Critique of Rail-Highway Grade Crossing Effectiveness Ratios and Resource Allocation Procedures

WILLIAM D. BERG

ABSTRACT

A significant portion of recent research on rail-highway grade crossings has focused on the development of effectiveness ratios for various improvement actions and the incorporation of accident history into accident prediction and resource allocation procedures. Data are presented to demonstrate that these procedures can introduce significant bias and lead to a misallocation of resources. A preferred procedure is to use a modeling approach to estimate expected accident reductions and to ignore the adjustment of predicted accident rate based on recent accident history. The latter adjustment is shown to represent a modeling of the regression-to-the-mean phenomenon and as such is not appropriate for resource allocation studies.

A significant portion of recent research on rail-highway grade crossings has focused on the development of effectiveness ratios for various safety improvement actions and the incorporation of accident history into accident prediction and resource allocation procedures (1-6). The purpose of this paper is to offer some additional theoretical and practical insights into the interpretation and application of effectiveness ratios and accident history in the selection and programming of crossing improvements. Data will be presented to demonstrate that the current recommended resource allocation procedures can lead (and presumably are leading) to a misallocation of resources for grade crossings. Finally, recommendations will be offered on the direction of future research on grade crossings.

BACKGROUND

The U.S. Department of Transportation (DOT) rail-highway grade crossing models for accident prediction and resource allocation were initially developed in the late 1970s (1,2). The accident prediction models are multivariate and were calibrated by using the national inventory and accident data base. Although constituting the largest sample of reliable data that had ever been available to researchers, the data base was nevertheless incomplete because of the lack of information on sight distance, highway approach speeds, and track circuit design.

The resource allocation model represents a cost-effectiveness type of decision-making methodology. The accident prediction models are used to estimate an accident rate for current conditions. Effectiveness ratios are then applied to predict the reduction in accidents expected to result from the implementation of a given upgrade in warning device. The larger the ratio of expected accident reduction per dollar cost of improvement, the more desirable is the proposed improvement. Recent research has emphasized the refinement of this basic methodology by incorporating accident history as an independent variable in the accident prediction models and by developing effectiveness ratios for various stratifications of the grade-crossing population (3-6).

Department of Civil and Environmental Engineering, 2204 Engineering Building, University of Wisconsin-Madison, Madison, Wisc. 53706.

EFFECTIVENESS RATIOS

Effectiveness ratios are defined as the percent reduction in accident rate for a grade crossing that has received a given upgrade in warning device. The ratios have been estimated by using standard statistical techniques (7,pp.150-187). However, what has not been widely recognized is that each effectiveness ratio is an estimate of a population ratio that represents the aggregate percent accident reduction over all grade crossings that receive a given upgrade in warning device. The 95 percent confidence intervals that are reported simply indicate that there is only a 5 percent likelihood that the true population ratio lies outside the stated range of values.

Caution must be exercised when an attempt is made to apply such ratios for the purpose of predicting the response of an individual grade crossing to a proposed improvement. Population ratios can be useful for policy studies in which the aggregate nationwide or statewide impact of a proposed program of capital improvements is of interest. In this application, the unique nature of each member of the population is not of concern. The confidence interval for the population ratio only indicates the variability of that ratio for the population as a whole. It does not reflect the variability in percent accident reduction for individual crossings.

The current interest in the estimation of population-based effectiveness ratios for various grade-crossing stratifications (single versus multiple tracks, fixed versus constant warning time track circuits, crossing angle, etc.) implies a recognition that the actual effectiveness of a warning device upgrade is dependent on the characteristics of the crossing that is being improved. The logical extension of this research would be the use of analysis of variance or multiple regression techniques and the incorporation of train and traffic volume data as a measure of the exposure to potential collisions. This, of course, is the approach used in the development of the DOT accident prediction models (1).

The use of population effectiveness ratios to predict the expected reduction in accidents for a given crossing improvement can be considered inappropriate for two reasons. First, the effectiveness ratios ignore the unique characteristics of a crossing, including exposure to collisions. Second, a direct estimate of accident reduction can be obtained by using the existing multivariate accident prediction models. The expected accident reduction would simply be calculated as the change in predicted accident rate assuming a specific warning device upgrade. The resulting estimated effectiveness would then account for the influence of those crossing characteristics included in the models.

Modeling Approach to Measuring Effectiveness

Two different sets of accident prediction models have been developed for use in the rail-highway resource allocation procedures. The original models (1), hereafter referred to as the old models, were calibrated by using the 1975 national inventory and accident data base. The predicted number of vehicle-train accidents per year is expressed as a function of the following variables:

- Type of warning device
- Average daily vehicular traffic
- Average trains per day
- Number of day through trains
- Number of main tracks
- Number of highway lanes
- Paved versus unpaved surface
- Urban population
- Highway functional classification

Recently, a new set of models (4,5) was developed by using the 1976 national inventory and accident data base. The principal differences between the two versions are that the new models include an exposure measure based on the product rather than the sum of train and traffic volumes as well as a new variable for maximum train speed.

Although no report has been published regarding the development and testing of the new models, a sensitivity analysis of the accident rates predicted by the old and new models reveals some substantial differences. To illustrate these differences, four typical grade-crossing environments were defined: a local rural road, a rural state trunk highway, a local urban street, and an urban arterial. The assumed characteristics of these hypothetical crossings are summarized in Table 1. The relationship between predicted accident rate and average daily traffic volume for a moderate train volume (five per day) at the four types of crossings is shown in Figures 1-4.

Figure 1 shows predicted accident rates for the assumed local rural road. The data reveal that the new models yield higher predicted accident rates than the old models. Two inconsistencies are also apparent. The new models predict a flashing-light accident rate that is greater than or equal to the crossbuck rate, and both of these rates are significantly greater than the crossbuck rate estimated by the old models. An additional observation is that the old models suggest that a flashing light is virtually as effective as a gate under low-exposure conditions. Although this may appear to be an inconsistency, it is not unreasonable for the assumed conditions.

With a rural trunk highway, the data in Figure 2 reveal that once again the new models yield higher predicted accident rates than the old models. The only significant inconsistency is that the new models predict a flashing-light accident rate that is about the same as the crossbuck rate predicted by the old models.

For the assumed typical local street in an urbanized area, it is demonstrated in Figure 3 that the old and new models produce similar results except in the case of flashing lights. For this condition, the new models predict accident rates that are substantially higher than those produced by the old models. Moreover, the new models suggest that a crossbuck is as effective, and sometimes more effective, than a flashing light. This is clearly inconsistent with both actual experience and intuition.

Finally, for the assumed urban arterial grade crossing, the data in Figure 4 again reveal that the new models predict a flashing-light accident rate that can exceed the crossbuck rate. The flashing-light accident rate predicted by the old models can approach and, under high train volumes, exceed the crossbuck rate. However, this apparent inconsistency may simply reflect that, at very high exposure levels, the predicted crossbuck accident rates represent unreliable extrapolations beyond the range of observable conditions in that such crossings are rarely equipped with anything less than automatic warning devices.

Because of the significant differences between the old and the new DOT accident prediction models, one set of models must be considered more reliable than the other. In the absence of any published documentation on the development and evaluation of the new models, it is concluded that the old models are superior because they do not exhibit the inconsistencies that characterize the new models.

Comparison of Predicted Accident Reductions

Notwithstanding the relative merits of the old versus new accident prediction models, the bias caused by the use of effectiveness ratios to estimate accident reductions as presently recommended by the U.S. Department of Transportation (4,5) can be illustrated by comparing these estimates with those that

TABLE 1 Assumed Grade-Crossing Characteristics

Characteristic	Local Rural Road	Rural Trunk Highway	Local Urban Street	Urban Arterial
Trains per day	1-25	1-25	1-25	1-25
Avg daily traffic	100-2,500	500-5,000	100-2,500	2,500-3,000
Maximum train speed (mph)	50	50	30	30
Day through trains (%)	50	50	50	50
Main tracks	1	1	1	1
Paved highway	Yes	Yes	Yes	Yes
No. of lanes	2	2	2	4
Urban population	NA	NA	100,000	100,000
Functional class	Local	Major collector	Local	Principal arterial

Note: NA = not applicable.

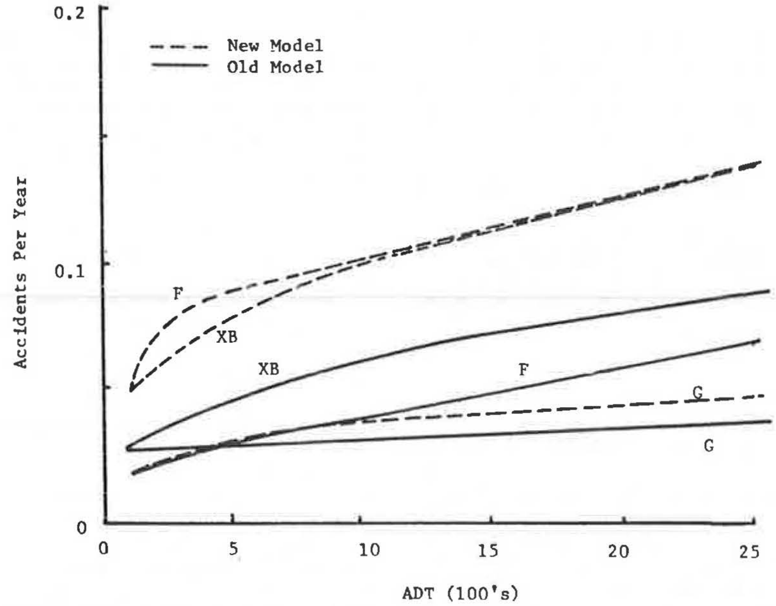

FIGURE 1 Predicted accident rate: local rural road, five trains per day (F = flashing signal, XB = crossbuck, G = gate).

would be obtained by a modeling approach. Using the same typical grade-crossing environments described in Table 1, expected accident reductions were calculated for crossbuck-to-flasher and crossbuck-to-gate upgrades. The modeling approach was performed by using both the old and the new DOT accident prediction models. The currently recommended effectiveness ratios (5) were applied to the "before" accident rate predicted by the new DOT accident prediction models.

For the assumed local rural road grade crossing, the data in Figure 5 reveal that the three methods yield significantly different predicted accident reductions. The use of effectiveness ratios over-

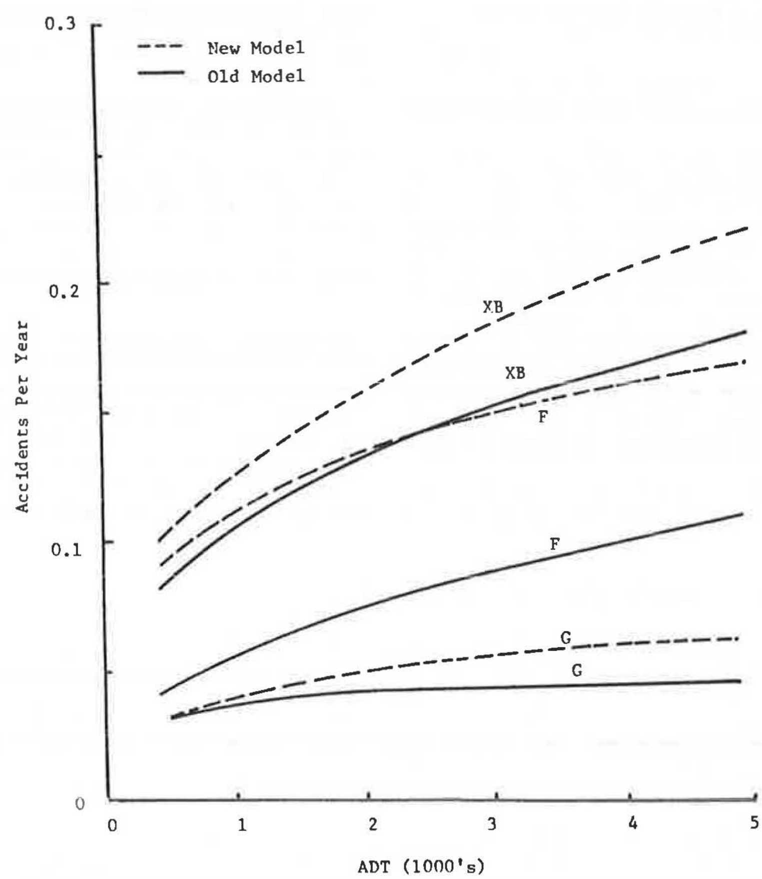

FIGURE 2 Predicted accident rate: rural trunk highway, five trains per day.

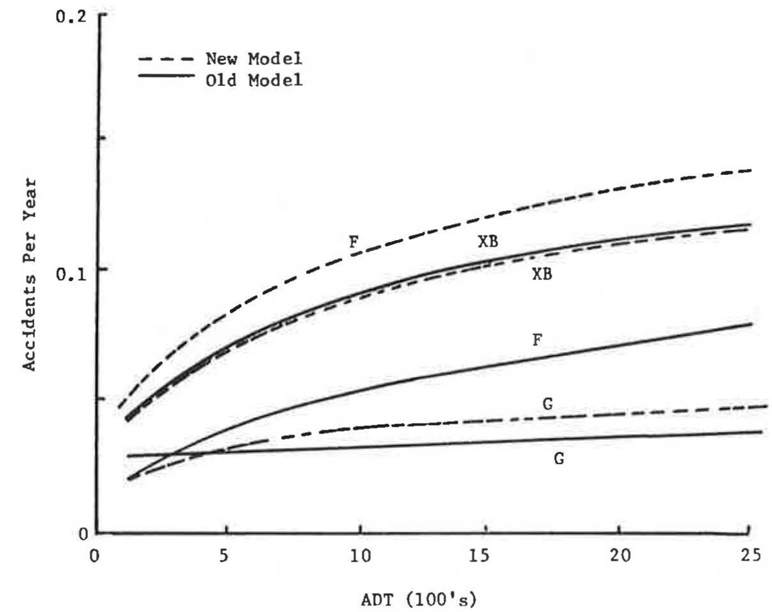

FIGURE 3 Predicted accident rate: local urban street, five trains per day.

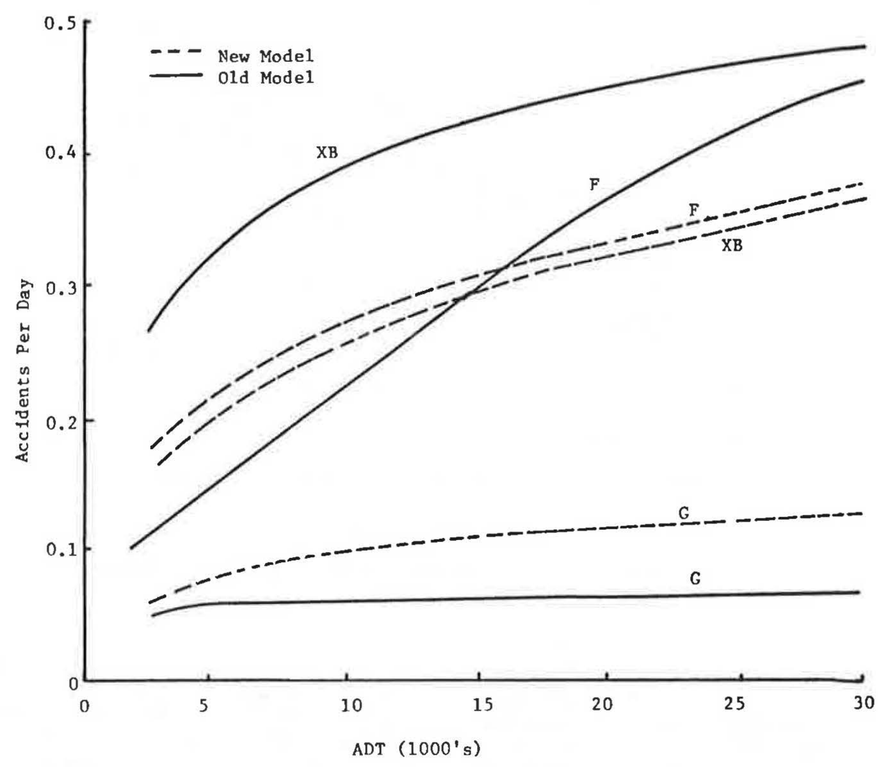

FIGURE 4 Predicted accident rate: urban arterial, five trains per day.

states the expected benefits of warning device upgrades compared by means of either modeling approach, especially for crossbuck-to-flasher upgrades. Use of the new models can also be observed to predict an increase in accidents following cross-buck-to-flasher upgrades at very low exposure crossings. This is a result of the inconsistencies associated with the new models, as previously discussed.

In Figure 6 the expected accident reduction that would result from flasher or gate upgrades on a typical rural trunk highway is shown. The use of effectiveness ratios again overstates the estimated benefits when compared with a modeling approach using either the old or the new DOT models. This is especially noticeable for the crossbuck-to-flasher upgrades. The effectiveness ratio procedure also indicates that the accident reduction expected with a gate upgrade is not much greater than that resulting from a flasher upgrade. The differences predicted by using a modeling approach are much greater for both the old and the new models, which is consistent with intuition.

The same pattern appears again in the case of the assumed local urban street grade crossing, as shown

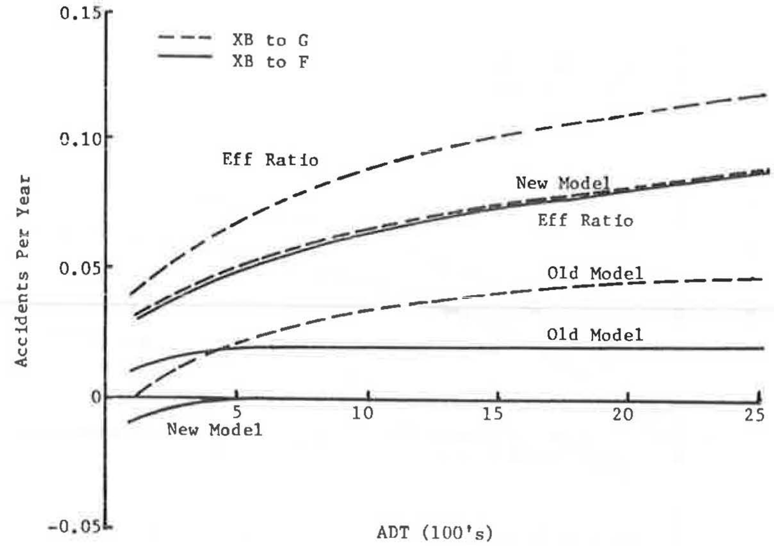

FIGURE 5 Predicted accident reduction: local rural road, five trains per day.

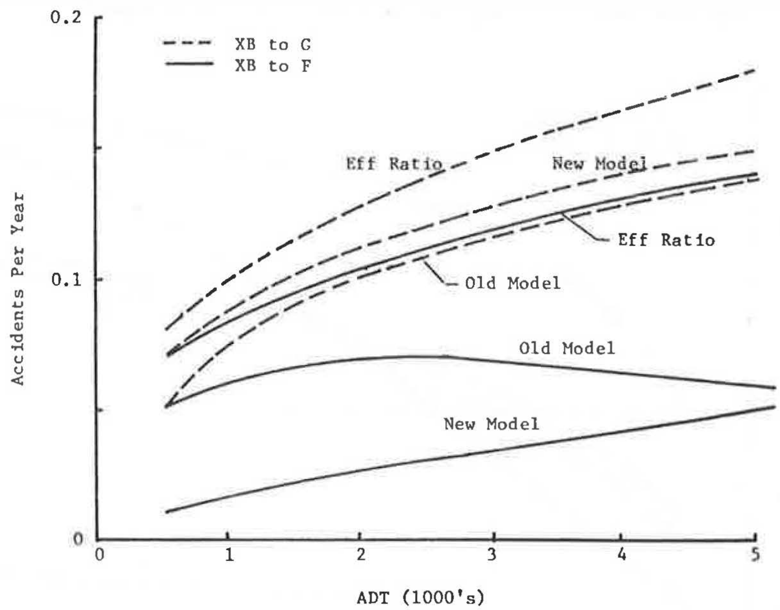

FIGURE 6 Predicted accident reduction: rural trunk highway, five trains per day.

in Figure 7. The problems associated with the accident prediction models are demonstrated by the negative accident reduction that is predicted following an upgrade from crossbucks to flashers. This obvious inconsistency does not occur when the old DOT models are applied.

Finally, for the case of the assumed urban arterial, the data in Figure 8 reveal substantial differences in the expected accident reductions following upgrades to either flashing lights or gates. Three inconsistencies are particularly noticeable. First, the effectiveness ratio procedure suggests that either a flasher or a gate upgrade would yield a comparable accident reduction. Second, the new DOT models imply that at low to moderate train volumes, an upgrade from crossbuck to flashing light would produce an increase in accident rate. Third, at high exposure levels, the old DOT models also predict an increase in accident rate following an upgrade to flashing lights. The inconsistency associated with the effectiveness ratio procedure is inherent in the method because of its insensitivity to specific crossing characteristics and its dependency on the new DOT models for an estimated "before" accident rate. The second inconsistency simply reinforces the observation that the new DOT models do not yield realistic estimates for a number of typical real-world conditions. The inconsistency noted with the old DOT models is probably due to the fact that, with very high exposure levels, it would be unusual to find a crossing that did not already have automatic warning devices. Under these conditions, a negative predicted accident reduction would be unreliable because the crossbuck model has probably been extrapolated to conditions beyond the range used for its calibration.

On the basis of the foregoing comparisons, it is evident that the currently recommended effectiveness ratio procedure does not yield accident rate reductions that are consistent with a modeling approach.

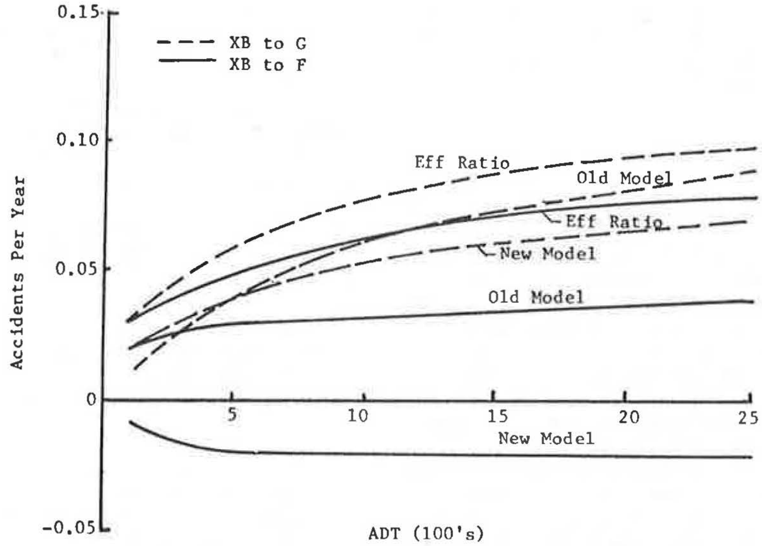

FIGURE 7 Predicted accident reduction: local urban street, five trains per day.

Moreover, use of the old versus the new DOT accident prediction models in the modeling procedure also produces inconsistencies. Although none of the three estimates necessarily provide an accurate reflection of real-world patterns, one of the methods must be considered superior to the other two.

Considering the effectiveness ratio procedure versus a modeling approach, it is concluded that a modeling approach to the prediction of accident reductions is preferred. The models are multivariate and can account for many of the unique characteristics of a grade crossing, especially train and traffic volumes. The effectiveness ratios do not have this capability. The effectiveness ratio procedure was also observed to predict comparable accident reductions for both flasher and gate upgrades under many conditions. This is tantamount to stating that flashers are almost as effective as gates, which is clearly inconsistent with empirical data.

If it is accepted that a modeling approach is preferred on the basis of both theoretical and empirical considerations, the remaining question is which set of accident prediction models should be used to estimate accident reductions resulting from

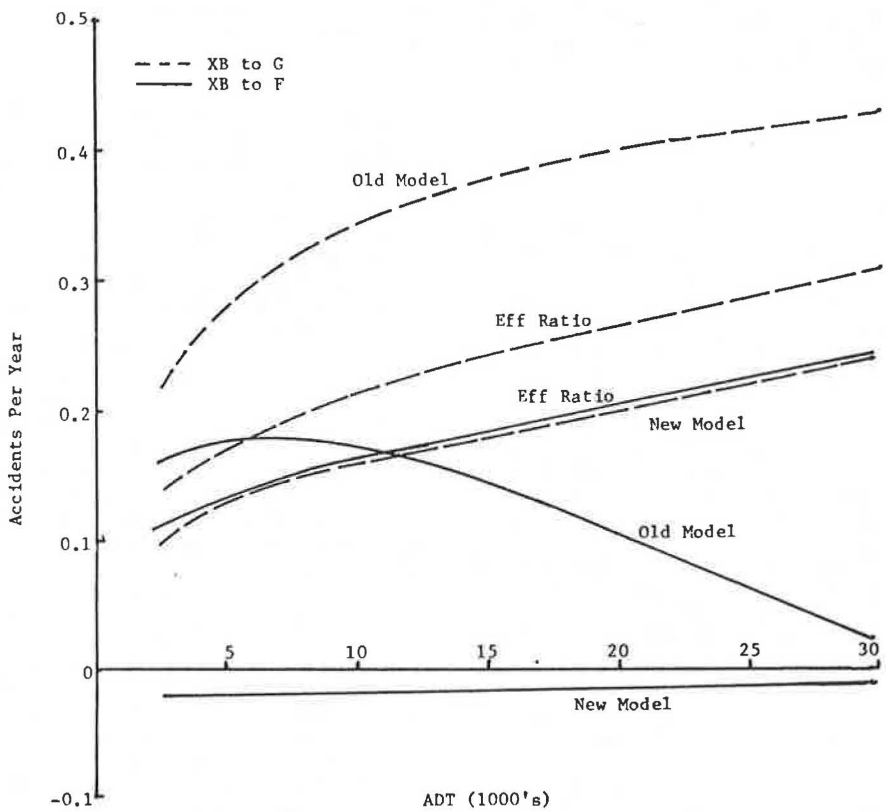

FIGURE 8 Predicted accident reduction: urban arterial, five trains per day.

warning-device upgrades. On the basis of the earlier comparison of the old and new DOT models, it is concluded that the old models (1) are more reliable and will produce more consistent results.

ACCIDENT HISTORY

Following the development of the old accident prediction models, DOT initiated work to incorporate accident history as a modifying factor to the basic accident prediction models. The premise was that the models were deficient because they did not base the estimation of hazard (i.e., the predicted accident rate) on actual accident history (1). In practical terms, this supposition is incorrect. Both the old and the new models were formulated and calibrated with accident history (aggregated by crossing years of experience) as the dependent variable.

Nevertheless, with the appearance of the new DOT accident prediction models came a procedure to adjust the model estimates on the basis of recent accident history. Rather than improving the reliability of the predicted accident rate reduction that would be input to the resource allocation model, the effect of this modification is to introduce a substantial bias, which can lead to a misallocation of resources. The fundamental problem with the method is that short-run rather than long-run accident predictions are being utilized. This violates a basic principle of engineering economics and resource allocation (8,9). When cost-effectiveness is used as the decision-making criterion in a capital budgeting exercise, the estimation of both effectiveness (i.e., accident reduction) and costs should be based on the economic life of the proposed investment. However, as noted in the original DOT research report (1,p.H-4), the objective of incorporating accident history in the accident prediction methodology was to determine the expected number of accidents that would occur during the next 1-year period. What should be of interest is the average number of accidents per year that would occur over an extended time period under the assumed conditions.

Regression to the Mean

The incorporation of accident history as a modifying factor to the basic accident prediction simply constitutes a method of estimating a phenomenon called regression to the mean (10). This is a pattern in which, when a random deviation from the mean occurs, it is expected that the next observation will be closer to the population mean. For example, if a 10 percent sample of grade crossings having the highest 3-year accident history is selected from a given region, then it would be found that, during the next 3 years, the accident experience at these crossings would be reduced even if no safety measures were implemented. Additional discussion and examples of regression to the mean may be found in a paper by Hauer and Persaud (10).

The accident history adjustment to the current DOT accident prediction models simply formalizes the phenomenon of regression to the mean through the use of Bayesian statistics. The net effect is to produce the following accident prediction methodology:

1. Examine the most recent accident history (5 years, if available) and calculate the accident rate for that time period;
2. Estimate the long-run expected accident rate for all similar crossings by using the DOT accident prediction model; and
3. Compare the historical accident rate with the expected long-run accident rate; if the most recent accident rate is greater than the mean (as estimated by the DOT models), then the accident rate during the next year should be lower (i.e., closer to the mean); if the most recent accident rate is less than the mean, then the expected rate during the next year should increase.

Clearly, for resource allocation purposes where investments are to serve productively for many years, it is the long-run accident rate that is of interest, not that which is expected to occur during the ensuing year.

An illustration of the phenomenon of regression to the mean is shown in Figure 9 for an assumed crossing having the following characteristics:

- Trains per day: 10
- Day through trains: 5
- Maximum train speed: 50 mph
- One main track
- Rural, paved, two-lane minor arterial
- Crossbucks

The basic predicted accident rate using the new DOT model is shown to increase with average daily traffic volume as expected. Assuming that there has been one accident during the most recent 5 years, the mean rate based on accident records is 0.2 accident per year as shown by the lower dashed line in Figure 9. If there had been two accidents during the 5-year period, the mean rate would have been 0.4 accident per year as shown by the upper dashed line. Application of the accident history modification to the basic rate results in the adjusted accident prediction curves. It is apparent that the net effect is to adjust the mean historical rate (dashed lines) closer to the predicted long-run accident rate. As shown by the arrows, these adjustments may be either increases or decreases.

Further demonstration of the phenomenon of regression to the mean, as well as the unreliability of population effectiveness ratios when applied to individual grade crossings, is shown in Figure 10. By means of data presented by Eck and Halkias (6), before-and-after data for upgrades from flashing lights to gates at 1,626 single-track crossings were stratified into two groups: those that had experienced an accident before the upgrade and those that had not. Figure 10 shows the before-and-after accident rates for both groups, as well as those for the total population.

Considering the 939 grade crossings that had no accidents during the period before the upgrade and assuming that no upgrade to gates had been made, it would be expected that the accident rate in the period after the upgrade would regress toward the mean of 0.24 accident per year (on the basis of the combined before-upgrade accident experience of all 1,616 crossings). If this were not so, then the before-and-after data for the 939 crossings would indicate that flashers are more effective than gates because the accident rate increased following the upgrade. Similarly, had gates not been installed at the 687 crossings that did experience an accident in the before-upgrade period, it would be expected that the accident rate during the after-upgrade period would regress downward toward the population average of 0.24 accident per year.

Finally, it is readily observed that when all 1,616 grade crossings are considered, the population effectiveness ratio for a flasher-to-gate upgrade is 0.71. However, at more than 50 percent of these

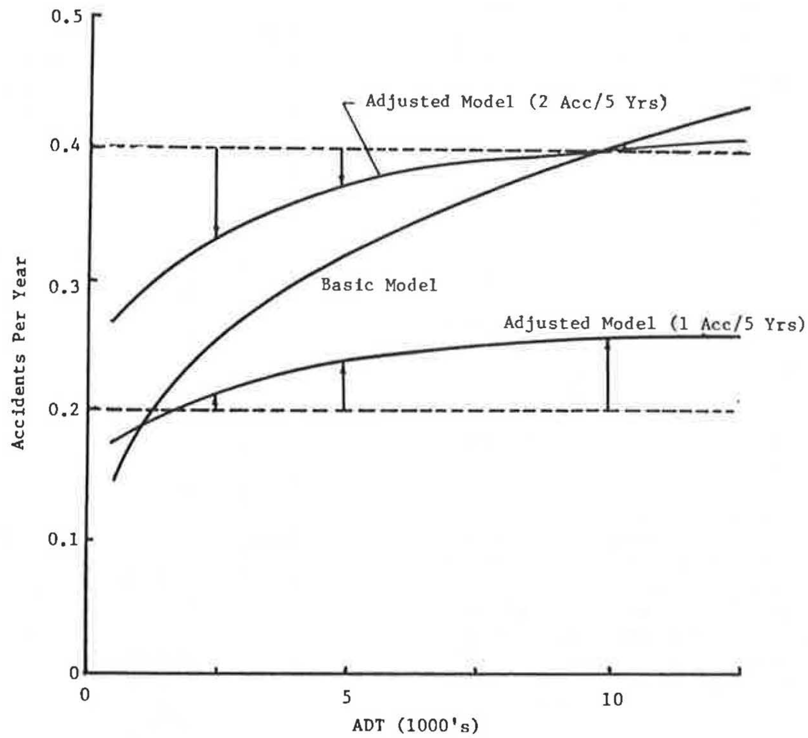

FIGURE 9 Influence of accident history on regression to the mean.

crossings (the 939 that had no accidents during the before-upgrade period), the actual upgrade effectiveness was negative. Clearly, application of a population ratio to a single crossing, or a subset of crossings, can lead to substantial short-run prediction errors.

Comparison of Hazard Rankings

The use of the accident history modification to the basic predicted accident rate can create a significant bias in the ranking of grade crossings in terms of relative hazard. For example, with data for 119

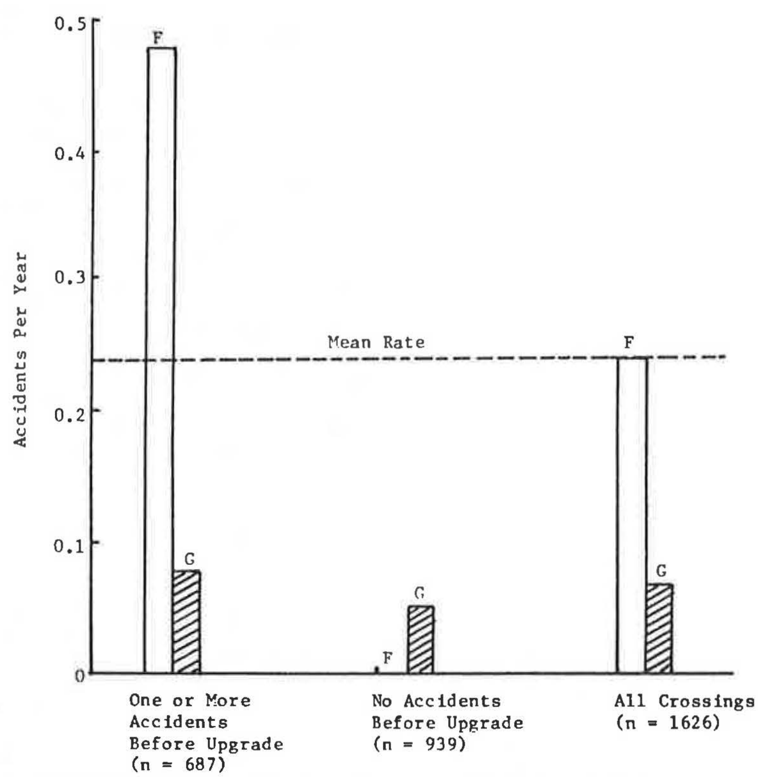

FIGURE 10 Influence of accident history on effectiveness of flasher-to-gate upgrades at single-track crossings.

grade crossings in three southeastern Wisconsin counties, three different sets of predicted accident rates for existing conditions were generated by applying the following models:

1. Old DOT accident prediction models (1),
2. New DOT accident prediction models with accident history adjustment (5), and
3. New DOT accident prediction models without accident history adjustment (5).

The crossings were then ranked within each set of predicted accident rates.

The three sets of ranked crossings were compared by selecting a decile from the list generated by using the old models and then determining the percentage distribution of these crossings over the various deciles in the remaining two distributions. In Figure 11 the results are shown for the comparison between the old models and the new models with the accident history adjustment. In Figure 12 the comparison between the old models and the new models without the accident history adjustment is shown. Both figures show the top three deciles from the rankings generated by the old models.

It is readily apparent that there is a significant difference between the three sets of rankings. The rankings resulting from the new models with the accident history adjustment are not at all similar to those generated by the old models. When the accident history adjustment is ignored, the differences in the rankings are reduced at the two highest deciles, but still remain substantial over the remaining deciles. The implication of these differences is that when the three modeling procedures are used to identify the most hazardous grade crossings in a given state or region, they will not identify the same set of crossings. Therefore, the models cannot be considered interchangeable, and one must be better than the other two. Because of the previously discussed inconsistencies associated with the new models, as well as the theoretical problems with the accident history adjustment, the old modeling procedure is considered to be the preferred method for ranking crossings on the basis of relative hazard.

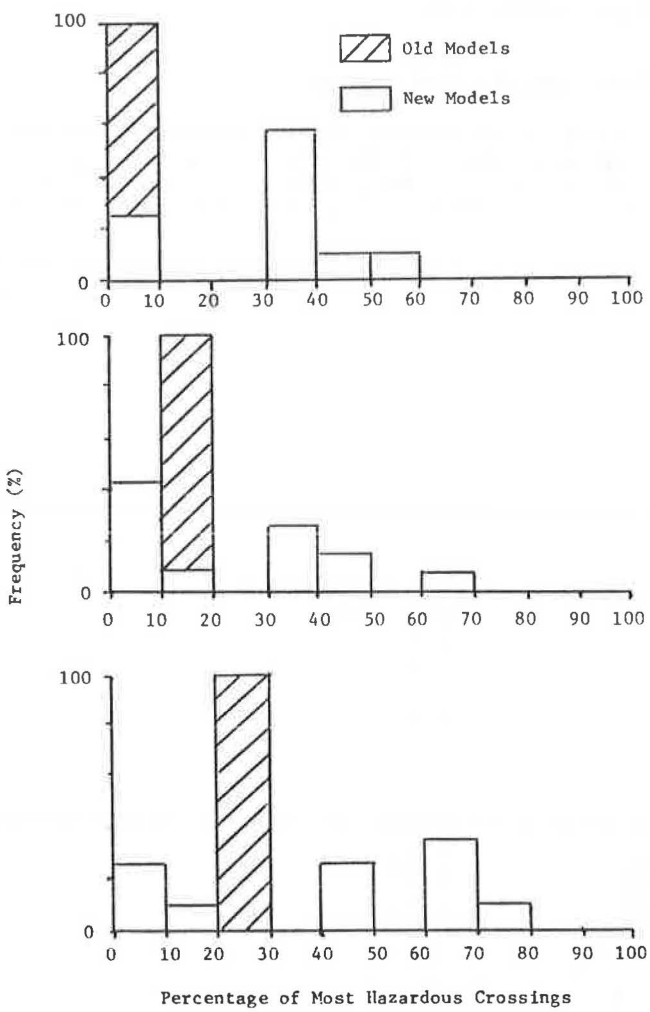

FIGURE 11 Comparison of hazard rankings: old models versus new models with accident history adjustment.

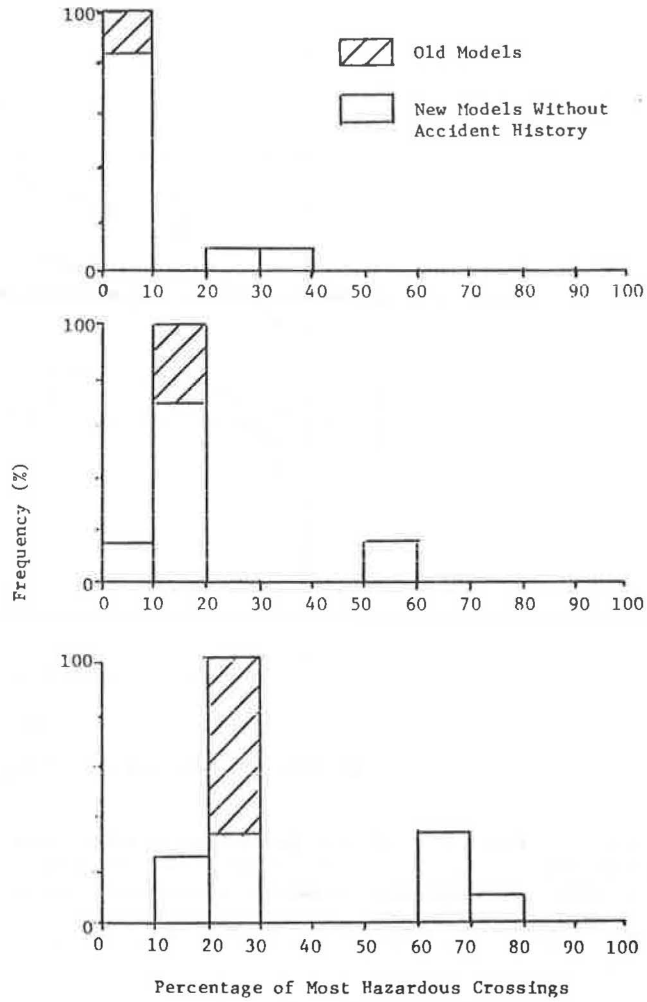

FIGURE 12 Comparison of hazard rankings: old models versus new models without accident history adjustment.

Bias in Resource Allocation

Use of the accident history adjustment in the calculation of predicted accident rates can produce significant bias in the selection and programming of grade crossing improvements. The extent of this bias will be demonstrated by using the previously described set of 119 grade crossings from southeastern Wisconsin and an assumed budget of $1,000,000.

The DOT resource allocation procedures (2,5) were first modified to accept accident reduction predictions generated by using the old DOT accident pre-

TABLE 2 Resource Allocation Results

					DOT Procedure					
		Old Models			With Accident History			No Accident History		
Identification No.	Existing Warning Device	Rank	Proposed Warning Device	Accidents Reduced per $1,000,000	Rank	Proposed Warning Device	Accidents Reduced per $1,000,000	Rank	Proposed Warning Device	Accidents Reduced per $1,000,000
176691X	F	1	G	17.37				1	G	6.47
176693L	F	2	G	12.49				2	G	5.64
176881B	XB	3	G	12.85				7	F	4.53
176703P	F	4	G	11.40				3	G	5.63
178978E	XB	5	G	8.09						
176723B	F	6	G	8.04				4	G	4.87
176729S	F	7	G	7.61				6	G	4.76
689838N	XB	8	F	7.25						
176721M	F	9	G	6.95	6	G	3.29	5	G	4.77
176704W	F	10	G	6.65	7	G	3.33	8	G	4.50
689861H	XB	11	G	6.29	19	G	2.43			
689854X	F	12	G	6.17	3	G	4.98	14	G	4.18
176727D	F	13	G	6.16				10	G	4.35
176732A	F	14	G	5.75				12	G	4.23
176865S	F	15	G	5.72				21	G	3.78
176720F	F	16	G	5.60				13	G	4.18
390521P	F	17	G	5.58	2	G	5.10	9	G	4.48
178927U	XB	18	F	5.31						
178936E	XB	19	F	5.31						
178945S	XB	20	F	5.31						
178968Y	XB	21	F	5.31						
178974C	XB	22	F	5.31						

Note: F = flashing light; XB = crossbuck; G = gate.

diction equations. A ranked list of crossing improvements was then created with the modified resource allocation model. The unmodified model was then run for the same set of crossings, both with and without the accident history adjustment. Accident reductions were estimated by using currently recommended effectiveness ratios.

The results from the three resource allocation exercises are summarized in Table 2. The grade-crossing improvements recommended using the old accident prediction models are ranked in order of decreasing cost-effectiveness. Both the current warning device and the proposed warning-device upgrade are also listed. Comparing these results with those generated by using the DOT procedure reveals substantial differences. The currently recommended procedure, which utilizes effectiveness ratios and accident history, shows only five crossings, which also appear in the base list. When accident history is ignored, there is better correspondence between the ranked improvements; however, significant differences still remain.

CONCLUSIONS

The application of cost-effectiveness procedures in the selection and programming of grade-crossing improvements should be widely encouraged. However, the use of effectiveness ratios and accident history as now recommended in the DOT resource allocation model (5) can introduce significant bias and lead to a misallocation of resources. A preferred procedure is to use the old DOT accident prediction models (1) to estimate expected accident reductions for alternative warning-device upgrades. The new DOT accident prediction models (5) can produce inconsistent results and are therefore not recommended. Adjusting predicted accident rates on the basis of recent accident history does not yield an estimate of the long-run average accident rate as is required in resource allocation procedures, and therefore is not recommended.

New research on the effectiveness of various safety measures at individual grade crossings should use before-and-after data for a sample of treatment and control sites (11). Various stratifications of crossing characteristics can be examined, and the statistical significance of the percent change in accident rate can be calculated. The statistical test of significance will reveal whether the estimated percent change can be reliably applied as a means of predicting potential accident reduction.

Additional research is now needed to permit incorporation of several new variables into the old DOT accident prediction models. These variables would include sight distance, track circuit design speed, presence of cantilevered versus post-mounted flashing light signals, and highway approach speeds. A recommended research approach would be to select a sample of grade crossings from the national inventory, collect the necessary field data at these crossings, and perform the appropriate statistical analyses. The goodness of alternative model formulations would be evaluated by testing the correlation between predicted and actual accident experience for clusters of ranked grade crossings. Each cluster might represent 5 percent of the total sample.

ACKNOWLEDGMENTS

This work was supported by the University of Wisconsin Engineering Experiment Station. Vinod K. Tuli assisted with computer programming.

REFERENCES

1. P. Mengert. Rail-Highway Crossing Hazard Prediction Research Results. Report FRA-RRS-80-02. Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., March 1980.
2. E. Farr. Rail-Highway Crossing Resource Allocation Model. Report FRA-RRS-001. U.S. Department of Transportation, April 1981.
3. J. Morrissey. The Effectiveness of Flashing Lights and Flashing Lights with Gates in Reducing Accident Frequency at Public Rail-Highway

Crossings 1975-1978. Report FRA-RRS-80-005. U.S. Department of Transportation, April 1980.
4. R. Coulombre, J. Poage, E. Farr, and J. Hitz. Summary of the Department of Transportation Rail-Highway Crossing Accident Prediction Formulas and Resource Allocation Model. Report DOT-TSC-FRA-82-1. Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., Sept. 1982.
5. J. Hitz and M. Cross. Rail-Highway Crossing Resource Allocation Procedure User's Guide. Report FHWA-IP-82-7. Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., Dec. 1982.
6. R.W. Eck and J.A. Halkias. Further Investigation of the Effectiveness of Warning Devices at Rail-Highway Grade Crossings. In Transportation Research Record 1010, TRB, National Research Council, Washington, D.C., 1985, pp. 94-101.
7. W.G. Cochran. Sampling Techniques, 2nd ed. J. Wiley and Sons, New York, 1963.
8. R. de Neufville and J. H. Stafford. Systems Analysis for Engineers and Managers. McGraw-Hill Book Company, New York, 1971.
9. R. Winfrey. Economic Analysis for Highways. International Textbook Company, Scranton, Pa., 1969.
10. E. Hauer and B. Persaud. Common Bias in Before-and-After Accident Comparisons and Its Elimination. In Transportation Research Record 905, TRB, National Research Council, Washington, D.C., 1984, pp. 164-174.
11. W.D. Berg. Experimental Design for Evaluating the Safety Benefits of Railroad Advance Warning Signs. Report FHWA-RD-79-78. FHWA, U.S. Department of Transportation, April 1979.

Discussion

John B. Hopkins*

I have known and respected William Berg for more than a decade and have often admired the depth and rigor of his work in a field of research not always characterized by those qualities. Indeed, we have in the past debated some of the points raised in his current paper. I regret that I must respectfully disagree with the conclusions he reaches. This position is based not on any flaw in the specific research or calculations he describes, but rather on differences in basic assumptions and perceptions concerning the topics discussed. These assumptions are inherently somewhat intuitive and subjective, but I suggest that my viewpoint may be closer to that of most practitioners in the field. The remarks that follow are intended not to criticize Berg's work but rather to illuminate the principal points that I consider to be at issue.

These comments are personal views based on my past activities and research in the field of crossing safety. They should not in any way be seen as necessarily reflecting or representing the position or viewpoint of the U.S. Department of Transportation or any of its components. I have not actually worked in this area for several years and was not personally involved in development of the DOT accident prediction and resource allocation models.

Berg first raises the issue of how warning-device effectiveness should be determined for estimating the potential benefits of upgrading crossing warnings. He argues that values based on accident prediction models are preferable to the before-and-after whole-population methods currently used. I agree completely that the effectiveness of specific devices is likely to depend on characteristics of the particular site and that ideally effectiveness would be considered to be a function of the several variables currently used in estimating accident probability: crossing angle, number of tracks, and so on. However, I strongly disagree that effectiveness values of this type can be determined by comparison of accident prediction equations for the same crossing with and without the proposed improvement.

Consider as an example the question of gate effectiveness. For a crossing with specified characteristics as found in the national inventory (rail and highway traffic, number of tracks, train speed, etc.), one can calculate the estimated accident probability with gates installed and with only crossbucks. Berg's view is that the ratio of those predictions (gates/crossbuck) is a good measure of the effectiveness of gates. However, this is true only if the characteristics on which the estimates are based (the inventory data) completely determine the hazard at each crossing, and it is widely accepted that they do not.

Many factors such as sight distance and obstructions, details of the type and distribution of rail and highway traffic, and visual distractions in the area can all have a strong impact on safety. It is to be expected, and is actually observed, that because of these effects, for any set of nominally identical crossings (in the sense of inventory characteristics) there is a distribution of actual hazard, with some crossings more dangerous than the predicted average value and some significantly safer. Historically these special (noninventory) factors, as well as actual accident experience, have played a major role in determining which crossings received upgrading and which did not. The strong likelihood is that for two crossings that have identical inventory characteristics except that one has gates and one does not, the gated crossing was--at the time of installation of gates--significantly more hazardous than the other. If indeed it has always (on average) been the more dangerous crossings that have received upgrading, it is inevitable that the ratio of gated to crossbuck accident prediction equations will underestimate gate effectiveness.

Consider the oversimplified but instructive hypothetical example of two crossings that are identical in terms of inventory entries but are so different in other ways (visibility or traffic distributions or whatever) that one has a "true" (if unknowable) hazard twice as great as that of the other. Subjectively aware of the greater accident potential, even if not able to quantify it, the authorities upgrade the more dangerous crossing with a device that has a "true" effectiveness of 50 percent--it reduces the number of accidents by half. Over a sufficiently long period of time, both crossings would then experience the same number of accidents. If this process of upgrading the more hazardous member of matched pairs went on over the entire crossing population for a long time, and one then sought to develop accident prediction equations for the two groups of crossings (with and without the 50 percent effective device), the two equations would be the

*Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass. 02142.

same for both groups, implying by Berg's reasoning that the device effectiveness was zero, rather than the actual 50 percent. Yet in a much more complex and convoluted way, this is just the process that created the real population of crossings on which the DOT (and all previous) accident prediction equations have been developed. Thus, I argue that ratios of accident prediction equations are virtually certain to understate warning-system effectiveness, and sometimes may even appear to defy logic and intuition by showing negative effectiveness. This effect is not a new one; it was apparent in the accident prediction work of Coleman and Stewart in 1972 (1), if not earlier. But, as discussed earlier, the paradox is easily resolved when one recognizes the inherent bias in the crossing populations from which the equations are developed.

Given that ratios of predicted accident rates cannot accurately estimate effectiveness, it becomes necessary to fall back on the simple effectiveness measures calculated from before-and-after data for relatively large populations. As noted earlier, I agree with Berg that it would be desirable to determine effectiveness as a function of the major relevant variables, but the large variation of crossing characteristics and the relatively small number of accidents so far have not provided a data base sufficient to achieve this goal and is unlikely to do so. Possibly the improved accident reporting system established during the last decade will in future years permit significant advances in this area. However, I also wish to defend the conventional population-based effectiveness numbers. They have been remarkably stable for more than 40 years (at least), with each new and more sophisticated study tending to confirm the previous findings. Given that accuracy of 5 to 10 percent is probably as much as is needed or meaningful (there are, after all, significant uncertainties and inaccuracies in all crossing-related data), I really doubt that the quality of decisions regarding crossing resource allocation would be significantly affected by more-nearly-perfect effectiveness data.

Arguing from the apparently greater inconsistencies in the newer model-derived effectiveness estimates, Berg states a strong preference for the older version of the DOT accident prediction models. Given my view that the models are virtually irrelevant to effectiveness measures, I do not find this argument to be convincing. My perception, as described earlier, is that the alleged inconsistencies (for example, between crossbuck and flashing-light crossings) are not flaws in the models but rather represent an entirely proper bias in the manner in which the crossings originally were chosen for upgrading. But more fundamentally, it should be noted that the newer accident prediction equations were developed by very much the same methodology as the earlier ones, differing primarily by drawing on a substantially larger and more robust data base. The statistical procedures and the underlying data are not questioned by Berg, and I am not clear what other grounds there can be for not accepting the newer results as being the best available, even if admittedly not perfect.

I will not attempt to deal with the issues raised concerning the statistical concept of regression to the mean and its relationship to the accident history adjustment to the prediction equations. This relates primarily to the testing and evaluation of the prediction equations, rather than to their conceptual validity. For me, the case is relatively simple. Most practitioners in the area of crossing safety have always believed that actual accident history should be considered in selecting crossings for improvement. This is more than simply an intuitive or subjective view. The crossing data base lacks information on some factors that are almost certainly relevant to accident probability and experience: sight distance, visual clutter or distractions, special traffic or geometric characteristics, and so on. The observed accident rate can, to some degree, serve as a surrogate measure for these factors. The accident history correction factor as used in the DOT model has a firm theoretical foundation (it is derived by using the same approach as insurance companies use to explore "accident proneness") and yields very reasonable results. That is, the relative weighting between accident experience and the unadjusted analytical model depends in a logical and intuitively reasonable way on the number of years for which history data are available, number of accidents, and the magnitude of the purely analytical value. The history-adjusted prediction thus incorporates actual safety experience in a consistent and rational manner, and is the first and only model I know of that does this. The alternative is to continue making this correction in a subjective fashion or to fall back on simpler but less valid rules of thumb.

In summary, I am in friendly disagreement with all of Berg's conclusions. In my opinion, effectiveness ratios based on whole-population before-and-after studies are the best estimate available, far preferable in theory and in fact to the use of ratios between accident prediction equations. I find the "new" DOT model to be a significant improvement over the old, because of the more robust data base on which it rests, and do not consider the apparent paradox of low implied warning-device effectiveness to be substantive. Finally, I see the accident history adjustment as a valuable addition to the standard model for cases in which accidents have been experienced or for theoretically high-accident crossings that actually have been accident free for several years.

As a concluding comment, I am moved to note that the whole area of accident prediction equations and hazard indices may be a horse sufficiently moribund as to warrant no further beating. The immense effort at improving crossings during the last two decades has virtually eliminated the high-hazard tail of the crossing distribution. I am very dubious that it makes any real difference which of the current models is used, because at the margin they all select among crossings that on average are roughly equivalent in hazard. Given the inevitable special factors affecting crossing safety and the inherent limitations of the data bases, as well as the large variations in improvement cost among crossings, further statistical elaboration is unlikely to yield any significant practical advance. I contend that research in accident prediction equations has met its goals and served its purpose. Although updating and refinement of the models should certainly continue, the basic objective--consistent, firmly founded models with accuracy adequate to the function--has been achieved, and further efforts are not likely to bring significant improvement in the cost-effectiveness with which resources are allocated. To the degree that research funds can be found in this area, they might better be devoted to examination of special attributes or situations--possibly not captured in the accident and inventory data or occurring too infrequently to show up in accident prediction models--that make certain crossings substantially more or less accident prone than the basic data and models suggest, and to the identification and evaluation of particular specialized countermeasures best suited to those cases.

REFERENCE

1. J. Coleman and G.R. Stewart. Investigation of Accident Data for Railroad-Highway Grade Crossings. In Transportation Research Record 611, TRB, National Research Council, Washington, D.C., 1976, pp. 60-67.

Author's Closure

The comments offered by Hopkins are welcomed and, it is hoped, will serve to highlight several of the fundamental concepts and issues that must be addressed by users of the DOT accident prediction and resource allocation procedures. Although Hopkins presents arguments and hypothetical situations to support his disagreement with the conclusions presented in the paper, I do not find his discussion persuasive and will briefly attempt to clarify the essential points made in the paper.

Hopkins argues that because the models do not directly account (i.e., include independent variables) for all relevant factors, their predictions will not account for the effect of these missing elements. This is correct. What should be emphasized, however, is the fact that effectiveness ratios do not account for the influence of any of the possible relevant factors. Therefore, compared with a multivariate modeling approach used in the old DOT models, effectiveness ratios constitute an even less sensitive indicator of the potential accident reduction that would result from a warning-device upgrade at a specific crossing.

With respect to consideration of the uncontrolled variables in the final decision about upgrading of warning devices, I do agree that these factors are important. It should be recognized that the resource allocation model simply generates a list of candidate improvement projects. Final recommendations should only be made following a field inspection, at which time the unique characteristics of the crossing can receive consideration.

Hopkins then theorizes that accident prediction models will always underestimate the effectiveness of a warning-device upgrade because, historically, the most hazardous crossings have received improvements first. Although it is certainly true that the worst crossings have traditionally been the first ones selected for improvement, this does not invalidate regression models that are calibrated with actual accident histories for the national inventory data base. In fact, both the old and new DOT models were adjusted so that the total number of predicted accidents for all crossings was equal to the actual total number of accidents reported to the Federal Railroad Administration. Furthermore, Hopkin's theory is of little consequence because the effectiveness ratio procedure for estimating accident reductions is directly dependent on the same accident prediction models of which he is critical. One simply begins with a "before" accident rate obtained either directly from the accident prediction models or alternatively by regressing the historical accident rate toward the mean rate as estimated by the accident prediction models. If the models are individually acceptable for establishing a "before" accident rate, there can be no logical reason for not using them jointly to estimate potential accident reductions.

Hopkins also suggests a hypothetical situation when all grade crossings have been improved so that all have the same accident rate. He furthermore suggests that accident prediction models developed at that point in time would show that all crossings would have the same predicted accident rate and that this in some way proves that accident prediction models underestimate warning-device effectiveness. In rebuttal, it should be noted that under Hopkin's scenario, the grade-crossing problem is finally solved because all crossings have the identical, and acceptably low, accident potential. Those improvements necessary to reach this goal can be readily determined by applying the old DOT models in an iterative procedure until every crossing has a predicted accident rate less than or equal to some threshold value.

Hopkins next refers to the grade crossing paradox, or the observation that some accident prediction models indicate that the accident rate will increase following an upgrade in warning devices. He suggests that the obvious inconsistency is due to a bias in the data base used to develop the models. This argument is simply not supported by the data presented in the paper. Although the new DOT models exhibit inconsistencies, the old DOT models do not. The old models were developed with 1975 data, whereas the new ones were developed with 1976 data. The differences in the models are due to the calibration process, not the data base. No explicit comments were made in the paper relative to the calibration process because the research reports that documented the procedures had been marked as nonreferable documents for internal DOT use only.

Regarding the accident history adjustment factor, Hopkins suggests that because insurance companies use regression to the mean, grade-crossing resource allocation procedures should be the same. However, the requirements of the insurance industry are significantly different from those associated with capital investment planning. The insurance industry wants to assign the next annual policy premium on the basis of the loss (or accident rate) expected during the same year, whereas engineering decisions must be based on consequences that occur over the economic life of the proposed improvement. The insurance company can alter its premium each year in response to actual accident experience, but a grade-crossing improvement is generally permanent once it is made. It is for this reason that the long-run expected accident rate should be used in resource allocation procedures.

Hopkins concludes by suggesting that it makes no practical difference what modeling procedure is used for allocating grade-crossing resources. I cannot agree because, as shown in Figure 11 and Table 2, the three basic procedures are more than marginally different. Therefore, one must be preferred, although even it may be imperfect.

In conclusion, I do not believe that Hopkins has offered reasonable rationale or empirical data to support his disagreement with the findings and conclusions presented in the paper. I expect that this debate will continue and hope that new insights will be gained and will in turn lead to more cost-effective use of grade-crossing safety resources.

Publication of this paper sponsored by Committee on Railroad-Highway Grade Crossings.