Quantitative prediction of the risk of heavy haul freight train derailments due to collisions at level crossings

S.G. Chadwick & C.P.L. Barkan
University of Illinois at Urbana-Champaign, Illinois, USA

M.R. Saat
Association of American Railroads, Washington, D.C., USA

ABSTRACT: The current methodology for prioritizing level crossing (LC) warning system upgrades and elimination in the United States (U.S.) focuses on the likelihood of collisions between highway and rail vehicles as well as highway user fatalities. However, these metrics do not encompass all LC risks. In particular, they do not consider the risk of derailment that LCs pose to trains, crews, and cargo (especially dangerous goods). Little previous research has considered this aspect of LC risk, although its impact is potentially severe. LCs have caused a number of train accidents in the U.S., including several that resulted in dangerous goods releases leading to injuries and fatalities. In this paper, we present a multi-factor statistical model that predicts the likelihood of a train derailment as a result of various LC parameters. The model was developed based on extensive data from the U.S. Department of Transportation's Federal Railroad Administration. It extends and formalizes previous work that identified factors leading to increased derailment likelihood for freight trains in LC collisions such as involvement of heavy highway vehicles (e.g. trucks/lorries) and higher train and motor vehicle speeds. The new model accounts for train and locomotive weight as additional factors to quantify derailment likelihood. The goal is development of a comprehensive understanding of the risk that level crossings pose to railroads and train operations.

1 INTRODUCTION

Highway-rail level crossing safety has been a topic of concern to railroads and the general public since the beginning of railroad construction. From 1991 to 2010, approximately 71,000 collisions occurred at public highway-rail level crossings in the United States (U.S.), including about 57,000 at publicly-accessible level crossings on mainline railroad tracks (FRA 2011a). Each collision has the potential to cause not only casualties to highway users, but also train passenger and crew casualties, property damage, or the release of dangerous goods. A number of serious level crossing collisions have occurred in recent years, resulting in casualties (NTSB 2015a, b, Associated Press 2015) and dangerous goods releases (NTSB 2014).

Since resources for level crossing improvements are finite, it is important to identify crossings that pose the greatest risk. Researchers and practitioners have devoted significant effort and resources to reducing risk to highway users. A variety of methods for modeling collision likelihood at level crossings have been developed, focusing on the risk trains pose to highway vehicles and their occupants, including the U.S. Department of Transportation Accident Prediction Model (FRA 1987; Ogden & Korve Engineering 2007) that is widely used in the U.S., and a variety of models developed to address limitations of that model (Benekohal & Elzohairy 2001, Austin & Carson 2002, Saccomanno et al. 2004, Oh et al. 2006, Washington & Oh 2006, Saccomanno et al. 2007). The results of these and other studies have led to improved level crossing warning systems, integration of level crossing operations with highway traffic signaling, public education programs such as Operation Lifesaver, and numerous other improvements in engineering and education (Mok & Savage 2005). These technologies and programs aim to reduce the number of casualties due to train-highway vehicle collisions, and the result has been a steady decline in the number of incidents and casualties over the past several decades.

Although the focus on level crossing safety has led to considerable improvements, one aspect has been largely overlooked – the risk that highway-rail level crossings pose to trains. Each year, 0.5 to 1% of level crossing collisions result in a train derailment. Even if a train does not derail, casualties can occur to passengers and crew aboard the train, and damage to the railroad track can result in lost service time and financial impacts. If the train does derail, there is additional potential for casualties among passengers and crew, as well as the risk of a release if the train is carrying dangerous goods. With increased interest in passenger rail transportation and the growth in transportation of hazardous materials such as crude oil, the importance of comprehensive understanding of the risk of level crossing collisions is more critical than ever.

This paper presents a statistical model that enables quantitative assessment of the relative risk of different crossings to cause a derailment. Such a model enables more informal allocation of safety resources to minimize risk due to level crossings. This model could ultimately be integrated into an overarching risk analysis framework that would consider all sources of risk at a level crossing.

2 METHODOLOGY

2.1 *Dataset*

The U.S. Department of Transportation's Federal Railroad Administration (FRA) maintains two databases that were used to build the dataset for this study: the Rail Equipment Accident/Incident (REA) database, and the Highway Rail Accident (HRA) database. Data for all U.S. mainline railroads (both freight and passenger) during the 20-year period 1991 through 2010 were used to develop the model. It was validated using data from 2011 through 2014.

The REA database collects data on any damage sustained by a train consist that exceeds a reporting threshold set by the FRA. This threshold periodically changes to account for inflation and other adjustments; as of 2011 it was set at $9,400. These data are reported to the FRA using the FRA F 6180.54 form, filed by railroads that experienced an incident meeting this criterion. It provides useful information about incidents, such as incident cause, number of cars or locomotives derailed, length of consist, type of track involved, and a number of other variables of interest.

The HRA database collects data concerning "any impact, regardless of severity, between a railroad on-track equipment consist and any user of a public or private crossing site" (FRA 2011b). All level crossing collisions are reported to the FRA regardless of the monetary value of damage caused. The data are reported using form FRA F 6180.57. The database contains a variety of information including data about the type of highway vehicle involved, speed of the train at collision, and environmental factors such as time of day and weather conditions.

2.2 *Statistical method*

The statistical model presented in this paper was developed using the LOGISTIC procedure in the Statistical Analysis Software (SAS) computer package. This procedure uses the method of maximum likelihood to fit a linear logistic regression model to binary response data (SAS Institute 2013). In this way, the relationship between explanatory variables and the outcome responses can be analyzed. For the case of level crossing incidents, for each incident record the output of the model is a value between 0 and 1 representing the probability of a derailment occurring. Logistic regression is generally discussed in terms of "events" and "non-events"; in this paper, a derailment is an event, and an incident in which no derailment occurs is a non-event.

When logistic regression is used on data that has many more non-events than events, the regression will produce a poor fit even though there are indications of strong statistical relationships in the data. The models predict non-events correctly at the expense of predicting events, since this reduces the error rate. In this way, the model predicts a large percentage of all events correctly, but has poor fit because it fails to predict most derailment events. This problem can be remedied using a modified form known as "rare events logistic regression" (RELR) (King & Zeng 2001, van den Eeckhaut et al. 2006). RELR corrects for the disproportionate number of non-events by selecting a random subset of non-events equal to 1 to 5 times the number of events. In this case, a dataset was created containing a number of randomly-selected, non-derailment events equal to twice the number of derailment events.

Dick et al. (2001) define this as a "retrospective" model, as opposed to a "prospective" model. The retrospective model makes predictions about past events using a subset of the data, consisting of some number of events and some number of non-events. The output of this retrospective model must be calibrated to more accurately represent the probability of a derailment occurring in the overall population. While the factor coefficients from the small data set are equally valid for the large data set, the intercept term needs to be adjusted in the prospective model to account for the average rate of events in the actual population (Scott and Wild 1986). This adjusted "prospective" model can then be used to make predictions about unseen data.

2.3 *Model variables*

Six variables were selected as part of the modelling process (Table 1). Vehicle speed (VS) is the speed the highway vehicle was traveling at the time of collision, while train speed (TS) is the speed of the train at collision. Highway vehicle size (LV) differentiates between large highway vehicles such as semi-trucks (lorries) and small highway vehicles such as automobiles.

The FRA databases differentiate between level crossings of different "incident type" (IT). There are two defined incident types: incidents where the train strikes the vehicle (TSV) and incidents where the vehicle strikes the train (VST). Due to factors including the interaction between the train's wheels and the rail, the effects of the other model factors differ significantly by incident type. This paper further distinguishes between incidents where the train strikes a stopped vehicle (TSV-S) and incidents where the train strikes a moving vehicle (TSV-M). About 43% of TSV incidents involve vehicles that are stopped on the crossing. It seemed plausible that this might mask the true effect of highway vehicle speed when conducting statistical regression. The same problem did not exist for VST incidents, since very few (less than five out of the whole database) trains involved in VST incidents are stopped at the time of collision.

Equipment class (EC) describes the type of rail vehicle that was struck by the highway vehicle. There are four types: freight car (FC), freight locomotive (FL), passenger car (PC), and passenger locomotive (PL).

Train length (TL) is the length of the train in number of rail vehicles. Rail vehicles include both locomotives and railcars.

Table 1. Definition of model variables

Variable Name	Definition	Variable Type	Range
VS	Highway vehicle Speed (mph)	Continuous	0-79
TS	Train speed (mph)	Continuous	0-106
LV	Was a large highway vehicle involved?	Binary	N if no; Y if yes
IT	Incident type	Categorical	VST, TSV-S, TSV-M
EC	Rail equipment class	Categorical	FC, FL, PC, PL
TL	Train length (rail vehicles)	Continuous	1-161

3 RESULTS

3.1 *VST incidents*

Of the derailment incidents reported in the REA database, 97 involved incidents in which the highway vehicle struck the train. To use RELR, 194 non-derailment incidents were randomly selected from the portion of the HRA database involving VST incidents. Combining 97 derailment and 194 non-derailment incidents creates a model dataset with a ratio of 1:2 events to non-events.

Initially, selection within the set of VST non-derailment incidents was done completely randomly. However, this resulted in selections that did not represent the true ratio of different rail vehicle types in the population because incidents involving passenger rail vehicles are so rare. Of the VST records, 30% involved a freight car, 64% involved a freight locomotive, 1% involved a passenger car, and 5% involved a passenger locomotive. Thus, 59 freight car incidents, 124 freight locomotive incidents, 2 passenger car incidents, and 9 passenger locomotive incidents were randomly selected to compile the model dataset of 194 non-derailment incidents. Repeating this process generated four different model datasets. Regression on each of them developed four models that performed similarly well and selected the same factors for the model, but one had the best fit statistics. This "best model" is:

$$p = \frac{1}{e^{-x}+1} \tag{1}$$

$$x_{VST} = -2.0204 + 0.0607\,VS + \begin{cases} 0, & LV = Y \\ -1.5459, & LV = N \end{cases} + \begin{cases} 1.8213, & EC = PC \\ 0.0648, & EC = FC \\ 0, & EC = PL \\ -1.3087, & EC = FL \end{cases} \tag{2}$$

This model provided the best fit to the data, with a Hosmer-Lemeshow (HL) goodness-of-fit test result of 0.7222. Values closer to 1 indicate good model fit, and values closer to 0 indicate poor fit. This model also has the ability to discriminate between derailment and non-derailment events, as measured by the area under the ROC curve. Generally, a model is considered to provide good discrimination if the ROC value is greater than 0.8. The area under the ROC curve for this result is 0.9011.

Additional performance statistics for this model are given in Table 2. For these values, the threshold value for predicting a derailment was a p value of 0.3. If the calculated value of p for a data point was greater than 0.3, it was classified as predicting a derailment, and if it was less than 0.3, it was classified as predicting no derailment.

Table 2. Performance statistics for retrospective model

Statistic	Value
Percent correct	81.8
Sensitivity	86.6
Specificity	79.5

In this model, the intercept term (b = -2.0204) is based on the average probability of a derailment for the RELR model dataset. This term needs to be adjusted in the prospective model to account for the average rate of derailment in the actual population of all level crossing collisions by altering the intercept term to account for the 20-year average likelihood of a VST derailment occurring. For the total VST population, the average derailment likelihood, $p_{all\ VST}$ can be calculated as:

$$p_{all\ VST} = \frac{97\ derailments}{7{,}040\ total\ events} = 0.0138 \tag{3}$$

The intercept term is modified to account for $p_{all\ VST}$ using the log-odds operator.

$$b_{VST} = b + \ln\left(\frac{p_{all\,VST}}{1-p_{all\,VST}}\right) \tag{4}$$

$$b_{VST} = -6.2912$$

Using the modified intercept term adjusts the probabilities predicted by the model to reflect the actual observed rate of derailments. Therefore, for all VST incidents, the final model is:

$$x_{all\,VST} = -6.2912 + 0.0607\,VS + \begin{cases} 0, & LV = Y \\ -1.5459, & LV = N \end{cases} + \begin{cases} 1.8213, & EC = PC \\ 0.0648, & EC = FC \\ 0, & EC = PL \\ -1.3087, & EC = FL \end{cases} \tag{5}$$

An ROC curve was generated by analysing the total population dataset with Equation 5 (Figure 1). The area under the ROC curve was equal to 0.9056. Additionally, model performance was quantified using the Brier score. This model had a Brier score of 0.0809; Brier scores closer to zero indicate better fit.

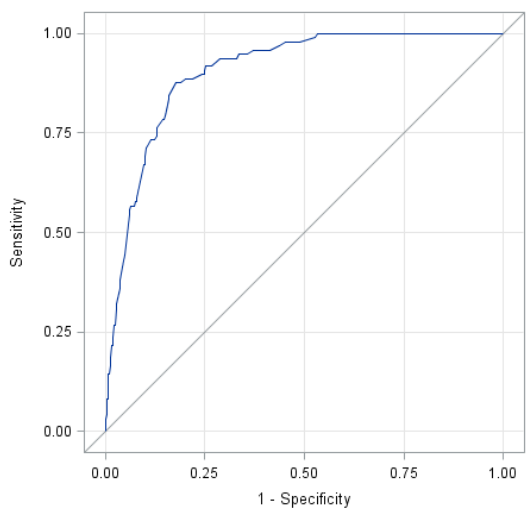

Figure 1. ROC Curve for dataset VST. Area under the ROC curve is equal to 0.9056.

In addition to these traditional techniques, the model was tested to see how it performed at ranking incidents by derailment likelihood, and whether this ranking corresponded to whether a derailment actually occurred. This technique has the advantage of being independent of the selected threshold value. To do this, all VST incidents in the HRA database were ranked by their $p_{all\ VST}$ value as calculated by the model, from least likely to most likely to derail. The dataset was divided into quintiles and the number of derailments in each quintile were counted (Table 3).

Table 3. Performance of VST model based on ranking

Quintile	Assigned Rank	Actual Derailments	Percent of Derailments
1	0 – 1,408	0	0
2	1,409 – 2,816	1	1.03
3	2,817 – 4,334	7	7.22
4	4,225 – 5,632	12	12.37
5	5,633 – 7,040	77	79.38

* Incidents in quintile 1 are least likely to derail, while incidents in quintile 5 are most likely to derail

Since approximately 80% of actual derailment incidents were ranked in the 5th quintile, the model does a good job of identifying derailment incidents. If, for example, level crossing decision makers ranked all crossings by derailment likelihood and chose to focus their efforts on the top 20%, they would likely capture 80% of all derailments.

3.2 *TSV-S incidents*

Of the derailment incidents reported in the REA database, 60 involved incidents where the train struck a stationary (VS = 0) highway vehicle. To use RELR, 120 non-derailment incidents were randomly selected from the portion of the HRA database involving TSV-S incidents. Combining 60 derailment and 120 non-derailment incidents results in a model dataset with a ratio of 1:2 events to non-events.

Selection within the set of TSV-S non-derailment incidents was random. The ratio of incidents involving freight and passenger rail vehicles was the same in the randomly selected development dataset as in the overall population. Approximately 11% of TSV-S incidents involved passenger trains. Unlike the VST case, it is not critical (and not possible) to differentiate between locomotives and railcars, because in TSV incidents less than a tenth of a percent (0.07%) involved a railcar. This is to be expected given that the vast majority of freight trains have a locomotive in the lead position. The dataset generation process was repeated to yield four different model datasets. Then a regression was run on each of them to develop four models. The four models performed similarly well and all selected the same factors, but one had the best fit statistics. This "best model" is:

$$x_{TSV-S} = -5.2729 + 0.0893\ \text{TS} + 0.0362\ \text{TL} - 0.00075\ \text{TS} \times \text{TL} + \begin{cases} 0, & LV = Y \\ -1.5459, & LV = N \end{cases} \tag{6}$$

This model had an HL test result of 0.8535 and an area under the ROC curve of 0.8688. Additional performance statistics for this model are given in Table 4 for a p value of 0.3.

Table 4. Performance statistics for retrospective model

Statistic	Value
Percent correct	78.2
Sensitivity	88.1
Specificity	73.3

The intercept term (b = *-5.2729*) needs to be adjusted in the prospective model to account for the average rate of derailment in the population of all grade crossing collisions. For the overall TSV-S population, the average derailment likelihood, $p_{all\ TSV\text{-}S}$ can be calculated as:

$$p_{all\ VST-S} = \frac{60\ derailments}{11{,}248\ total\ events} = 0.0053 \tag{7}$$

The intercept term is modified to account for $p_{all\ TSV\text{-}S}$ using the log-odds operator.

$$b_{TSV-S} = b + \ln\left(\frac{p_{all\ TSV-S}}{1-p_{all\ \mathrm{TSV-S}}}\right) \tag{8}$$

$$b_{TSV-S} = -10.5065$$

For all TSV-S incidents, the model is:

$$x_{all\ TSV-S} = -10.5065 + 0.0893\ \mathrm{TS} + 0.0362\ \mathrm{TL} \\ -0.00075\ \mathrm{TS} \times \mathrm{TL} + \begin{cases} 0, & LV = Y \\ -1.5459, & LV = N \end{cases} \tag{9}$$

The area under the ROC curve for the total population dataset was 0.8790, which is considered good discrimination (Figure 2). The model has a Brier score of 0.0892 indicating good fit.

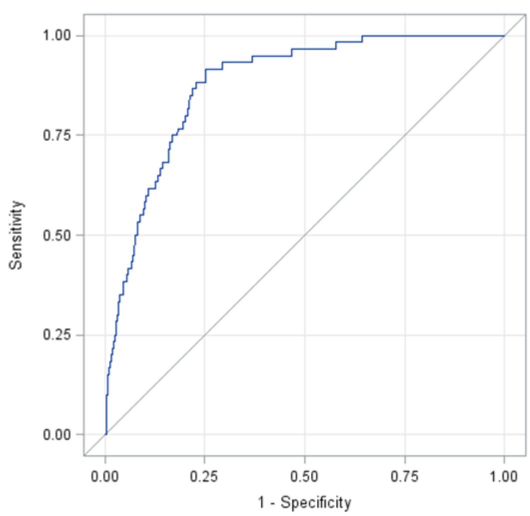

Figure 2. ROC Curve for dataset TSV-S. Area under the ROC curve is equal to 0.8790.

All TSV-S incidents were ranked by their $p_{all\ TSV\text{-}S}$ value as calculated by the model, from least likely to most likely to derail. The dataset was divided into quintiles to determine how many actual derailments occurred in each quintile (Table 5).

Table 5. Performance of TSV-S model based on ranking

Quintile	Assigned Rank	Actual Derailments	Percent of Derailments
1	0 – 2,261	0	0
2	2,262 – 4,522	1	1.67
3	4,523 – 6,783	2	3.33
4	6,784 – 9,044	10	16.67
5	9,045 – 11,305	47	78.33

* Incidents in quintile 1 are least likely to derail, while incidents in quintile 5 are most likely to derail

Since approximately 80% of actual derailment incidents were ranked in the 5th quintile, the model does a good job of identifying derailment incidents.

3.3 *TSV-M incidents*

Of the derailment incidents reported in the REA database, 114 involved incidents where the train struck a moving (VS > 0) highway vehicle. To use RELR, 228 non-derailment incidents were selected from the portion of the HRA database involving TSV-M incidents. Combining 114 derailment and 228 non-derailment incidents gave a model dataset with a ratio of 1:2 events to non-events.

Selection within the set of TSV-M non-derailment incidents was random. The ratio of incidents involving freight and passenger rail vehicles was the same in the randomly selected development dataset as the overall population. Approximately 11% of TSV-M incidents involved passenger trains. As with TSV-S incidents, it is not critical to differentiate between locomotives and railcars. The dataset generation process was repeated to generate four different model datasets, then a regression was run on each of them to develop four models. The four models performed similarly well and selected the same factors for the model, but one had the best fit statistics. This "best model" is:

$$x_{TSV-M} = -3.2144 + 0.0243\ \mathrm{VS} + 0.0233\ \mathrm{TS} \\ + \begin{cases} 0, & LV = Y \\ -2.2628, & LV = N \end{cases} \tag{10}$$

This model provided the best fit to the data, with an HL goodness-of-fit test result of 0.9152 and an area under the ROC curve of 0.8438. Additional performance statistics for this model are given in Table 6 for a p value of 0.3.

Table 6. Performance statistics for retrospective model

Statistic	Value
Percent correct	74.1
Sensitivity	98.2
Specificity	61.5

The intercept term (b = *-3.2144*) needs to be adjusted in the prospective model to account for the average rate of derailment in the actual population of all grade crossing collisions. For the total TSV-M population, the average derailment likelihood, $p_{all\ TSV\text{-}M}$ can be calculated as:

$$p_{all\ TSV-M} = \frac{114\ derailments}{15{,}027\ total\ events} = 0.0076 \qquad (11)$$

The intercept term is modified to account for $p_{all\ TSV\text{-}M}$ using the log-odds operator.

$$b_{TSV-M} = b + \ln\left(\frac{p_{all\,TSV-M}}{1-p_{all\,TSV-M}}\right) \qquad (12)$$

$$b_{TSV-M} = -8.0882$$

For all TSV-M incidents, the model is:

$$x_{all\ TSV-M} = -8.0882 + 0.0243\ \mathrm{VS} + 0.0233\ \mathrm{TS} + \begin{cases} 0, & LV = Y \\ -2.2628, & LV = N \end{cases} \qquad (13)$$

The area under the ROC curve created by evaluating all records with $p_{all\ TSV\text{-}M}$ was 0.8625, considered good discrimination (Figure 3). The model has a Brier score of 0.1038 indicating good fit.

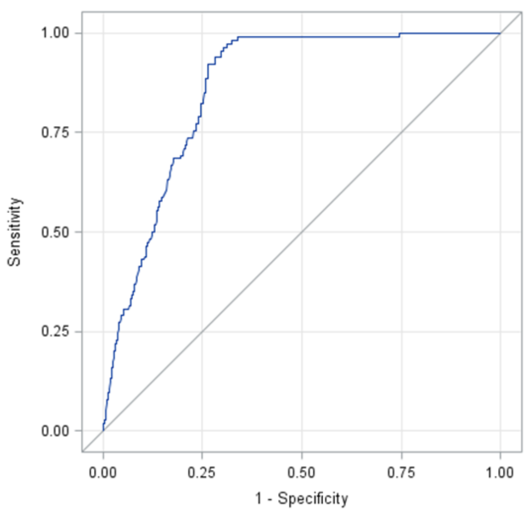

Figure 3. ROC Curve for TSV-M. Area under the ROC curve is equal to 0.8625.

TSV-M incidents in the model development dataset were ranked by their $p_{all\ TSV\text{-}M}$ value calculated by the model, from least likely to most likely to derail. The dataset was divided into quintiles and the actual derailments occurring in each quintile were counted (Table 6).

Table 7. Performance of TSV-M model based on ranking

Quintile	Assigned Rank	Actual Derailments	Percent of Derailments
1	0 – 3,006	0	0
2	3,007 – 6,012	1	0.88
3	6,013 – 9,018	0	0
4	9,019 – 12,024	0	0
5	12,025 – 15,027	113	99.12

* Incidents in quintile 1 are least likely to derail, while incidents in quintile 5 are most likely to derail

Since nearly 100% of actual derailment incidents were ranked in the 5th quintile, the model can be said to do a good job of identifying derailment incidents.

3.4 *Model validation*

To verify that the models developed based on data from 1991 to 2010 were valid for incidents outside the study period, data between 2011 and 2015 were tested to see where derailment incidents would be ranked. These results show that the model performs as well for more recent incidents as it did for incidents in the development dataset (Table 8).

4 DISCUSSION

4.1 *Interpretation of model terms*

Considered together, the models presented above contain five terms that indicate the effects of different vehicle and accident characteristics. A sixth characteristic, incident type, is accounted for by developing the three separate models. The SAS LOGISTIC procedure, using stepwise selection, chose three independent variables for the VST model, three independent variables and an interaction term for the TSV-S

Table 8: Performance metrics for validation dataset (2011-2015)

	VST (n = 1,150)		TSV-S (n = 2,578)		TSV-M (n = 2,553)	
AUC	0.9014		0.8935		0.8762	
Brier	0.0521		0.0855		0.0825	
Quintile	Actual Derailments	Percent Derailments	Actual Derailments	Percent Derailments	Actual Derailments	Percent Derailments
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	1	4.34	0	0	1	3.85
4	5	21.74	3	21.43	6	23.08
5	17	73.92	11	78.57	19	73.07

model, and three independent variables for the TSV-M model.

4.1.1 *VST incidents*

The first term in the VST model (Equation 5), $0.0607\, VS$, indicates that the speed of the vehicle at collision affects derailment likelihood. As vehicle speed increases, the probability of derailment also increases.

The second term in the model, $\begin{cases} 0, & LV = Y \\ -1.5458, & LV = N \end{cases}$, indicates that the type of highway vehicle involved in the collision affects derailment likelihood. If the highway user is a small vehicle (*LGVEH* = *N*) then this term assumes a value of -1.5458; if the highway user is a large vehicle (*LGVEH* = *Y*) then the term disappears. This means that, all else equal, a collision where the highway user is a large vehicle is more likely to result in a derailment.

The third and final term in the model, $\begin{cases} 1.8213, & EC = PC \\ 0.0648, & EC = FC \\ 0, & EC = PL \\ -1.3087, & EC = FL \end{cases}$, shows the effect of equipment class on derailment likelihood. As the coefficients become more positive, derailment likelihood will increase. This means that incidents involving passenger cars are more likely to result in derailment than those involving freight rail, which in turn are more likely to result in derailment than those involving passenger locomotives, which in turn are more likely to result in derailment than those involving freight locomotives. This trend is what one would expect if lighter rail equipment is more likely to derail than heavier rail equipment.

It should be noted that the confidence intervals of the estimates for freight railcars and passenger locomotives overlap, meaning it is statistically uncertain if there is a difference between these two equipment classes. This overlap was observed in all four of the candidate models, suggesting it is not an artefact of the dataset. The overlap is probably explained by the fact that freight cars vary widely in weight. A loaded freight car can weigh five times as much as an empty one. When unloaded, the average freight car is lighter than the average passenger locomotive, but the opposite is true for loaded freight cars. Unfortunately, the HRA database does not track whether the railcar was loaded or empty. Therefore, it was not possible to distinguish between loaded and empty freight cars.

4.1.2 *TSV-S incidents*

The first term in the TSV-S model (Equation 9), $0.0893\, TS$, indicates that the speed of the train at collision affects derailment likelihood. As train speed increases, the probability of derailment also increases.

The second term in the model, $0.0362\, TL$, indicates that there is a relationship between train length and derailment likelihood. As the length of the train increases, so does derailment likelihood.

The third term in the model, $-0.00075\, TS \times TL$, is an interaction effect between train speed and train length. This indicates that at higher train speeds, the likelihood of derailment will decrease with longer train length; or, alternatively, for longer trains, the likelihood of derailment will decrease with higher speeds.

The final term in the model, $\begin{cases} 0, & LV = Y \\ -1.5733, & LV = N \end{cases}$, indicates that the type of highway vehicle involved in the collision affects derailment likelihood. If the highway user is a small vehicle (LGVEH = N) then this term assumes a value of -1.5733; if the highway user is a large vehicle (LGVEH = Y) then the term disappears. *Ceteris paribus*, a collision where the highway user is a large vehicle is more likely to result in a derailment.

4.1.3 *TSV-M incidents*

The first term in the TSV-M model (Equation 13), $0.0243\, VS$, indicates that the speed of the vehicle at collision affects derailment likelihood. As vehicle speed increases, the probability of derailment also increases.

The second term in the model, 0.0233 TS, indicates that the speed of the train at collision affects derailment likelihood. As train speed increases, the probability of derailment also increases.

The final term in the model, $\begin{cases} 0, & LV = Y \\ -2.2628, & LV = N \end{cases}$, indicates that the type of highway vehicle involved in the collision affects derailment likelihood. If the highway user is a small vehicle (LGVEH = N) then this term assumes a value of -2.2628; if the highway user is a large vehicle (LGVEH = Y) then the term disappears. This means that, *ceteris paribus*, a collision where the highway user is a large vehicle is more likely to result in a derailment.

4.2 *Model limitations*

As with any model, these findings are limited by the quantity and quality of data available. Derailments due to grade crossing collisions are uncommon events. Development of a reasonably-sized dataset of accidents required use of 20 years of data during which there were 399 verified derailments due to grade crossing incidents.

Overall, the quality of the data are good. There are some errors and inconsistencies between the REA and HRA databases, but in general it is a simple matter to identify and correct these errors. There are sufficient data that incomplete or internally-inconsistent records can be dropped if they cannot be corrected.

5 CONCLUSIONS

This paper explored the development of a set of models to predict derailment rates for both trains at highway-rail level crossings using logistic regression analysis. Three regression models were ultimately developed based on incident type – one each for incidents where the vehicle strikes the train, incidents where a train strikes a stopped vehicle, and incidents where a train strikes a moving vehicle. Results show that, other than incident type, five factors are important to derailment prediction: highway vehicle type, highway vehicle speed, train length, rail equipment type, and train speed. The key factors varied for each of the three regression models in ways that are consistent with expectations given the physical forces for each incident type. This model could ultimately be integrated into an overarching risk analysis framework that would consider all sources of risk at a level crossing.

6 ACKNOWLEDGMENTS

This research has been supported by ABS Consulting, the Association of American Railroads, and the National University Rail (NURail) Center, A U.S. DOT Tier-1 OST-R University Transportation Center. The first author has additionally been supported by the Eisenhower Transportation Fellowship Program, and a CN Graduate Research Fellowship in Railroad Engineering. The authors would like to thank Steve Laffey at the Illinois Commerce Commision for insight and feedback.

7 REFERENCES

Associated Press 2015. Passenger Train Collides With Truck in North Carolina. *The New York Times*. <http://www.nytimes.com/2015/03/10/us/passenger-train-collides-with-truck-in-north-carolina.html?_r=1> Accessed March 2016.

Austin, R.D. & Carson, J.L. 2002. An alternative accident prediction model for highway-rail interfaces. Accident Analysis & Prevention 34, 31-42.

Benekohal, R.F. & Elzohairy, Y.M. 2001. A New Formula for Prioritizing Railroad Crossings for Safety Improvement. In: Proceedings of the Institution of Transportation Engineers Annual Meeting, Chicago, IL.

Dick, C.T. 2001. *Factors Affecting the Frequency and Location of Broken Railway Rails and Broken Rail Derailments.* M.S. Thesis, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana.

Federal Railroad Administration (FRA) 1987. Rail-Highway Crossing Resource Allocation Procedure, 3rd ed. U.S. Department of Transportation, Washington, DC.

Federal Railroad Administration (FRA) 2011a. Download Accident Data On Demand. http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx. Accessed March 2011.

Federal Railroad Administration (FRA) 2011b. *FRA Guide for Preparing Accident/Incident Reports*, FRA, U.S. Department of Transportation.

King, G. & Zeng, L. 2001. Logistic Regression in Rare Events Data. Society for Political Methodology.

Mok, S.C. & Savage, I. 2005. Why has safety improved at rail-highway grade crossings? Risk Analysis 25 (4), 867-881.

National Transportation Safety Board (NTSB) 2014. Highway Accident Report: Highway-Railroad Grade Crossing Collision: Rosedale, Maryland. < https://app.ntsb.gov/doclib/reports/2014/HAR1402.pdf> Accessed March 2016.

National Transportation Safety Board (NTSB) 2015a. Crash Summary Factual Report: Valhalla, NY. < http://dms.ntsb.gov/public/58000-58499/58036/583617.pdf> Accessed March 2016.

National Transportation Safety Board (NTSB) 2015b. Preliminary Report: HWY15MH006 (Grade Crossing Collision in Oxnard, California). < http://www.ntsb.gov/investigations/AccidentReports/Reports/HWY15MH006_preliminary.pdf> Accessed March 2016.

Ogden, B.D. & Korve Engineering 2007. Railroad-Highway Grade Crossing Handbook – Revised 2nd Edition. Office of Safety Design, Federal Highway Administration, Washington, DC.

Oh, J., Washington, S.P. & Nam, D. 2006. Accident prediction model for railway-highway interfaces. Accident Analysis & Prevention 38 (2), 346-356.

Saccomanno, F.F., Fu, L. & Miranda-Moreno, L.F. 2004. Risk-Based Model for Identifying Highway-Rail Grade Crossing Blackspots. Transportation Research Record 1862, 127-135.

Saccomanno, F.F., Park, P.Y.-J. & Fu, L. 2007. Estimating countermeasure effects for reducing collisions at highway-railway grade crossings. Accident Analysis & Prevention 39, 406-416.

SAS Institute Inc. 2013. The LOGISTIC Procedure. *SAS Help and Documentation*.

Scott, A.J. & Wild, C.J. 1986. Fitting logistic models under case-control or choice based sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.170-182.

Van Den Eeckhaut, M., T. Vanwalleghem, et al. 2006. Prediction of Landslide Susceptibility Using Rare Events Logistic Regression: A Case-Study in the Flemish Ardennes (Belgium). *Geomorphology* 76.

Washington, S. & Oh, J. 2006. Bayesian methodology incorporating expert judgment for ranking countermeasure effectiveness under uncertainty: Example applied to at-grade railroad crossings in Korea. Accident Analysis & Prevention 38, 234-247.