Highway-Rail Grade Crossing Incident Consequence Analysis

Submitted for consideration for presentation and publication at the Transportation Research Board 97th Annual Meeting

Submission Date: 1 August, 2017

Paper Number: 18-06474

Lijun Zhang[1], Samantha Chadwick, and Christopher P. L. Barkan

Rail Transportation and Engineering Center
Department of Civil and Environmental Engineering
University of Illinois at Urbana-Champaign
205 N. Mathews Ave, Urbana, IL, 61801

Lijun Zhang[1]	Samantha Chadwick	Christopher P. L. Barkan
(217) 419-2694	(508) 667-0259	(217) 244-6338
lzhng110@illinois.edu	schadwi2@illinois.edu	cbarkan@illinois.edu

4,031 words + 6 Figures + 7 Tables = 7,281

[1] Corresponding author

ABSTRACT

Highway rail grade crossing safety has improved considerably in the past two decades. Among the reasons is investment in more effective warning systems. Such upgrades are costly and resources are limited so decisions on which crossings should receive these investments should be made so that they will have the greatest impact on improving safety. A key tool for prioritizing grade crossings warning system upgrades are analysis tools that are used to develop quantitative metrics of the likelihood of incidents given the characteristics of a crossing. These tools have received a great deal of research attention and part of the success in reducing grade crossing incidents can be attributed to their use. Less research has been devoted to development of a comprehensive, quantitative approach to measure the consequences of grade crossing incidents. Consequences are the other element of risk analysis so development of more accurate metrics for this aspect of grade crossing incidents will improve our understanding of grade crossing risk. This paper describes a statistical approach using US DOT data and other information to quantitatively predict consequences in terms of several metrics and under a variety of circumstances.

INTRODUCTION

Researchers have developed various models to estimate collision and derailment likelihood at railroad grade crossings (4). One of the earliest is the Peabody-Dimmick formula developed in 1941. It can be used to predict the number of incidents at a grade crossing over a 5-year period based on annual average daily traffic (AADT), average daily train traffic, protection factor, and a smoothing factor (5). Another method is the New Hampshire model, which calculates a hazard index that can be used to predict the expected number of collisions at a crossing (6). These models can also be improved to account for local characteristics. Benekohal (7) developed a model to predict collision likelihood at grade crossings in urban areas in Illinois. The most common model used for incident likelihood prediction in the U.S. is the U.S. Department of Transportation (DOT) Accident Prediction Model (5, 8, and 9).

Models have also been developed for grade crossing incident consequence analysis, such as casualty and property damage assessment. The U.S. DOT developed an equation to estimate the probability of an injury or a fatality in an accident (9). In terms of crashworthiness, Martinez et al. (10) developed a model based on full-scale collision testing to predict crash patterns, while Samavedam and Kasturi (11) improved previous models to more accurately predict damage to locomotives. Saccomanno at el. (12) developed a consequence prediction model to define grade crossing severe spots based on weighted sum of different types of consequence, including injuries, fatalities, and property damage to obtain a crossing collision consequence score. The Australian Level Crossing Assessment Model (ALCAM) model provides an overview of grade crossing consequences, whose risk score is affected by infrastructure, exposure, and consequence factors (13).

Previous research provided comprehensive models to estimate incident likelihoods and predictions, but it is important to also assess the consequences of grade crossing incidents. The goal of this research is to develop a more comprehensive grade crossing incident consequence severity model, that considers injuries, fatalities, property damage, and traffic delay for both highways and railways, and derailment factors. The paper involves data analytics, including Akaike information criterion (AIC), Bayesian information criterion (BIC), and shrinkage methods for consequence model variable selection, multiple regressions for prediction and regression trees for classification.

DATA SOURCE AND METHODOLOGY

This consequence model involves two types of variables: incident-specific variables and grade-crossing specific variables (Table 1). The DOT U.S. Federal Rail Administration (FRA) records annual train incident statistics in the Rail Equipment Accident or Incident (REA) database and the Highway Rail Accident or incident (HRA) database, and these were both used to obtain incident-specific variables. The two databases contain different complementary information. The REA database records data for all incidents that exceed a certain monetary threshold. It provides information such as train speed and number of cars derailed (14). The HRA database contains all grade-crossing collisions regardless of the damages and it provides some other incident-specific information such as highway vehicle speed and train speed at collision (14). The third database is FRA Grade Crossing Inventory, that includes information on the conditions of each grade crossing, such as AADT, highway class, and type of crossing.

TABLE 1: Variables Classified by Incident or Crossing Specific

Incident Specific	Crossing Specific
	AADT
derailment	total trains per day
train speed	train type
vehicle speed	warning device
accident type	track class
vehicle size	highway class
	type of crossing

A methodology was developed to combine these three databases to obtain inputs for the consequence model, including information on train derailment such as train speed, highway vehicle speed, incident type, etc, as well as information on the response variables, including casualties, property damage, hazardous material release, delay on highway, cost of railway lost time, maintenance, and repair, etc.. Freight train and passenger train incidents were first separated, and then the data were further divided into derailment and non-derailment incidents. Some unreliable and incomplete records were reviewed based on narrative information and omitted accordingly. After the data cleanup, there were 38,240 data points for freight train incidents and 513 data points for passenger train incidents from 1996 to 2015. For property damage analysis, only incidents above the REA reporting threshold had data, leaving 2,459 data points for freight train incidents, and 503 data points for passenger train incidents.

Consequence Model Methodology

The objective of the consequence model is to predict consequence severity for highway-rail grade crossing incidents based on various characteristics using previous incident data to determine probability for the occurrence of certain conditions (Figure 1).

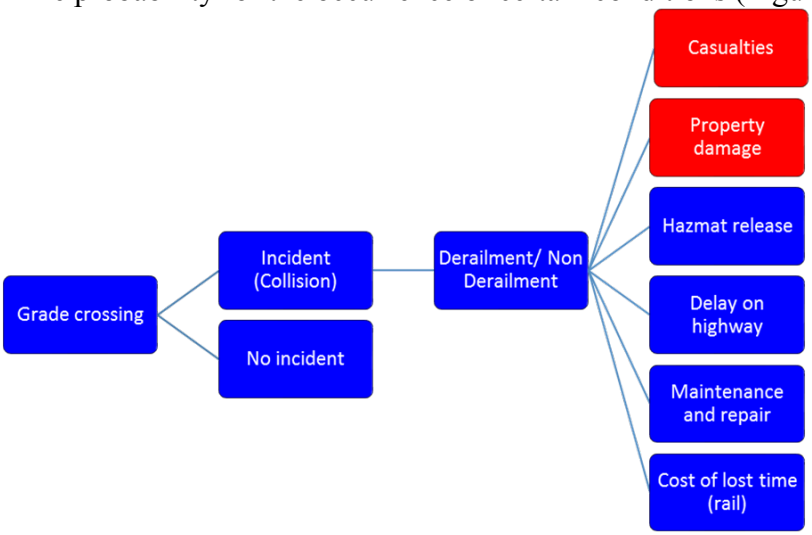


FIGURE 1. Consequence Model Framework

Incidents are separated by collision and derailments. For each condition, there are two basic prediction factors for consequence severity, probability (Pr) and consequence (C) measurements.

The final severity ranking is based on the product of two predictors, Pr×C. The consequence indicates the outcomes, such as casualties and property damage. For each outcome, there is an individual regression model for prediction based on the significant variable selection using mathematical and statistical methods, for both probability and consequence. For example, when a grade crossing incident occurs and results in a derailment, casualty analysis would be defined as follows:

$$\Pr(I) \times \Pr(D|I) \times \Pr(C|D|I) \times Casualty$$

Where $\Pr(I)$ represents the probability of an incident occuring at a grade crossing, $\Pr(D|I)$ represents the probability of a derailment if an incident occurs, $\Pr(C|D|I)$ represents probability of casualties when derailment incident occurs, Casualty means the number of actual casualties. Property damage analysis would be defined as:

$\Pr(I) \times \Pr(D|I) \times \Pr(\$DMG|D|I) \times Cost$

Where $\Pr(I)$ represents probability of an incident at a grade crossing, $\Pr(D|I)$ represents probability of derailment if an incident occurs, $\Pr(\$DMG|D|I)$ represents probability of property damage if an incident occurs with derailment, and Cost is the total amount of damage in dollars.

Therefore, in order to develop this consequence model, both probability and consequence predictions are needed. It is necessary to understand if each variable will influence probability or consequence or both, and Table 2 demonstrates the variables included in different predictions.

TABLE 2. Variables Classified by Their Impact on Probability or Consequence

Probability	Consequence	Probability & Consequence
AADT	train type	speed (train & vehicle)
total trains per day		track class
		highway class
		vehicle size

The variables AADT and total trains per day only affect the probability model because they influence the likelihood of an incident, but not the consequence severity. Average passenger trains per day influences consequences because passenger and freight trains differ in the resultant casualties and property damage. Train speed and vehicle speed are related to track class and highway class, so they will affect both probability and consequence. Vehicle size influences both probability and consequence because vehicle dynamic characteristics influence incident likelihood, and it can also cause different damage and casualties. “Warning device” reflects different warning systems for each grade crossing, so it has a significant effect on probability, and will be examined to determine if it also has a relationship with consequence.

Previous research has been done to predict the probability of an incident at a grade crossing, $\Pr(A)$. In this consequence model, the USDOT accident prediction model (6) was used, and the general expression of the formula is as follows:

$$a = K \times EI \times MT \times DT \times HP \times MS \times HT \times HL$$

$$B = \frac{T_0}{T_0 + T} * a + \frac{T}{T_0 + T} * \frac{N}{T}$$

$$A = \begin{cases} 0.7159B(for\ passeive\ devices\ ) \\ 0.5292B(for\ flashing\ lights) \\ 0.4921B(for\ gates) \end{cases}$$

Where:
a = initial collision prediction
K = constant
EI = factor for exposure index
MT = factor for number of main tracks
DT = factor for number of trains per day
HP = factor for highway paved
MS = the factor for maximum timetable speed
HT = factor for highway type
HL = factor for number of highway lanes
B = adjusted accident frequency value
T0 = formula weighting factor
A = normalized accident frequency value.

Chadwick's (4) model can be used to obtain derailment likelihood $\Pr(D|A)$. The remaining elements to be investigated are $\Pr(C|D,I)$, $Casualty$, $Pr(\$DMG|D,I)$, and $Cost$.

Casualty Analysis

Highway-rail grade crossing incidents have a higher number of casualties, including fatalities and injuries than other type of incidents. Although the number of grade crossing incidents has been decreasing in recent years (Table 3), a large number of fatalities and injuries still occur. Therefore, understanding the impact of further improvements would have on incident casualty severity is important. Several studies investigated fatalities and injuries, one of which resulted in the U.S. DOT's model predicting the probability of injury or fatality in a grade crossing incident. It includes factors for maximum timetable speed, number of tracks, train traffic, and crossing type (9).

$$P(IA|A) = \frac{1 - P(FA|A)}{1 + CI \times MS \times TK \times UR}$$

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

Where:
CI = formula constant = 4280
CF = formula constant = 695
MS = factor for maximum timetable train speed
TK = factor for number of tracks
TT = factor for through trains per day
TS = factor for switch trains per day
UR = factor for urban or rural crossing

As for injury severity analysis, existing research on motor-vehicle crashes has developed multinomial logit, nested logit, and ordered probit models, demonstrating the effect of each variable (15). Eluru et al. analyzed incidents at private grade crossings, and Haleem (16) developed a mixed logit model to define the probability of have at least casualties, including fatalities and injuries. The unconditioned mixed logit probability is estimated as follows:

$$P_{jn} = \int \frac{\exp(\beta_j X_{jn})}{\sum_j \exp(\beta_j X_{jn})} f(\frac{\beta}{\theta}) d\beta$$

Where β_j= vector of parameters to be estimated for crash injury severity category j (i.e., injury or fatality)

X_{jn}= vector of independent variables for crash injury severity category for crash n

$f(\frac{\beta}{\theta})$ = the density function of β and θ is the vector of parameters for the assumed distribution (e.g., mean and variance for the normal distribution).

TABLE 3. Number of Grade Crossing Incidents, 1996 – 2015

	Number of Incidents			
	Derailment		Non-derailment	
	Passenger	Freight	Passenger	Freight
1996	4	17	264	3,239
1997	2	22	279	2,915
1998	1	17	253	2,584
1999	1	18	305	2,566
2000	1	16	296	2,706
2001	1	21	266	2,377
2002	1	15	244	2,270
2003	2	16	240	2,204
2004	0	12	249	2,294
2005	2	14	229	2,269
2006	2	20	248	2,134
2007	1	14	244	2,005
2008	5	11	239	1,758
2009	1	9	263	1,325
2010	2	14	253	1,439
2011	2	12	180	1,457
2012	0	14	137	1,469
2013	2	11	151	1,491
2014	0	12	199	1,627
2015	1	11	170	1,461
Total	31	296	4,709	41,590

More specifically, there are three casualty types in grade crossing incidents: railroad employee, passenger, and highway user. The latter includes users of automobiles, buses, trucks, motorcycles, bicycles, pedestrians, and other transport modes. For passenger train incidents between 1996 and 2015, the majority of casualties were passengers, with 333 and an average of 10.74 per incident (Table 4a). For freight train derailment incidents, the majority of casualties were highway users, with 208 and an average of 0.7 per incident. Highway users had the highest casualties, followed by railroad employees.

TABLE 4. (a) Grade Crossing Derailment Accident Average Casualty Summary and (b) Grade Crossing Non-Derailment Accident Average Casualty Summary

(a)

	Casualties for Derailment Incidents (327 incidents)							
	Passenger (31 incidents)				Freight (296 incidents)			
	Hw User	RR Emp	Pass	Total	Hw User	RR Emp	Pass	Total
Total	23	83	333	439	208	173	0	381
Average per Incident	0.74	2.68	10.74	14.16	0.70	0.58	0.00	1.29

(b)

	Casualties for Non-Derailment Incidents (46,299 incidents)							
	Passenger (4,709 incidents)				Freight (41,590 incidents)			
	Hw User	RR Emp	Pass	Total	Hw User	RR Emp	Pass	Total
Total	2,722	372	1,015	4,109	19,211	922	0	20,133
Average per Incident	0.58	0.08	0.22	0.87	0.46	0.02	0.00	0.48

Highway user casualties per incident are fairly similar for passenger and freight train incidents. For passenger train incidents, highway users are now the highest in frequency and average casualties. For freight train incidents, highway users remain the highest frequency and average casualty per incident.

The results show that derailments have the greatest impact on the number of casualties. It is also necessary to separate freight and passenger train incidents because the latter have considerably more train occupants. To compare this difference, Figure 2 illustrates the casualty frequency for highway users and train occupants, the latter includes both railroad employees and passengers. The four charts show each combination of train and incident type. There are few observable differences between the different types of casualties and they display similar distributions. There are two types of grade crossing incident, “train struck vehicle (TSV), and vehicle struck train (VST)” (4), and this also affects casualties. After a series of regression tests, in addition to derailment and incident type, train speed and vehicle speed were also significant variables for casualty prediction.

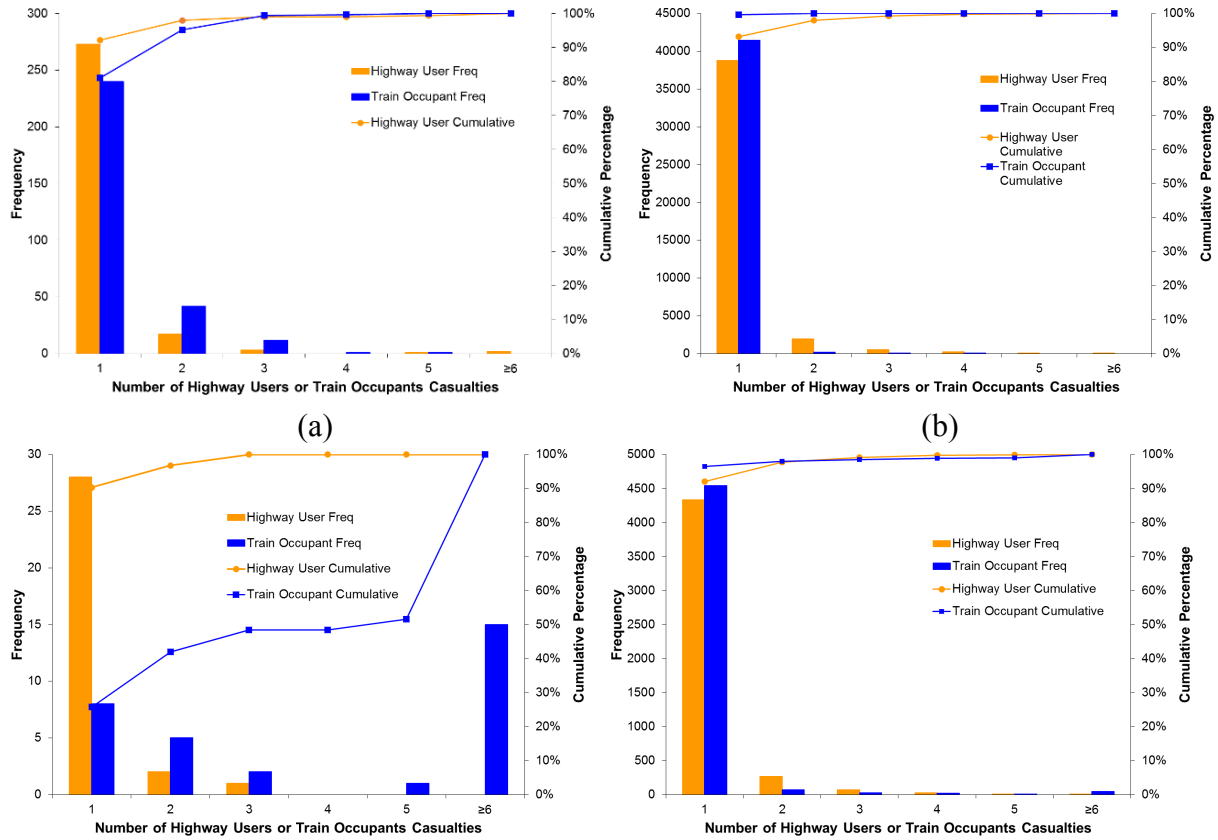


Figure 2. (a) Number of Casualties of Freight Train Derailment Accidents, (b) Number of Casualties for Freight Train Non-Derailment Accidents, (c) Number of Casualties for Passenger Train Derailment Accidents, and (d) Number of Casualties for Passenger Train Non-Derailment Accidents

Based on the data for the past 20 years, several regression models were compared to find the relationship between casualties and dependent variables, including simple linear regression, linear regression with transformed variables, and Poisson regression. The data showed that more cars derailed at higher speed does not necessarily indicate more casualties; instead, instances with one or two cars derailed had the highest casualties. Because the casualty data were not quite smooth based on the selected variables; regression tree was selected for prediction (17). Regression tree is a recursive partitioning process that splits data into branches using algorithms and can be used to represent decisions visually and explicitly (18). It acts as a predictive model using classification method, number of terminal nodes, and complexity parameter for tree selection. Based on the cost-complexity criterion including number of terminal nodes and complexity parameter, the lowest variance tree is selected as the optimal solution. The algorithm minimizes sum of squares error as shown below:

$$SSE = \sum_{j=1}^{J} \sum_{i \in R_j} \left(y_i - \hat{y}_{R_j} \right)^2$$

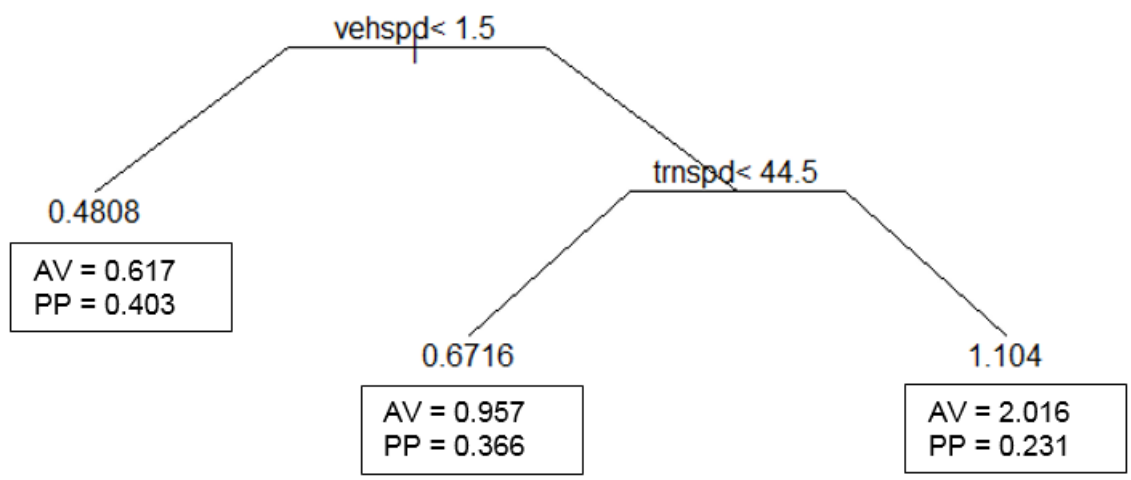
Where:

R_j=$region\ for\ partition\ j$

$\hat{y}_{R_j}$=$mean\ for\ training\ observations\ in\ j\ partition$

After selecting variables for freight and passenger train grade crossing incident data individually, freight train data were separated into four types: TSV (train struck vehicle) with derailment, VST (vehicle struck train) with derailment, TSV without derailment, and VST without derailment. Passenger train data were separated into just two types: TSV and VST, due to the smaller amount of data available for derailment incidents (only 31 incidents in the past 20 years). To make the casualty data smoother with less noise, the actual number of casualties was transformed into log (casualty +1) as the response variable.

The casualty prediction for the six different scenarios were calculated using the regression tree analysis (Figure 3). The algorithms for constructing regression trees work from top downward, by choosing the significant predictor variables at each step that best splits the set of items. Derail, trnspd, vehspd, vehtyp, AV, and PP each represent the number of derailed cars, train speed, vehicle speed, vehicle type, average casualties, and proportion of observations. For instance, in Figure 4a, for TSV freight train incidents with derailment, vehicle speed is the first split. When vehicle speed is less than 1.5 mph, which is nearly stopped, the average casualty is 0.4808. When the vehicle is moving (vehicle speed greater than 1.5), train speed is a contributing variable, and when train speed is higher than 44.5 mph, the average number of casualties is larger than when train speed is less than 44.5 mph. For the VST freight-train incidents without derailment, the conditions are more complicated. The passenger train incidents shown in Figure 4e and 4f, number of cars derailed and vehicle type were significant variables. All the cross validated error rates were less than 50%, and close to 25%, which indicates the good fit of these regression tree models.

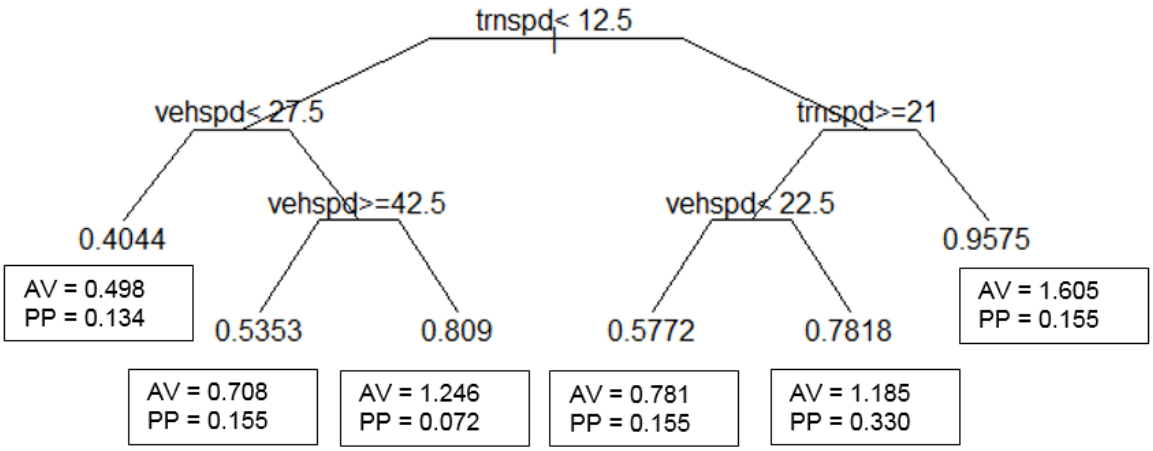


Data Points = 186

Log(casualty+1) = F (trnspd + vehspd + vehtyp)

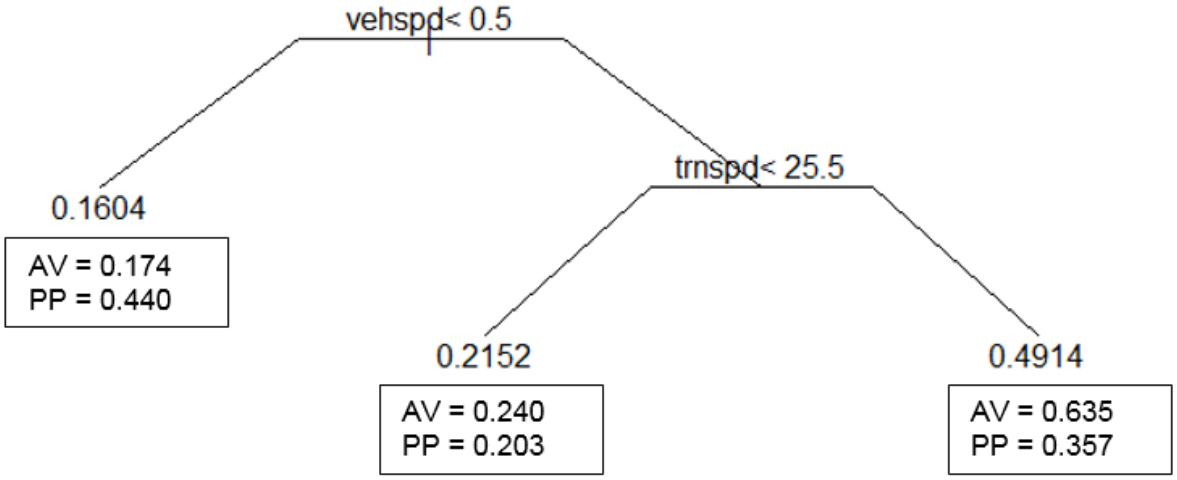
Training error rate = 0.19107(19.1%) Cross-validated error rate = 0.29777(29.8%)

(a)



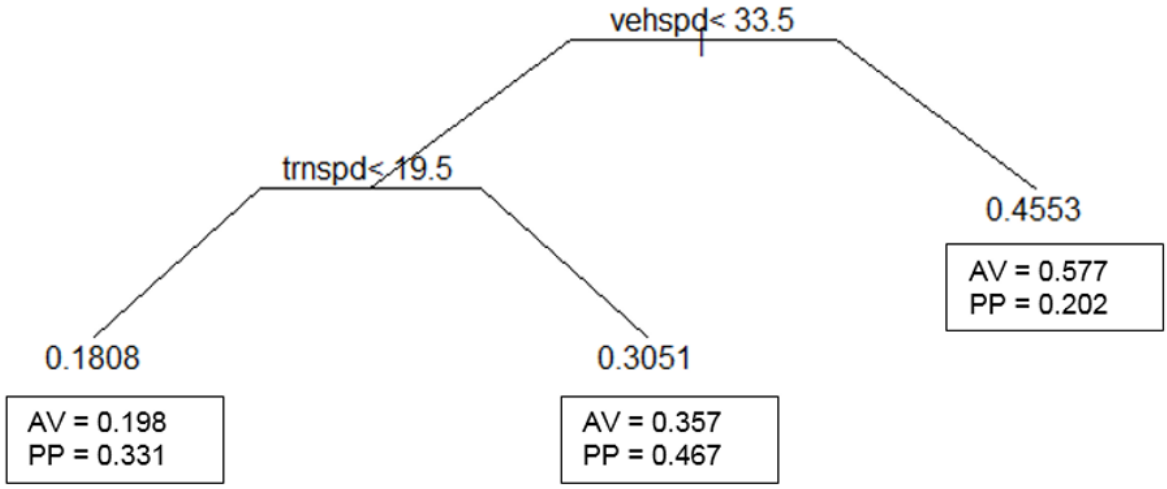
Data Points = 97
Log(casualty+1) = F (trnspd + vehspd + vehtyp)
Training error rate = 0.18290(18.3%) Cross-validated error rate = 0.21436(21.4%)

(b)



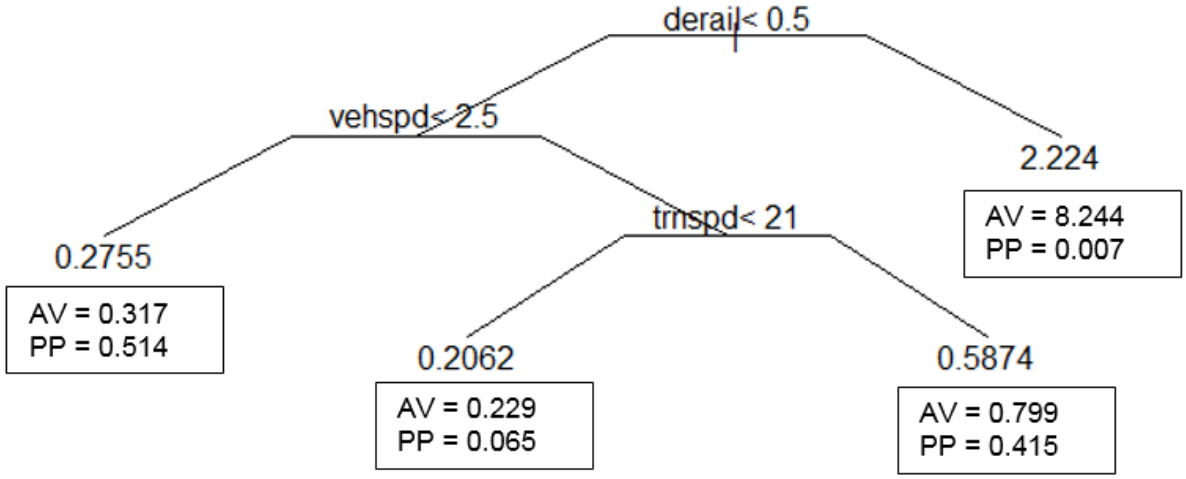
Data Points = 30,725
Log(casualty+1) = F (trnspd + vehspd + vehtyp)
Training error rate = 0.14908(14.9%) Cross-validated error rate = 0.17209(17.2%)

(c)



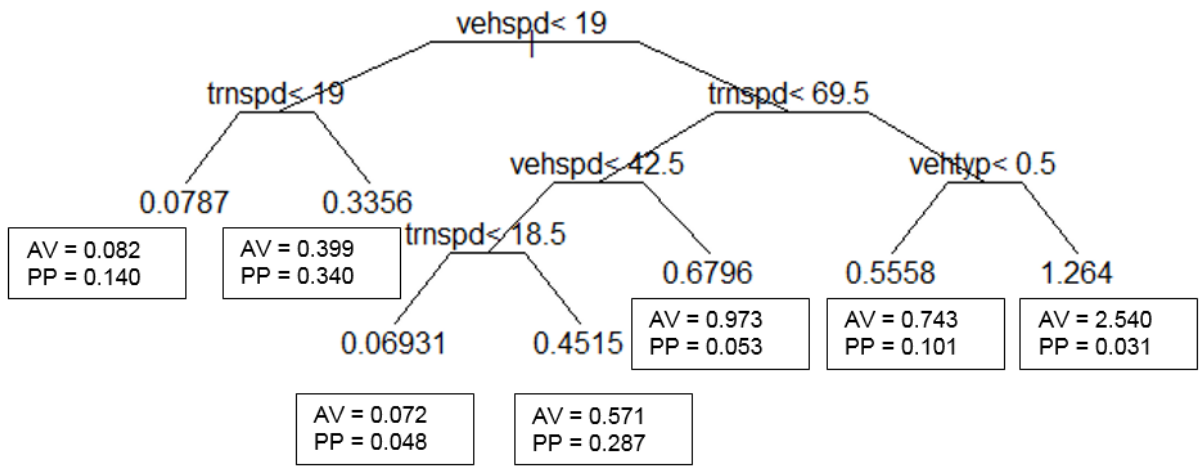
Data Points = 7,192
Log(casualty+1) = F (trnspd + vehspd + vehtyp)
Training error rate = 0.15746(15.7%) Cross-validated error rate = 0.16702(16.7%)

(d)



Data Points = 3,409
Log(casualty+1) = F (derail + trnspd + vehspd + vehtyp)
Training error rate = 0.23981(24.0%) Cross-validated error rate = 0.28799(28.8%)

(e)



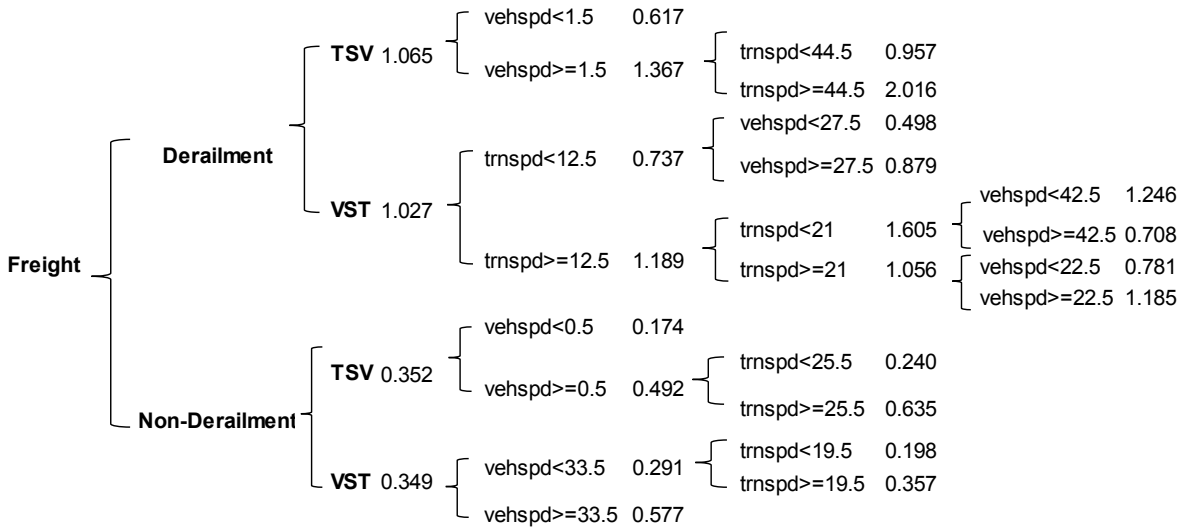
Data Points = 415
Log(casualty+1) = F (derail + trnspd + vehspd + vehtyp)
Training error rate = 0.24476(24.5%) Cross-validated error rate = 0.29954(30.0%)

(f)

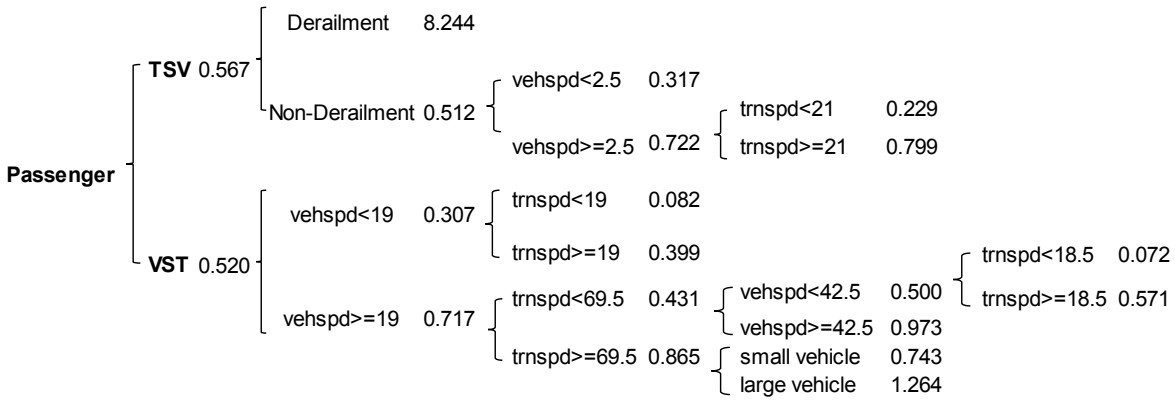
FIGURE 3. (a) Casualties Prediction for Freight Train Derailment Incidents (TSV), (b) Casualties Prediction for Freight Train Derailment Incidents (VST), (c) Casualties Prediction for Freight Train Non-Derailment Incidents (TSV), (d) Casualties Prediction for Freight Train Non-Derailment Incidents (VST), (e)Casualties Prediction for Passenger Train Incidents (TSV), and (f) Casualties Prediction for Passenger Train Incidents (VST)

Where derail = number of derailed cars; trnspd = train speed; vehspd = vehicle speed; vehtyp= vehicle type; AV = average casualties; PP = proportion of observations.

After analyzing the results from regression tree models, weighted average casualties were calculated for each condition. For each mixed condition, the product of the sum of average casualty and the proportion of each sub-condition, represents the average casualties. Figure 5a and 5b show the weighted average casualties prediction for freight and passenger train grade crossing incidents. The average number of casualties in each condition was used to define casualty severity when analyzing casualty consequence in the overall consequence model.



(a)



(b)

FIGURE 4. (a) Weighted Average Casualties Prediction for Freight Train Incidents and (b) Weighted Average Casualties Prediction for Passenger Train Incidents

TABLE 6. Poisson Regression Modeling of Casualty (Passenger Train)

Data	Regression	Dep. Var.	Indep. Var.	Coefficient	Standard Error	P value	Total P value	R Square
Passenger Train (513)	Poisson	Casualty	severity0	59.795	15.565	0.000	0.0000	0.5781
			Severity1	63.828	15.566	0.000		
			Severity2	59.069	16.114	0.000		
			Severity3	omitted	omitted	omitted		
			trnspd	0.837	0.200	0.000		
			vehspd	0.001	0.004	0.756		
			typacc	-0.598	0.167	0.000		
			trnspd_severity0	-0.822	0.200	0.000		
			trnspd_severity1	-0.828	0.200	0.000		
			trnspd_severity2	-0.772	0.210	0.000		
			trnspd_severity3	omitted	omitted	omitted		
			vehspd_severity0	0.026	0.007	0.000		
			vehspd_severity1	omitted	omitted	omitted		
			vehspd_severity2	omitted	omitted	omitted		
			vehspd_severity3	omitted	omitted	omitted		
			Constant	-61.111	15.564	0.000		

**Severity0: non-derailment, Severity1: [1,5] derailed cars, Severity2: [6-10] derailed cars, Severity3: [10,20] derailed cars*

$$Casualty_i = \exp\left(\beta_0 + \beta_1 trnspd + \beta_2 vehspd + \beta_3 acctyp + \beta_4 severity + \beta_5 trnspd\_severity + \beta_6 vehspd\_severity\right)$$

TABLE 7. Linear Regression Modeling of Casualty (Freight Train)

Data	Regression	Dep. Var.	Indep. Var.	Coefficient	Standard Error	P value	Total P value	R Square
Freight Train (38,240)	Linear	Casualty	derail	0.090	0.018	0.000	0.0000	0.081
			trnspd	0.011	0.000	0.000		
			vehspd	0.126	0.000	0.000		
			typacc	-0.068	0.011	0.000		
			trnspd_derail	-0.001	0.000	0.070		
			vehspd_derail	0.000	0.000	0.330		
			Constant	0.122	0.016	0.000		

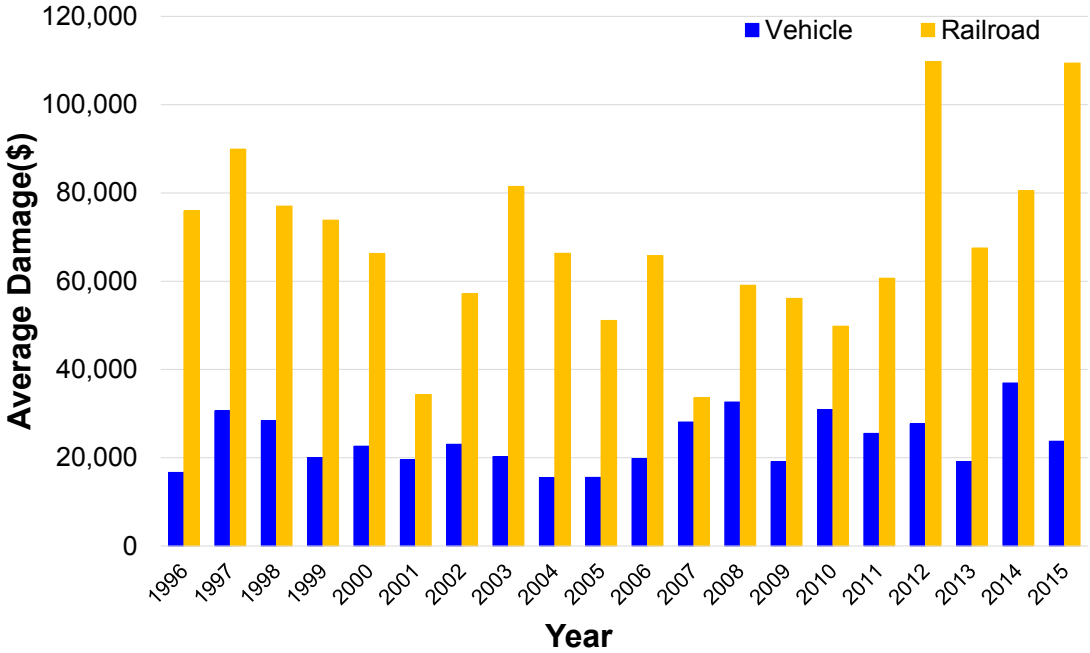
$$Casualty_i = \beta_0 + \beta_1 trnspd + \beta_2 vehspd + \beta_3 acctyp + \beta_4 derail + \beta_5 trnspd\_derail + \beta_6 vehspd\_derail$$

Property Damage Analysis

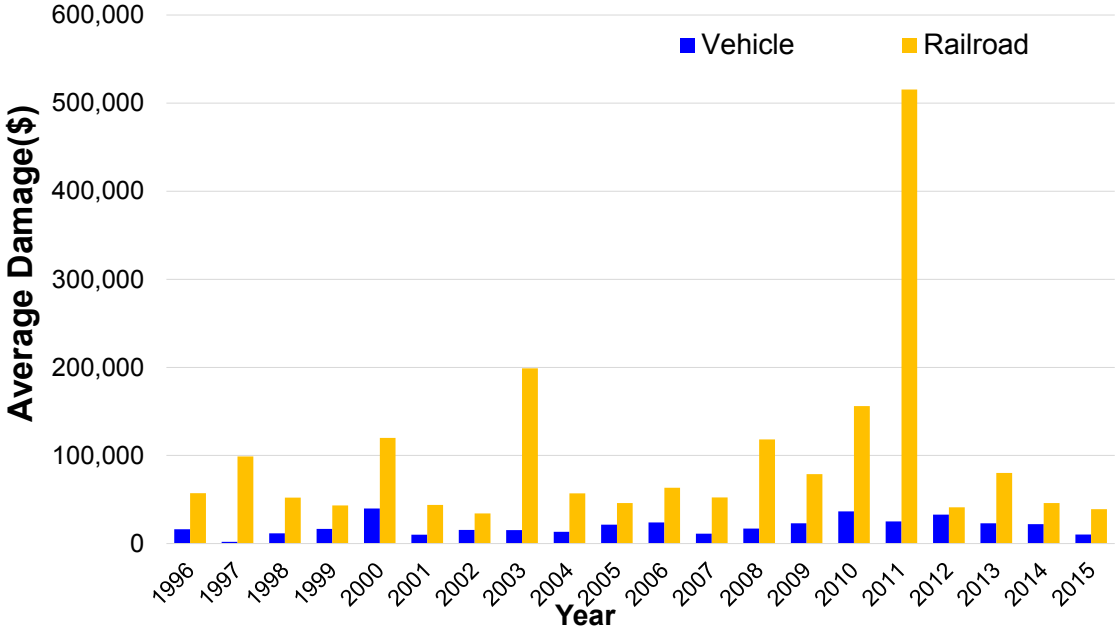
Property damage is inevitable when an incident occurs; therefore, this should also be considered when analyzing grade crossing incident consequence. Property damage for grade crossing incidents differs from damages for other types of railroad incidents, because it includes both vehicle damage and railroad damage. Instead of using mechanical tests and simulation models, actual damage values from the incident reports are analyzed. Vehicle damage data are available from the HRA database, and railroad rolling stock and infrastructure damage data are from the REA database (1).

The average railroad and vehicle property damage for grade crossing incidents were plotted for 1996 to 2015 (Figure 5). Average damage is the total damage divided by number of incidents in that year. Railroad average damages are higher than average vehicle damage, and for passenger train incidents, the ratio is more than freight train (Figure 5a and 5b). The reason is that freight and passenger trains have certain differences in their characteristics. Passenger train equipment is typically more expensive than freight, and passenger trains travel faster than freight, with the corresponding potential to cause greater damages. After statistical comparison

of the damage distributions, freight and passenger train incidents were separated for the damage analysis.



(a)

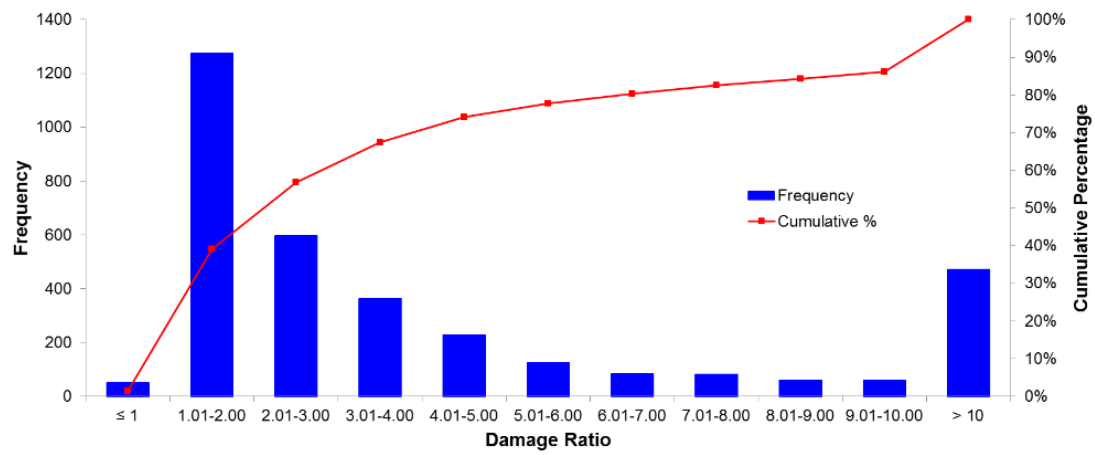


(b)

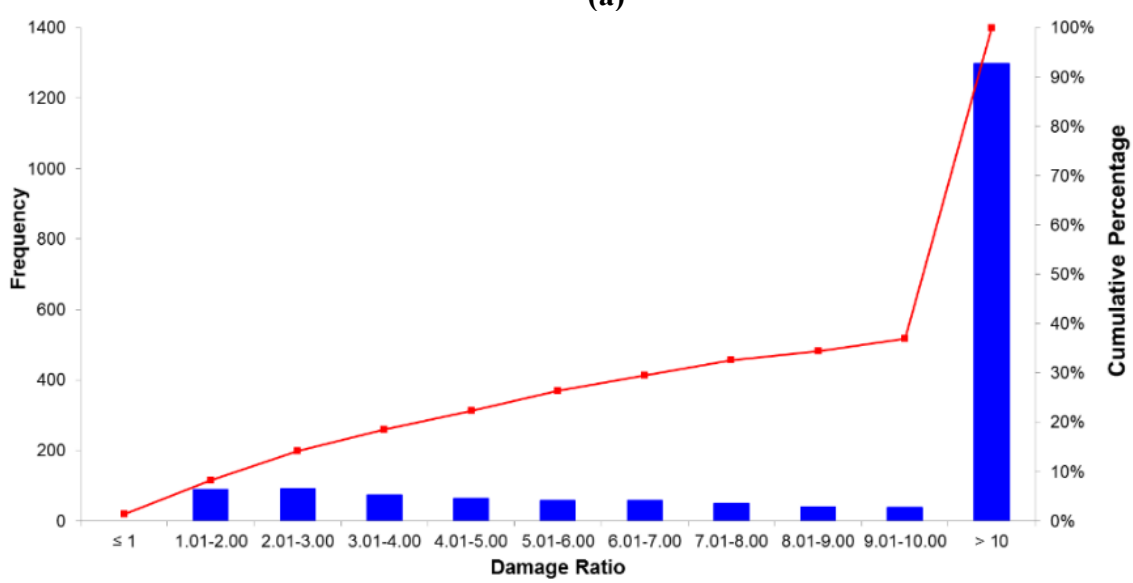
FIGURE 6. (a) Property Damage of Freight Train Incidents at Grade Crossings and (b) Property Damage of Passenger Train Incidents at Grade Crossings

Railroad Property Damage

Due to the FRA REA reporting damage threshold, railroad property damage data below the damage threshold are not recorded. Distribution fitting methods were used to estimate the missing data based on its probability density function. The distribution of railroad property damage was graphed for the most recent five-year period (2011-2015). The damage distributions were quite similar; however, the reporting threshold is periodically adjusted for inflation, so the damage ratio was used. The damage ratio is the damage divided by the threshold for each year. The damage distribution for grade crossing incidents was compared to the distribution for mainline freight train broken-rail derailments (Figure 6a and 6b). The difference is evident; grade crossing damages tend to much less costly than broken-rail derailments. The Log-normal distribution had the best fit for grade crossing incidents (Table 5).



(a)



(b)

FIGURE 6. (a) Railroad Property Damage Ratio Distribution for Grade Crossing Incidents, and (b) Railroad Property Damage Ratio Distribution for Mainline Freight Train Broken Rail Incidents

TABLE 5 Results of Railroad Property Damage Ratio Distribution Fitting

	Gamma	Weibull	Lognormal	Normal	Logistic	Exp
AIC	21311	20344	18303	34752	27980	22103
BIC	21324	20356	18315	34765	27992	22109

Variable Selection

To set up a model for prediction, the first step is to choose significant variables. For variable selection, different variables are optimized by different transformation. Derailment factor is represented by the number of derailed cars per incident for freight train incidents. Zero cars derailed indicates a non-derailment incident, while for passenger train incidents, the Freedman-Diaconis optimal bin size was used to determine derailment severity for each incident based on Inter Quantile Range:

$$h = 2\frac{IQR(x)}{n^{1/3}} \quad (19)$$

Where h = bin size

IQR = interquartile range of the data

n = the number of observations in the sample x

The two types of scenarios existing in grade crossing incidents, TSV and VST, are defined as a binary variable called accident type. Type of vehicle is also a binary variable, to represent small vehicle (automobile and pick-up truck) and large vehicle (truck, truck-trailer, van, bus and school bus). Train speed and vehicle speed, as recorded in the database were used in regression models. In addition, interaction variables between train speed, vehicle speed, and derailment were investigated to increase correlation and selectively omitted to avoid multicollinearity. To simplify the regression model, two measures were used to assess the variable selection for the goodness of fit: Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), estimated as follows:

$$AIC = -2LL + 2K$$

$$BIC = -2LL + \ln(N)K$$

Where LL = maximum likelihood estimator

K = number of parameters in the model

N = number of observations.

Based on grade crossing incident data for railroad property damage, BIC criteria picks a smaller model than AIC, and shows a higher R-squared value and lower mean squared error (MSE). To improve the prediction model accuracy and reduce the prediction error, a method commonly used in machine learning known as the shrinkage method is applied in the BIC-selected model. LASSO (least absolute shrinkage and selection operator) was used for variable selection. The objective of LASSO is:

$$\min \ \frac{1}{N}\left|\left|y - X\beta\right|\right|^2 + \alpha|\beta|$$

Where y = response

X = covariate matrix

α = shrinkage penalty

β = estimator for each variable

$$|\beta| = \sum |\beta_i|$$

N = number of observations.

After optimizing the regression models, the final results are shown in Table 6 for freight and passenger train incidents. There are 2,459 data points for freight train incidents, and 503 data points for passenger train incidents. The best model for freight train incidents was the linear regression model, while the best for passenger train incidents was the Poisson regression model. Judging from the p-value and R-squared value, railroad property damage shows a strong relationship with the variables involved in estimation.

TABLE 6. (a) Linear Regression Modeling Results of Railroad Damage Ratio (Freight Train) and (b) Poisson Regression Modeling Results of Railroad Damage (Passenger Train)

(a)

Data	Regression	Dep. Var.	Indep. Var.	Coefficient	Standard Error	P value	Total P value	R Square
Freight Train (2,459)	Linear	RRdmg_R	derail	-1.282	0.423	0.002	0.0000	0.6056
			trnspd	-0.005	0.026	0.837		
			vehspd	0.031	0.327	0.350		
			typacc	-1.013	1.322	0.443		
			trnspd_derail	0.152	0.008	0.000		
			vehspd_derail	0.031	0.007	0.000		
			Constant	5.626	1.819	0.002		

$$Casualty_i = \beta_0 + \beta_1 trnspd + \beta_2 vehspd + \beta_3 acctyp + \beta_4 derail + \beta_5 trnspd\_derail + \beta_6 vehspd\_derail$$

(b)

Data	Regression	Dep. Var.	Indep. Var.	Coefficient	Standard Error	P value	Total P value	R Square
Passenger Train (503)	Poisson	RRdmg_R	Severity0	-2.932	2.894	0.311	0.0000	0.4419
			Severity1	-0.580	2.897	0.841		
			Severity2	-5.343	3.151	0.090		
			Severity3	omitted	omitted	omitted		
			trnspd	0.011	0.038	0.764		
			vehspd	-0.032	0.002	0.000		
			typacc	1.152	0.038	0.000		
			typveh	0.742	0.016	0.000		
			trnspd_severity0	0.007	0.038	0.848		
			trnspd_severity1	0.006	0.038	0.876		
			trnspd_severity2	0.079	0.042	0.063		
			trnspd_severity3	omitted	omitted	omitted		
			vehspd_severity0	0.048	0.003	0.000		
			vehspd_severity1	omitted	omitted	omitted		
			vehspd_severity2	omitted	omitted	omitted		
			vehspd_severity3	omitted	omitted	omitted		
			Constant	0.802	2.894	0.782		

**Severity0: non-derailment, Severity1: [1,5] derailed cars, Severity2: [6-10] derailed cars, Severity3: [11,20] derailed cars*

$$RRdmg\_R_i = \exp\left(\beta_0 + \beta_1 trnspd + \beta_2 vehspd + \beta_3 acctyp + \beta_4 typveh + \beta_5 severity + \beta_6 trnspd\_severity + \beta_7 vehspd\_severity\right)$$

Vehicle Property Damage

There was no significant relationship between the vehicle property damage and variables selected statistically, so average damages for each type of vehicles were used for vehicle damage prediction (Table 7). Vehicle damages in passenger train incidents were typically higher than freight train incidents.

TABLE 7. Average Vehicle Damage ($) for Grade Crossing Incidents

	Freight Train	Passenger Train
Automobile	3,690	5,130
Truck	7,242	11,746
Truck-trailer	14,250	25,773
Pick-up truck	5,028	7,189
Van	4,430	5,940
Bus	10,893	36,000
School bus	2,663	20,000
Motorcycle	3,339	2,686
Other motor veh	9,769	15,088

Traffic Delay Analysis

Delay of train operation and highway traffic were also considered as part of g the consequences of a grade crossing incident. These incidents interrupt the traffic operation and generate additional costs due to the delay. For highway traffic delay, theoretical development is more extensive, and the Webster uniform delay model can be applied to estimate stopped highway traffic delay (20). The impact of grade crossing incidents on highway traffic is similar to stopped vehicular traffic delay, therefore, this model can be applied. Morales (21) also developed a method to estimate highway traffic delay in short-term closure condition, using figures to calculate total delay.

For railway traffic delay, there are two different scenarios: incidents with a derailment and those without derailments. For non-derailment incidents, the total train delay is typically one to two hours, which in some circumstances could affect following trains. This condition is similar to a reportable track defect, when a slow order may be implemented. The delay model can be depicted by plotting cumulative number of trains processed over time for both normal operation and disrupted operation when a grade crossing incident occurs using Lovett et al.'s method (22). If a derailment occurs due to an incident, the train delay may be 24 hours or more, which means rail traffic is completely stopped. Lovett et al. (22) developed a mathematical equation to calculate traffic delay for an initial period where the line is closed to railway traffic. This can be adapted to analyze grade crossing derailment incident impact on railway traffic operation.

Conclusion

This paper describes development of a statistical model to quantitatively estimate consequences of grade crossing incidents, with different measurements, such number of casualties, cost, and delay time. Because these consequence metrics have differing units, severity criteria need to account for this. The U.S. DOT (23) has investigated use of economic value to assess consequences for casualties and time of traffic delay, measured in dollars. The work described here represents the most systematic, comprehensive approach to developing a consistent approach to date for quantifying all types of grade crossing consequences that the authors are aware of.

More detailed models can be developed to analyze different grade crossing incident consequences using the same criterion. Additional input variables could be added to the consequence model, and each condition categorized in more detail. The results of grade crossing incident consequence studies will enable more comprehensive information to guide risk reduction investments to reduce the number of casualties at grade crossings, and can also provide insight regarding safer grade crossing design.

References

1. Federal Railroad Administration, U.S. Department of Transportation, Washington, D.C.Accident Query. FRA Office of Safety Analysis. http://safetydata.fra.dot.gov/OfficeofSafety, 2016.

2. Peck, S.M., Anya A.C., and Miriam K. Private Highway-Rail Grade Crossing Safety Research and Inquiry, Volume I. Federal Railroad Administration, Washington, D.C. 2010.

3. National Transportation Safety Board (NTSB). Railroad Accident Report NTSB/RAR-02/01. Washington D.C. 1998.

4. Chadwick, 2017. Quantitative Analysis of Train Derailments Due to Highway-Rail Grade Crossing Incidents. Doctoral Thesis, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

5. R.D. Austin, and J.L. Carson. An Alternative Accident Prediction Model for Highway-Rail Interfaces. *Accident Analysis and Prevention*, Vol 34, 2002.

6. Chadwick, S., N. Zhou, and M.R. Saat. Highway-Rail Grade Crossing Safety Challenges for Shared Operations of High-Speed Passenger and Heavy Freight Rail in the U.S. *Safety Science*, Vol 68, 2014.

7. Benekohal, R.F. and Y.M. Elzohairy. A New Formula for Prioritizing Railroad Crossings for Safety Improvement. *In: Compendium of Technical Papers for the 71th Annual Meeting of ITE in Chicago*, 2001.

8. Faghri, A., M.J. Demetsky, and Virginia Highway & Transportation Research Council. Evaluation of Methods for Predicting Rail-Highway Crossing Hazards. Virginia Highway & Transportation Research Council, Charlottesville, VA, 1986.

9. Ogden, B.D., and Korve Engineering. Railroad-Highway Grade Crossing Handbook – Revised. Office of Safety Design, Federal Highway Administration. 2007.

10. Martinez, E., D. Tyrell, and J. Zolock. Rail Car Impact Tests with Steel Coil: Car Crush. *In: Proceedings of ASME/IEEE Joint Railroad Conference*, 2003.

11. Samavdam, G. and K. Kasturi. Full-Scale Locomotive Dynamic Crash Testing and Correlations: Locomotive Consist Colliding with Steel Coil Truck at Grade Crossing (Test 3). Federal Railroad Administration Office of Railroad Policy and Development, Washington, DC, 2011.

12. Saccomanno, F.F., L. Fu, L.F. Miranda-Moreno. Risk-based Model for Identifying Highway-Rail Grade Crossing Blackspots. *In Transportation Research Record: Journal of the Transportation Research Board,* No. 1862, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 127–135.

13. Australian Transport Council. The Australian Level Crossing Assessment Model: ALCAM in Detail, 2010.

14. Federal Railroad Administration, U.S. Department of Transportation, Washington, D.C. FRA Guide for Preparing Accident/Incident Reports, 2011.

15. Kim, J.K., G.F. Ulfarsson, V.N. Shankar, and F.L. Mannering. A Note on Modeling Pedestrian-Injury Severity in Motor-Vehicle Crashes with the Mixed Logit Model. *Accident Analysis and Prevention*, Vol. 42, 2010.

16. Haleem, K. Investigating Risk Factors of Traffic Casualties at Private Highway-Railroad Grade Crossings in the United States. *Accident Analysis and Prevention,* Vol. 95, 2016.

17. Li-Yen, Chang. and H.W. Wang. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis and Prevention*, Vol 38, 2006.

18. Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning*, 2nd ed. Springer, New York, NY, USA.

19. Tippett, L.H.C. On the Extreme Individuals and the Range of Samples Taken from a Normal Population. *Biometrika*, Vol 17, 1925.

20. Roess, P.R., W.R. McShane, and E.S. Prassas. 2004, *Traffic Engineering*, Prentice Hall, Upper Saddle River, New Jersey.

21. Morales, J.M. Analytical Procedures for Estimating Freeway Traffic Congestion. *Institute of Transportation Engineers Journal*, Vol 57, 1987.

22. Lovett, A.H., C.T. Dick, and C.P.L. Barkan. Predicting the Cost and Operational Impacts of Slow Orders on Rail Lines in North America. Working Paper. University of Illinois at Urbana-Champaign, Urbana, 2017.

23. U.S. Department of Transportation. Guidance on Treatment of the Economic Value of a Statistical Life in U.S. Department of Transportation Analyses. U.S. DOT, Washington, D.C., 2016.