PB85107167

FHWA-RD-83/092
DOT-TSC-FHWA-84-1

Accident Severity Prediction Formula for Rail-Highway Crossings

E.H. Farr
J.S. Hitz

Transportation Systems Center
Cambridge MA 02142

July 1984
Final Report

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation

Prepared for

Federal Highway Administration
Office of Research, Development and Technology

Federal Railroad Administration
Office of Safety

Technical Report Documentation Page

1. Report No. FHWA/RD-83/092	2. Government Accession No.	3. Recipient's Catalog No. PB85 107167
4. Title and Subtitle Accident Severity Prediction Formula for Rail-Highway Crossings		5. Report Date July 1984
		6. Performing Organization Code DTS-74
7. Author(s) E. H. Farr, J. S. Hitz		8. Performing Organization Report No. DOT-TSC-FHWA-84-1
9. Performing Organization Name and Address U.S. Department of Transportation Research and Special Programs Administration Transportation Systems Center Cambridge, MA 02142		10. Work Unit No. (TRAIS) HW-321; RR-333
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Highway Administration Office of Research, Development, and Technology Washington DC 20590	U.S. Department of Transportation Federal Railroad Administration Office of Safety Washington DC 20590	13. Type of Report and Period Covered Final Report FY 1983
		14. Sponsoring Agency Code HSR-30
15. Supplementary Notes		

16. Abstract This report describes the development of formulas which predict the severity of accidents at public rail-highway crossings. They employ the previously developed DOT accident prediction formula, U.S. DOT-AAR National Rail-Highway Crossing Inventory, and the FRA accident files. With these new formulas used in the DOT Resource Allocation Procedure, information will be available to assist in making better decisions about where to install motorist warning devices to further increase crossing safety for a given level of funding.

Established statistical techniques are used to develop two formulas: one that estimates the number of fatal accidents per year at a crossing and one that estimates the number of injury accidents per year at a crossing. It was found that the factors in the inventory that significantly influence fatal accident severity, given that an accident occurred, were maximum timetable train speed, the number of through trains per day, the number of switch trains per day, and the urban-rural location. For injury accident severity, given that an accident occurred, the significant factors were maximum timetable train speed, the number of tracks, and the urban-rural location.

The performance of these severity formulas is discussed and calculated results are presented.

i

17. Key Words Rail-Highway Crossing Railroad Accidents Accident Severity Accident Prediction Grade Crossing		18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

PREFACE

The DOT Rail-Highway Crossing Resource Allocation Procedure, developed at the U.S. Department of Transportation's Transportation Systems Center (TSC), employs an accident prediction formula. In an attempt to improve the effectiveness and usefulness of the resource allocation procedure, the present study was undertaken to incorporate a quantitative measure of severity into the accident prediction formula.

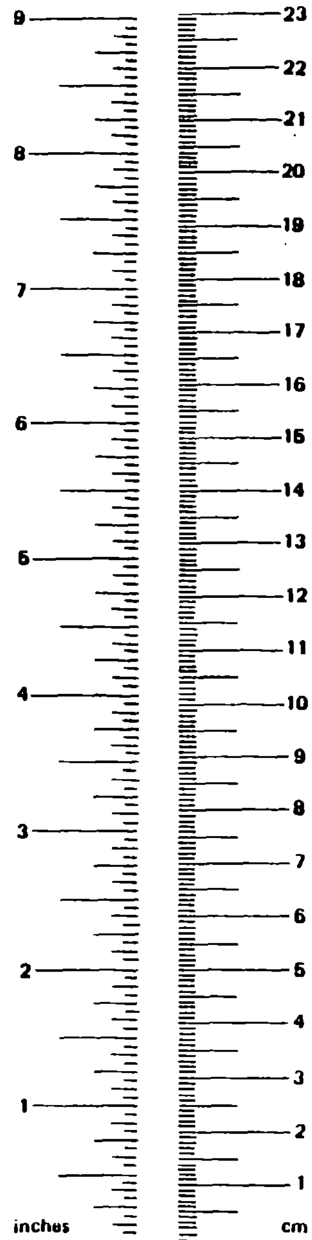
The original work on the resource allocation procedure and the present study were sponsored jointly by the Federal Highway Administration's (FHWA) Offices of Research, Development, and Technology and the Federal Railroad Administration's (FRA) Office of Safety. The authors express their appreciation to Janet Coleman, FHWA, and Bruce George, FRA, for their technical contributions to the study. Development of the accident severity formula at TSC was the responsibility of Dr. Edwin Farr and John Hitz. The statistical procedure was designed by Dr. Peter Mengert. Mary Cross was responsible for providing systems support to the project.

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in. = 2.54 cm (exactly). For other exact conversions and more detail tables see NBS Misc. Publ. 286. Units of Weight and Measures. Price $2.25 SD Catalog No. C13 10 286.



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F

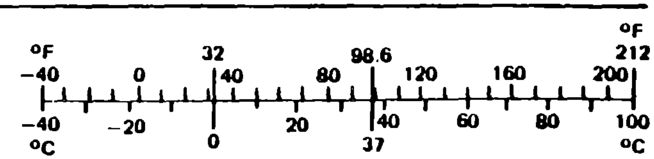


TABLE OF CONTENTS

SUMMARY

This report describes the development of formulas which predict the severity of accidents at public rail-highway crossings. They employ the previously developed DOT accident prediction formula, U.S. DOT-AAR National Rail-Highway Crossing Inventory (The Inventory), and the FRA accident files. With these new formulas used in the DOT Resource Allocation Procedure, information will be available to assist in making better decisions about where to install motorist warning devices that will further increase crossing safety for a given level of funding.

Established statistical techniques are used to develop two formulas: one that estimates the number of fatal accidents per year at a crossing and one that estimates the number of injury accidents per year at a crossing. It was found the factors in The Inventory that significantly influence fatal accident severity, given that an accident occurred, were maximum timetable train speed, the number of through trains per day, the number of switch trains per day, and the urban-rural location. For injury accident severity, given that an accident occurred, the significant factors were maximum timetable train speed, the number of tracks, and the urban-rural location.

The performance of these severity formulas is discussed and calculated results are presented.

Preceding page blank

1. INTRODUCTION

1.1 PURPOSE

This report documents a study to develop a formula for predicting the severity of accidents at rail-highway crossings. The resulting formula is to be incorporated into the DOT Rail-Highway Crossing Resource Allocation Procedure (Ref. 1 and 2).

1.2 BACKGROUND

The Highway Safety Acts of 1973 and 1976 and the Surface Transportation Assistance Acts of 1978 and 1982 provide Federal funding authorizations to States specifically for safety improvement projects at public rail-highway crossings. These safety improvements frequently consist of the installation of active motorist warning devices such as flashing lights or gates. To promote the effective use of Federal funds for these safety projects, the U.S. Department of Transportation has developed a procedure for States and railroads to assist in planning rail-highway crossing safety programs. This procedure, the DOT Rail-Highway Crossing Resource Allocation Procedure (DOT Procedure), determines crossing safety improvements that result in the greatest accident reduction benefits based on consideration of predicted accidents at crossings, the costs and effectiveness of safety improvement options, and budget limits (Ref. 1 and 2).

Two analytical methods have been developed as part of the DOT Procedure. Their development followed completion of a joint U. S. DOT-AAR (Association of American Railroads) National Rail-Highway Crossing Inventory (hereafter referred to as The Inventory), which numbered and collected inventory information for all public and private crossings in the United States (Ref. 3). The first analytical method included in the DOT Procedure is the DOT accident prediction formula, which computes the expected number of accidents at crossings based on information available in The Inventory and crossing accident data files. The second analytical method is a resource allocation model designed to rank candidate crossings for improvement on a cost-effective basis and recommend the type of warning device to be installed.

The current effort is motivated by the recognition that rail-highway crossing accidents are not equally severe. In 1981 there were a total of 8,546 rail-highway crossing accidents (Ref. 4). Of these accidents 5,761 had no casualties, 2,224 had injuries only, and 561 involved fatalities. Thus, 67 percent of the accidents had no measurable casualty severity while only 6.6 percent had the highest level of severity in terms of having fatalities. This unequal distribution of severity among crossing accidents makes it important, but difficult, to discern those crossings which are likely to have high severity accidents. Use of safety improvement funds based on a prioritization of crossings by predicted accidents, as performed by the current DOT Procedure, could be significantly different than one based on predicted accident severity.

1.3 CONCEPT OF ACCIDENT SEVERITY FORMULA

The traditional approach to risk analysis (Ref. 5) views safety risk as the product of two independent factors: (1) the frequency of accident occurrence, and (2) the severity or consequences of accident occurrence. The product of these two factors for a given hazardous situation provides the total safety risk for that hazard. For example, a rail-highway crossing with a predicted accident frequency of 0.5 accidents per year and a predicted accident severity of 0.2 fatalities per accident poses a total safety risk of 0.1 fatalities per year. This dichotomy of safety risk into accident frequency and severity components is particularly appropriate for the current effort since one of the components, the DOT accident prediction formula, currently exists. Under this concept the proposed severity prediction formula would be used with the current accident prediction formula to provide a prediction of total safety risk as follows:

$$R = A \times S \tag{1-1}$$

where: R = risk of a crossing measured in expected casualties per year

A = predicted accident frequency from the current DOT accident prediction formula

S = predicted accident casualties from the severity prediction model.

A major benefit of this approach is that the current DOT accident prediction formula will remain unchanged and can be used either with or without the severity formula as desired.

Under the current effort two severity formulas are to be developed: one formula to predict fatality severity and another to predict injury severity. These formulas are to provide their predictions on the basis of crossing characteristics as described in The Inventory.

2. APPROACH

The approach for development of the severity prediction formulas can be generally described in terms of a series of tasks. The first task involved the selection of specific measures of severity to be quantified by the formulas. Severity measures considered for the fatality formula included: fatal accidents per accident (i.e. accidents involving at least one fatality within a year), fatalities per accident, and fatalities per vehicle occupant per accident. Similar measures were considered for the injury severity formula: injury accidents per accident (i.e. accidents involving injuries but not fatalities), injuries per accident and injuries per vehicle occupant per accident. The selection of severity measures is discussed in Section 3.1.

The next task was to identify factors describing the characteristics of crossings which are potentially useful in predicting accident severity. To accomplish this, histograms were developed relating the measures of severity to values of all factors that were thought to contribute to the severity of crossing accidents. Based on a review of the histograms, those factors which showed a strong correlation with the measures of severity were identified for possible inclusion in the severity formulas. The selection of factors in this way is discussed in Section 3.2.

The severity formulas were developed in the next task as described in Section 3.3. The statistical technique used in developing the formulas was the same as that used previously in developing the accident prediction formula (Ref. 6). This regression procedure is referred to as the logistic discriminant approach which employs an iterative weighted regression technique that is a modification of a method described in Cox (Ref.7).

The last task in development of the severity formulas was to evaluate their performance as discussed in Section 4. The evaluations involved: calculating accident severity from the formulas for typical values of the formula factors and comparing results with severity characteristics obtained from the histograms described in Section 3.2, comparing predicted accident severity with actual observed accident severity, comparing the ability of the severity formulas to rank crossings by predicted severity with random rankings, and using the DOT Procedure for sample applications with and without the severity formulas and comparing results.

3. SEVERITY PREDICTION FORMULA DEVELOPMENT

3.1 SELECTION OF SEVERITY MEASURES

The proposed concept for use of the severity formulas dictates that severity will be measured in terms of consequences, given an accident occurred (see Section 1.). The severity measures must therefore be expressed in terms of consequences per accident. Furthermore, rail-highway crossing accidents have three basic dimensions of severity which can be measured: fatalities, injuries and property damage. For these dimensions of severity the following measures were therefore proposed for the severity formulas:

1. Fatality Severity:

   - Fatalities per accident
   - Fatal accidents per accident
   - Fatalities per vehicle occupant per accident

2. Injury Severity:

   - Injuries per accident
   - Injury accidents per accident
   - Injuries per vehicle occupant per accident

3. Property Damage Severity:

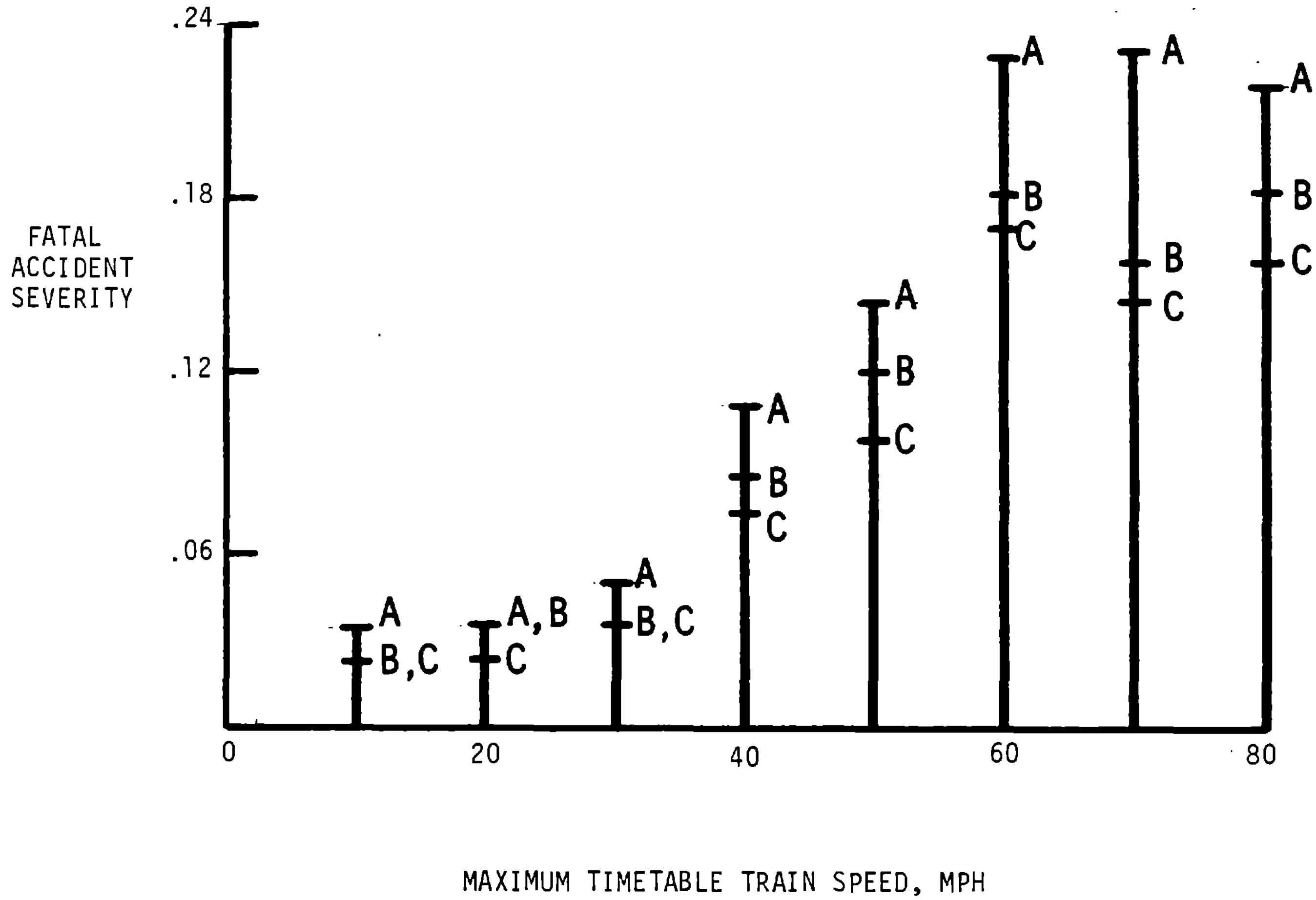
   - Dollars property damage per accident

The current effort concentrated on developing formulas for fatality and injury severity since they are most readily associated with safety risk. A property damage severity formula may be developed as part of a later effort but will not be discussed further here.

For purposes of this study, the following definitions are provided for fatal and injury accidents: a fatal accident is an accident in which at least one fatality occurred within a year independent of injuries or property damage; an injury accident is an accident in which there were no fatalities and at least one injury occurred independent of property damage.

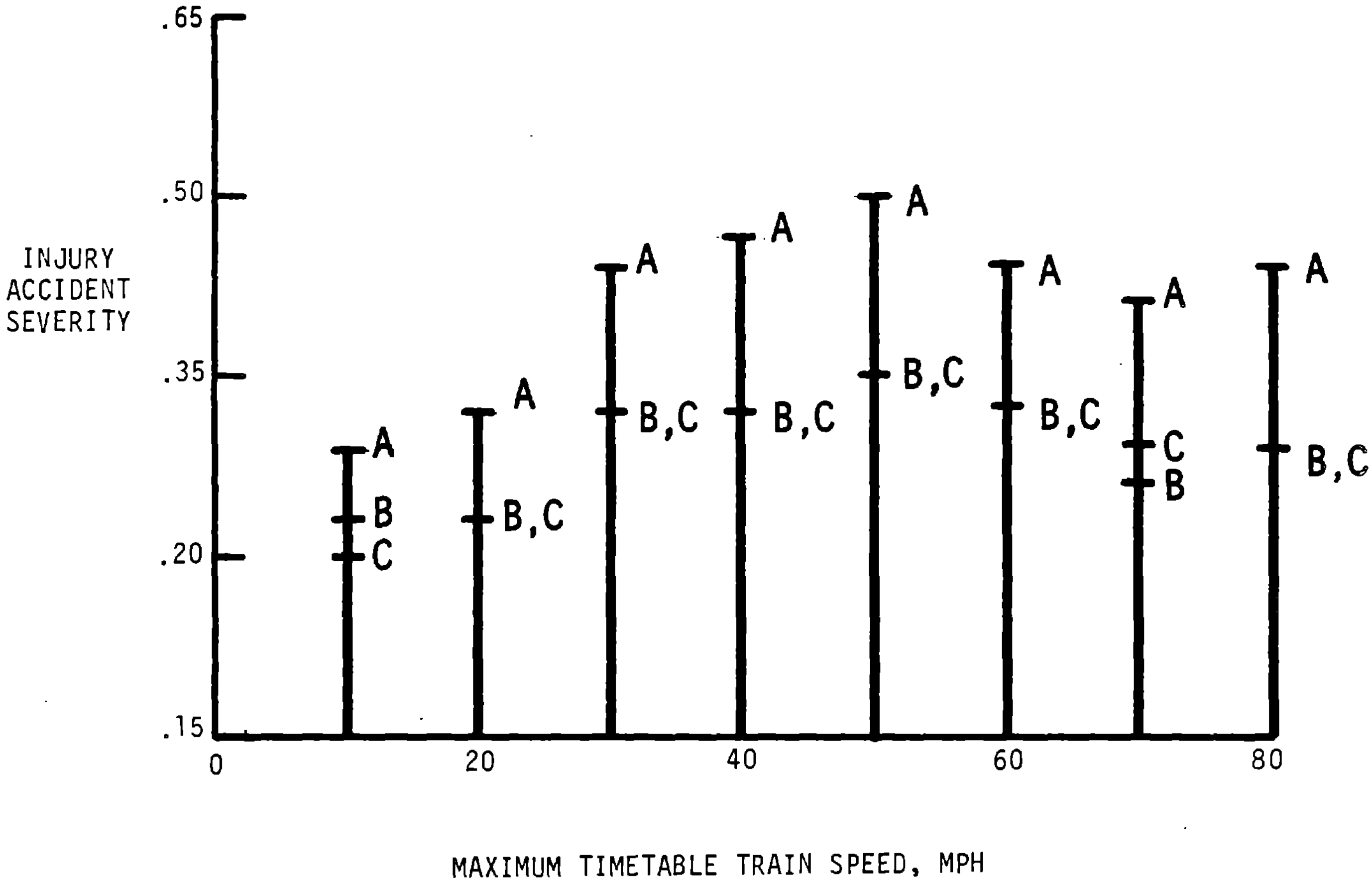
The number of fatalities per accident and the number of injuries per accident were originally proposed as the severity measures. These measures are somewhat dependent on the number of vehicle occupants at the time of the accident, however, and this tends to be a random factor. It would be more appropriate to adjust the number of fatalities or injuries by the number of vehicle occupants. Two additional measures of severity were therefore proposed to accomplish this: (1) fatalities and injuries per occupant per accident (these measures normalize casualties by the number of vehicle occupants), and (2) fatality and injury accidents per accident (these measures indicate the probability of producing one or more casualties regardless of the number of vehicle occupants). For a given level of severity, fatalities and injuries per accident should yield the largest value of the alternative measures since total casualties are counted. Fatalities and injuries per occupant per accident should yield the lowest value since total casualties are divided by total occupants. Fatal and injury accidents per accident should be of intermediate value since only the first casualty is counted but it is not divided by the number of casualties.

To assist in evaluating the fatality and injury severity measures, histograms were developed as shown in Figures 3-1 and 3-2. These histograms relate average values of the measures, calculated from accident records, to accidents grouped by intervals of maximum train speed. This permits a review of how the measures vary as a function of a factor (maximum timetable train speed) previously shown to be correlated with accident severity (Ref. 4 and 8). It should be noted that maximum timetable train speed is a crossing characteristic included in The Inventory and is used here as a surrogate for actual train speed at the time of an accident.



A = FATALITIES PER ACCIDENT
B = FATAL ACCIDENTS PER ACCIDENT
C = FATALITIES PER OCCUPANT

FIGURE 3-1. COMPARISON OF FATALITY SEVERITY MEASURES



A = INJURIES PER ACCIDENT
B = INJURY ACCIDENTS PER ACCIDENT
C = INJURIES PER OCCUPANT

FIGURE 3-2. COMPARISON OF INJURY SEVERITY MEASURES

A review of the histograms in Figure 3-1 shows that the three fatality measures vary with train speed in the same general manner. All three increase with train speed to about 60 mph beyond which they remain relatively constant. This is intuitive since, beyond some high value of severity, fatalities can no longer increase. As originally surmised, values for fatalities per accident are higher than fatal accidents per accident which, in turn, are higher than fatalities per occupant per accident. The shape of the histograms for the three measures are generally the same, however, suggesting that either measure could be used with similar results. Given the general compatibility of the measures, fatal accidents per accidents was chosen as the measure of fatality severity since it avoids the complexities of dealing with vehicle occupants. This measure can be restated, in statistical terms, as the probability of a fatal accident given an accident.

The same arguments as above can be stated for the selection of injury accidents per accident as the measure of injury severity. Again, in statistical terms, this measure can be restated as the probability of an injury accident given an accident. It is of interest to note from Figure 3-2 that the shape of the injury severity histograms increase and then decrease with increasing train speed. This is also intuitive since, beyond some severity threshold, casualties will increasingly become fatalities rather than injuries. This characteristic of the injury measure (failure to monotonically increase with severity) presents problems, however, both in development of the fomula, as discussed in Section 3.3, and its use for purposes of resource allocation.

With regard to resource allocation, the shape of the injury severity function can result in a crossing with a high actual severity rating having a predicted injury severity equal to or less than a crossing with a low actual severity rating. Resource allocation priorities based on predicted injury severity can therefore produce less than optimal results. For this reason, the preferred measure for resource allocation purposes is fatal accidents per accidents by itself or possibly used with injury accidents per accidents to produce a total casualty index.

3.2 SELECTION OF SEVERITY FACTORS

Development of the severity formulas started with identification of factors which correlate with the severity measures and are thus potential predictors of severity. All crossing characterstic factors in The Inventory were systematically reviewed to identify those correlated with the severity measures. To accomplish this, histograms were developed relating average values of the measures calculated for accidents grouped by intervals of the factor in question. The factors evaluated in this way are listed below.

Number of Day Thru Trains
Number of Night Thru Trains
Number of Day Switch Trains
Number of Night Switch Trains
Maximum Timetable Train Speed
Number of Main Tracks
Number of Other Tracks
Warning Device
Type of Development
Highway Paved
Crossing Angle
Crossing Surface
Number of Lanes
Functional Class of Highway
Urban/Rural Crossing
Annual Average Daily Traffic (AADT)
Percent Trucks

A typical example of a histogram for one factor, maximum timetable train speed, is shown in Figure 3-3. The histogram relates train speed to both severity measures being considered, fatal accidents per accidents (F) and injury accidents per accident (I). The train speed factor was the strongest predictor of fatal accident severity of all the factors on the above list. This is consistent with results obtained by

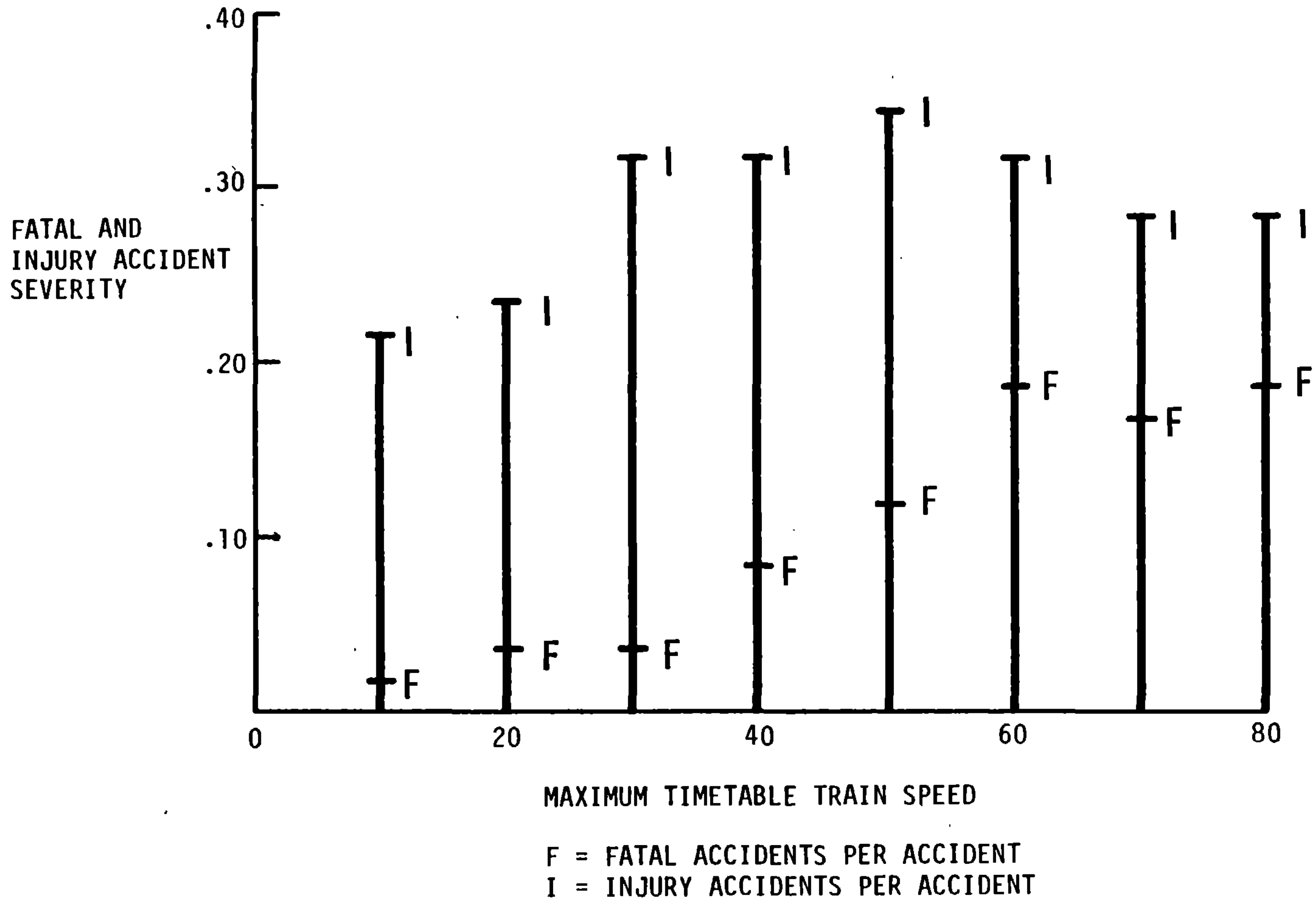


FIGURE 3-3. COMPARISON OF FATALITY AND INJURY-RELATED SEVERITY MEASURES

Coleman and Stewart in an earlier study of crossing accident data (Ref. 8). Note that the fatality severity measure increases monotonically with increasing train speed while the injury measure increases and then decreases.

Histograms were also constructed relating the severity measures to two factors. Examples of these two-dimensional histograms are shown in Figures 3-4 and 3-5. Figure 3-4 shows the frequency of fatal accidents as a function of maximum timetable train speed and the urban/rural location of the accidents (crossings). Figure 3-5 shows the frequency of injury accidents for the same factors. In both cases maximum timetable train speed and urban/rural location appear to be significant factors; i.e., severity generally increases with train speed but is less for urban accidents than for rural accidents.

As a result of reviewing the histograms the following factors were identified as potentially useful in predicting fatality and injury severity:

- Maximum Timetable Train Speed
- Urban/Rural Crossing
- Number of Main Track
- Number of Other Tracks
- Number of Thru Trains
- Number of Switch Trains

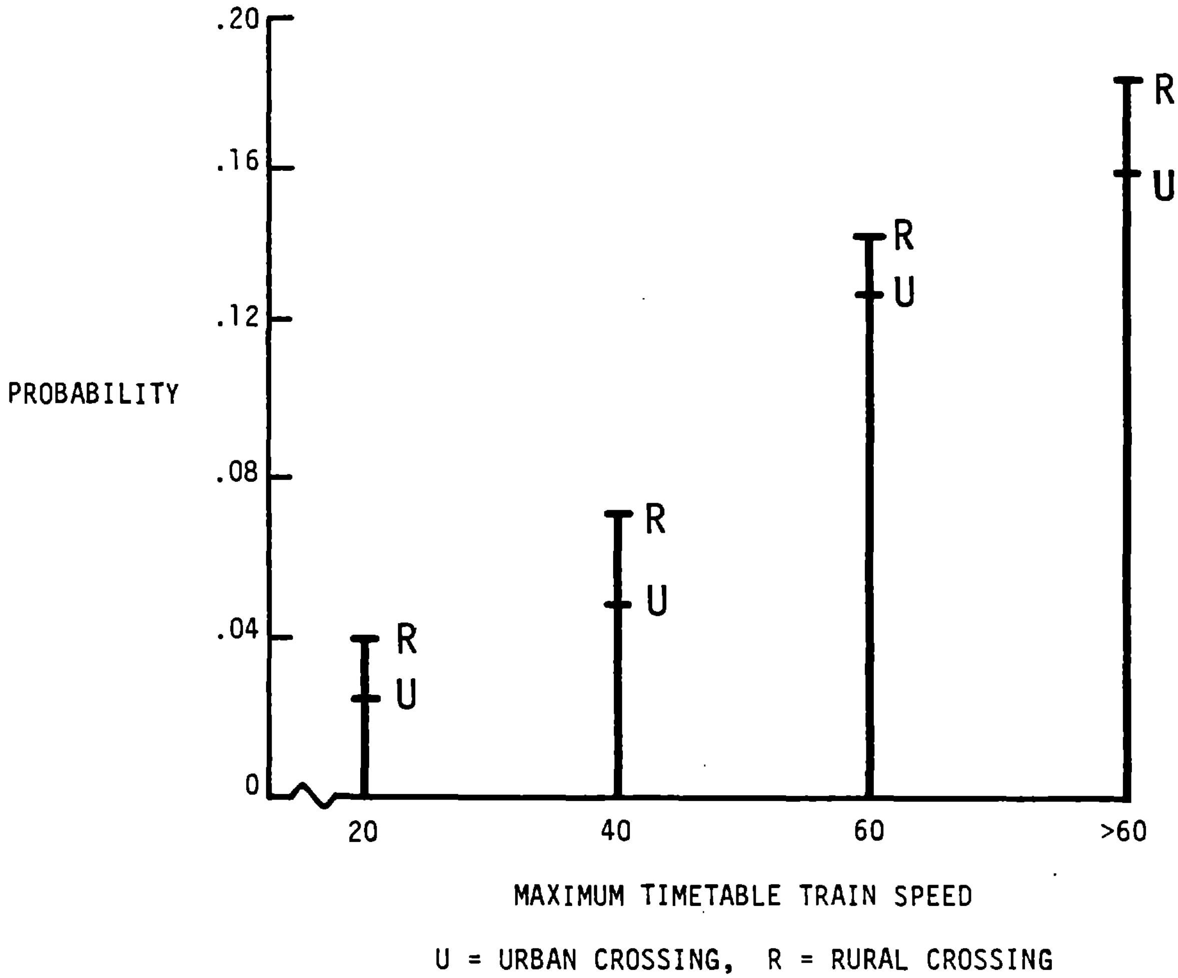


FIGURE 3-4. PROBABILITY OF FATAL ACCIDENT BY URBAN/RURAL CROSSING AND SPEED

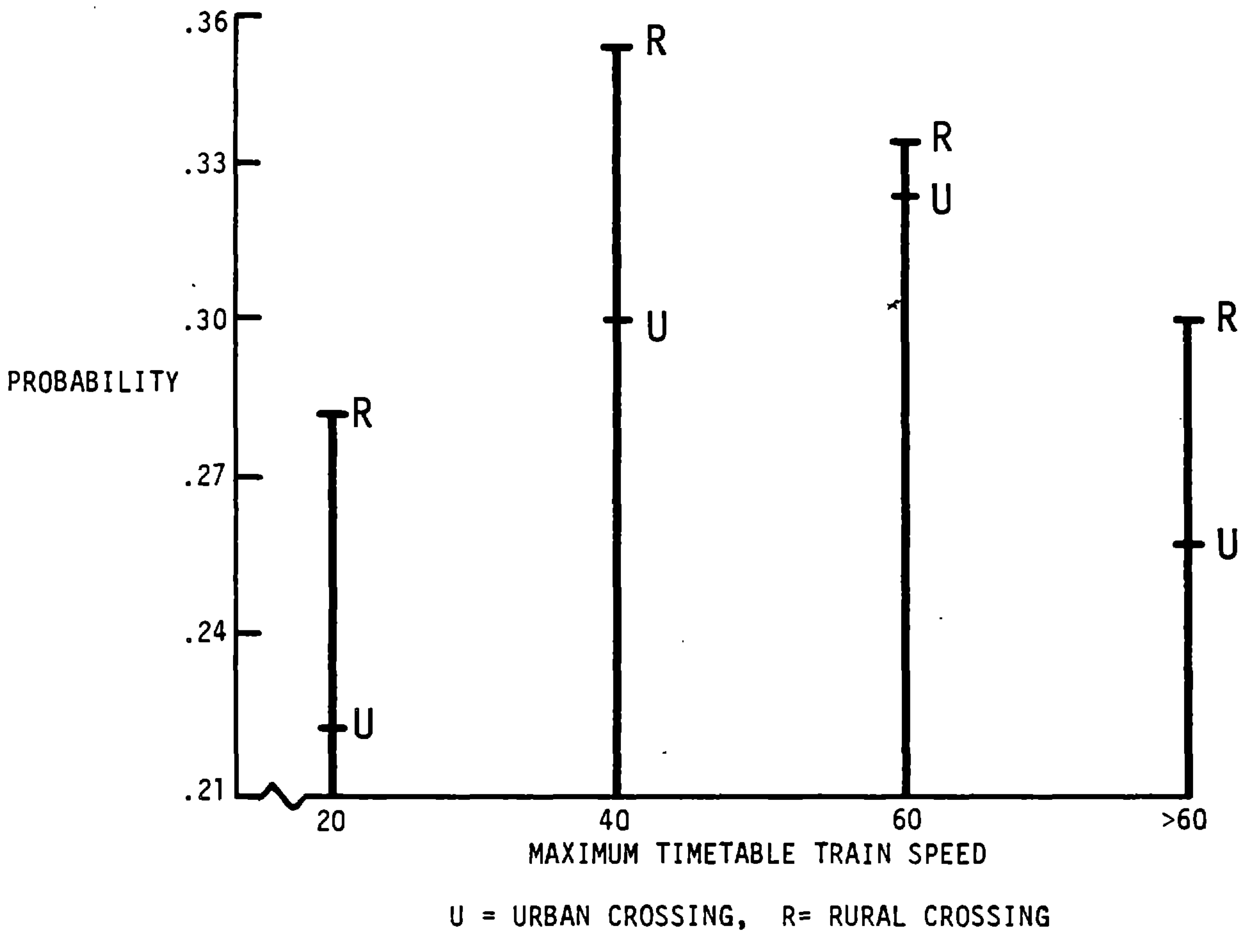


FIGURE 3-5. PROBABILITY OF INJURY ACCIDENT BY URBAN/RURAL CROSSING AND SPEED

3.3 SUMMARY OF FORMULA DEVELOPMENT

The analytical objective of this phase of the study was to develop formulas which will predict the probability of a fatal accident given an accident, P(FA|A), and the probability of an injury accident given an accident, P(IA|A). From these two formulas the safety risk expressed in terms of expected number of fatal accidents, R_f, and injury accidents, R_i per year at a crossing can be determined from:

$$R_f = A \times P(FA|A) \tag{3-1}$$

$$R_i = A \times P(IA|A) \tag{3-2}$$

where:

R_f and R_i = fatality and injury measures of safety risk for the crossing as described in Section 1.3.

A = the expected number of accidents per year at the crossing from the DOT accident prediction formula.

The analytic character of the fatal accident probability function, P(FA|A), relative to observed data can be seen in Figure 3-6. This graph is a frequency plot of the observed ratio of fatal accidents to total accidents versus maximum timetable train speed. The function P(FA|A) is represented by the dashed line which is a best fit to the observed data points connected by the solid line. Of course, the severity formula is multivariate and, hence, the dashed line for P(FA|A) would be a multidimensional "surface".

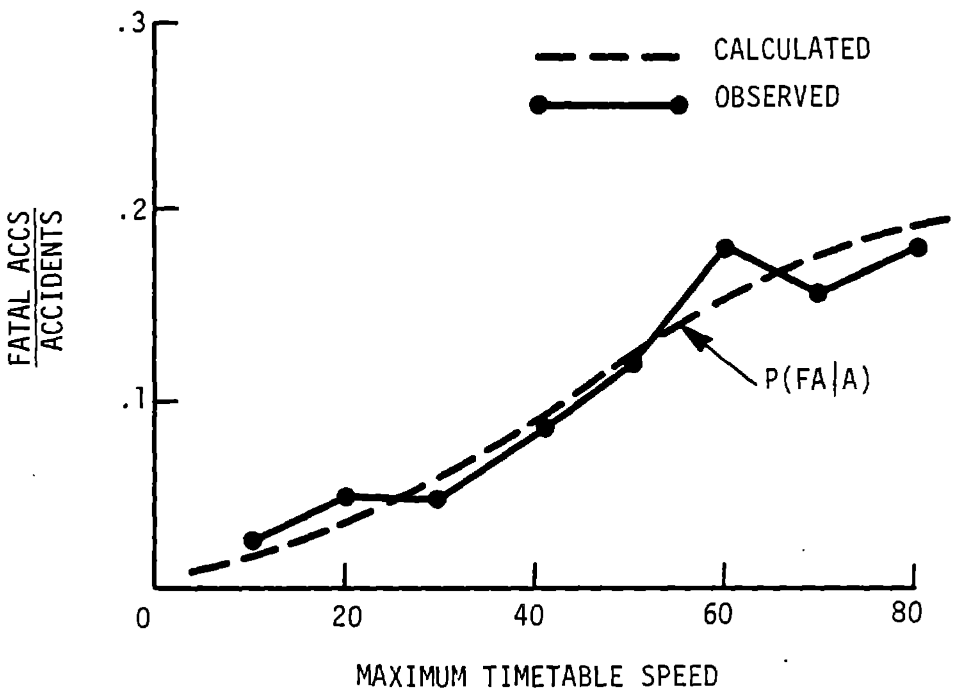


FIGURE 3-6. TYPICAL PLOT OF OBSERVED FATAL ACCIDENT FREQUENCY AND CALCULATED VALUES P(FA|A)

The analytic character of the injury accident probability function P(IA|A) relative to observed data can be seen in Figure 3-7. This graph is a frequency plot of the observed ratio of injury accidents to total accidents versus the same variable, maximum timetable train speed. In this case, the function P(IA|A) does not increase monotonically with severity as the fatal accident function does. However, the particular regression procedure used to develop the severity formulas (see Appendix A) involved fitting a monotonic function to the observed data. The required formula for predicting injury accident probability could, therefore, not be obtained directly from the regression analysis. This problem was overcome by limiting the accident data to non-fatal accidents. A formula was then developed from the regression analyses that predicted the probability of an injury accident given that a non-fatal accident occurred, P(IA|NFA). The formula for P(IA|NFA) is, as required, a monotonically increasing function of severity. Having obtained the formula for P(IA|NFA), the desired formula for P(IA|A) was then obtained from the following relationship:

Probability of an injury accident given an accident	=	Probability of an injury accident given a non-fatal accident	x	Probability of an non-fatal accident given an accident	
P(IA|A)	=	P(IA|NFA)	x	P(NFA|A)	(3-3)

where:

P(NFA|A) = 1 - P(FA|A) (3-4)

and

P(FA|A) = fatal accident probability formula obtained earlier,

hence,

P(IA|A) = P(IA|NFA) x (1 - P(FA|A)) (3-5)

In performing the regression analyses, the observed data for the dependent variable were assigned only two values. In the case of the fatal accident formula these values were +1 for a fatal accident and -1 for a non-fatal accident. For the injury accident formula the values assigned were +1 for an injury accident and -1 for a non-injury accident. The data used for the analyses was for the years 1978, 1979 and 1980. The regression analyses produced non-linear formulas for the dependent variable f, from the fatal accident data, and i, from the injury accident data.

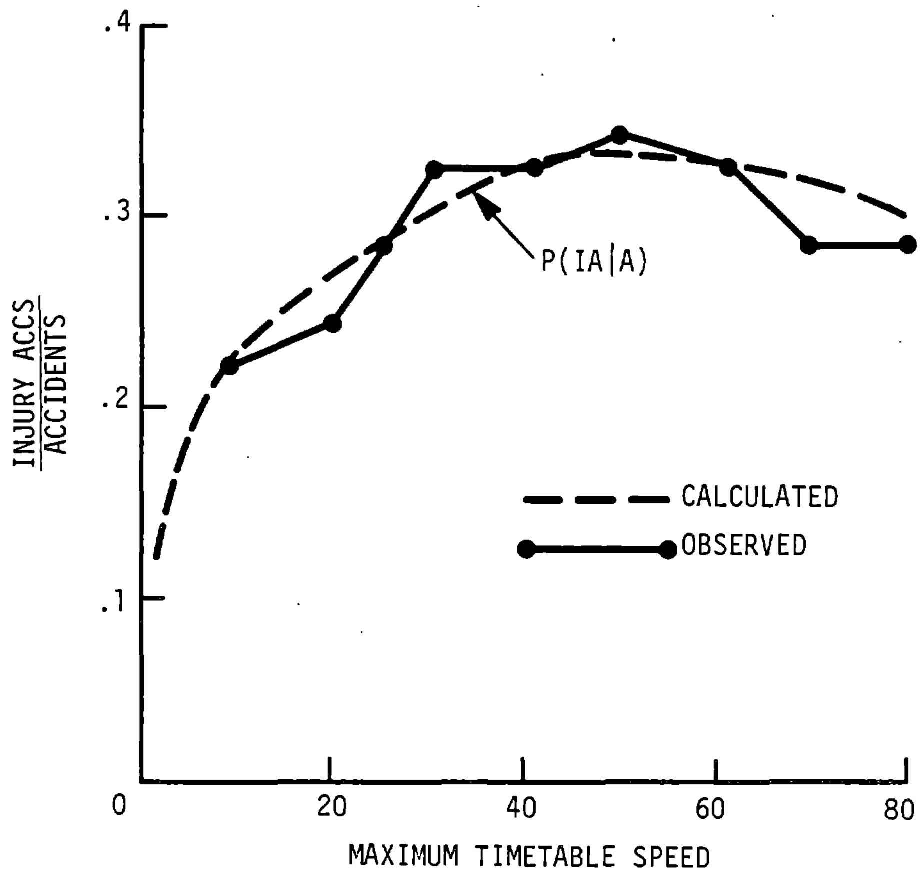


FIGURE 3-7. TYPICAL PLOT OF OBSERVED INJURY ACCIDENT FREQUENCY AND CALCULATED VALUES OF P(IA|A).

The resulting regression formulas produced values for f and i primarily between +1 and -1, for typical values of the independent variables, since the observed data were assigned only those values. For extreme values of the independent variables, however, f and i can be considered to have values from $+\infty$ to $-\infty$. Large values of f and i correspond to a high probability of a fatal or injury accident and vice versa. The desired values for f and i, however, are between 0 and 1 as required by the probability functions P(FA|A) and P(IA|A). The formulas for f and i, therefore, had to be transformed into probability functions. To accomplish this the following transformation was made to f to obtain the desired fatal accident probability formula:

$$P(FA|A) = 1/(1+e^{-2f}) \qquad (3\text{-}6)$$

A review of Equation 3-6 will show that P(FA|A) will have the desired values between 0 and +1 for all values of f between $+\infty$ and $-\infty$.

For the injury accident formula, the probability of an injury accident given a non-fatal accident, P(IA|NFA) was obtained first:

$$P(IA|NFA) = 1/(1 + e^{-2i}) \qquad (3\text{-}7)$$

The desired probability of an injury accident given an accident, P(IA|A), was then obtained by substituting Equations 3-6 and 3-7 into Equation 3-5 as described above.

The above discussion has provided an overview of the strategy involved in obtaining the required formulas for predicting fatal accident and injury accident probabilities. A more detailed discussion of the regression analysis is presented in Appendix A.

3.4 RESULTING SEVERITY PREDICTION FORMULAS

The resulting formulas for predicting the probabilities of fatal accidents and injury accidents can be expressed in terms of several factors which are combined by simple mathematical operations. Each factor in the formulas represents a characteristic of the crossing as described in The Inventory. The probability of a fatal accident given an accident, $P(FA|A)$, is expressed as:

$$P(FA|A) = 1/(1 + CF \times MS \times TT \times TS \times UR) \qquad (3\text{-}8)$$

where: CF = formula constant = 695
MS = factor for maximum timetable train speed
TT = factor for thru trains per day
TS = factor for switch trains per day
UR = factor for urban or rural crossing

The probability of an injury accident given an accident, $P(IA|A)$, is expressed as:

$$P(IA|A) = \left[1 - P(FA|A)\right]/(1 + CI \times MS \times TK \times UR) \qquad (3\text{-}9)$$

where: $P(FA|A)$ = probability of a fatal accident, given an accident, obtained from Equation 3-8.

CI = formula constant = 4.280
MS = factor for maximum timetable train speed
TK = factor for number of tracks
UR = factor for urban or rural crossing

The equations for calculating values of the crossing characteristic factors are listed in Table 3-1 for the fatal accident probability formula and Table 3-2 for the injury accident probability formula. To simplify use of the formulas, the values of the crossing characteristic factors have been tabulated for typical values of crossing characteristics. These values are to be found in Tables 3-3 and 3-4 for the fatal accident and injury accident probability formulas, respectively. An inspection of the factor value tables shows the relative influence of the various factors on accident severity. In the case of fatal accident severity (Table 3-3) maximum timetable train speed has factor values which range over two orders of magnitude while the other factor values range over less than one order of magnitude. Maximum timetable train speed, therefore, has a much stronger influence on fatal accident severity than the number of trains or the trains or the urban-rural location of the crossing. For injury accident severity (Table 3-4) the number of tracks has a slightly greater influence on severity than maximum timetable train speed. The urban-rural location of the crossing has the least influence on injury accident severity.

TABLE 3-1. EQUATIONS FOR CROSSING CHARACTERISTIC FACTORS FOR FATAL ACCIDENT PROBABILITY FORMULA

Fatal Accident Probability Formula: $P(FA|A) = 1/(1 + CF \times MS \times TT \times TS \times UR)$

CROSSING CHARACTERISTIC FACTOR	EQUATION FOR CROSSING CHARACTERISTIC FACTOR
Formula Constant	$CF = 695$
Maximum Timetable Train Speed Factor	$MS = ms^{-1.074}$
Thru Trains Per Day Factor	$TT = (tt + 1)^{-0.1025}$
Switch Trains Per Day Factor	$TS = (ts + 1)^{0.1025}$
Urban - Rural Crossing Factor	$UR = e^{0.1880ur}$

where:

ms = maximum timetable train speed, mph

tt = number of thru trains per day

ts = number of switch trains per day

ur : urban crossing = 1, rural crossing = 0

TABLE 3-2. EQUATIONS FOR CROSSING CHARACTERISTIC FACTORS FOR INJURY ACCIDENT PROBABILITY FORMULA

Injury Accident Probability Formula: $P(IA|A) = \left[1 - P(FA|A)\right]/(1 + CI \times MS \times TK \times UR)$

CROSSING CHARACTERISTIC FACTOR	EQUATION FOR CROSSING CHRACTERISTIC FACTOR
Fatal Accident Probability	$P(FA|A)$, see Equation 3-8 and Table 3-1
Formula Constant	$CI = 4.280$
Maximum Timetable Train Speed Factor	$MS = ms^{-0.2334}$
Number of Tracks Factor	$TK = e^{0.1176tk}$
Urban - Rural Crossing Factor	$UR = e^{0.1844ur}$

where:

ms = maximum timetable train speed, mph

tk = total number of tracks at crossing

ur: urban crossing = 1, rural crossing = 0

TABLE 3-3. FACTOR VALUES FOR FATAL ACCIDENT PROBABILITY FORMULA

Fatal Accident Probability Formula: $P(FA|A) = 1/(1 + CF \times MS \times TT \times TS \times UR)$

FORMULA CONSTANT CF	MAXIMUM TIMETABLE TRAIN SPEED	MS	THRU TRAINS PER DAY	TT	SWITCH TRAINS PER DAY	TS	URBAN-RURAL CROSSING	UR
695.0	1	1.000	0	1.000	0	1.000	0 (rural)	1.000
	5	0.178	1	0.931	1	1.074		
	10	0.084	2	0.894	2	1.119	1 (urban)	1.207
	15	0.055	3	0.868	3	1.152		
	20	0.040	4	0.848	4	1.179		
	25	0.032	5	0.832	5	1.202		
	30	0.026	6	0.819	6	2.221		
	40	0.019	7	0.808	7	1.238		
	50	0.015	9	0.790	9	1.266		
	60	0.012	10	0.782	10	1.279		
	70	0.010	20	0.732	20	1.366		
	80	0.009	30	0.703	30	1.422		
	90	0.008	40	0.683	40	1.464		
	100	0.007	50	0.668	50	1.497		

TABLE 3-4. FACTOR VALUES FOR INJURY ACCIDENT PROBABILITY FORMULA

Injury Accident Probability Formula: $P(IA|A) = \left[1 - P(FA|A)\right]/(1 + CI \times MS \times TK \times UR)$

FATAL ACCIDENT PROBABILITY, P(FA!A)	FORMULA CONSTANT CI	MAXIMUM TIMETABLE TRAIN SPEED	MS	TOTAL NUMBER OF TRACKS	TK	URBAN-RURAL CROSSING	UR
See Equation 3-8 and Tables 3-1 & 3-3	4.280	1	1.000	0	1.000	0 (rural	1.000
		5	0.687	1	1.125	1 (urban)	1.202
		10	0.584	2	1.265		
		15	0.531	3	1.423		
		20	0.497	5	1.800		
		25	0.472	6	2.025		
		30	0.452	7	2.278		
		40	0.423	8	2.562		
		50	0.401	9	2.882		
		60	0.385	10	3.241		
		70	0.371	15	5.836		
		80	0.360	20	10.507		
		90	0.350				
		100	0.341				

3.5 Use of Severity Prediction Formula

A sample application of the fatal and injury accident severity formula for a typical crossing is provided to demonstrate their use. Characteristics of the sample crossing are listed below in Table 3-5.

TABLE 3-5. CHARACTERISTICS OF SAMPLE CROSSING

CHARACTERISTIC	VALUE
Maximum Timetable Train Speed, mph	40
Thru Trains Per Day	10
Switch Trains Per Day	5
Total Number of Tracks (main plus other)	2
Urban-Rural Location	Rural

To determine the probability of a fatal accident given an accident at the sample crossing, Equation 3-8 is used. Values for the factors in the fatal accident severity formula (Equation 3-8) can be computed from the equations listed in Table 3-1 or looked up in Table 3-3. Using the look-up table, the following factor values are found for the crossing characteristics specified:

$CF = 695.0$
$MS = 0.019$
$TT = 0.782$
$TS = 1.202$
$UR = 1.000$

Substituting the factor values into the fatal accident probability formula yields:

$$P(FA|A) = 1/(1 + CF \times MS \times TT \times TS \times UR)$$

$$= 1/(1 + 695.0 \times 0.019 \times 0.782 \times 1.202 \times 1.000)$$

$$= .075 \text{ (the probability of a fatal accident given an accident)}$$

To determine the probability of an injury accident given an accident, at the same sample crossing, Equation 3-9 is used. Values for the factors in Equation 3-9 can be obtained from the equations listed in Table 3-2 or from Table 3-4. Using the look-up table, the following factor values are found for the characteristics of the sample crossing:

$P(FA|A) = .075$ (from fatal accident severity formula)
$CI = 4.280$
$MS = 0.423$
$TK = 1.265$
$UR = 1.000$

Substituting the factor values into the injury accident probability formula yields:

$$P(IA|A) = [1 - P(FA|A)]/(1 + CI \times MS \times TK \times UR)$$
$$= (1 - .075)/(1 + 4.280 \times 0.423 \times 1.265 \times 1.000)$$
$$= 0.281 \text{ (the probability of an injury accident given an accident)}$$

4. SEVERITY FORMULA PERFORMANCE

To illustrate characteristics of the fatal and injury severity formulas, the two functions P(FA|A) and P(IA|A) are plotted as a function of maximum timetable train speed in Figure 4-1. The figure contains five individual plots which show how the functions change when one of the other four factors which influence severity (thru trains, switch trains, tracks and urban-rural location) is varied. The values of the factors are shown on the individual plots.

Several observations can be made regarding the characteristics of the functions. The probability of a fatal accident given an accident P(FA|A) increases as a nearly linear function of timetable train speed. Changes in the number of thru and switch trains or the urban-rural location of the crossings does not have a major influence on fatal accident severity.

The probability of an injury accident given an accident P(IA|A) increases as a nonlinear function of timetable train speed. Injury accident severity generally increases rapidly with timetable train speed and then remains relatively constant beyond 40 mph. The function actually decreases at high speeds under certain conditions as previously predicted from observation of actual accident data (see Figure 3-7). The number of tracks at the crossing has a significant influence on the function (injury accident severity decreases with the number of tracks); however, the urban-rural location has only a minor influence.

The performance of the severity formulas was evaluated using two methods: (1) comparing predicted versus actual severity for sample sets of accidents, and (2) comparing their ability to rank accidents by severity versus a random ranking. Results of the first evaluation are summarized in Table 4-1. Using 1978, 1979 and 1980 data, the severity formulas were used to predict the number of fatal and injury accidents for sets of accidents which occurred in 1981. The predictions were then compared with actual accident records for the same set of accidents. The set of accidents considered were selected from the top of a list of accidents ranked by predicted severity. As Table 4-1 demonstrates, the severity prediction formulas compare well with observed data. For example, the first row shows that, for the top 100 accidents in 1981, the

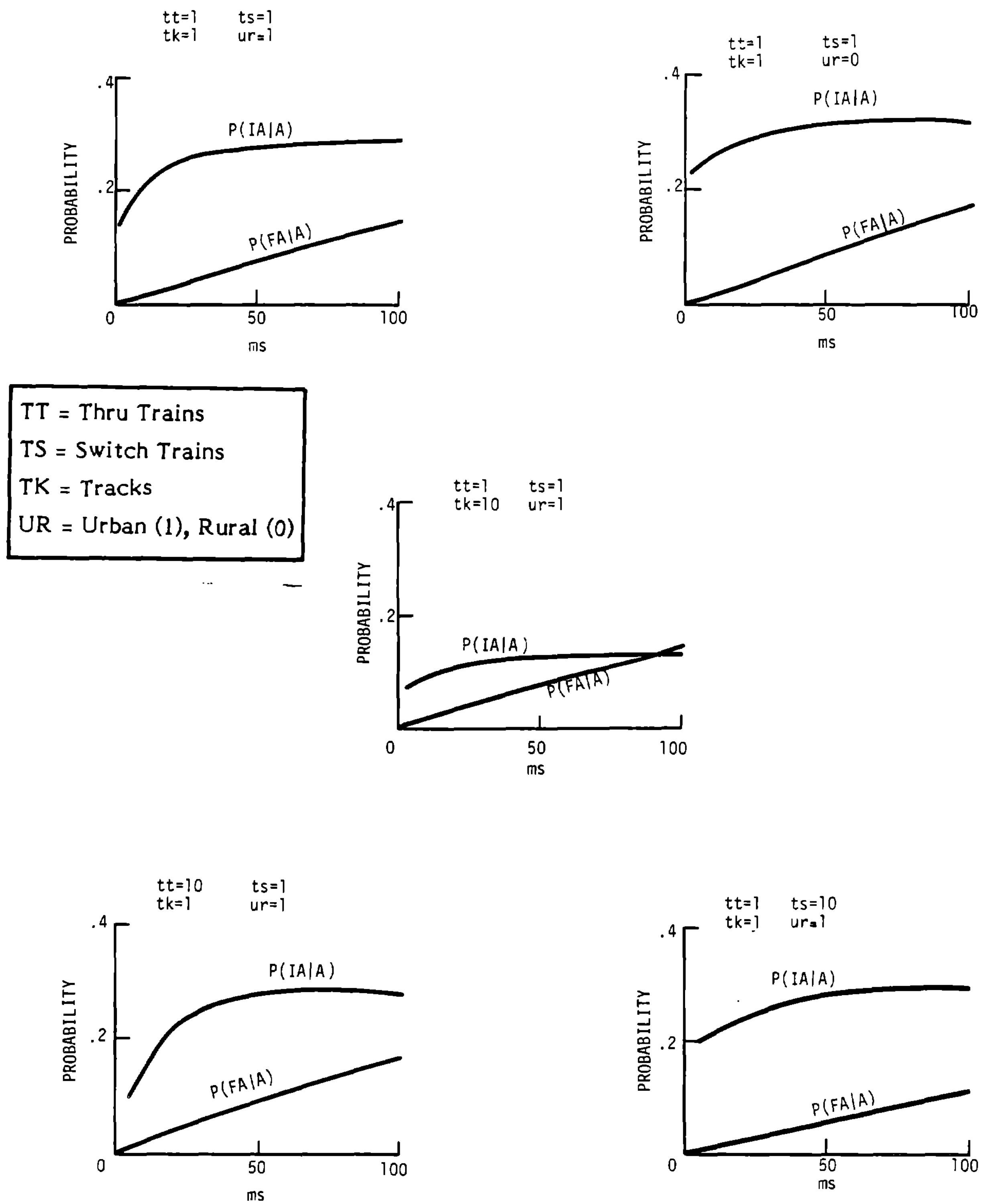


FIGURE 4-1. TYPICAL PLOTS OF PROBABILITY OF FATAL ACCIDENTS P(FA|A) AND PROBABILITY OF INJURY ACCIDENTS P(IA|A) AS A FUNCTION OF TIMETABLE TRAIN SPEED ms.

formulas predicted 18.2 fatal accidents versus 13 actual and 31.3 injury accidents versus 42 actual. It should be noted that the predicted severity values represent expected long-term annual rates and should be used with caution when estimating severity at individual crossings, particularly for a short-term period.

TABLE 4-1 PREDICTED VERSUS ACTUAL ACCIDENT SEVERITY

NUMBER OF RANKED ACCIDENTS	PREDICTED FATAL ACCIDENTS	ACTUAL FATAL ACCIDENTS	PREDICTED INJURY ACCIDENTS	ACTUAL INJURY ACCIDENTS
100	18.2	13	31.3	42
500	79.3	76	154.2	171
1000	142.6	145	305.9	348
7934	511.9	539	2018.5	2192

Results of the second evaluation of the severity formulas are based on the premise that, for accidents properly ranked by predicted severity, those at the top of the list (the most severe) should have a higher than average number of actual fatal and injury accidents. On the other hand, accidents at the top of a randomly ranked list should have only an average number of actual fatal and injury accidents. The ratio of actual accident severity for a set of accidents ranked by predicted severity to actual accident severity for the same size set of randomly ranked accidents is a measure of the prediction formula's ability to identify potentially severe accidents. This measure is referred to as the "power factor" for the prediction formula.

The power factors for the fatal and injury formulas for sets of accidents, ranked by predicted severity, are shown in Table 4-2. The table indicates, for example, that for the top 100 ranked accidents the power factors for the fatal and injury formulas are 1.91 and 1.57, respectively. This means that the top 100 accidents ranked by the formulas have 1.91 and 1.52 times the number of fatal and injury accidents, respectively, as a randomly selected set of 100 accidents. Similar comparisons are made for the top 500 and 1000 accidents. The results all show that the fatal and injury severity formulas are quite effective in predicting accident situations which tend to be more severe than the average.

TABLE 4-2. RANKING PERFORMANCE OF SEVERITY FORMULAS

NUMBER OF RANKED ACCIDENTS	FATAL SEVERITY FORMULA POWER FACTORS*	INJURY SEVERITY FORMULA POWER FACTORS*
100	1.91	1.52
500	2.24	1.24
1000	2.13	1.26

* Actual severity for ranked group of accidents/actual severity for randomly selected group of accidents.

REFERENCES

1. Coulombre, R., et al., Summary of the Department of Transportation Rail-Highway Crossing Accident Prediction Formula and Resource Allocation Model, Report No. DOT-TSC-FRA-82-1, (Washington, D.C.: Federal Railroad Administration, September, 1982).

2. J. Hitz and M. Cross, Rail Highway Crossing Resource Allocation Procedure User's Guide, No. FHWA-1P-82-7, (Washington, D.C.: Federal Railroad Administration, December 1982).

3. J. Hitz, ed., Summary Statistics of the National Railroad-Highway Crossing Inventory for Public at Grade Crossings, No. FRA-RPD-78-20, (Washington, D.C.: Federal Railroad Administration, September 1978).

4. Federal Railroad Administration, Rail-Highway Crossing Accident/Incident and Inventory Bulletin No. 4, 1981, (Washington, D.C.: U.S. Department of Tranpsortation).

5. National Transportation Safety Board, Risk Concepts in Dangerous Goods Tranpsortation Regulations, Washington, D.C., 1971.

6. P. Mengert, Rail-Highway Crossing Hazard Prediction Research Results, No. FRA-RRS-80-02, (Washington, D.C.: U.S. Department of Transportation, March 1980).

7. D. R. Cox, Analysis of Binary Data, Halstead Press/John W. Stey, New York, 1970.

8. J. Coleman and G. R. Stewart, Investigations of Railroad-Highway Grade Crossing Accident Data, Transportation Research Record, No. 611, Washington, D.C., National Academy of Sciences, 1976.

9. E. Farr, Rail-Highway Crossing Resource Allocation Model, No. FRA-RRS-81-001, (Washington, D.C.: U.S. Department of Transportation, April 1981.

10. J. Heisler, and J. Morrissey, Rail-Highway Crossing Warning Device Life Cycle Cost Analysis, FRA-RRS-80-003, (Washington, D.C.: U.S. Department of Transportation, 1980).

11. J. Morrissey, the Effectiveness of Flashing Lights and Flashing Lights with Gates in Reducing Accident Frequency at Public Rail-Highway Crossings, FRA-RRS-80-005, (Washington, D.C.: U.S. Department of Transportation, March 1980).

APPENDIX: DETAILED DESCRIPTION OF PROCEDURE FOR DEVELOPMENT OF FATAL AND INJURY ACCIDENT SEVERITY FORMULAS

A.1 General Formulation - Fatal Accident Formula

In describing this analysis, a fatal accident will be denoted by +1 and a non-fatal accident by -1. With these observed values as the dependent variable, a model is developed to produce a value f, which will be an approximation to -1 or +1 for any given crossing. This value of f is used to get P(FA|A) by substituting in the formula:

$$P(FA|A) = 1/(1 + e^{-2f})$$

For analysis purposes, f can be considered to fall in the range $-\infty < f < \infty$. P(FA|A) is then seen to be in the range $0 < P(FA|A) < 1$, with small values of P(FA|A) occurring for negative values of f.

The procedures used in determining the desired formula involves a certain amount of judgment, based on experience in analyzing rail-highway crossing accidents, along with rather sophisticated analytic techniques. The analytic technique uses logistic discriminants and is a modification of a method described by Cox (Ref. 7). The procedure consists of the following five steps:

1. From the histograms of Section 3.2, determine which factor or factors are the most significant and the functional form relating probability of fatal accident to the factor or factors selected. This becomes the basic function which is used in subsequent steps in determining the formula. In this case one factor stood out above the rest: train speed, which is listed as "maximum timetable speed" in The Inventory. Denoting train speed as ms, it was determined from the histograms that log ms was the best functional form for this factor.

2. Calculate the coefficients A and B in the speed formula: A + B log ms. This is done by the iterative process outlined in step 5.

3. Determine which factors other than the basic speed formula are potentially significant and deserve further analysis. Also determine the functional form of these factors. This step is a judgment process involving the use of the histograms of Section 3.2. The factors selected become candiates for inclusion in the final fomula. In the present case the factors are:

$\log \dfrac{tt + 1}{ts + 1}$, ur, ur x log ms, log (tt + 1), log (ts+1), where:

tt = Number of thru trains per day

ts = Number of switch trains per day

$$ur = \begin{cases} 1 \text{ if urban crossing} \\ 0 \text{ if rural crossing} \end{cases}$$

The log functions are for base 10.

4. This step is called "selection regression". The factors from step 3, are selected sequentially from a list ranked in order of decreasing amount by which the sum of squares of errors is reduced in a regression formula. As each factor is selected, a t-value is calculated and the factor is accepted for the final formula if it's t-value is greater than four (in absolute value) and its presence does not significantly lessen the t-values of the previous factors. If a factor fails either of these tests, it is not included in the final formula. The factors surviving these tests are:

   log ms, $\log \dfrac{tt + 1}{ts + 1}$, and ur.

5. This step determines the coefficients of the final formula, which is of the form:

$$f = C_0 + C_1 \log ms + C_2 \log \frac{tt + 1}{ts + 1} + C_3 ur.$$

These coefficients are determined by a process called "logit analysis" (Ref. 6). This is an iterative process in which a first approximation to the values of the coefficients $C_0^{(1)}$, $C_1^{(1)}$, $C_2^{(1)}$ and $C_3^{(1)}$ are obtained by ordinary least squares using +1 and -1 for values of the dependent variable f. An iterative equation, used to get successively more accurate values of the coefficients, is

$$f^{(i)} = C_o^{(i)} + C_1^{(i)}\log ms + C_2^{(i)}\log \frac{tt+1}{ts+1} + C_3^{(i)}ur$$

Nonlinear functions applied to the dependent variable are introduced:

$$U(f) = \operatorname{sech}^2 f\ (\tanh f/f)$$

$$V(f) = \operatorname{sech}^2 f\ (f/\tanh f)$$

To obtain coefficients at step i + 1, given the coefficients at step i, new variables are introduced using the above nonlinear transformations:

$$(\log ms)^{(i)} = (\log ms)\sqrt{U(f^{(i)})}$$

$$\left(\log \frac{tt+1}{ts+1}\right)^{(i)} = \left(\log \frac{tt+1}{ts+1}\right)\sqrt{U(f^{(i)})}$$

$$(ur)^{(i)} = ur\sqrt{U(f^{(i)})}$$

$$Y^{(i)} = Y\sqrt{V(f^{(i)})}$$

where Y is +1 for a fatal accident and -1 for a non-fatal accident. These transformations use the calculated value $f^{(i)}$ obtained by using the coefficients at step i and the factors evaluated for each crossing (i.e., accident). The new coefficients are obtained by ordinary least squares using the transformed equation:

$$\hat{Y}^{(i)} = C_o^{(i+1)} + C_1^{(i+1)}(\log ms)^{(i)} + C_2^{(i+1)}\left(\log \frac{tt+1}{ts+1}\right)^{(i)} + C_3^{(i+1)}(ur)^{(i)}$$

and minimizing $\sum(\hat{Y}^{(i)} - Y^{(i)})^2$, where the summation is over all the accidents. The iteration is continued until the differences in the coefficients for two successive steps are less than some predetermined amount.

A.2 General Formulations - Injury Accident Formula

The analysis for the injury formula is basically the same as for the fatality formula. An injury accident given a non-fatal accident will be denoted by +1 and a non-injury non-fatal accident by -1. An injury model is constructed to produce a value i, which will be an approproximation to -1 or +1 for any given crossing. This value of i is used to get P(IA|NFA) by substituting in the formula:

$$P(IA|NFA) = 1/(1 + e^{-2i})$$

The five step procedure outlined in the previous section is followed with only slight variation:

1. Same as before.

2. Same as before.

3. The potentially significant factors, other than the speed formula, selected as candidates for the final injury formula are:

   $$tt,\ ts,\ \log \frac{tt + 1}{ts + 1}$$, main tracks, other tracks, total tracks (tk),

   flashing lights (yes/no), gates (yes/no), highway paved (yes/no), ur, lanes, functional class, ur x log ms.

4. The factors surviving these tests for the injury formula are log ms, tk, and ur.

5. The final formula is of the form:

   $$i = D_0 + D_1 \log ms + D_2 tk + D_3 ur.$$

   In this case i is +1 for an injury accident and -1 for a non-injury non-fatal accident. The coefficients D_0, D_1, D_2, D_3 are determined in the same way as the previous case.

A.3 Selection Regression - Fatal Accident Formula

To provide a better understanding of the procedure used in developing the formulas and to provide a quantitative measure of the relative significance of the different factors, some results are provided in this section. These results are for step 4 of Section A.1 dealing with selection regression.

The speed formula calculated in step 2 of Section A.2 is -3.626 - 1.471 log ms. The sum of squares of errors reduced by each of the factors in step 3 are:

TABLE A-1. SUM OF SQUARES OF ERRORS REDUCED - FATALITY FACTORS

FACTOR	SUM OF SQUARES OF ERRORS REDUCED
Speed Formula	424.04
$\text{Log } \frac{tt + 1}{ts + 1}$	4.68
ur	2.47
ur x log ms	.50
Log tt	.083
Log ts	.002

From this table it is seen that train speed has a much larger sum of squares of errors reduced than the other factors.

The t-values for four successive steps in the selection regression process are shown below:

TABLE A-2. t-VALUES FOR FATALITY FORMULA FACTORS

Factor	t-value Step 1	t-value Step 2	t-value Step 3	t-value Step 4
Speed Formula	64	53	52	31
$\text{Log}\ \dfrac{tt + 1}{ts + 1}$		6.7	6.3	6.4
ur			-4.9	-2.7
ur x log ms				2.2

Step 3 is the last step in which the t-values are greater than four in absolute values and at the same time the previous t-values have not deteriorated significantly. Therefore, the three factors at step 3, train speed, log (tt + 1/ts + 1), and ur are selected for the final formula.

For step 5, since the speed formula is a linear function of log ms, the factor log ms is substituted for the speed formula and a new coefficient for log ms is calculated.

A.4 Selection Regression - Injury Accident Formula

The speed formula calculated in step 2 of Section A.2 is -.8855 + 2679 log ms.

Since so many other factors were entered into selection regression for the injury formula than for the fatality formula, two sets of factors had to be analyzed because of a limitation on the computer program. After these two mutually exclusive sets of factors were analyzed, the strongest candidate factors were selected from each set to form a final candidate set to be analyzed. These final factors along with their sum of squares of errors reduced are shown in the following table:

TABLE A-3. SUM OF SQUARES OF ERRORS REDUCED - INJURY FACTORS

FACTOR	SUM OF SQUARES OF ERRORS REDUCED
ur	42.9
Total Tracks (tk)	24.5
Speed Formula	23.7
Gates	14.9
Lanes	10.9
Log ts	6.1
Functional Class	3.9
Log (tt + 1/ts + 1)	.3

In comparison to the sum of squares for the fatality formula, the speed formula for the injury formula is not nearly so strong. In fact, ur and tk have a larger sum of squares.

The t-values for four successive steps in the selection regression process are shown below:

TABLE A-4. T-VALUES FOR INJURY FORMULA FACTORS

FACTOR	T-VALUE STEP 1	T-VALUE STEP 2	T-VALUE STEP 3	T-VALUE STEP 4
ur	-8.0	-6.9	-4.5	-3.4
tk		-6.0	-6.4	-5.5
Speed Formula			5.9	-7.0
Gates?				-4.7

In step 4, the t-values for ur has deteriorated significantly and even though the t-value for Gates? is greater than four in absolute value, it was decided to reject this factor. Therefore, the three factors at step 3, ur, tk, and speed formula, are selected for the final formula.

As in the fatality formula for step 5, log ms is substituted for the speed formula and a new coefficient for log ms is calculated.

A.5 Final Accident Severity Formulas

After the coefficients are determined by step 5, the final formulas are:

Fatality Formula

$$f = -3.272 + 1.236 \log ms - .09415\, ur + .1180 \log \frac{tt + 1}{ts + 1}$$

$$P(FA|A) = 1/(1 + e^{-2f})$$

Injury Formula

$$i = -.7267 + .2688 \log ms - .09221\, ur - .05881\, tk$$

$$P(IA|NFA) = 1/(1 + e^{-2i})$$

$$P(IA|A) = \left[1/(1 + e^{-2i})\right] / \left[1 - 1/(1 + e^{-2f})\right]$$