National University Rail Center - NURail
US DOT OST-R Tier 1 University Transportation Center

NURail Project ID: (NURail2013-UIUC-R07)

Railroad Grade Crossing Micro-Level Safety and Risk Analysis - Phase 2 (Evaluation of Safety Risk at Highway Rail Grade Crossings)

By

Jacob Mathew
Jmathew7@illinois.edu

Rahim (Ray) F. Benekohal
rbenekoh@illinois.edu

Sai Sravya Polavarapu
polavrp2@illinois.edu

Civil and Environmental Engineering
University of Illinois Urbana Champaign

23-11-2018

Grant Number: DTRT13-G-UTC52 (Grant 2)

DISCLAIMER

Funding for this research was provided by the NURail Center, University of Illinois at Urbana - Champaign under Grant No. DTRT13-G-UTC52 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research & Technology (OST-R), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL SUMMARY

Title

Railroad Grade Crossing Micro-Level Safety and Risk Analysis - Phase 2 (Evaluation of Safety Risk at Highway Rail Grade Crossings)

Introduction

The objective of this project is to define and quantify risk at highway rail grade crossings. Even though there are models available in the literature to predict the number of accidents at a crossing location this alone is not sufficient to quantify the risk at a location. This study defines the risk at a crossing by considering the number and severity of the accidents that occurred at the crossing. The risk at a railroad grade crossing should be calculated weighing in the severity of each accident at the crossing.

This project evaluates the effectiveness of the USDOT severity prediction formulas by comparing the number of fatal/injury accidents predicted by the formula to what is observed in the field. Furthermore, this study explores if any additional information about the crossing could be used to apply corrections to the formulas to improve the prediction capability of the severity equations.

Approach and Methodology

The project defines risk at a crossing by considering the severity (likelihood of an accident being Fatal, Injury or PDO) of each accident that occurred at the crossing and by weighing them based on the accident severity.

The severity of the accidents at each crossing was calculated using the FRA formulas to evaluate the risk at a crossing. The authors explored the variables, “Maximum Timetable Train Speed” and “Crossing Surface” to apply corrections to these USDOT severity prediction equations. An exhaustive search approach was used to determine the required corrections to be applied to each of the newly identified variables to improve the predictive power of the severity prediction formulas.

This approach was carried out on the accident dataset (between the years 2011 to 2015) and inventory dataset from Illinois. The corrections identified were also validated for generality on data from other states (Texas, Iowa, South Carolina and Pennsylvania).

Findings

The major findings/results of the study are as follows

1. Identification of new variables ("Maximum Timetable Train Speed" and "Crossing Surface") which can improve the predictive power of USDOT severity equations.
2. Identification of corrective factors for each of the newly identified variables to improve the prediction power of the USDOT severity equations.
3. Development of a methodology to evaluate the risk for a group of railroad grade crossings.
4. A step by step procedure to apply corrections and evaluate the risk for a group of railroad grade crossings.

Conclusions

The study recognized that the corrections to the USDOT severity prediction equations has the potential to improve its severity prediction capacity. New variables and their suggested correction factors that improved to the USDOT severity prediction equations were identified in this study. The study also developed a step by step methodology to evaluate the risk for a group of railroad grade crossings

Recommendations

This study has shown that the USDOT severity equations has the potential to improve its predictive powers. Further research into the USDOT severity equations using new datasets (census dataset) may further improve the predictive power of the equations and should be explored.

Other recommendations include using the improved severity prediction formulas (using an excel toolkit) to evaluate the risk at highway rail grade crossings. The step by step procedure discussed in this study can be translated into an excel calculator for ease of use by a practitioner.

Publications

None

Primary Contact

Principal Investigator
Rahim (Ray) F. Benekohal
Professor
Civil and Environmental Engineering
University of Illinois Urbana Champaign
rbenekoh@illinois.edu

Other Faculty and Students Involved

Jacob Mathew
Graduate Research Assistant
Civil and Environmental Engineering
University of Illinois Urbana Champaign
jmathew7@illinois.edu

Sai Sravya Polavarapu
Graduate Research Assistant
Civil and Environmental Engineering
University of Illinois Urbana Champaign
polavrp2@illinois.edu

NURail Center
217-244-4999
nurail@illinois.edu
http://www.nurailcenter.org/

Table of Contents

Introduction

This project studies the accidents at Highway Rail grade crossings with the objective to develop a methodology to define and quantify risk at the crossings. The grade crossings are analyzed on a macroscopic scale to compute a value for the risk at a crossing. Even though there are models available in the literature to predict the number of accidents at a crossing location (USDOT accident prediction formula (1), Zero Inflated Models (2,3) etc.), the prediction of number of accidents alone is not sufficient to quantify the risk at a location. This is because the predicted values given by these models do not consider the severity of the accident that occur at a crossing. The computed value of risk should be able to relatively rank the crossings considering the severity of the accidents in case of occurrence of an accident at the crossing location.

The risk at a railroad grade crossing should be calculated weighing in the severity of each accident at the crossing. Equations to compute the severity prediction value are available in the Highway Rail Grade Crossing Handbook (1). This project evaluates the effectiveness of the severity prediction formulas by comparing the number of fatal/injury accidents predicted by the formula to what is observed in the field. Furthermore, this study explores if any additional information about the crossing could be used to apply corrections to the formulas to improve the prediction capability of the severity equations. The determined corrections are validated using a separate dataset.

This study uses the most recent grade crossing accident and inventory datasets that were downloaded from the Federal Railroad Administration's (FRA) website. This was done to ensure that the most up to date data was used in this study. Data from the state of Illinois was used to evaluate the severity prediction equations and determine necessary corrections to improve the equations. Data from four other states (Iowa, Pennsylvania, Texas and South Carolina) was used to validate the corrections. These 4 states were selected because they were spread across the continental United States. The researchers also ensured that the filters do not wipe out most of the data points from those states. (States like California and Washington were not selected for this reason).

Improved accident risk predictions at railroad grade crossing will allow for more efficient resource allocation for crossing upgrades, enhancing the investment of resources to maximize the risk reductions.

This report is for a NURail study titled: Railroad Grade Crossing Micro-Level Safety and Risk Analysis – Phase 2. Phase 1 of the study resulted in a report titled "Micro and Macro Level Safety Analysis at Railroad Grade Crossings, NURail, March 2016 (15) by Medina, Shen, and Benekohal.

Literature Review

A literature review was conducted to summarize the existing state of knowledge regarding risks of accidents at highway rail grade crossings. Studies by various researchers done to develop accident prediction and severity models at grade crossing locations are listed in this section. This review of this literature also helped in listing out traffic and site related variables that were identified as significant contributors to accidents at grade crossings.

The study by Austin et. al. (4) developed a new negative binomial model to predict the accident frequency at grade crossings. The study characterized the variables which proved to be significant in affecting highway rail grade crossing accident frequencies into three: traffic characteristics, roadway characteristics and crossing characteristics. The traffic characteristics include number of nightly (not total) through trains, aadt, number of main track lines, number of traffic lanes and maximum timetable train speed. The roadway characteristic which was identified was the crossing surface. The study also explored other roadway characteristics like development type, roadway geometry, site obstructions and these variables did not have a significant effect on the grade crossing accident frequency.

The study by Saccomanno et. al (5) presents a risk based approach to identify Highway Rail Grade Crossing Blackspots in the Ontario region in Canada. The study developed a negative binomial model to fit the collision data and introduced a weighted consequence score that represent the combined collision severity. The weights that were used in the study were obtained from insurance claims. The average costs of different collision consequences were to obtain weights for each severity level: Fatality: $2,710,000/fatality, Injuries: $65,590/injury and Average Property Damage: $61,950/train collision. The weight for a PDO accident was set as 1 and the weights for the other severity levels were scaled based on the accident costs. The authors considered the total risk as the product of accident frequency and the expected consequence.

McCollister et. al (6) used logistic regression in their study to model the probability of injuries and fatalities at highway rail grade crossing. The study identified accident history and traffic congestion as the most significant variables. Other significant variables identified in this study include number of through trains at night, number of switching trains during daytime and maximum speed on a section of a track. Another conclusion made in this study was that trucks were 60% less likely to be involved in a crossing accident than passenger automobiles.

Hu et. al (7) recognizes that accident frequency and severity must be simultaneously addressed to access the accident risk at a highway rail crossing. The study used a generalized logit model with stepwise variable selection to identify explanatory variables that were significant in predicting the severity of accidents at railroad grade crossings in Taiwan. The study identified that the number of daily trains, highway separation, number of daily trucks, obstacle detection device and approaching crossing markings significantly affected the levels of accident severity at a railroad grade crossing.

The study by Berrado et. al (8) developed a framework for risk management in the railway sector and gives an illustration of the framework on Moroccan level crossings. The framework involves several activities including hazard identification, risk analysis, treatment and control. The risk analysis at a crossing is considered in two components: the estimation of the frequency of the accidents and the respective consequence.

Hao et.al. (9) applied an ordered probit model to identify determinants of driver injury severity at highway rail grade crossings for various control measures in the United States. The study found the factors: schedule factor, weather condition, visibility, vehicle speed, vehicle type, train speed, driver’s biographic information (age and gender), area type, pavement and traffic to be significantly associated with higher injury levels. The findings of this study are consistent with the findings by previous researchers as well.

Based on these studies discussed above, the risk of a crossing in this report is defined by considering both the frequency and the severity of the accidents at a crossing. The following section gives a mathematical expression to capture this definition of risk.

Risk at a Crossing

Accident risk at a crossing could be evaluated as weighted sum of the severity. This is mathematically expressed as

$$\begin{aligned}Risk\ at\ a\ crossing\ location = {} & Relative\ severity\ of\ fatal\ accidents * Number\ of\ fatal\ accidents + \\ & Relative\ severity\ of\ injury\ accidents * Number\ of\ injury\ accidents + \\ & Relative\ severity\ of\ PDO\ accidents * Number\ of\ PDO\ accidents\end{aligned}$$

Depending on relative severity of a fatal/injury accident with respect to a PDO accident, different severity scales could be chosen to evaluate the accident risk at a crossing.

The relative severity of a fatal/injury accident with respect to a PDO accident could be determined in several ways. Different approaches are available in the literature. The National Safety Council (10) gives the accident costs based on the severity of accidents and a ratio of these severity costs can be the basis of a severity scale (Scale 1 in Table 1). Geurts et. al.(11) uses a different scale to define priority values for black spots on highways (Scale 4 in Table 1). The Ministry of Transportation and Communications in Taiwan considers 1 injury to be equivalent to 0.368 fatalities, while the British Office of Rail Regulation equalized one fatality in a railroad crash as 10 serious injuries or 200 minor injuries (12). Saccomanno et. al (5) equated one fatalities to 44 injuries to yield a crossing collision score. **Table 1** below gives 4 different severity scales that an engineer can use to determine the risk at a crossing.

Table 1: Relative weights of accident Severity

Scale	Fatal	Injury	PDO
1	367.14	21.42	1
2	200	10	1
3	44	5	1
4	5	3	1

The scales give an indication of the relative severity that the user assigns to a fatal accident as compared to an injury accident or a PDO accident. For example, if an engineer chooses to use the monetary costs of each type of severity of accident as given in the National Safety Council report, he/she can choose scale one. Similarly, scale 2 values injury twice that of scale 3 at the same time, scale 2 values a fatal accident at nearly 5 times that of scale 3. If the engineer has no preference, he can choose a "default" value given in scale 4.

DOT Models for Accident Severity

The equations for accident severity prediction are given in "Summary of the dot rail-highway crossing resource allocation procedure-revised" (13). The probability of fatal accident given an accident, P(FA|A), is expressed as in equation (1).

$$\mathrm{P}\,(\mathrm{FA} \mid \mathrm{A}) = \frac{1}{1+\mathrm{KF}*\mathrm{MS}*\mathrm{TT}*\mathrm{TS}*\mathrm{UR}} \qquad (1)$$

Where

KF = 440.9

MS = (maximum timetable train speed in mph)$^{-0.9981}$

TT = (number of thru trains per day + 1)$^{-0.0872}$

TS = (number of switch trains per day + 1)$^{0.0872}$

UR = exp(0.3571) for urban crossing and 1 for rural crossing

The probability of an casualty accident given an accident, P(CA|A), is expressed as in equation (2). Note that the difference between the number of causality accidents and the number of fatal accidents gives the number of injury accidents.

$$P\,(CA \mid A) = \frac{1}{1+KC*MS*TK*UR} \qquad (2)$$

Where

KC = 4.481

MS = (maximum timetable train speed in mph)$^{-0.343}$

TK = exp (0.1153 *total number of tracks at crossing)

UR = exp (0.2960) for urban crossing and 1 for rural crossing

Corrections to DOT Severity Prediction

This study explored if the use of variables available in the FRA database could improve the widely accepted FRA severity prediction outcome. A fatal (casualty) prediction model should be able to relatively rank the grade crossings such that the top ranked crossings are the ones with high number of fatal (casualty) accidents. The following subsection describes the data used in this study followed by the description of the methodology used.

Data Source and Data Cleaning

A dataset from the state of Illinois was selected to try and see if variables from the FRA database could be used to adjust the FRA predicted severity values to identify the top crossing with both fatal and casualty accidents.

This study uses two databases maintained by the FRA (14). They are

1. Highway—Rail Crossing Inventory Database
   This database gives data regarding the characteristics of the crossings including the traffic and train volume, crossing surface characteristics etc.
2. Highway—Rail Crossing Accident Database
   This database gives information regarding each accident including their severity at the grade crossing.

Inventory data and accident data for 5 years (2011 to 2015) was downloaded from the FRA website. Both the databases were combined based on the crossing ID and this database was used in developing the severity model. This combined database was cleaned based on the filters mentioned in **Table 2.**

Table 2: Filters Applied to Database

	Variable Name	Variable Description	Filter	Comments
1	TypeXing	Crossing Type	Keep only ‘3’	‘3’ stands for ‘Public’
2	PosXing	Crossing Position	Keep only ‘1’	‘1’ stands for ‘At Grade’
3	Aadt	AADT Count	Remove missing values and 1	

4	AadtYear	AADT Year	Remove every year less than or equal to 2000	
5	ReasonID	Reason for Update	Remove “15” and “16”	15 stands for New crossing 16 stands for closed crossing
6	Xangle	Smallest Crossing Angle	Remove missing values	
7	HwySpeed	Highway Speed Limit	Remove ‘NULL’	
8	SpselIDs	Train Detection	Remove missing values	
9	HwyPved	Is Roadway/Pathway Paved?	Remove missing values	
10	MaxTtSpd	Maximum Timetable Speed	Remove ‘NULL’ and values = 0	

Filters on the accident data were also applied. This includes,

1. Select only those accidents along the main line at the crossing (This information is available under the “TypTrk” variable under the accident database)
2. Select only motor vehicle accident and remove any train-pedestrian accidents (This information is available under “TypVeh” variable)
3. Remove all accidents which has missing track class or track class marked as “X” (This information is available under “TrkClas” variable)

Further checks were also made to ensure that none of the variables had any missing values. After applying the above filters, the database had 7151 locations and 374 accidents.

Variables Used for Severity Correction

The analysis started with evaluating 16 different crossing engineering factors like maximum timetable train speed, train volume, aadt, traffic lanes and so on. The driver characteristics are not included in this study. Using the current format of the DOT equations and the same variables used by the DOT, regression analysis using the new data from the state of Illinois was not able to discover any better models. (Please note that this doesn't mean that there are no better models, but rather the use of new data on the same variables couldn't improve the models).

At this point, rather than building a new model from scratch, it was decided to adjust the current DOT severity prediction model to improve its prediction power. The study identified two variables which had the potential to improve the severity prediction models: maximum timetable train speed and crossing surface. The following subsection gives a brief description of the two variables considered.

MaxTtSpd (Maximum Timetable Train Speed)

This variable gives the maximum timetable train speed at the crossing. The normalizations tried on this variable are listed below. The variable was discretized as shown in the **Table 3** below.

Table 3: MaxTtSpd Categories

MaxTtSpd	Factor for MaxTtSpd Category
=10	-3
11-20	-2
21-30	-1
31-40	0
41-50	1
51-60	2

61-70	3
>70	4

XSurfaceIDs (Crossing Surface)

This variable gives information on the type of surface at the crossing. The crossing surface variable was categorized into 5 as shown in **Table 4** below.

Table 4: Crossing Surface Categories

Crossing Surface	**Factor for Crossing Surface Category**
Unconsolidated	-2
Timber	-1
Asphalt	0
Concrete	1
Rubber	2

Methodology Development

Additive corrections were carried to the FRA severity prediction value using the variables mentioned in the previous subsection. The adjusted severity prediction value is calculated as

$$P'(FA|A) = P(FA|A) \pm \frac{k}{x_f} P(FA|A) \pm \frac{l}{y_f} P(FA|A) \tag{3}$$

$$P'(CA|A) = P(CA|A) \pm \frac{k}{x_i} P(CA|A) \pm \frac{l}{y_i} P(CA|A) \tag{4}$$

$$P'(PDO|A) = 1 - P'(CA|A) \tag{5}$$

Where

P'(FA|A) = adjusted fatal accident prediction value given an accident

P'(CA|A) = adjusted casualty accident prediction value given an accident

P'(PDO|A) = adjusted PDO accident prediction value given an accident

K = Factor for Maximum Timetable Train Speed category as given in Table 2

L = Factor for Crossing Surface category as given in Table 3

X_f = Factor for Maximum Timetable Train Speed correction for fatal prediction

Y_f = Factor for Crossing Surface correction for fatal prediction

X_c = Factor for Maximum Timetable Train Speed correction for casualty prediction

Y_c = Factor for Crossing Surface correction for casualty prediction

The values X_f, Y_f, X_c, and Y_c were identified using the filtered database from Illinois (described above). An exhaustive search approach was used to correct the DOT severity predictions with the objective of ranking more severe crossings higher on the list. The range of the factor for the correction for the two variables were chosen such that the probability is never negative and the effect of correction factor is not nullified or insignificant.

Ranges of X and Y

The ranges of X_f and Y_f (factors for fatal accident correction) are selected from 4 to 26 while the range of X_c and Y_c (factors for casualty accident correction) range from 4 to 42.

Each variable was tried independently to identify if it could improve the predictions over the DOT severity equations. The crossings ranked per the FRA fatal accident prediction value identify 9 fatal accident locations among the top 48 locations (There were 48 locations with at least 1 fatal accident in the Illinois dataset). None of the top 48 crossings had multiple fatal accidents in the Illinois dataset. These values were used as a base to compare the values given by the adjusted severity values. The results of the adjustments using MaxTtSpd is given in **Table 5**.

Table 5: Correction to Fatal Accident Prediction Values using MaxTtSpd

X_f	Fatal locations (Base: 9)	Multiple Accident Locations (Base: 0)	X_f	Fatal locations (Base: 9)	Multiple Accident Locations (Base: 0)
+4	8	0	-4	5	2
+6	8	0	-6	8	2
+8	8	0	-8	9	2
+10	8	0	-10	9	0
+12	8	0	-12	8	0
+14	9	0	-14	7	0

+16	8	0	-16	7	0
+18	8	0	-18	7	0
+20	8	0	-20	7	0
+22	8	3	-22	7	0
+24	8	4	-24	7	0
+26	8	3	-26	7	0

From **Table 5**, it is seen that none of the correction values could adjust the model to identify more number of fatal accident locations than what the FRA severity equation could identify. But, X_f values of -6 and -8 could identify two crossings which had multiple accidents while also identifying at least 8 fatal accident locations. The values for X_f close to -6 and -8 which didn't significantly deteriorate the model were also chosen to see if the inclusion of the second variable could improve the adjusted model. Based on this, the search space for X_f was reduced to -6 to -12.

A similar approach was taken for the Crossing Surface variable (Y_f), independent of X_f, to see if it could improve the predictions over the FRA severity equations. **Table 6** gives the results for adjustments using Crossing Surface.

Table 6: Correction to Fatal Accident Prediction Values using Crossing Surface

Yf	Fatal locations (Base: 9)	Multiple Accident Location (Base: 0)	Yf	Fatal locations (Base: 9)	Multiple Accident Location (Base: 0)
+4	12	3	-4	8	0
+6	13	2	-6	9	0

+8	11	2	-8	7	0
+10	12	2	-10	6	0
+12	11	1	-12	6	0
+14	11	0	-14	6	0
+16	11	0	-16	6	0
+18	11	0	-18	6	0
+20	10	0	-20	7	0
+22	11	1	-22	7	0
+24	10	0	-24	7	0
+26	10	0	-26	8	0

The surface category improved both the total number of predicted location with accidents as well as fatal accident locations with multiple accidents using positive corrections for Y_f in the range of 4 to 14.

A similar methodology is adopted to adjust the casualty accident prediction value. The crossings ranked per the DOT casualty accident prediction value identify 94 casualty accident locations among the top 166 locations. 13 of these top 166 crossings had multiple accidents in the Illinois dataset. These values were used as a base to compare the values given by the adjusted casualty prediction values.

Table 7 gives the results using the adjusted injury severity values.

Table 7: Correction to Casualty Accident Prediction Values using MaxTtSpd

Xc	Number of Casualty Locations (base=94)	Multiple Accident Locations (Base 13)	Xc	Number of Casualty Locations (base=94)	Multiple Accident Locations (Base 13)
4	93	16	-4	84	7
6	93	15	-6	82	6
8	93	15	-8	85	6
10	93	15	-10	86	6
12	93	15	-12	90	7
14	97	13	-14	90	9
16	96	13	-16	87	9
18	95	13	-18	86	9
20	94	13	-20	86	10
22	94	12	-22	86	10
24	94	12	-24	86	10
26	94	12	-26	84	12
28	93	12	-28	85	9
30	93	12	-30	85	11
32	94	12	-32	85	10
34	93	12	-34	87	9
36	93	11	-36	86	10
38	94	12	-38	87	11

40	94	12	-40	88	10
42	94	11	-42	87	10

From **Table 7**, adjustment of casualty accident prediction value using the variable maximum timetable train speed shows significant improvement in the number of injury accident locations identified the range of Xi: positive X_i from 14 to 42

Table 8 gives the results for adjustments using Crossing Surface

Table 8: Correction to Casualty Accident Prediction Values using Crossing Surface

Yc	Number of Casualty Locations (base=94)	Multiple Accident Locations (Base 13)	Yc	Number Casualty Locations (base=94)	Multiple Accident Locations (Base 13)
4	89	24	-4	86	5
6	88	24	-6	87	7
8	91	20	-8	86	7
10	91	20	-10	86	7
12	91	21	-12	85	8
14	91	17	-14	87	7
16	92	17	-16	83	8
18	93	16	-18	84	7
20	95	16	-20	85	8

22	93	15	-22	84	8
24	92	15	-24	85	9
26	93	16	-26	85	8
28	93	14	-28	86	9
30	94	14	-30	86	10
32	93	13	-32	87	10
34	92	13	-34	87	10
36	91	13	-36	86	10
38	93	14	-38	86	10
40	93	14	-40	90	10
42	94	12	-42	90	10

From **Table 8**, it is seen that positive corrections for surface category with factor Y_i in range of 18 to 42 improved the predicted number locations significantly.

Adjustments to the FRA severity prediction values using both the variables: maximum timetable train speed and crossing surface, are tried using the reduced search spaces as identified for each of the variables. All possible permutations of $X_{f \text{ or } i}$ and $Y_{f \text{ or } i}$ from the reduced search space were tried. **Table 9** gives the results using adjusted fatal accident prediction values using both the variables and how they compare to the unadjusted values.

Table 9: Correction to Fatal Accident Prediction Values using both variables

Correction for maximum timetable train speed factor (X_f)	Correction for crossing surface factor (Y_f)	Fatal Locations (Base: 9)	Multiple Accident Location (Base: 0)
-6	4	11	3
-6	6	10	4
-6	8	10	5
-6	10	9	4
-6	12	9	4
-6	14	7	3
-8	4	13	3
-8	6	13	3
-8	8	13	4
-8	10	12	4
-8	12	12	5
-8	14	9	3
-10	4	13	3
-10	6	13	4
-10	8	12	4
-10	10	11	3
-10	12	11	3
-10	14	10	2

-12	4	13	3
-12	6	14	4
-12	8	12	4
-12	10	10	2
-12	12	10	1
-12	14	10	1

From the 24 combinations, the factors capable to improve total predicted locations, multiple accident locations and maintain consistency are selected. Three values are selected for X_f : -8, -10, -12 and one value for Y_f was selected: +6. These are highlighted in **Table 9.**

Similar adjustments were tried on the FRA injury accident prediction value using both the variables. **Table 10** gives the results using adjusted injury accident prediction values using both the variables and how they compare to the unadjusted values.

Table 10: Correction to Casualty Accident Prediction Values using both variables

Xc	Yc	Number of Casualty Locations (Base 94)	Multiple Accident Locations (Base 13)
14	18	89	17
14	20	89	17
14	22	90	17
14	24	91	16

14	26	91	17
14	28	91	16
14	30	91	16
14	32	93	16
14	34	92	16
14	36	91	15
14	38	91	15
14	40	92	16
14	42	92	17
16	18	90	18
16	20	90	18
16	22	93	16
16	24	94	16
16	26	93	16
16	28	93	16
16	30	93	15
16	32	93	15
16	34	93	15
16	36	93	13
16	38	93	14
16	40	93	13

16	42	94	15
18	18	92	17
18	20	93	16
18	22	93	16
18	24	93	16
18	26	93	16
18	28	93	16
18	30	93	16
18	32	93	15
18	34	93	15
18	36	94	16
18	38	94	14
18	40	94	14
18	42	94	14
20	18	91	17
20	20	93	16
20	22	93	16
20	24	93	16
20	26	93	16
20	28	93	16
20	30	93	16

20	32	93	16
20	34	94	16
20	36	94	15
20	38	94	14
20	40	94	13
20	42	94	13
22	18	91	17
22	20	93	16
22	22	93	16
22	24	93	16
22	26	93	16
22	28	93	16
22	30	93	16
22	32	93	16
22	34	94	16
22	36	96	15
22	38	95	14
22	40	95	14
22	42	95	14
24	18	92	17
24	20	93	16

24	22	93	16
24	24	93	16
24	26	93	16
24	28	94	15
24	30	95	15
24	32	94	14
24	34	94	14
24	36	95	15
24	38	94	14
24	40	95	14
24	42	95	14
26	18	91	17
26	20	93	16
26	22	94	16
26	24	94	16
26	26	94	16
26	28	94	15
26	30	94	15
26	32	94	14
26	34	95	14
26	36	95	14

26	38	95	14
26	40	95	14
26	42	95	14
28	18	93	16
28	20	92	17
28	22	93	16
28	24	94	16
28	26	95	15
28	28	94	15
28	30	96	14
28	32	96	14
28	34	96	14
28	36	96	14
28	38	96	14
28	40	95	14
28	42	96	14
30	18	92	17
30	20	94	16
30	22	94	16
30	24	95	15
30	26	94	15

30	28	96	14
30	30	96	14
30	32	96	14
30	34	96	14
30	36	96	14
30	38	96	14
30	40	95	14
30	42	95	14
32	18	92	17
32	20	93	17
32	22	94	16
32	24	94	15
32	26	95	16
32	28	95	15
32	30	96	14
32	32	96	14
32	34	96	14
32	36	96	14
32	38	96	14
32	40	95	14
32	42	95	14

34	18	93	17
34	20	93	17
34	22	94	15
34	24	95	16
34	26	95	15
34	28	95	14
34	30	96	14
34	32	96	14
34	34	96	14
34	36	96	14
34	38	96	14
34	40	96	14
34	42	96	14
36	18	93	17
36	20	93	17
36	22	94	15
36	24	95	16
36	26	95	15
36	28	96	15
36	30	95	15
36	32	96	14

36	34	96	14
36	36	96	14
36	38	95	14
36	40	96	14
36	42	95	14
38	18	93	17
38	20	93	17
38	22	94	16
38	24	95	16
38	26	95	15
38	28	95	15
38	30	96	15
38	32	96	14
38	34	96	14
38	36	96	14
38	38	95	14
38	40	96	14
38	42	95	14
40	18	93	17
40	20	94	17
40	22	95	16

40	24	95	16
40	26	95	15
40	28	95	14
40	30	96	15
40	32	95	15
40	34	96	14
40	36	96	14
40	38	96	14
40	40	95	14
40	42	95	14
42	18	93	17
42	20	94	16
42	22	94	16
42	24	95	16
42	26	95	16
42	28	96	15
42	30	96	15
42	32	95	15
42	34	96	14
42	36	96	14
42	38	96	14

42	40	96	14
42	42	95	14

Table 10 shows that adjustments using both the variables can improve the total number of injury locations and the number of multiple accident locations identified. From **Table 10** different combinations of X_c and Y_c are selected. Three different X_c values (30, 32, 34) and five different Y_c values (34, 36, 38, 40, 42) were selected. This results 15 different combinations.

The adjusted value for casualty accident prediction value is used to calculate the adjusted PDO accident prediction value. This value is calculated as

$$P'(PDO|A) = 1 - P'(CA|A) \quad (6)$$

Using the 15 casualty correction values identified, the PDO prediction values are calculated. These values are evaluated against the unadjusted PDO accident prediction value. This is done to ensure that the corrections do not deteriorate the number of PDO locations identified. Now as the correction factors are obtained, they are used to evaluate the impact of corrections on PDO prediction (1-P'(CA|A)) for all the selected factors of X_c and Y_c for casualty in the above step. The PDO accidents location predicted are obtained and tabulated in **Table 11**.

Table 11: Correction to PDO Accident Prediction using both variables

Xc	Yc	Number of PDO Locations (base 102)
30	34	104

30	36	104
30	38	104
30	40	104
30	42	104
32	34	104
32	36	104
32	38	104
32	40	104
32	42	104
36	34	104
36	36	104
36	38	104
36	40	104
36	42	104

From **Table 11**, the fifteen sets of casualty corrections identified previously showed improvement in identifying PDO locations.

The researchers added additional filtering on the data set before validating the correction factors identified above. These additional filters are mentioned below.

1. Remove all crossings where total train (sum of Daylight Thru Trains, Night Time Thru Trains and Total Switching Trains) is 0.
2. Remove crossings with surface type marked as “Metal”, “Composite” or “Others” (This information is available under the variable “XSurfaceIDs”)

The researchers also chose to use the accident data for the years 2012-2016 for validation of the identified corrections. Also, crossings with missing values for the variables, including Highway Speed and Maximum Timetable Train Speed were removed.

After the addition of these filters, the database was reduced as follows.

Table 12: Crossings and Severity of Accidents in Data for Validation

State	Number of Xings	Number of Xings with Acc	Number of Accidents	Fatal	Injury	PDO
Illinois	6424	332	375	47	131	197
Pennsylvania	2380	136	154	11	47	96
South Carolina	2062	152	176	9	61	106
Texas	5500	385	490	40	148	302
Iowa	2992	135	145	17	41	87

Validation of Corrections

Using the database described above, the validation of the corrections identified were carried out. **Table 13** below shows the validation results. For the validation of the casualty corrections, the number of casualty locations were used (and not the injury locations as used in the methodology development). This is to ensure that the casualty corrections identified improves the casualty locations and not just the injury locations.

Table 13: Number of Fatal/Casualty Accidents Identified after applying corrections

Fatal Corrections		Fatal Locations Predicted
Illinois		
X_f	Y_f	Base:8
-12	6	10
-10	6	10
-8	6	10
Pennsylvania		
X_f	Y_f	Base:4
-12	6	3
-10	6	3
-8	6	2
South Carolina		
X_f	Y_f	Base:1
-12	6	1
-10	6	1
-8	6	1
Texas		
X_f	Y_f	Base:4
-12	6	5

-10	6	5
-8	6	4
Iowa		
X_f	Y_f	Base:6
-12	6	4
-10	6	3
-8	6	3

Table 13 shows the number of fatal locations predicted on the new dataset. It can be seen from the table that the recommendations made in this report also has the potential to improve the number of fatal/casualty locations in the newly filtered dataset.

The number of fatal locations identified using the corrected DOT fatal prediction values showed improvement in Illinois and in Texas using correction sets 1 and 2. The fatal correction set 3 did not show any improvement in any state except Illinois. Correction set 1 and 2 are selected for fatal corrections.

Tables 14 – 18 shows the results of applying the identified casualty corrections in each of the 5 states in the validation dataset.

Table 14: Validation Results using Casualty Correction on Illinois

Xc	Yc	Number of Casualty Locations Identified (Base = 94)	Number of PDO Locations Identified (Base = 103)
30	30	96	103
30	32	96	103
30	34	96	103
30	36	96	103
30	38	97	104

32	30	96	103
32	32	96	103
32	34	96	103
32	36	96	103
32	38	97	103
34	30	96	102
34	32	96	103
34	34	96	103
34	36	96	102
34	38	96	103

Table 15: Validation Results using Casualty Correction on Pennsylvania

Xc	Yc	Number of Casualty Locations Identified (Base = 27)	Number of PDO Locations Identified (Base = 58)
30	30	25	57
30	32	25	57
30	34	24	57
30	36	24	57
30	38	24	57
32	30	25	57
32	32	25	57
32	34	25	57

32	36	25	57
32	38	25	57
34	30	25	57
34	32	25	57
34	34	25	57
34	36	25	57
34	38	25	57

Table16: Validation Results using Casualty Correction on South Carolina

Xc	Yc	Number of Casualty Locations Identified (Base = 29)	Number of PDO Locations Identified (Base = 69)
30	30	27	66
30	32	27	66
30	34	27	66
30	36	27	66
30	38	27	66
32	30	27	66
32	32	27	66
32	34	27	66
32	36	27	66
32	38	27	66

34	30	27	66
34	32	27	66
34	34	27	66
34	36	27	66
34	38	27	66

Table 17: Validation Results using Casualty Correction on Texas

Xc	Yc	Number of Casualty Locations Identified (Base = 85)	Number of PDO Locations Identified (Base = 179)
30	30	85	172
30	32	85	174
30	34	84	175
30	36	85	175
30	38	85	175
32	30	84	174
32	32	85	175
32	34	85	176
32	36	83	175
32	38	83	176
34	30	85	175
34	32	86	175
34	34	86	175

34	36	86	175
34	38	84	175

Table 18: Validation Results using Casualty Correction on Iowa

Xc	Yc	Number of Casualty Locations Identified (Base = 28)	Number of PDO Locations Identified (Base = 57)
30	30	27	57
30	32	27	58
30	34	27	58
30	36	27	58
30	38	27	58
32	30	26	57
32	32	26	57
32	34	27	57
32	36	27	58
32	38	27	58
34	30	26	56
34	32	26	57
34	34	26	57
34	36	26	57
34	38	26	57

From **Tables 14 – 18** we can see that the application of casualty corrections enabled us to identify more number of casualty locations in the states of Illinois and Texas. The values of Xc ,Yc pairs identified for Illinois are 30,38 and 32,38. The respective values identified for Texas are 34,32 , 34,34 , 34,36. On further inspection, the number of PDO locations identified in the Illinois dataset used in validation (Table 14), is better for the pair 30,38.

From these results, it is safe to say that the corrections using Maximum Timetable Train Speed and the Crossing Surface at a grade crossing location has the potential to improve the number of locations identified in the dataset. A DOT engineer would have to identify adjustments required to make the corrections on the dataset that he/she is working on.

Recommended Correction Values

Tables 19 shows the recommended values of X_f,Y_f and X_c, Y_c. For the fatal correction, the authors recommend values -12,6 or -10,6. For casualty corrections, the authors recommend using values 30,38 or 34,34. These values are chosen from the validation dataset in Illinois and Texas.

Table19: Recommended Correction Values

	X_f	**Y_f**
1	-12	6
2	-10	6
	X_c	**Y_c**
1	30	38
2	34	34

An analyst can apply this methodology to identify a correction factor for his/her state or choose a correction set that would improve the number of fatalities and/or causalities predicted in the state and use the adjusted severity prediction value to evaluate risk at crossings.

Applying Corrections

A step by step procedure to apply the corrections identified is discussed in this section.

1. Find the number of locations within the dataset with fatal accidents. This is called N_F
2. Find number of locations that within the dataset that had an accident. This is called M
3. Sort the M crossings with accidents in descending order using
    a. $P^0(FA|A)$ calculated using equation (3).
    b. $P^1(FA|A)$ calculated based on fatal correction set 1.
    c. $P^2(FA|A)$ calculated based on fatal correction set 2.
4. In crossings sorted in step 3, find the number of fatal accident locations in the top N_F locations. These are called N_{F0}, N_{F1}, and N_{F2}
5. Choose the correction set that corresponds to the highest number among N_{F0}, N_{F1}, and $N_{F2.}$ The probability corresponding to the highest number is called P'(FA|A). This value is used to calculate the risk at a crossing in Equation 7.
6. Find the number of locations within the dataset with casualty accidents. This is called N_C
7. Sort the M crossings with accidents in descending order using
    a. $P^0(CA|A)$ calculated using equation (3).
    b. $P^1(CA|A)$ calculated based on casualty correction set 1.
    c. $P^2(CA|A)$ calculated based on casualty correction set 2.
8. In crossings sorted in step 7, find the number of fatal accident locations in the top N_C locations. These are called N_{C0}, N_{C1}, and N_{C2}
9. Choose the correction set that corresponds to the highest number among N_{C0}, N_{C1}, and $N_{C2.}$ The probability corresponding to the highest number is called P'(CA|A). This value is used to calculate the risk at a crossing in Equation (7).
10. Calculate P'(PDO|A) as 1-P'(CA|A). This is also used in Equation (7).

The **Figure 1** below shows the step by step procedure that a DOT engineer can follow to decide if he/she chooses to apply a correction or not.

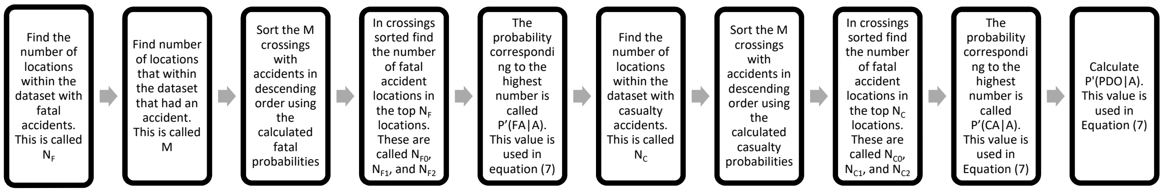

Figure 1: Stepwise procedure to select correction factor for fatal prediction

The engineer can look at the number multiple accident locations identified in the top locations in the case of a tie. A similar step can be adopted independently to decide which of the two correction sets to use (or to not apply corrections) for casualty corrections and PDO corrections.

To evaluate the risk at a crossing, a user should select a severity scale (given in **Table 1**). This selection could depend on the user's preferences and/or local conditions in the region of the crossings being considered. The risk at a crossing should be computed as

$$\begin{aligned} Risk\ at\ a\ crossing &= Relative\ severity\ of\ Fatal\ Accident * P'(FA|A) \\ &\quad + Relative\ severity\ of\ Injury\ Accident * (P'(CA|A) - P'(FA|A)) \\ &\quad + Relative\ severity\ of\ PDO\ Accident * P'(PDO|A) \end{aligned} \tag{7}$$

Conclusions and Recommendations

This report discusses a methodology to define and quantify risk at a railroad grade crossing. The predicted number of accidents alone at a crossing is insufficient to determine the risk at a crossing and hence the severity of accidents at the crossing is also considered in this study. The DOT severity equations are used as a starting point in this study and the variables crossing surface and maximum timetable train speed at a crossing are used to refine the severity prediction values.

This study used the most recent data from the state of Illinois to identify corrections for prediction of fatal and casualty accidents at a location. An exhaustive search approach come up with the corrective values to improve the prediction power of the equations. The identified corrections were tried out in other states spread across continental United States to validate the results.

The comparisons between the adjusted severity prediction values and the unadjusted severity prediction values show that the adjustments have the potential to improve the prediction power of the DOT severity equations. The adjusted DOT severity prediction equations could identify more fatal/casualty accidents among the top crossings than the unadjusted DOT severity prediction equations in certain states evaluated. Based on the analysis conducted, recommendations for correction values are also made.

This study also developed a stepwise procedure to evaluate a group of crossings. The procedure is designed to help an engineer to decide which of the recommended correction values to use.

Recommendations and Future Work

Based on the above discussion, we can see that the corrections to the FRA severity prediction equations has the potential to improve its severity prediction capacity. This improved severity prediction formulas should be used to evaluate the risk at a crossing using the expression considering the relative weights of accident severity discussed in the earlier section “Risk at a Crossing”. An engineer can follow the methodology discussed in this report to evaluate the risk at the crossing.

Using a bigger dataset with data from more states could aid in fitting a better prediction model. Furthermore, more datasets could be added giving further information about the crossing that could affect the model, i.e. census data etc.

The study identified sets of adjustments for DOT casualty prediction formula. The adjusted severity prediction equations could be incorporated into an excel calculator.

References

1. FHWA. Railroad-Highway Grade Crossing Handbook - Revised Second Edition [Internet]. 2007. Available from: http://safety.fhwa.dot.gov/xings/com_roaduser/07010/sec05.htm

2. Lee J, Park D, Nam D. Analyzing the Relationship Between Grade Crossing Elements and Accidents. J East Asia Soc Transp Stud. 2005;6:3658–68.

3. Medina JC, Benekohal RF. Macroscopic Models for Accident Prediction at Railroad Grade Crossings. Transp Res Rec J Transp Res Board [Internet]. 2015;2476:85–93. Available from: http://trrjournalonline.trb.org/doi/10.3141/2476-12

4. Austin RD, Carson JL. An alternative accident prediction model for highway-rail interfaces. Accid Anal Prev. 2002;34(1):31–42.

5. Saccomanno, Frank Fu, Liping Miranda-Moreno L. Risk-Based Model for Identifying Highway-Rail Grade Crossing Blackspots. Transp Res Rec. 2004;1862(1862):127–35.

6. McCollister GM, Pflaum CC. A model to predict the probability of highway rail crossing accidents. Proc Inst Mech Eng Part F J Rail Rapid Transit. 2007;221(3):321–9.

7. Hu SR, Li CS, Lee CK. "Investigation of key factors for accident severity at railroad grade crossings by using a logit model." Saf Sci [Internet]. 2010;48(2):186–94. Available from: http://dx.doi.org/10.1016/j.ssci.2009.07.010

8. Berrado A. A Framework for Risk Management in Railway Sector: Application to Road-Rail Level Crossings. Open Transp J [Internet]. 2011;5(1):34–44. Available from: http://benthamopen.com/ABSTRACT/TOTJ-5-34

9. Hao W, Daniel J. Severity of Injuries to Motor Vehicle Drivers at Highway-Rail Grade Crossings in the United States. Transp Res Rec J Transp Res Board [Internet]. 2013;2384:102–8. Available from: http://trrjournalonline.trb.org/doi/10.3141/2384-12

10. National Safety Council. Estimating the costs of unintentional injuries, 2011. 2013;(C):23–5. Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Estimating+the+Costs+of+Unintentional+Injuries#5%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Estimating+the+costs+of+unintentional+injuries#5

11. Geurts, Karolien, and Geert Wets. "Black spot analysis methods: Literature review." (2003).

12. Hu, Shou-Ren, Chin-Shang Li, and Chi-Kang Lee. "Assessing casualty risk of railroad-grade crossing crashes using zero-inflated poisson models." Journal of transportation engineering137.8 (2010): 527-536.

13. Farr EH. Summary of the dot rail-highway crossing resource allocation procedure-revised. 1987.

14. Office of Railroad Safety F. Federal Railroad Administration Office of Safety Analysis [Internet]. 2016. Available from: https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/downloads/downloads.aspx

15. Juan C. Medina, Shiyu Shen, and Rahim F. Benekohal “Micro and Macro Level Safety Analysis at Railroad Grade Crossings, NURail, Revised March 2016.