Development and Application of a Railroad-Highway Accident-Prediction Equation

Robert A. Lavette, Florida Department of Transportation

This paper reports the development of an accident-prediction equation for train-vehicle collisions at railroad-highway grade crossings that can be used as the basis for the establishment of a priority order for signal improvements. Most of the quantitative and physical factors in the grade-crossing environment were included. Of the 6000 public grade crossings in Florida, 1140 on state roads were used as the study base. The accident-prediction model was developed by the use of a stepwise regression analysis and three unconventional statistical techniques: (a) the analysis of the plots of the residuals, which indicated that a transformation was required (with the transformation of the dependent variable to a logarithmic form, the plot of the residuals was reasonably symmetric); (b) the observed interaction between the independent variables, which resulted in the use of dummy variables, particularly those for active (warning devices) times daily traffic and number of trains; and (c) a bias in the accident prediction that was introduced by the use of logarithms and eliminated by use of a nonlinear least squares adjustment. The accident-prediction model had a multiple correlation of 0.43. The independent variables in the model were the traffic, number of trains, vehicle speeds, train speeds, number of lanes, and presence of warning devices. The accuracy of the accident-prediction equation was demonstrated by comparisons of actual accidents to predicted accidents. The actual number of train-vehicle accidents in 1975 was 70 percent of the number predicted by the model. In 1975, the total number of accidents remained unchanged from that in 1974, but the number of train-vehicle accidents decreased 22 percent.

This paper presents a method for developing an accident-prediction equation for train-vehicle collisions at railroad-highway grade crossings and illustrates the benefits of such an equation to a transportation agency responsible for establishing a priority order for signal improvements. A stepwise regression analysis and three statistical techniques (transformation of data, use of dummy variables, and transformation of the accident-prediction model to its original scale) not previously employed in the development of an accident-prediction model were used.

In July 1972, the Florida Legislature authorized the Florida Department of Transportation (FDOT) to

> determine and adopt a program . . . for the construction cost of projects for the elimination of hazards of rail-highway crossings. Every railroad company . . . shall, upon reasonable demand and notice from the DOT, install, maintain, and operate at such a crossing an automatic flashing light signal, the design of which shall be approved by the FDOT.

The grade crossings on state-maintained roads had been previously inventoried. After the 1972 legislation, the remaining 5000 public grade crossings on local streets were also surveyed. The bases for the data collected were those physical factors that influence accidents (2). The field survey, which was conducted jointly by the FDOT and the railroad companies in 1973, was also part of the Federal Railroad Administration-Association of American Railroads national crossing survey. (This survey confirmed the earlier survey of grade crossings on state-maintained highways.)

The first step in the development of an accident prediction model for rail-highway grade crossings was a review of the existing statistical data and the previous publications on the subject and led to the following conclusions (1):

1. It is imperative that the data base be as accurate as possible. Previous studies did not indicate its verification.
2. Previous reports did not take advantage of all of the statistical techniques available, such as analysis of the residuals and dummy variables.
3. The theory of linear statistical models could be applied conveniently to the available data.
4. Many variables involved in a train-vehicle accident were not included in the data; thus, any model selected would have considerable inherent variation in the number of accidents at a particular crossing.

STATISTICAL ANALYSIS OF THE DATA

Data

The main object was the determination of the relative influences of selected physical factors on the number of train-vehicle collisions at rail-highway grade crossings in Florida. The data analyzed were limited to those features that FDOT could modify and for which complete data were available. For example, there were no data available for the analysis of driver behavior, the optical effectiveness of the railroad signals, or driver-traffic characteristics. A complete listing of the data collected and analyzed is shown below.

1. Maximum posted train speed in miles per hour,
2. Average number of trains per day,
3. Highway system (e.g., Federal-Aid Primary),
4. Rural or urban location of crossing,
5. Road facility (e.g., arterial),
6. Number of lanes,
7. Posted crossings speed limit in miles per hour,
8. Average daily traffic in units of 1000 vehicles/d,
9. Warning device, type 1 (crossbucks, flashing lights, and such),
10. Warning device, type 2 (illumination or stop sign),
11. Minimum approach distance in feet (sight distance to crossing),
12. Parallel road characteristics (within 61 m of track),
13. Minimum clear sight distance in feet (triangle to train),
14. Quadrants with minimum clear sight distance,
15. Maximum clear sight distance in feet,
16. Quadrants with maximum clear sight distance (I, II, III, or IV),
17. Year most recent protection device was installed,
18. Rate of accidents for 5-year period,
19. Number of accidents in 1967,
20. Number of accidents in 1968,
21. Number of accidents in 1969,
22. Number of accidents in 1970,
23. Number of accidents in 1971, and
24. Total number of accidents for 5-year period.

(The model derived in this paper is designed for U.S. customary units; therefore the variables in the tables and equations are not given in SI units.)

The data sources included the FDOT annual inventory and traffic counts (average annual daily traffic) of all state-maintained roads and the field inventory, which confirmed the physical data, including the measurement of approach and triangle (quadrant) sight distances. The train speeds and the number of trains per day were obtained from railroad company timetables and verified by station masters.

Since a careful review showed that the accident history for 1967 (25 percent fewer accidents than in other years) was unreliable, only the data for the years 1968 through 1971 were used in the study. Of the 1155 state-maintained crossings, 1140 were selected for use. Although there were data available for 225 city-maintained crossings, these were on one railroad line with the same number of trains per day, and with no variation in an important independent variable, its regression coefficient could not be determined. Therefore, that sample was not further investigated.

Tables of the basic characteristics of the state-maintained crossings were compiled. Good distribution was obtained for train speed, vehicle speed, traffic counts, and land use (urban, rural, or municipality with a population of less than 5000 persons). The number of trains per day was predominantly in the lower range: Of 1140 crossings, 622 had fewer than 5 trains/d and 1048 had fewer than 15 trains/d. Only 10 crossings had more than 30 000 vehicles/d. (The FDOT is not confident of the accident predictions for crossings with traffic counts above 30 000 and those with fewer than 1 train/d, since the predictions appeared too high.)

Basic Model

Only those statistical methods not discussed in the previous report (1) are fully discussed in this report.

Analysis of the Residuals

The multiple regression model shown below implies that the relation between the dependent variable (y_j) and the ith dependent variable is linear (1).

$$y_j = \beta_0 + \beta_1 x_{1j} + \ldots + \beta_k x_{kj} + \epsilon_j \qquad (j = 1, \ldots, N) \qquad (1)$$

where

y_j = observed number of accidents at crossing j for a particular time period,
x_{kj} = value of kth characteristic for crossing j (assumed constant for the specified time period),
β_k = regression coefficient for kth crossing characteristic, and
ϵ_j = unexplained residual variation.

The usual statistical assumptions for a regression model are that the errors ϵ_j have a mean of zero and a constant variance, and are uncorrelated and normally distributed. These assumptions are never exactly satisfied in real life.

Again, if y_j, the independent variable, is the number of accidents per year, the variance of y_j will likely depend on the level of the x_{ij}, i.e., in effect on $\sum_{i=1}^{k} \beta_i x_{ij}$. The reason for this is, in part, that the dependent variable must have at least Poisson variation, and the variance of the Poisson distribution is equal to the mean. To stabilize the variance, a transformation is needed and the type of transformation will be suggested by an examination of the residuals. . . .The original form of the dependent variable y_j was the raw number of accidents for the period observed at crossing j. On the basis of the Poisson nature of accidents at a crossing, regressions involving the raw score (y_j) were quickly discarded. Plots of the residuals ($y_j - \hat{y}_j$) against the independent variables such as the average number of trains per day, indicated that the distribution was highly skewed and some transformation was in order.

The same was true for independent variables such as train speed and average daily traffic.

Data Transformations

A square-root transformation was attempted prior to the logarithmic transformation since this would stabilize the residual variance at 0.25 for a Poisson random variable, but the plot of the residuals ($y_j^{1/2} - \hat{y}_j^{1/2}$) against the average daily traffic unfortunately exhibited considerable skewness (1).

The square-root model produced a higher multiple correlation than the model finally adopted (R = 0.4528 against R = 0.4285 respectively), but the increase was not judged to outweigh the advantage of the symmetric distribution of residuals produced by the logarithmic model.

There may be some question as to why none of the regressions reported here explained more than 20 percent (100 R^2) of the variance while those reported by Coleman and Stewart (8) explained from 63 to 78 percent. This may be due to the fact that, besides the differences in sample size and area covered, Coleman and Stewart were fitting to grouped data so that their R^2 figures pertain to variances of group means. The variance of individual crossings within groups is not considered, although it would have a substantial effect on the variability of a prediction for a single crossing.

From the report by van Belle, Meeter, and Farr (1),

The residual mean square, after fitting the square-root model, was 0.288 as compared to the theoretical variance of 0.25. Since the variance about the mean (for the square-root model) was 0.359, a significant reduction in variance was produced by fitting the model, but a substantial amount of variation (0.288 versus 0.250) remained unexplained.

Logarithmic transformations on independent variables that had large coefficients of variation were also positively skewed.

The transformation $\ln(y_j + a)$ was selected with the value of a determined from a plot of the residuals. Initial values of a tested were 1, ½, and . . . finally 0.04.

The residuals were reasonably symmetric for a = 0.04. An additional reason for this particular choice of a is that an analysis on an annual basis using $\ln[(y_j/4) + 0.01]$ would differ from the present analysis by only a constant.

Partitioning the Sample

Besides the investigation of several transformations, the data were partitioned into two samples by the rural-urban as well as the active-passive dichotomies. This approach resulted in equations which fitted the data as closely (in the sense of multiple R) as the approach finally adopted: i.e., the dummy variable techniques. The latter was used because dummy variables allow the selective interaction of variables.

However, partitioning the sample produces separate estimates for all of the parameters in both samples, whether significantly different or not. Also,

dummy variables allow the examination of more combinations of dichotomies without reducing the sample size.

Finally, the splitting of the sample on the basis of the active-passive warning devices created the problem of crossings whose warning devices were improved during the study period. These crossings had to either be discarded or be allocated into both the active and passive groups, which would artificially increase the sample size.

Influence of Sight Distances

One aspect of the partitioned samples that was not present in the final model is the introduction of independent variables of sight distance in relation to the required stopping sight distance. These two variables were the ratio of the available approach sight distance to the crossing to the desirable required stopping sight distance and the ratio of the quadrant clear sight distance (the sight triangle of the approaching train measured along the road) to the desirable required stopping sight distance (3). The quadrant clear sight distance is similar to that described by Schoppert and Hoyd (2) and is the distance from the tracks at which a line of sight to the approaching train would be obstructed.

The data were partitioned into rural versus urban area crossings and also into active versus passive warning-device crossings to observe the effect of sight distance. Those crossings at which the warning devices were modified during the study period were eliminated from the active versus passive partition. The approach sight distance variable was significant only for the urban partition with an F-value of 3.8 (multiple R was 0.46). In Florida, with its flat terrain, there are very few crossings with restricted approach sight distances, particularly if the stopping distances for dry pavement are used. The required stopping sight distance is based on a wet pavement condition.

The effect of the clear sight distance variable on the partitioned models was significant. The F-values for the rural and passive partitions were 7.9 and 5.7 respectively (the multiple Rs were 0.40 and 0.51 respectively). The critical F-value at the 0.05 significance level was 3.84, whereas the F-values for the urban and active partitions were only 0.2 and 1.4 respectively. The low highway speeds in urban areas and the fact that most crossings in these areas have active (train-activated) warning devices mean that sight distances have minimum significance in urban areas or at crossings with active warning devices. These models did not fully use the analyses of residuals or dummy variables. In retrospect, two mistakes were made in analyzing the sight-distance data. One was that dummy variables should have been used for the active versus passive condition. The other was that, since it was unusual for wet pavement to be a contributing factor in a train-vehicle collision, the required stopping distances for dry pavement should have been used (for study purposes only, not for design practices). Thus, if properly analyzed with the use of dry pavement conditions, the clear sight distance would be a significant variable, but, in Florida, the approach sight distance would not be significant.

Of the 1140 grade crossings surveyed, 95 percent had at least one quadrant in which driver vision was obstructed, whereas the sight distance to the crossing itself was obstructed for only 9 percent of the crossings. These sight distances were calculated and surveyed by using the required stopping distance on wet pavements (plus a perception and reaction time of 2.0 s) (3).

The independent variable used to determine the significance of quadrant sight distance was

$$\text{ratio } C = (\text{minimum clear sight distance}/\text{required stopping sight distance}) \times 10 \tag{2}$$

For a dry pavement condition, not only is the denominator (the required stopping sight distance) reduced but also the hypotenuse and base of the sight triangle are substantially reduced. The field survey was conducted by observing a point on the track that formed the angle of the base and hypotenuse of the sight triangle. The base distance along the track (the critical approach distance) is the distance a train travels at its maximum allowable speed during the time that it takes a passenger vehicle to stop. Thus, if the passenger-vehicle stopping time were reduced, the base distance along the track would also be reduced. Since the minimum clear sight distance is the distance along the highway from the track to a point where the driver cannot view a train at its critical approach distance, the number of crossings where the minimum clear sight distance equals the required stopping sight distance would be increased.

Selection of Dummy Independent Variables

To account for the effects of automatic train-warning devices on the other independent variables, categories of basic warning devices were established as shown below.

Category	Variable: Flashing Lights (PD211)	Gates (PD27)	Advance Light (PD29)	Active-Passive Code
Passive	0	0	0	A = 0
Active 1 (flashing lights)	1	0	0	A = 1
Active 7 (flashing lights and gates)	1	1	0	A = 1
Active 9 (A7 plus traffic signal preempted)	1	0	1	A = 1

Thus, PD211 (flashing lights) is nonzero only when there are active warning devices; PD27 (flashing lights and gates) is nonzero only when there is gate protection. The regression coefficients associated with these variables indicate the additional reduction in ln(number of accidents + 0.04) when a particular automatic warning device is present (1). For example, the regression coefficient associated with PD211 estimated the additional reduction in the dependent variable due to an active warning device. All three variables were included in a regression equation with coefficients b_1, b_7, and b_9 if active devices 1, 7, and 9 were present at this crossing. Thus, a priori, the coefficients are expected to be negative (warning devices should decrease the number of accidents), and b_7 should be greater than b_9 since the most protection should produce the greatest decrease.

One problem associated with the data was that the warning devices at some of the crossings were modified during the study period. The assumption was made that... such modifications occurred at the midpoint of the year and the nominally 0-1 variables for these crossings were coded as follows:

	Year of Modification: 1968	1969	1970	1971	
1(Active)	$\frac{7}{8}$	$\frac{5}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	0(Passive)

Thus, if a crossing were modified from passive (dummy code 0) to active (code 1) in the year 1970, it was assumed to be active for 3/8 of the 4-year study and was coded with this value.

Similar dummy variables were established for the rural versus urban categories. The 25 dummy variables listed in the table below were derived from the basic variables listed in Table 1.

Variable	Description	Variable	Description	Variable	Description
15	A x 1	24	R x 1	32	R x 13
16	A x 2	25	R x 2	33	R x A x 1
17	A x 6	26	R x 4	34	R x A x 2
18	A x 7	27	R x 5	35	R x A x 9
19	A x 8	28	R x 9	36	R x A x 10
20	A x 9	29	R x 10	37	R x A x 11
21	A x 10	30	R x 11	38	R x A x 12
22	A x 11	31	R x 12	39	R x A x 13
23	A x 12				

During the development of the regression model, considerable attention was given to the interactions among the independent variables.

Variables can interact, that is, their joint effect on the dependent variable could be markedly different from the sum of their individual effects. For example, the effect on the accident rate of adding active warning devices . . . varies as the average daily traffic varies. To allow for this, additional independent variables were constructed by multiplying the active-passive dummy variable A (flashing lights and flashing lights and gates) and the rural-urban dummy variable R by other independent variables. These variables are denoted by A x 1, A x 2, . . . , R x 1, R x 2, etc.

For example, A × 1 is the interaction of a kind of automatic warning device with variable 1 (ln of maximum posted speed). The actual selection of the variables and the interactions entered into the program was often the result of knowledge of the grade-crossing environment and not necessarily the result of the stepwise regression procedure. Some variables, such as the crossing speed limit, can be altered, but others, e.g., the location of the crossing, cannot. In this kind of situation, it is not helpful to say that urbanization causes more accidents at grade crossings.

Although certain interaction variables were forced into the regression program to observe their effect, they did not improve the final model. For example, when the variable A × 1 (ln of the posted maximum train speed) was forced into the model, it had an acceptable F-value, but another variable, A × 2 (ln of the number of trains per day) dropped out.

Accident-Prediction Model

The final stepwise regression analysis, after 20 analyses, involved the 39 independent variables listed above and in Table 1. The standard error of estimate for the final regression was 1.52 and the multiple correlation was 0.43. Eight independent variables, shown in Table 2, had F-ratios greater than 7. The critical F-value at the 0.05 significance level is 3.84; however, the increase in predictive variance precluded the addition of other predictor (independent) variables. The model selected was

$$\begin{aligned}\text{predicted } \ln(y + 0.04) = & -8.0757 + 0.4368\,[\ln(\text{ADT})] \\ & - 0.1440[\text{A} \times \ln(\text{ADT})] \\ & + 0.3178[\ln(\text{maximum train speed})] \\ & + 0.4838[\ln(\text{number of trains per day})] \\ & - 0.3180[\text{A} \times \ln(\text{number of trains per day})] \\ & + 0.3870[\ln(\text{crossing speed limit})] \\ & + 0.2249(\text{A} \times \text{number of lanes}) \\ & - 0.4662(\text{PD27}) \end{aligned} \tag{3}$$

where y is the total number of accidents for 1968-1971 (1).

Four of the eight variables are expected to be involved in any model for accident prediction at grade crossings: daily traffic volume, maximum train speed, number of trains per day and crossing speed limit. All of these independent variables have positive regression coefficients.

Hence, they are positively correlated with the number of accidents at a grade crossing.

The other four independent variables involve the nature of the warning device at a crossing. The first of these A × ln (ADT), i.e., the effect of an active rather than a passive warning device, varies with the level of the traffic volume. In particular, for crossings with passive signing, the predicted ln(y + 0.04) is increased by 0.4368 for each unit increase in ln(ADT), whereas for crossings with active warning devices, the predicted ln(y + 0.04) is increased by (0.4368 - 0.1440) for each unit increase in ln(ADT). This is because the variable [A × ln(ADT)] is nonzero only for crossings with active warning devices. The interpretations for the other interactions are similar.

Some of the active versus passive dummy variables are highly correlated, such as A × 10 (crossing speed limit) and A × 2 (number of trains per day) or A × 9 (number of lanes) and could have been substituted for each other with little effect on the predictive accuracy of the fitted equation.

Transforming the Accident Prediction

The use of the logarithmic transformation for the dependent and independent variables gives a statistical model that more closely satisfies the least squares assumption. This model also provides a method for scheduling crossing improvements that is based on accident prediction on a logarithmic prediction scale. When the logarithmic form of the model (y = 0.04) was transformed to the original scale, a substantial negative bias was introduced. To obtain an unbiased transformation, Beauchamp and Olsen (5) used a complicated procedure to derive estimates for the mean of a lognormal variable that depends on a single independent variable, but a simpler approach was used here. The objectives were that the sum of the predicted accident rates equal the sum of the actual number of accidents; that all predictions be nonnegative; and, subject to this, that the predictions should satisfy a least squares property. Thus, for i = 1, . . ., 1140, x_i = 4-year total accidents at crossing i, $\hat{y}_i$ = least squares estimate of rate obtained from crossing i [in the ln(x + 0.04) scale], and $\hat{x}_i$ = predicted x_i in the original scale.

Let

$$\hat{x}_i = \exp(\alpha\beta\hat{y}_i) \tag{4}$$

where we should obtain $\Sigma\hat{x}_i = \Sigma x_i$. Let T = Σx_i; then

$$\exp(\alpha) = T/\Sigma(\exp\beta\hat{y}_i) \tag{5}$$

so that the estimator now depends only on β, i.e.,

$$\hat{x}_i = [T/\Sigma(\exp\beta\hat{y}_i)]\ (\exp\beta\hat{y}_i) \tag{6}$$

A value of β is chosen to minimize S where

$$S = \left\{x_i - [T/\Sigma(\exp\beta\hat{y}_i)]\ (\exp\beta\hat{y}_i)\right\}^2 \tag{7}$$

A computer program written to evaluate S gave values of α = 1.109 and β = 0.968 (1).

Table 1. First-order variables used in stepwise regression analysis of accidents at crossings on state-maintained roads.

Variable	Description	Mean	Standard Deviation
1	ln of posted maximum train speed[a]	3.373	0.673
2	ln of number of trains per day[a]	1.482	0.891
3	PD211 [flashing lights (warning device)]	0.504	0.470
4	PD27 (flashing lights and gates)	0.127	0.323
5	PD29 (flashing lights, gates, and preemption)	0.007	0.081
6	Rural versus urban, category 2, small municipality	0.161	0.386
7	Rural versus urban, category 3, urban characteristics	0.072	0.258
8	Rural versus urban, category 1, rural	0.503	0.500
9	Number of lanes	2.493	0.993
10	ln of crossing speed limit[a]	3.698	0.332
11	ln of average daily traffic[a]	7.715	1.503
12	Ratio C[b]	2.733	2.645
13	Ratio D[c]	9.592	1.450
14	ln of (total accidents - 0.04)[a]	-2.120	1.677

[a] Logarithms are to base e.
[b] Ratio C = coefficient of impaired view of approaching train = 10 × minimum clear sight distance/required stopping sight distance.
[c] Ratio D = coefficient of impaired view of a crossing protection device = 10 × minimum approach distance/required stopping sight distance.

Table 2. Final stepwise regression analysis: variables retained, regression coefficients, and F-values (state-maintained roads only).

Step	Variable	Entered Description	Final Regression Coefficient	F-Value	R-Correlation
1	11	ln(avg daily traffic)[a]	0.436 8	139.1	0.280 8
2	22	A(active) × 11 (avg daily traffic)	-0.144 0	16.9	0.343 0
3	1	ln(maximum train speed)[a]	0.317 8	15.6	0.379 5
4	2	ln(number of trains per day)[a]	0.483 8	30.1	0.395 1
5	16	A × 2(number of trains per day)	-0.318 0	7.3	0.406 9
6	10	ln(highway speed limit)[a]	0.387 0	7.8	0.415 7
7	20	A(active) × 9 (A × number of lanes)	0.224 9	9.0	0.421 8
8	4	PD27 (flashing lights and gates)	-0.466 2	7.9	0.428 5
—	—	intercept (b_0)	-8.075 71	—	—

Notes: Dependent variable is ln(number of accidents + 0.04); multiple correlation R = 0.4285; standard error of estimate = 1.52; F-ratio = 31.793; and critical F-value = 3.84 at 0.05 confidence level.

[a] Logarithms are to base e.

MODIFICATION OF MODEL

Modification of Regression Coefficients

The regression model selected indicated that when the coefficient for the number of lanes was entered into the model (0.225 × number of lanes) its effect offset the value of the gates (-0.466). From an engineering standpoint, it is obvious that gates of adequate length will drastically reduce the sight restrictions at multilane roads, where a driver's view of signal lights, for example, could be obstructed by a truck in an adjacent lane. However, two-lane roads are the normal condition and should not increase the risk when active warning devices are present. Thus, the number of lanes was allowed to affect the model only when there were more than two lanes and to not affect the model when gates were present. To offset this change, the coefficient for gates was reduced from -0.466 to -0.233.

Sight distances—the ability of the driver to view the approaching train and to see the warning signs or flashing lights—definitely are part of train-vehicle accident prediction and consequently are part of the accident-prevention environment, and these independent variables were significant when the data were partitioned into rural versus urban or active versus passive categories.

The coefficients for C (clear sight distance) and D (approach sight distance) in the stepwise regression models apply only to passive signing for C, and only to crossings in urban areas for D. These coefficients were used as a guide to derive the following terms:

$$C = 0.33 - (\text{minimum clear sight distance}/\text{required stopping sight distance}) \times 10 \times 0.123 \quad (8)$$

When C is less than zero, this term is not used.

$$D = 0.28 - (\text{minimum approach distance}/\text{required-stopping sight distance}) \times 10 \times 0.028 \quad (9)$$

Examination of the clear sight distance term shows that it increases the accident prediction only when the distance from the track at which the driver can first view an approaching train is no more than one-fourth of the required stopping sight distance (3) on wet pavement. Of the 1140 grade crossings examined, 65 percent had at least one quadrant in which this occurred. However, the minimum approach distance term increases the accident prediction whenever the approach (sight) distance (the ability to see the crossing) is less than the required stopping sight distance. But, only 9 percent of the crossings had any sight-distance restriction to the crossing.

Calculations of the reduced clear sight-distance triangle made by using the stopping distance for dry pavement and assuming that the obstacle restricting the view of the approaching train was 4.6 m (15 ft) from the edge of the travelway showed that, if the minimum clear sight distance were three-fourths of the required stopping sight distance under wet pavement conditions, then there would be sufficient sight distance available for dry pavement conditions. Therefore, of the 1140 crossings examined, 60 had adequate clear sight distance on wet pavements and 125 had adequate clear sight distance on dry pavements.

The variables for restricted approach sight distances (D) and the restricted clear (triangle) sight distances (C) were included so that, when the actual regression coefficients were obtained at a later date, new terms would not have to be added to the existing computer programs. The final model used by FDOT later proved to be satisfactory, and the planned subsequent regression analysis was not undertaken.

Accident-Prediction Equations

By using the final stepwise regression model and the modifications discussed below, two equations were established. The first (Equation 10) calculates the accident potential (t_p) for 4 years at grade crossings with only passive signing. The second (Equation 11) calculates the accident potential (t_a) for grade crossings with active warning devices.

$$\begin{aligned} t_p = {} & -8.075 + 0.318 \ln S_t + 0.484 \ln T + 0.437 \ln A \\ & + 0.387 \ln V_v + [0.28 - 0.28(\text{MASD/RSSD})] \\ & + [0.33 - 1.23(\text{MCSD/RSSD})] \\ & + 0.15 \quad \text{(if no crossbucks)} \end{aligned} \quad (10)$$

$$y = [\exp(0.968 t_p + 1.109)]/4 \quad (10a)$$

$$\begin{aligned} t_a = {} & -8.075 + 0.318 \ln S_t + 0.166 \ln T + 0.293 \ln A \\ & + 0.387 \ln V_v + [0.28 - 0.28(\text{MASD/RSSD})] \\ & + 0.225(L - 2) - 0.233 \quad \text{(if gates)} \end{aligned} \quad (11)$$

$$y = [\exp(0.968t_a + 1.109)]/4 \quad (11a)$$

where

A = vehicles per day or annual average daily traffic,
L = number of lanes,
MASD = actual minimum stopping sight distance along roadway,
MCSD = clear sight distance (ability to see approaching train along the roadway, recorded for the four quadrants established by the intersection of the railroad tracks and road),
RSSD = required stopping sight distance on wet pavement,
S_t = maximum speed of the train,
T = yearly average of the number of trains per day,
t_a = ln of the predicted number of accidents in the 4-year period at crossings with active protection,
t_p = ln of the predicted number of accidents in the 4-year period at crossings with passive protection,
V_v = posted vehicle speed limit unless geometrics dictate a lower speed, and
y = predicted number of accidents per year at crossing.

(The variable [0.33 - 1.23(MCSD/RSSD)] is omitted if the crossing is protected by flagmen or the calculation is less than zero, the variable [0.28 - 0.28(MASD/RSSD)] is omitted if sight restriction is due to a parallel road, and the variable (L - 2) is omitted when there are gates.)

Adjustment for Accident History

The stepwise regression model is a reasonable accident predictor for each grade crossing, which admittedly would be biased by the introduction of the accident history of a crossing. It is also possible that the phenomenon of regression toward the mean may mean that a crossing that has two or three accidents in 1 year may not have any more until it reaches its actual predicted accident rate. However, the accident history can be used as an adjustment to compensate for some of the failings of the accident predictor. The need for an accident-history adjustment was based on the following:

1. The present stepwise regression model explains only 18 percent of the accidents that occur (the multiple correlation R was 0.43) because human failure is involved in over 90 percent of them. Although it would be possible to increase the multiple correlation by taking into account different driver profiles at various crossings, it would be impossible to collect such data.
2. Accident histories are used by engineers to identify deficient systems, and in the event of a lawsuit, it would be difficult to explain in court why the accident history at a particular location had been ignored.

Thus, an accident-history adjustment equation that would increase but never decrease the accident predictor was used. This adjustment for accident history is calculated only when the accident history is greater than the accident prediction.

$$y' = y(H/yP)^{1/2} \quad (12)$$

where

y = accident prediction based on the regression model,
y′ = accident prediction adjusted for accident history,
H = number of accidents for a 6-year history or since the year of the last improvement, and
P = number of years of the accident-history period.

The accident-history adjustment has not been a major factor in determining the most hazardous grade crossings. Of the 98 crossings with the highest accident prediction, 61 were not affected by the accident history. The use of accident history has, however, helped to identify grade crossings with unique problems that were not identified in the accident-prediction model.

USE OF ACCIDENT-PREDICTION EQUATIONS

Selection of Grade-Crossing Improvement Projects

A simple method of rating each grade crossing from 0 to 90 was derived from the accident-prediction model. This method, entitled the safety index, was used to rank each grade crossing. A safety index of 70 is considered safe (no further improvement is necessary): A grade crossing with an accident prediction of 0.05 or one accident every 20 years would have a safety index of 70. It is not economical to provide active warning devices at grade crossings having lower accident-prediction indexes. A safety index of 60, or one accident every 9 years, would be considered marginal.

Each grade crossing is assigned a statewide priority number based on the safety index, i.e., the grade crossing with the lowest safety index would be assigned priority one. If there were no fund limitations, the selection of grade crossings for an improvement program would be simplified. However, the funds for the program, which are received primarily from the 1973 and 1976 Highway Safety Acts, are divided between Federal-Aid routes and off-system routes. Since these funds have become available, FDOT has scheduled 125 grade crossings on Federal-Aid routes for improvement at a cost of $5.8 million. As of June 1, 1977, 90 of them had been completed. When the total 125 are completed (this does not include the 330 urban streets that were recently added to the Federal-Aid system), all grade crossings on Federal-Aid routes that had a safety index of less than 70 will have automatic warning devices.

However, there will still be the major problem of the 4460 grade crossings on off-system routes. At the beginning of the program, 4150 of these had only passive warning signs; 400 did not even have crossbucks. FDOT has scheduled 130 of them for improvement, and as of June 1, 1977, 80 had been completed.

Reduction of High-Accident Sections

The regression analysis showed that reducing train or vehicle speeds or both reduces the probability of accidents. Also, since the accident prediction increases the logarithm of the number of vehicles, the same number of vehicles using fewer grade crossings will reduce the probability of accidents. Thus, the closing of any grade crossing will decrease the accident probability; i.e., the accident prediction for the one crossing with the combined traffic will be less than the sum of the accident prediction for the two crossings. Of course, if the traffic from a closed crossing is diverted to a grade crossing with automatic warning devices, the probability of an accident will be reduced even more.

A computer program that compares the sum of the accident predictions on each track to a statewide average for a particular category of track was developed. Each combination of urban versus rural and type of train is a different category. For example, if the track is in a small municipality and both freight and passenger trains use it, the category will be town and passenger. The formula used to select a high-accident (abnormal track section is

$$\lambda p = \lambda c + K(\lambda c/T)^{1/2} - 1/2T \tag{13}$$

where

λp = critical accident potential per mile of track,
λc = average accident potential per mile for the category of track being tested,
T = natural logarithm of the average number of trains per day, and
K = constant.

The magnitude of K determines the level of statistical significance and controls the number of track sections that should be investigated. The K-value is 2.567, which means that the probability of the accident prediction on the track section selected being abnormal is 99 percent.

Any section of track where the sum of the accident predictions is higher than λp should be examined to reduce the accident prediction. Any of the following actions can be taken: (a) close unnecessary crossings, (b) install automatic warning devices, (c) reduce train speeds, (d) reduce highway speeds, and (e) construct grade separations. The action(s) taken should depend on the feasibility and benefit-cost studies. The reduction of accidents as one of the benefits is based on the reduced accident predictions for the crossings affected.

Effects of Grade-Crossing Improvements on Accident Reduction

In 1974, Governor Reubin Askew committed FDOT to a massive railroad-highway grade-crossing improvement program. The goal was to reduce fatalities by 50 percent (from 90 to 45) by improving 20 percent (1200) of the 6000 grade crossings in the state in 6 years. The improvement of the grade crossings on the primary highway system had been under way, and this was increased by the use of Emergency Highway Safety Funds. In 1974, after the implementation of the Highway Safety Act of 1973, the rate of grade-crossing improvements was doubled to 120/year, and many grade crossings not on the system were included.

The analysis of the effect of the crossing improvements on the 1974 and 1975 accident rate is complicated by concurrent events, such as the late 1973 oil embargo, the 88.5-km/h (55-mph) speed limit, and a decrease in the 1974 vehicle operating speed. (In 1975, vehicle operating speeds increased but remained below 1973 operating speeds.) Also, the number of vehicle-kilometers traveled varied from 94 800 million in 1973 to 98 000 million in 1974 to 106 000 million in 1975 (53 300 million, 61 900 million, and 62 200 million vehicle-miles traveled respectively), and the number of train movements decreased approximately 10 percent. However, some interesting comparisons still can be made.

Three groups of grade crossings that had no signal improvements in 1974 and 1975 were analyzed. According to the accident-prediction model, these crossings had the highest potential for accidents. The results of the analysis are shown below.

Group	Number in Group	Accident Prediction	1975 Accidents	Number Without Accidents (1969 to 1974)	First Accident in 1975
Highest	98	46.5	34	26	9
Second highest	99	36.9	24	38	5
Third highest	100	25.4	18	35	2

The broad rankings produced by the accident-prediction model are borne out by the 1975 accident experience. For the three groups, the accident experience ranged from 65 to 73 percent (average 70 percent) of the accident prediction. The prediction model is based on the 1968 to 1972 accident (preenergy crisis) data and 1973 vehicle speed limits (maintained for accident predictions), which may be one of the explanations as to why the predictions are higher than the 1975 experience. Only 25 percent of those accident predictions were affected by the accident history.

An examination of the statewide accident trends shows that the number of train-vehicle collisions decreased from 498 in 1974 to 390 in 1975 (22 percent), although the total number of accidents remained unchanged. Among the 108 train-vehicle accident decrease, only 13 can be attributed directly to the installation of automatic warning devices. This leaves a 19 percent decrease in accidents from 1974 that is still unexplained. In 1976, train-vehicle collisions were reduced another 15 percent to 330 (the total number of accidents decreased 4 percent). From 1974 to 1976, the number of fatalities due to train-vehicle collisions decreased from 75 to 55 (25 percent).

These statistics indicate that those crossings having higher accident predictions experienced the most accidents. The accident experience of those crossings without previous accident experience also was proportional to their accident predictions. Also, the accident-prediction model was within 15 to 30 percent of the actual accident experience even with the current downward trend in train-vehicle collisions.

The effect of the installation of automatic warning devices on the number of subsequent accidents was further analyzed. Of those grade crossings modified between July 1, 1974, and October 30, 1975, only 30 were based on the accident-prediction model. (All but 3 of these were installed after July 1, 1975.) However, even these limited results are encouraging. The following results were achieved from the analysis of two groups of 100 grade crossings each.

Crossings	Accident Prediction: Without Modification: One Year	Accident Prediction: Without Modification: Post Installation	Accident Prediction: With Modification: Post Installation	1975 Accidents
Modified	24	17	5	4
Unmodified	25	–	–	18

Since the 100 modified grade crossings included those that were modified during 1975, the accident prediction was adjusted downward to include only the time period after the installation of the crossing warning devices. Only those accidents that occurred after the installation were counted. During this period, 5 accidents were expected, but only 4 occurred, which agrees with the control group of unmodified crossings that experienced only 72 percent of their predicted number of accidents.

Thus it appears that the accident-prediction model consistently predicts accidents to within 15 to 30 percent of their actual occurrence, and the reduction in accidents after the installation of automatic warning devices is as expected. The grade crossings that are modified in fiscal year 1976 will provide better data, since three-fourths of them were selected on the basis of the accident-prediction model. During this period, 43 of the 100 most hazardous grade crossings will be modified.

CONCLUSION

The accident-prediction model can be effectively used to develop a grade-crossing improvement program. It identifies groupings of crossings (with or without the accident-history adjustment) that can be expected to experience the most accidents if they are not modified, and the accident experience after modification has been in reasonable agreement with that predicted.

ACKNOWLEDGMENTS

This report could not have been developed without the efforts of G. van Belle, D. Meeter, and W. Farr, who prepared the report, Influencing Factors for Rail-Highway Grade-Crossing Accidents in Florida. In addition, the efforts of Meeter in editing the many rewrites of this report and in providing needed technical assistance are greatly appreciated. Appreciation is also given to Winifred Bailey, who suffered through the typing and editing of the many revisions, and to Wyndal Hand, who assisted with the data.

The changes to the model and the subsequent conclusions as to the sight-distance factors and the effects of gates are those of the author and are not necessarily those of the above-mentioned consultants.

REFERENCES

1. G. van Belle, D. Meeter, and W. Farr. Influencing Factors for Rail-Highway Grade-Crossing Accidents in Florida. Accident Analysis and Prevention, Vol. 7, 1975, pp. 103-112.
2. D. Schoppert and D. Hoyd. Factors Influencing Safety at Highway-Rail Grade Crossings. NCHRP, Rept. 50, 1969.
3. A Policy on Design Standards for Stopping-Sight Distances. American Association of State Highway and Transportation Officials, Sept. 2, 1971.
4. J. B. Hopkins and F. R. Holmstrom. Toward More Effective Grade-Crossing Lights. Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., 1975.
5. J. Beauchamp and J. Olsen. Estimates for the Mean and Variance of Lognormal Distribution Where the Mean Is a Function of an Independent Variable. Eastern Deciduous Forest Biome Study, Oak Ridge National Laboratory, Technical Rept. 71-101, 1972.
6. Manual on Uniform Traffic-Control Devices. American Association of State Highway and Transportation Officials and U.S. Department of Transportation, 1971.
7. W. J. Dixon, ed. BMD Biomedical Computer Programs. Univ. of California Press, Berkeley, 1971.
8. J. Coleman and G. Stewart. Accident and Accident-Severity Prediction Equations. Proc., National Conference on Railroad-Highway Crossing Safety, Washington, D.C., 1974.

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

Visual Performance of Drivers During Rainfall

Ron S. Morris, John M. Mounce, Joe W. Button, and Ned E. Walton, Texas Transportation Institute, Texas A&M University

This paper reports an investigation of the effect of rain on the visual performance of drivers. The degradation of static visual acuity in terms of visual angle, detection probability, and legibility as a function of rain intensity was determined by experiments that used a rainfall simulator that produced artificial rain. The significant findings include the following: (a) Water on the windshield is the primary factor accounting for reduced visual performance, (b) visual degradation in the daytime with windshield wipers in operation appears to be a linear function of the rain rate with normal drop sizes, (c) during nighttime conditions, drop size is a significant factor in reducing visual performance (smaller drops are a more serious problem than is the rain rate), (d) wiper speeds above 50 CPM do not improve visual performance, (e) without windshield wipers, visual performance is reduced to levels that are unacceptable for driving (equivalent to visual acuity greater than 20/200) at rain rates greater than 2.5 cm/h (1 in/h), and (f) the effective rain rate can be determined from the vehicle velocity, the terminal velocity of the drop, the rake angle of the windshield, and the actual rain rate.

The factor of visibility during adverse weather has been largely neglected by the highway transportation industry. There are at least two reasons for this: These are that the problems associated with driver visibility have been underestimated and that objective measurements of the effects of wet weather on the visual performance of drivers are difficult to obtain. Thus, there have been very few developments designed specifically to assist the automobile driver in the performance of visual tasks during adverse weather (1).

EQUIPMENT AND METHODOLOGY

The objective tests used in this research determined the effects of selected, controlled intensities of artificial (simulated) rainfall on the visual performance of drivers relative to visual acuity, target detection, recognition, and legibility. These tests were also designed to assess the improvement to driver visibility afforded by windshield wipers at various cyclic rates. All of the tests were conducted on overcast days to more closely simulate actual rain conditions. To eliminate the effects of wind on the paths of the falling drops, the tests were