accidents in Two-Year Period Before Treatment	No. of Sites in Group	Apparent Treatment Effectiveness (%)	Unbiased Estimate of Treatment Effectiveness (%)
≥4	7	74	63
≥3	15	76	63
≥2	27	66	52
≥1	36	64	60

Column 3 is based on the simple (biased) before-and-after comparison; column 4 gives the unbiased estimate of treatment effectiveness. Two points deserve mentioning. First, as has been argued earlier, the apparent treatment effectiveness is an overestimate. Second, the unbiased estimates are quite consistent with the finding of 57 percent reduction based on the more elaborate controlled experiment.

SUMMARY AND DISCUSSION

We have shown that when systems are selected for remedial treatment because of their poor accident history, simple before-and-after comparisons tend to overestimate the safety effectiveness of the treatment due to the regression-to-the-mean phenomenon. Several real-world data sets have been used to show that this phenomenon is real and important. A method for purging the bias from simple before-and-after comparisons has been suggested and summarized by three equivalent estimation rules. The performance of these rules for the data sets introduced earlier has been examined.

The estimation rules are based on analytical considerations described fully elsewhere (5). The same results were published earlier by Robbins (6).

Within the limits of accuracy influenced by the random variations of the data and the fact that the before conditions are never identical to the after conditions, the validity of the estimation rules is supported by the findings. It should be noted that when estimation is based on a small number of accidents, the accuracy is correspondingly limited.

To improve estimation accuracy when conclusions must be based on the comparison of a relatively small number of accidents, an empirical Bayes approach has been suggested (7,8). The performance of the two alternative methods of estimation has been examined in a different context (9). There it has been concluded that for large sample sizes and under certain other conditions, the estimation-rule approach is preferable.

A similar examination in a safety context has not yet been performed.

ACKNOWLEDGMENT

This publication was supported by the Insurance Institute for Highway Safety. The opinions, findings, and conclusions expressed are ours and do not necessarily reflect the views of the Institute.

REFERENCES

1. Sir Francis Galton. Typical Laws of Heredity. Proceedings of the Royal Institute, Vol. 8, Feb. 9, 1877, pp. 282-301.
2. R.D. Helliar-Symons. Yellow Bar Experimental Carriageway Markings: Accident Study. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Rept. 1010, 1981.
3. U. Brüde and J. Larsson. Regression-to-the-Mean Effect: Some Empirical Examples Concerning Accidents at Road Junctions. Proc., OECD Seminar on Short-Term and Area-Wide Evaluation of Safety Measures, Organization for Economic Cooperation and Development, Netherlands, April 1982.
4. 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.
5. H. Robbins. Prediction and Estimation for the Compound Poisson Distribution. Proceedings of the National Academy of Sciences, Vol. 74, No. 7, July 1977, pp. 2670-2671.
6. D.F. Jarrett, C. Abbess, and C.C. Wright. Bayesian Methods Applied to Road Accident Blackspot Studies: Some Recent Progress. Proc., OECD Seminar on Short-Term and Area-Wide Evaluation of Safety Measures, Organization for Eco-

nomic Cooperation and Development, Netherlands, April 1982.
7. 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.
8. D.G. Morrison and D.C. Schmittlein. Predicting Future Random Events Based on Past Performance, Management Science, Vol. 27, No. 9, Sept. 1981.

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