Nebraska Rail Crossing Safety Research

F I N A L R E P O R T

Dr. Aemal Khattak

Yashu Kang, Ph.D. Candidate

Dr. Huiyuan Liu

Department of Civil and Environmental Engineering

Nebraska Transportation Center

University of Nebraska-Lincoln

Sponsored By

Nebraska Department of Transportation and the U.S. Department of Transportation Federal Highway Administration

December 31, 2020

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
SPR-P1 (19) M091		

4. Title and Subtitle	5. Report Date
Nebraska Rail Crossing Safety Research	December 01, 2020
	6. Performing Organization Code
7. Author(s) Dr. Aemal Khattak, Yashu Kang, Dr. Huiyuan Liu	**8. Performing Organization Report No.** 26-1121-0032-001
9. Performing Organization Name and Address UNL Nebraska Transportation Center 2200 Vine St. PO Box 830851 Lincoln, NE 68583-0851	**10. Work Unit No.** **11. Contract or Grant No.** SPR-PI (19) M091
12. Sponsoring Agency Name and Address Nebraska Department of Transportation 1500 Nebraska 2 Lincoln, NE 68502	**13. Type of Report and Period Covered** Final report 07/2018-12/2020 **14. Sponsoring Agency Code**

15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. Abstract

The research objectives of this project were to update Nebraska Department of Transportation (NDOT) 1999 Nebraska Accident Prediction Model for Highway-Rail Grade Crossings (HRGCs) and to develop guidelines using Lancaster County Nebraska HRGCs for improving safety at urban gated HRGCs that are not designated quiet zones but are in the vicinity of quiet zone crossings.

FRA crash and HRGC inventory data were utilized for estimation of the new model after inventory information on 742 HRGCs was updated. HRGC crashes for 2008-2018 period were used for model estimation while 2019 HRGC crashes were used for model prediction validation. After consideration of several different model formulations, a Poisson regression model with scaled parameters was selected as the 2020 Nebraska HRGC Crash Prediction Model.

Lancaster County HRGCs consistency assessment was performed using Federal Railroad Administration's (FRA) Quiet Zone Calculator to analyze gated non-quiet zone HRGCs that are in proximity of designated quiet zone HRGCs. The general guidance on achieving a more consistent driving experience at such HRGCs is to consider the use of Supplemental Safety Measures including the use of mountable medians with reflective traffic channelization devices (vertical panels or tubular delineators) or non-traversable curb medians with or without channelization devices at non-quiet zone gated HRGCs that are in proximity of established quiet zones. A complete update of the statewide HRGC inventory is recommended to remove errors and missing values from the existing database.

17. Key Words	18. Distribution Statement
Rail crossing, prediction model, crash, safety	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	83	

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

Table of Contents

Disclaimer vi
Abstract vii
Chapter 1 Introduction 1
1.1 Background 1
1.2 Objectives 3
1.3 Research Outline 3
Chapter 2 Literature Review 5
2.1 Peabody-Dimmick Formula 6
2.2 New Hampshire Hazard Index 7
2.3 NCHRP 50 Accident Prediction Model 7
2.4 US DOT Accident Prediction Model 8
2.5 Connecticut DOT Hazard Ranking Index 8
2.6 Florida DOT Safety Hazard Index 9
2.7 Missouri DOT Exposure Index 10
2.8 North Carolina DOT Investigative Index 10
2.9 Texas DOT Priority Index 11
2.10 FRA's New Model for HRGC Accident Prediction and Severity 11
Chapter 3 Modeling Background 16
3.1 Poisson Regression Model 16
3.2 Zero-inflated Model 17
Chapter 4 Data Collection 21
4.1 FRA HRGC Inventory Database 21
4.2 FRA HRGC Crash Database 22
4.3 Field Validation of the FRA HRGC Database 23
4.4 Database for Updated Crash Prediction Model 31
4.5 Descriptive Statistics 31
Chapter 5 HRGC Crash Prediction Model Estimation 37
5.1 Analysis of Accident Prediction Models Based on Various Criteria 38
5.1.1 Over-dispersion test 38
5.1.2 Candidate model performance 39
5.1.3 The 2020 Nebraska HRGC Crash Prediction Model estimated coefficients 46
5.2 Interpreting Regression Output 47
5.3 Empirical Bayes Prediction Adjustment 49
Chapter 6 Crossing Consistency Assessment 52
6.1 Methodology for Consistency Assessment 52
6.2 HRGC Consistency Assessment 52
6.3 Recommendations for Lancaster County HRGCs 56
Chapter 7 Summary and Recommendations 58
7.1 Summary 58
7.2 Recommendations 59
Acknowledgments 61
Appendix A 62
Appendix B 68
Appendix C 73
References 80

List of Figures

Figure 3.1 Simulated Zero-inflated Poisson Distribution 19
Figure 4.1 HRGC Filtration Process for Lancaster County 24
Figure 4.2 Results of HRGC Filtration Process for Lancaster County 25
Figure 4.3 Data Correction Example, Crossing 064112B 27
Figure 4.4 An Example of an Abandoned Crossing (083524P) 28
Figure 4.5 Corrected or Added Values of Each Variable for Gage County 29
Figure 4.6 GIS Database for HRGCs in Lancaster County 30
Figure 4.7 Road inventory and land use data 30
Figure 4.8 Histogram of Highway Rail Grade Crossings by AADT (natural logarithm) 32
Figure 4.9 Histogram of Highway Rail Grade Crossings by Number of Through Trains 33
Figure 4.10 Distribution of Highway Rail Grade Crossings by Highway Classification 34
Figure 4.11 Distribution of Highway Rail Grade Crossings by Highway Function Classification 35
Figure 4.12 Distribution of Highway Rail Grade Crossings by Number of Traffic Lanes 36
Figure 5.1 Predicted Crashes by Presence of Crossbucks 49
Figure 5.2 Predicted Crashes by Presence of Crossbucks with EB Adjustment 51
Figure 6.1 Buffers of Varying Size Around Existing Quiet Zone HRGCs in Lancaster County . 53

List of Tables

Table 2.1 Usage of Different HRGC Safety Assessment Methods 6
Table 4.1 Summary of Corrections and Added Missing Values from Field Validation............... 26
Table 4.2 Summary of Corrections and Added Missing Values Using NDOT's PathWeb System 26
Table 5.1 Performance of Candidate Nebraska Crash Prediction Models................................... 42
Table 5.2 Deviance Residuals.. 46
Table 5.3 Estimated Coefficients of the 2020 Nebraska Crash Prediction Model 46
Table 6.1 Results of Buffers of Varying Size and Corresponding HRGC Crash Histories.......... 54
Table 6.2 Suggested Improvements for Consideration at Selected Lancaster County HRGCs ... 57
Table C.1 Performances of all candidate models. Trained on 2008-2016 data and validated on 2017 data (observed crash frequency was 18). 74
Table C.2 Performances of all candidate models. Trained on 2008-2017 data and validated on 2018 data (observed crash frequency was 35). 77

Disclaimer

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. The contents do not necessarily reflect the official views or policies of the Nebraska Department of Transportations nor the University of Nebraska-Lincoln. This report does not constitute a standard, specification, nor regulation. Trade or manufacturers' names, which may appear in this report, are cited only because they are considered essential to the objectives of the report.

The United States (U.S.) government and the State of Nebraska do not endorse products or manufacturers. This material is based upon work supported by the Federal Highway Administration under SPR-1(19) (M091). Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Federal Highway Administration.

NOTE: This report preferentially uses the term 'crash' to refer to a vehicular/train collision resulting in property damage and/or injuries and fatalities. However, the term 'accident' is also used when referring to legacy items (e.g., US DOT Accident Prediction Model) or when referencing or quoting published literature.

Abstract

The research objectives of this project were to update Nebraska Department of Transportation (NDOT) 1999 Nebraska Accident Prediction Model for Highway-Rail Grade Crossings (HRGCs) and to develop guidelines using Lancaster County Nebraska HRGCs for improving safety at urban gated HRGCs that are not designated quiet zones but are in the vicinity of quiet zone crossings.

FRA crash and HRGC inventory data were utilized for estimation of the new model after inventory information on 742 HRGCs was updated. HRGC crashes for 2008-2018 period were used for model estimation while 2019 HRGC crashes were used for model prediction validation. After consideration of several different model formulations, a Poisson regression model with scaled parameters was selected as the 2020 Nebraska HRGC Crash Prediction Model.

Lancaster County HRGCs consistency assessment was performed using Federal Railroad Administration's (FRA) Quiet Zone Calculator to analyze gated non-quiet zone HRGCs that are in proximity of designated quiet zone HRGCs. The general guidance on achieving a more consistent driving experience at such HRGCs is to consider the use of Supplemental Safety Measures including the use of mountable medians with reflective traffic channelization devices (vertical panels or tubular delineators) or non-traversable curb medians with or without channelization devices at non-quiet zone gated HRGCs that are in proximity of established quiet zones. A complete update of the statewide HRGC inventory is recommended to remove errors and missing values from the existing database.

Chapter 1 Introduction

1.1 Background

Highway-rail crossings are junctions between the rail and the highway network where the two meet. More than 97% of these crossings are at the same level (at-grade) in the US; such crossings are commonly referred to as highway-rail grade crossings (HRGCs). While trains have the right-of-way at HRGCs, every year there are a number of reported crashes when motor vehicles and other highway users fail to yield the right-of-way to trains. Motor-vehicle involved crashes at railroad crossings are invariably more severe compared to crashes on the rest of the surface transportation network mainly due to train involvement. In 2019, the number of crashes reported in the US at HRGCs was 2,220 resulting in 294 fatalities; fatal crashes were 13.24% of total reported incidents (Federal Railroad Administration 2020). During the same year, Nebraska accounted for 29 crashes at HRGCs involving 6 fatalities and 18 non-fatal injuries; fatal crashes were 17.24% of total reported crashes.

Rail crossing safety models based on reported crash data have provided an understanding of crash phenomenon at HRGCs, identifying associated factors in an attempt to improve safety, and for ranking competing rail crossings for safety improvement resource allocations. The Nebraska Department of Transportation (NDOT) currently utilizes the 1999 Nebraska Accident Prediction Model (HNTB, 1999) for rail crossings to identify and rank crossings that may need scrutiny and perhaps subsequent safety improvements. Developed by the Midwest Research Institute (under contract to HNTB Corp.) in 1999, this crash prediction model was based on 5-year rail crossing crashes and inventory data from September 1993 through August 1998. It updated the previously used 1973 Nebraska Department of Roads (NDOR) Hazard Index, which was a modified version of the NCHRP Report 50 Formula (NCHRP Report 50, 1968). The model over-predicts (about 10%), and results may not be optimal as many changes have occurred in terms of train and motor vehicle traffic, crash trends, and rail crossing inventory information since its adoption. Other state DOTs have recently updated their rail crossing crash prediction models or are in the process of doing so. Given the newly available statistical modeling approaches and the availability of a relatively large dataset, the hope is that the updated model will outperform the existing NDOT Nebraska Accident Prediction Model for rail crossings.

Furthermore, recent crashes reported at urban rail crossings in Nebraska call for a review of motor vehicle driver expectancy in terms of installed supplemental safety measures (e.g., 6-inch high mountable barriers along roadway centerlines to prevent passing around crossing gates). Installation of supplemental safety measures or alternative safety measures is an FRA requirement when public agencies apply for Quiet Zone designation (crossings where trains are not required to sound horns). For example, some crossings in Lincoln, Nebraska are Quiet Zones, but other proximate crossings are not designated as such. This creates a situation where drivers may expect supplemental safety measures at all crossings and their expectations violated when using crossings not designated as Quiet Zones. An example is the August 18, 2017 crash at S. Folsom St. (Lincoln, Nebraska) crossing (USDOT ID: 083044D) that claimed the lives of two high school students. The victim in this crash attempted to pass around the lowered crossing gates while an Amtrak train was on its way toward the crossing. The presence of a barrier along the roadway centerline (a supplemental safety measure) would likely have prevented this crash. Therefore, there may be merit in installing supplemental safety measures at select urban crossings that are not Quiet Zones but have crossings designated as Quiet Zones in the general vicinity.

1.2 Objectives

There were two objectives for this research: 1) to update NDOT's 1999 Nebraska Accident Prediction Model for rail crossings using the latest crash and rail crossing inventory data, and 2) to develop guidelines for improving safety (via uniformity of driver expectations) at urban rail crossings that are not designated quiet zones but are in the vicinity of existing quiet zone crossings. HRGCs located in Lancaster County, Nebraska were candidates for the second objective.

It was hoped that a newly developed crash prediction model that will outperform the 1999 Nebraska Accident Prediction Model for rail crossings thereby allowing for more informed decisions regarding resource allocation for rail crossings. Guidelines for improving safety of urban crossings that are not quiet zone crossings will enable Nebraska public agencies to improve public safety and reduce possible liability from crashes at HRGCs.

1.3 Research Outline

This research comprised of five tasks; the first was a meeting with the project Technical Advisory Committee (TAC) to discuss the research approach and review of published literature on rail crossing safety conducted with an emphasis on crash prediction models for rail crossings. Chapter 2 of this report presents a summary of the reviewed publications pertinent to this research. Chapter 3, the methodology, provides details about the statistical techniques utilized in this research. Chapter 4 presents research efforts regarding data acquisition and average annual daily traffic (AADT) data update, including a 12-year (2008-2019) crash data set and the public crossing inventory from FRA. While some AADT data were out-of-date, the research team provided updated AADT values. Chapter 5 presents estimated statistical models on the expected number of HRGC crashes per year in Nebraska. Various factors were taken into consideration with regards to their effects on crash occurrence at rail crossings, such as crossing characteristics, exposure measures, land use, etc. Chapter 6 provides an assessment of installed supplemental safety measures at urban crossings in Lancaster County that are not designated as quiet zones. Lastly, major findings from this research and conclusions are presented in Chapter 7. Guidelines on improving safety through installing supplemental safety measures at urban rail crossings are provided in Chapter 7 as well.

Chapter 2 Literature Review

The latest guidance on HRGCs including safety engineering treatments are available in the Highway-Rail Crossing Handbook 3rd Edition (Ogden and Cooper, 2020). Besides providing general information on HRGCs, this handbook also summarizes current best practices and provides options for safety enhancements at HRGCs. It provides guidance on how existing standards and recommended practices may be applied in developing safe and effective treatments for HRGCs.

The US Department of Transportation (DOT) Accident Prediction Model is a widely used hazard ranking model, currently used in 19 states for HRGC hazard ranking. Many states (e.g., Texas, Florida) have assessed the adequacy of HRGC hazard ranking models and/or developed new statistical models for hazard ranking. Other states, including Illinois and Missouri, have undertaken similar research studies but DOT staff reported the results of the studies could not be practically applied and therefore were not adopted (Sperry et al. 2017). Recent models developed for Florida and Texas utilize more modern statistical analysis for predicting crash frequency at a grade crossing. States such as North Carolina are moving toward an economic analysis model of hazard ranking to incorporate the US DOT model in a more comprehensive economic analysis of the grade crossing. Table 2.1 gives a summary of those models (Sperry et al. 2017).

Table 2.1 Usage of Different HRGC Safety Assessment Methods

Formula/Method	**Number of States**	**Percent of States**
US DOT Accident Prediction Model	19	38%
State-Specific Formula or Method	11	22%
None/No Formula Mentioned	11	22%
New Hampshire Hazard Index	5	10%
Multiple Formulas	2	4%
NCHRP 50 Accident Prediction Model	1	2%
Peabody-Dimmick Formula	1	2%
Total All States	50	100%

2.1 Peabody-Dimmick Formula

The earliest rail crossing crash prediction model was the Peabody Dimmick formula, which was published in 1941 and used extensively through the 1950s (Peabody and Dimmick 1941). It was based on five-year crash data reported at rural crossings in 29 states; the formula is:

$$A_5 = 1.28 * \frac{(v^{0.170})(T^{0.151})}{p^{0.171}} + K \tag{2.1}$$

where A_5 is the expected number of crashes at a rail crossing in five years, v is the AADT, T represents the average daily through trains, p is a protection coefficient (indicating presence of warning devices) and K is an additional parameter determined from a graph. The formula utilized

AADT and the number of through trains to measure crash exposure but does not take into account the temporal distribution of roadway and rail traffic.

2.2 New Hampshire Hazard Index

The New Hampshire Index is given by (Ogden 2007):

$$\mathrm{HI} = (V)(T)(P_f) \tag{2.2}$$

where HI is hazard index, V is the AADT, T represents the average daily through trains and P_f represents a protection factor (indicating the presence of warning devices). The basic formulation of the New Hampshire Index is based on AADT and train traffic. Several states developed their own hazard index formulae by using different values for P_f and adding other factors, such as train speed, highway speed, population, sight distance, number of tracks, surface condition, alignment, presence of nearby intersections, etc.

2.3 NCHRP 50 Accident Prediction Model

The National Cooperative Highway Research Program (NCHRP) Report 50 (Ogden 2007) reported the NCHRP Hazard Index for rail crossing assessment; it has the following form:

$$\mathrm{EA} = (A)(B)(CTD) \tag{2.3}$$

where EA is expected crash frequency, *A* is vehicles per day factor (provided in tabular format as a function of vehicles per day), *B* is a protection factor indicative of warning devices present at a crossing and *CTD* is the current trains per day at the crossing. According to Austin and Carson

(2002), no formal definition of urban and rural areas accompanied the Index and significantly different crash predictions were possible by switching between urban and rural values.

2.4 US DOT Accident Prediction Model

The US DOT Accident Prediction Model was more comprehensive than previous models with the following form:

$$\mathrm{a} = (K)(EI)(DT)(MS)(MT)(HP)(HL)(HT) \tag{2.4}$$

where *K* is a constant, EI the exposure index factor, DT is the day through trains, MS the max train speed, MT the number of main tracks; HP the highway paved factor, HL the highway lanes factor and HT is the highway type factor.

The FRA has developed additional tools and resources to make the US DOT Accident Prediction Model more accessible to users by way of its GradeDec.net evaluation tool (US Department of Transportation 2018) and the Web Accident Prediction System (Federal Railroad Administration 2020)

Besides some updates in the 1980s, the model structure of the US DOT Accident Prediction Model has not changed substantially since its initial development in the mid-1970s. The latest version was developed in 1986 by removing a variable for highway functional classification (Hitz 1986).

2.5 Connecticut DOT Hazard Ranking Index

This hazard index was first mentioned in the Connecticut Railway-Highway Crossing Program 2014 Annual Report (Connecticut Department of Transportation 2015).

$$HI = \frac{(T+1)*(A+1)*AADT*PF}{100} \tag{2.5}$$

where HI is Calculated Hazard Index, T is Train Movements per day, A is the number of vehicle/train crashes in the last 5 years, $AADT$ is annual average daily traffic and PF is protection factor.

2.6 Florida DOT Safety Hazard Index

In 2014, FDOT updated its hazard ranking index which was developed by researchers at Florida State University (Niu et al. 2014). This is a hybrid crash prediction model/Hazard index.

Logit model:

$$\begin{aligned} \mathrm{t} = -8.896 + 0.780 * Risk + 0.020 * MTS \\ + 0.014 * HWSPD + 1.023 * Track \\ + 0.965 * Lane - 0.540 * Flash \end{aligned} \tag{2.6}$$

Prediction model

$$\mathrm{P} = \exp(t) / [1 + \exp(t)] \tag{2.7}$$

Adjustment for Acc. History

$$P^* = \mathrm{P}\sqrt{\frac{H}{P*Y}} \tag{2.8}$$

Safety Index

$$\mathrm{I} = 90 + \left(1 - \sqrt{\frac{P^*}{MaxP}}\right) - 5 * \left(log_{10}(B+1)\right) * F \tag{2.9}$$

where $Risk = log(Train) * AADT$, $Train$ is a yearly average of the number of trains per day, $AADT$ is annual average daily traffic, MST is maximum timetable speed, $HWSPD$ is posted vehicle speed limit, $Track = log(main\ tracks + other\ tracks)$, $Lane$ is the number of highway lanes, $Flash$ is dummy variable for the presence of flashing lights, Y is predicted the number of crashes per year at crossing adjusted for history, H is the number of crashes at crossings during history period, P is the number of years of crash history period, I is safety index value, $MaxP$ is the maximum value of incident prediction, B is the number of school buses at crossing, and F is a variable for warning devices.

2.7 Missouri DOT Exposure Index

This index was developed in 2003 (Qureshi et al. 2003)

Passive Crossings: $$\mathrm{EI} = \mathrm{TI} + \mathrm{SDO}(\mathrm{TI}) \tag{2.10}$$

Active Crossings: $$\mathrm{EI} = \mathrm{TI} \tag{2.11}$$

where TI is traffic index, $\mathrm{TI} = \frac{(VM*VS)[(FM*FS)+(PM*PS)+(SM*10]}{10000}$, SDO is sight distance obstruction factor, $\mathrm{SDO} = \frac{Required\ sight\ distance - Actual\ sight\ distance}{Requuired\ sight\ distance}$, VM is annual average daily traffic, VS is vehicle speed, FM is daily freight train movements at a crossing, FS is freight train speed, PM is daily passenger train movements at a crossing, PS is passenger trains speed and SM is daily switching movements at a crossing.

2.8 North Carolina DOT Investigative Index

This index was described in the North Carolina Railway-Highway Crossing Program 2014 Annual Report (North Carolina Department of Transportation 2015). This index was initially developed in the 1970s and updated in the 1980s.

$$\mathrm{TI} = \frac{PF*ADT*TV*TSF*TF}{160} + \left(70 - \frac{A}{Y}\right)^2 + SDF \tag{2.12}$$

where PF is protection factor, ADT is average daily traffic, TV is daily train volume, TSF is train speed factor= $\frac{Maximum\ train\ speed}{50} + 0.8$, TF is track factor, A is number of crashes over history period, Y is number of years in crash history, and SDF is the sight distance factor = $\frac{sum(SDF_n)}{4} * 16$.

2.9 Texas DOT Priority Index

This index was first developed in 2013 (Weissmann et al. 2013) and revised in 2015. It's a state-specific hybrid crash prediction model, given by:

$$\mu = \exp\,[-6.9240 + PF + (0.2587 * HwyPaved) - (0.3722 * UrbanRural) + (0.0706 * TrafLane) + (0.0656 * TotalTrack) + (0.0022 * ActualSD) + (0.0143 * MaxSpd) + (0.0126 * MinSpd) + \left(1.0024 * log_{10}(TotalTrn + 0.5)\right) + \left(0.4653 * log_{10}(AADT)\right) - (0.2160 * NearbyInt) + (0.0092 * SpdLmt)] \tag{2.13}$$

where μ is the predicted number of crashes per year, *PF* is protection factor, $HwyPaved$ is dummy variable, *UrbanRural* is dummy variable, *TrafLane* is the number of roadway lanes, *TotalTrack* is the total number of tracks at a crossing, *ActualISD* is actual stopping sight distance for approach, *MaxSpd* is maximum typical train speeds, *MinSpd* is minimum typical train speeds for switching, *TotalTrn* is total daily trains, *AADT* is annual average daily traffic, *NearbyInt* is dummy variable representing nearby intersections, and *SpdLmt* is roadway speed limit on approach..

2.10 FRA's New Model for HRGC Accident Prediction and Severity

The FRA published an update to its accident prediction model (Brod and Gillen, 2020) to support grade crossing management by enabling more accurate risk ranking of HRGCs, more rational allocation of resources for public safety improvements and the ability to assess the statistical significance of variances in the measured risk. The model is based on the zero-inflated negative binomial (ZINB) regression along with the Empirical Bayes (EB) method that accounts for crash history while correcting for "regression to the mean" bias. A multinomial logistic (MNL) regression was utilized for the crash severity component having fatal, injury, and property damage only as the crash outcomes. The new ZINB regression model has the following equations (Brod and Gillen, 2020); the ZINB count model is given by:

$$\mathrm{N}_{CountPredicted} = e^{[\beta_0+\beta_1\cdot \mathrm{lExpo}+\ \beta_2\cdot \mathrm{D}_2+\beta_3\cdot \mathrm{D}_3+\beta_4\cdot \mathrm{RurUrb}+\beta_5\cdot \mathrm{XSurfID2s}+\beta_6\cdot \mathrm{lAadt}+\beta_7\cdot \mathrm{lMaxTtSpd}]}$$

The ZINB zero-inflated model is given by:

$$\mathrm{P}_{InflatedZero} = \frac{\mathrm{z}}{1+\mathrm{z}}$$
$$\mathrm{z} = e^{[\gamma_0+\gamma_1\cdot \mathrm{lTotalTrains}]}$$

The ZINB combined model is given by:

$$\mathrm{N}_{Predicted} = \mathrm{N}_{CountPredicted}\cdot\left(1-\mathrm{P}_{InflatedZero}\right)$$

Where:

$N_{CountPredicted}$	Predicted accidents of count model (data for left-hand side of regression are counts of accidents at crossings in 5-year period 2014–2018)
$P_{InflatedZero}$	The probability that the grade crossing is an "excess zero"
$N_{Predicted}$	Predicted accidents after accounting for excess zeroes
lExpo[1]	Exposure, equal to average annual daily traffic times daily trains
D_2	If warning device type is lights =1, 0 otherwise
D_3	If warning device type is gates =1, 0 otherwise
	(note: if both D_2 and D_3 are zero, then warning device type is passive)
RurUrb	If Rural = 0, if Urban = 1
XSurfID2s	Timber = 1, Asphalt = 2, Asphalt and Timber OR Concrete OR Rubber = 3, Concrete and Rubber = 4
lMaxTtSpd[1]	Maximum timetable speed (integer value between 0 and 99)
lAadt[1]	Average annual daily traffic
lTotalTrains[1]	Total number of daily trains

[1]These variables have been transformed as follows: $lx = \log(1+\alpha x)$, where x is the original variable and α is a factor. The factor α was selected so that for the median value of x, $\ln(1+\alpha x) = \ln(x)$

The estimated coefficients are as follows (Table 4.1 in Brod and Gillen, 2020):

ZINB regression count model coefficients (negative binomial with log link)

Variable	Estimate	Std. Error	Z value	Pr(>|z|) (p-value)	Significance Code
(Intercept)	-8.35922	0.32079	-26.059	< 2e-16	***
lExpo	0.19023	0.02866	6.638	3.18e-11	***
D2	-0.28478	0.04806	-5.926	3.10e-09	***
D3	-0.85770	0.04089	-20.976	< 2e-16	***
RurUrb	0.39346	0.03162	12.444	< 2e-16	***
XSurfaceID2s	0.13182	0.01715	7.686	1.52e-14	***
lMaxTtSpd	0.68760	0.68760	22.702	< 2e-16	***
lAadt	0.10626	0.10626	3.511	0.000446	***
Log(theta)	-0.25934	.08867	-2.925	.003447	**

ZINB regression zero-inflation coefficients (binomial with logit link)

Variable	Estimate	Std. Error	z-value	Pr(>|z|) (p-value)	Significance Code
(Intercept)	1.17084	0.19001	6.162	7.19e-10	***
lTotalTr	-1.01088	0.08452	-11.961	< 2e-16	***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The MNL crash severity model utilized grade crossing characteristics and modeled the probabilities of fatal, injury, and property damage-only crashes. Fatal crashes were selected as the reference category and the MNL estimated the probabilities of the other two categories relative to the reference category. The crash severity model equations were as follows.

Injury crash (relative to fatal crash):

$$ln\left(\frac{\mathrm{P}(acctype = injury \mid \mathrm{A})}{\mathrm{P}(acctype = fatal \mid \mathrm{A})}\right) =$$

$$\beta_{20} + \beta_{21} \cdot \mathrm{lMaxTtSpd} + \beta_{22} \cdot \mathrm{lTrains} + \beta_{23} \cdot \mathrm{RuralUrban} + \beta_{24} \cdot \mathrm{D_2}$$

Property damage crash (relative to fatal crash):

$$ln\left(\frac{\mathrm{P}(acctype = PDO \mid \mathrm{A})}{\mathrm{P}(acctype = fatal \mid \mathrm{A})}\right) =$$

$$\beta_{30} + \beta_{31} \cdot \mathrm{lMaxTtspd} + \beta_{32} \cdot \mathrm{lTrains} + \beta_{33} \cdot \mathrm{RuralUrban} + \beta_{34} \cdot \mathrm{D_2}$$

Where:

P(acctype = fatal | A)	The probability of a fatal accident given an accident A
P(acctype = injury | A)	The probability of an injury accident given an accident A
P(acctype = PDO | A)	The probability of a PDO accident given an accident A
lMaxTtSpd	Natural log of the maximum (rail) timetable speed at the crossing
lTrains	Natural log of the total number of daily trains at the crossing
RuralUrban	1 if crossing is in a rural (non-urban) environment, 0 if in urban
D2	Has value 1 if warning device type is lights, 0 otherwise

The estimated coefficients were as follows (Table 4.1 in Brod and Gillen, 2020).

Regression Output – Accident Severity				
Part A – For a given accident, probability of an injury accident (relative to a fatal accident)				
Variable	**Estimate**	**Std. Error**	**z-value**	**Pr(|z|)>0**
Intercept	5.248627	0.355109	14.78032	0
lMaxTtSpd	-0.92544	0.097943	-9.44876	0
lTrains	-0.28326	0.042458	-6.6716	2.53e-11
RuralUrban	-0.27408	0.072886	-3.76042	0.00017
D2	0.489354	0.141041	3.469598	0.000521
Part B – For a given accident, probability of a PDO accident (relative to a fatal accident)				
Variable	**Estimate**	**Std. Error**	**z-value**	**Pr(|z|)>0**
Intercept	6.957135	0.339015	20.52161	0
lMaxTtSpd	-1.23128	0.092907	-13.2528	0
lTrains	-0.22114	0.039411	-5.61125	2.01e-08
RuralUrban	-0.24085	0.067191	-3.58462	0.000338
D2	0.330487	0.135769	2.434192	0.014925

Forecasts for injury severity can then be obtained by using the standard equations for multinomial models.

Chapter 3 Modeling Background

This chapter presents background information on two types of models that are prevalent for count data such as yearly crashes at HRGCs: Poisson and the Zero Inflated Poisson/Negative Binomial model.

3.1 Poisson Regression Model

The nature of crash frequency is non-negative integers or count data and the widely adopted approach has been the Poisson regression model (Miaou and Lum, 1993). Poisson model is a parametric model in which the crash occurrence Y follows a Poisson distribution, which can be described mathematically:

$$Y \sim Poisson(\mu(.)) \tag{3.1}$$

Where μ is the model parameter. So, the probability of variable Y taking integer values 1, 2, 3,… can be represented as:

$$P\{Y = y\} = \frac{e^{-\mu}\mu^{y}}{y!} \tag{3.2}$$

$$E(Y) = var(Y) = \mu \tag{3.3}$$

Where the mean $E(Y)$ and variance $var(Y)$ are equal. Thus, the probability of zero is:

$$P\{Y = 0\} = e^{-\mu} \tag{3.4}$$

As the Poisson model became the basis of many studies, its variants also gained popularity due to the limitations of simple Poisson models. For example, the Negative Binomial/Poisson-Gamma model can handle over-dispersion which occurs when mean of response variable is much higher than the variance while it violates the basic Poisson model assumptions (Milton and Mannering, 1998). In a negative binomial distribution with parameters μ and α, the mathematical form is as follows:

$$P\{Y = y\} = \frac{\Gamma(y+\alpha^{-1})}{y!\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\mu+\alpha^{-1}}\right)^{\alpha^{-1}}\left(\frac{\mu}{\mu+\alpha^{-1}}\right)^{y} \tag{3.5}$$

$$P\{Y = 0\} = (1 + \alpha\mu)^{-1/\alpha} \tag{3.6}$$

$$E(Y) = \mu \tag{3.7}$$

$$var(Y) = \mu(1 + \alpha\mu) \tag{3.8}$$

Where a quadratic function of the mean for $\alpha > 0$, equivalent to the Poisson variance if $\alpha = 0$.

Furthermore, Lord and Mannering (Lord and Mannering, 2010) pointed out a variety of potential data and methodological issues in crash frequency analyses that have been identified in existing literature, including over-dispersion, under-dispersion, unobserved temporal and spatial correlation, low sample-mean and small sample size, crash-type correlation, fixed parameters, etc. These issues could lead to erroneously specifying analytical models and hence misleading inferences if not addressed properly.

3.2 Zero-inflated Model

Another set of models is zero-inflated Poisson and negative binomial models, designed to deal with a significant proportions of a response variable taking zero values or more zeros than one would expect in conventional count data scenario. The formulas for zero-inflated Poisson model is as follows, including a parameter π:

$$P\{Y = 0\} = \pi + (1 - \pi)e^{-\mu} \tag{3.9}$$

$$E(Y) = (1 - \pi)\mu \tag{3.10}$$

$$var(Y) = (1 - \pi)\mu(1 + \mu\pi) \tag{3.11}$$

On the other hand, a zero-inflated negative binomial model is formulated as follows:

$$P\{Y = 0\} = \pi + (1 - \pi)(1 + \alpha\mu)^{-1/\alpha} \tag{3.12}$$

$$E(Y) = (1 - \pi)\mu \tag{3.13}$$

$$var(Y) = (1 - \pi)\mu(1 + \mu(\pi + \alpha)) \tag{3.14}$$

Where if $\alpha = 0$ the model is equal to a zero-inflated Poisson model.

This model was used to model crash frequency. As the crash frequency is count data (non-negative integer), and crash occurrence at HRGC is a relatively rare event, the data is considered exhibiting over-dispersion and excess zero. The zero-inflated Poisson (ZIP) Model assumes that data distribution is a combination of Poisson distribution and logit distribution, which fits the circumstance of this research. Figure 3.1.1 simulates 500 samples that follow a zero-inflated Poisson distribution.

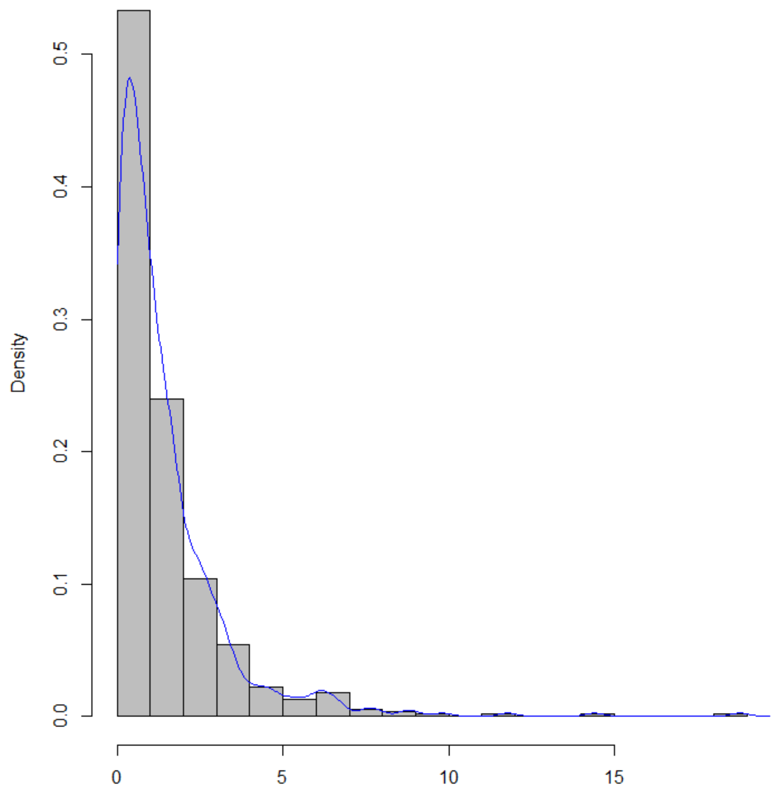


Figure 3.1 Simulated Zero-inflated Poisson Distribution

As can be seen from this figure, the distribution is a skewed Poisson distribution with large amount of data equal to zero. Therefore, to describe the distribution, the ZIP model contains two parts: a Poisson model, which is responsible for predicting non-negative value, and a logit model for predicting excess zeros. The ZIP model can be expressed as:

$$\mathrm{P}(y_i = 0) = p_i + (1 - p_i) * e^{-\mu_i} \tag{3.15}$$

$$\mathrm{P}(y_i = m, for\ m > 0) = (1 - p_i) * \frac{\mu_i^{y_i}}{m!} e^{-\mu_i} \tag{3.16}$$

$$p_i = \frac{\lambda_i}{1+\lambda_i} \tag{3.17}$$

$$\log(\mu_i) = \mathrm{X}\beta \tag{3.18}$$

$$logit(p_i) = log\left(\frac{p_i}{1-p_i}\right) \tag{3.19}$$

where pi is the logistic link function defined by equation (3.17), μ_i is the Poisson component defined by equation (3.4). As can be seen, the ZIP model splits the possibility of response values into two scenarios: equation (3.1) describes the scenario when the count is equal to zero, while equation (3.2) generates count values by a Poisson model when the count is not zero.

The coefficients can be estimated by solving its maximum likelihood function. The likelihood function can be expressed as:

$$\begin{aligned} L = & \sum_{for\ i\ if\ y_i=0} log(\lambda_i + e^{-\mu_i}) \\ & + \sum_{for\ i\ if\ y_i>0} [y_i log(\mu_i) - \mu_i - log(m!)] \\ & - \sum_i log(1+\lambda_i) \end{aligned} \tag{3.20}$$

Because it is often observed in crash data that many locations have no occurrence of crash, by splitting roadway segments into crash-free and crash-prone categories, zero-inflated models have been frequently considered in research (Shankar et al., 1997; Lee and Mannering, 2002; Lord et al., 2007). Critics have argued that the crash-free state has a long-term mean equal to zero, this model cannot properly reflect the crash-data generating process (Malyshkina and Mannering, 2010). Similarly, various other count data models were considered over the years including the Gamma model, the negative binomial-Lindley model, Conway-Maxwell-Poisson model, and so on.

Chapter 4 Data Collection

This chapter provides detailed information on the data utilized throughout this research study. Safety data regarding rail crossings from multiple sources were collected and integrated for analysis, including HRGC inventory database and crash history data extracted from publically-available FRA data, Railroad Inventory Management System (RIMS) obtained from NDOT, Lancaster roadway inventory database and land use data obtained from City of Lincoln, Nebraska. A significant number of database variables were manually inspected and verified during field visits such as roadway speed limit, pavement type, land use, etc.

4.1 FRA HRGC Inventory Database

According to the Federal-Aid Policy Guide (FAPG 924.9(a) (1)), each state should maintain "a process for collecting and maintaining a record of crash, traffic, and highway data, including, for railroad-highway grade crossings, the characteristics of both highway and train traffic" (U.S. Department of Transportation, 1991). National Highway-Rail Crossing Inventory Reporting Requirements also states that, "in order for the Crossing Inventory to serve as an effective database, States and railroads need to exchange information with each other and promptly update the crossing data records as changes occur". Thus, FRA collects from each state and maintains a database on HRGCs for the entire US.

Updates to HRGC inventory data are usually provided by the local coordinators and submitted using FRA-approved forms. These forms have specifications for different field names and value assignments. Authorized users must submit new values for specific field names accordingly. The field names, filed description and values used in this study are attached in Appendix A, which conformed to the FRA HRGC inventory database. Because reporting updates for the inventory database does not necessarily require verification from other agencies, data for some fields may not be updated regularly, such as AADT and train traffic volumes. This could lead to outdated or erroneous data, which could affect crash predictions by models based on the database. Accuracy issues in the FRA crossing inventory database raise concerns for states and railroad companies. In addition, FRA provides geospatial resources to the public on rail networks, including data on HRGCs, Amtrak stations, etc. Spatial information of a given crossing is denoted by latitude and longitude in the database.

Various fields are useful when integrating crossing inventory data with crash data, such as crossing ID, state, county, nearest city name, etc. The inventory database also provides details for the train traffic traversing a crossing: total daylight thru trains, total night time thru trains, total transit trains, number of main tracks, number of siding tracks, number of yard tracks, number of transit tracks, average passenger train count per day, etc. Variables with regards to safety measures include presence of signs/signals, number of crossbuck assemblies, number of stop signs, number of yield signs, number of bells, flashing lights, channelization devices/medians, gate configuration, etc. The FRA inventory database also provides information on the crossing highway, such as number of traffic lanes crossing rail track, pavement type, highway functional classification, street or road name, posted highway speed limit, etc.

4.2 FRA HRGC Crash Database

Title 49 Code of Federal Regulations (CFR) Part 225 (US GPO, 2006) requires reporting of railroad-related crashes to the FRA. Specifically, FRA has made efforts to build several databases to gather information on evaluating railroad safety, including: train crash database, trespasser crash database, rail equipment crash database, highway rail crossing crash database, railroad casualty database, etc. FRA uses the reported crash data to summarize a yearly report on crashes that involve the impact of a train with a roadway user. If a crash is involved with railroad signal failure or grade crossing failure, railroad companies are required to provide more details along with the crash report form. Furthermore, FRA requires various forms with regards to different scenarios, such as Form FRA F 6180.55 for injury and illness and Form FRA F 6180.57 for Highway-Rail Accident/Incident, etc.

The fields available in the crash database consist of a series of categories, such as crash information, crossing information, train information, environmental factors, highway characteristics, etc. For instance, the crash information includes time of crash, AM or PM, injury severity outcome, number of injuries or fatalities of roadway users, number of injuries or fatalities of railroad employees, number of injuries or fatalities of train passengers, etc. Environmental factors at the time of crash consist of temperature, weather conditions, lighting conditions, etc. Train information includes number of cars, number of locomotives, type of train, train speed, etc. Additionally, other important factors such as release of hazardous materials are also included. Textual descriptions of crashes can also be provided in the reporting form. Appendix B provides the FRA HRGC crash database fields.

4.3 Field Validation of the FRA HRGC Database

As part of Lancaster County HRGC consistency analysis, the research team validated the information contained in the FRA HRGC inventory database with HRGCs in the field. HRGCs were taken into consideration if they were public, at-grade, and operational. The research team visited public rail crossings in Lancaster County and compared field conditions with those of the database; corrections were made to any erroneous records in the database as well as missing values added when available in the field. This inventory validation effort was then extended to Cass, Douglas, Gage, Jefferson, Otoe, Saline, Sarpy and Saunders counties. The selection of these additional eight counties was based on railroad network considerations, urban/rural nature of a county, proximity to the University of Nebraska-Lincoln, and availability of funds in the project.

Figure 4.1 illustrates the HRGC filtration process and the FRA HRGC inventory database variables used for Lancaster County. A similar process and the same variables were used for HRGC filtration in other counties. Figure 4.2 graphically illustrates the results of HRGC filtration process for Lancaster County. For this county, there were 565 rail crossings in the FRA HRGC database; however, exclusion of private, elevated (grade-separated), and closed HRGCs resulted in the selection of 112 HRGCs. Field visits to the selected HRGCs revealed that seven HRGCs were either missing or relocated thereby resulting in 105 Lancaster County HRGCs that were field-verified.

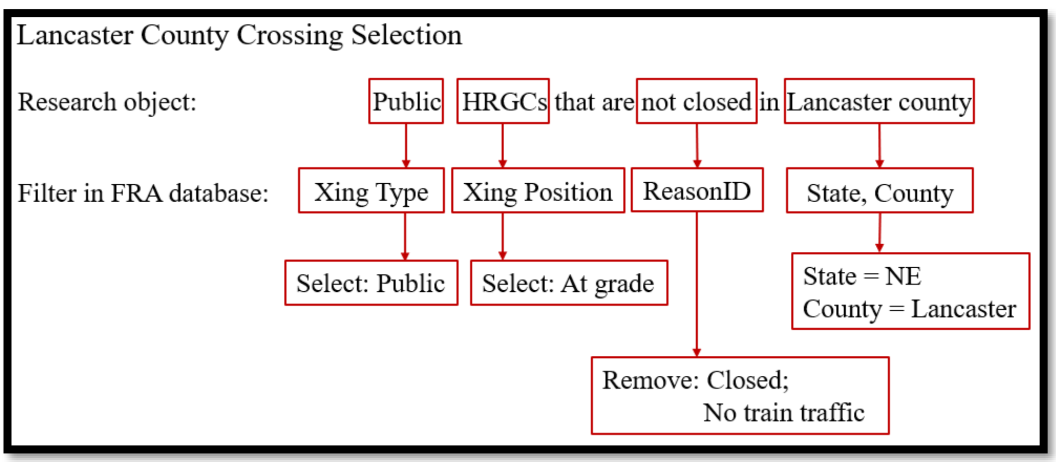


Figure 4.1 HRGC Filtration Process for Lancaster County

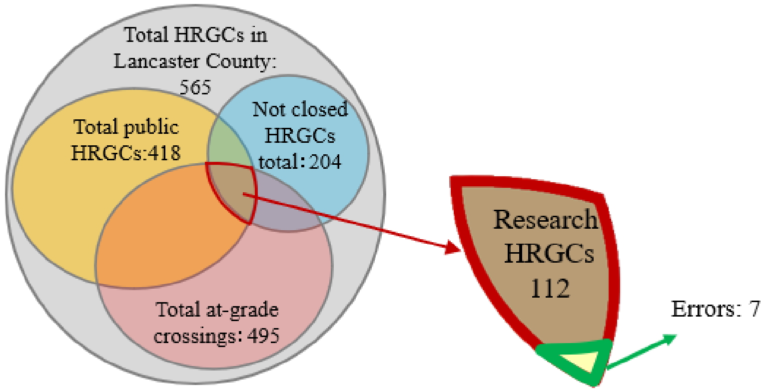


Figure 4.2 Results of HRGC Filtration Process for Lancaster County

For each field-visited HRGC, a total of 53 database variables were checked and digital pictures of the HRGC obtained. Any incorrect values in the database were corrected per field conditions as well as missing values added when they were available in the field. Table 4.1 presents a summary of the corrections and missing value additions for the nine Nebraska counties from field visits. In aggregate, 539 HRGCs were field-investigated and 27 (5.0%) were found to be either abandoned (non-operational), private (listed as public in the database) or altogether non-existent. This effort resulted in 2,241 values to be corrected and 1,732 missing values to be added giving an average of 7.4% of the database values that were changed at each HRGC.

Table 4.1 Summary of Corrections and Added Missing Values from Field Validation

County	Number of Corrected Values	Number of Missing Values Added	HRGCs Visited	Abandoned/Non-existent/Private HRGCs	Percent Corrected and Added Missing Values
Lancaster	376	657	112	7	9.2
Cass	307	83	55	2	7.1
Douglas	286	108	67	3	5.9
Gage	115	347	41	4	11.3
Jefferson	174	25	46	2	4.3
Otoe	285	46	79	4	4.2
Saline	119	37	38	0	4.1
Sarpy	144	59	25	2	8.1
Saunders	435	370	76	3	10.6
Total	2241	1732	539	27	7.4

During the spring 2020 COVID-19 shutdown, the research team relied on the NDOT's PathWeb system to validate the FRA HRGC database. This photo-based system is focused on state highways and therefore, HRGCs located only on the state highways could be checked. Table 4.2 presents a summary of the corrections and missing value additions using the PathWeb system. This effort identified 6 (2.9%) HRGCs that were either abandoned, private, or altogether non-existent. The number of corrected values was 670 while 109 missing values were added to the database for an average of 3.8% of the database values changed at each HRGC.

Table 4.2 Summary of Corrections and Added Missing Values Using NDOT's PathWeb System

State Highway system	Number of Corrected Values	Number of Missing Values Added	HRGCs Inspected	Abandoned/Non-existent/Private	Percent Corrected and Added Missing Values
PathWeb 2019	670	109	203	6	3.8

Database variables that frequently contained incorrect information included: the number of crossbucks, number of yield signs or stop signs, number of advance warning signs, presence of channelization devices, crossing surface type, approach surface type and highway speed limit. Figure 4.3 presents an example of the inconsistency between the FRA HRGC inventory database and field conditions at crossing 064112B in terms of presence of yield sign, pavement type, approach surface type and pavement marking. Figure 4.4 shows an example of a crossing (crossing ID 083524P) that was abandoned but is still in the FRA HRGC inventory database.

Figure 4.3 Data Correction Example, Crossing 064112B

Figure 4.4 An Example of an Abandoned Crossing (083524P)

The numbers of corrected or added missing values for various variables were recorded for each county. For instance, figure 4.5 shows the numbers of corrected or added values for different inventory variables in Gage County. The variables with high incorrect values were HwynrSig (does nearby highway intersection have traffic signals), Bkl_FlashPost (mast-mounted flashing lights: back lights), and Sdl_FlashPost (mast-mounted flashing lights: side lights).

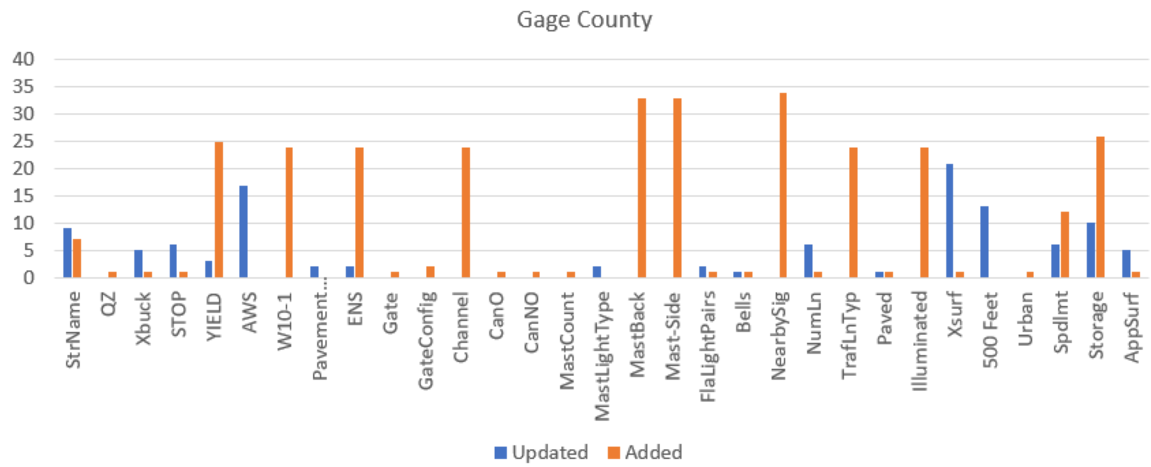


Figure 4.5 Corrected or Added Values of Each Variable for Gage County

In summary, the combined effort of field visits and use of the PathWeb system resulted in inventory verification of 742 HRGCs in Nebraska; in total 2,911 values were corrected and 1,841 missing values were added to the HRGC inventory database while 33 HRGCs were identified that were either abandoned, private listed as public or altogether non-existent. An Excel file containing the original and corrected/added values and a GIS database (including the HRGC digital pictures) using ESRI's ArcMap software were created for handover to NDOT (fig. 4.6). In addition, the Lancaster Roadway Inventory Database and land use data from City of Lincoln supplemented the GIS as shown in figure 4.7. This was then used for the HRGC consistency analysis.

According to the FRA HRGC inventory database, there are 2,863 public, at-grade, operational crossings in Nebraska. With 742 HRGCs validated via a combination of field visits and NDOT's PathWeb system, 2,121 HRGCs are remain in need of inventory information validation.

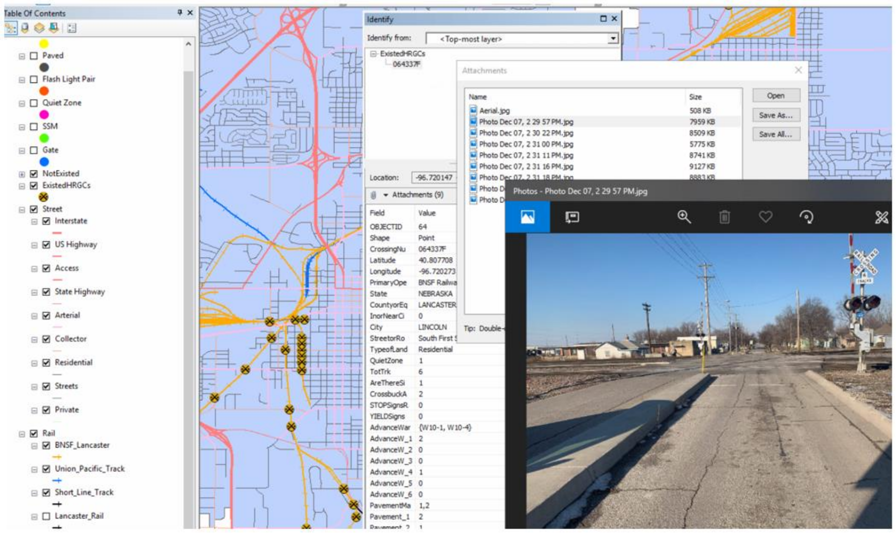


Figure 4.6 GIS Database for HRGCs in Lancaster County

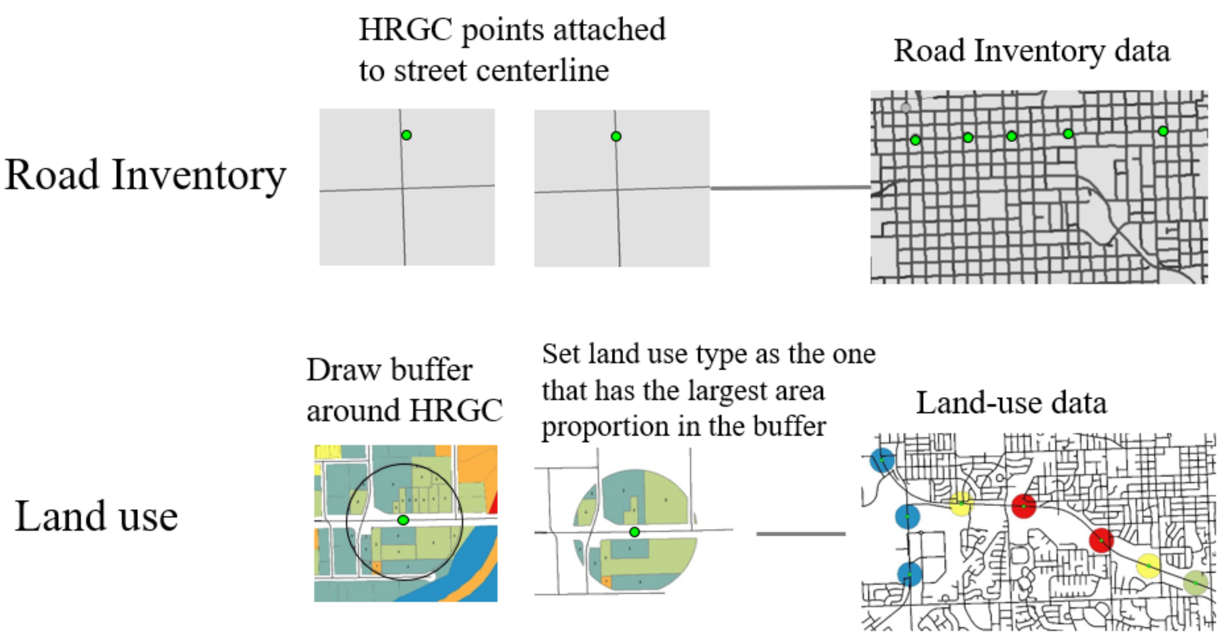


Figure 4.7 Road inventory and land use data

4.4 Database for Updated Crash Prediction Model

The corrected crossing inventory database records were appended to the HRGC crash database by using the unique crossing IDs available in the two databases to create a combined database for crash prediction modeling. The FRA HRGC crash database contained crash history data from 2008 to 2019 on Nebraska HRGCs. For model estimation, the yearly number of reported crashes for each HRGC was considered an observation. Using this framework, 393 observations were associated with crashes. Of these, 224 (57.0%) observations were crashes with no injuries, 124 (31.6%) observations with injuries and 45 (11.4%) observations involved fatal crashes. Model parameter estimation was based on 2008-2018 crash plus inventory data while the 2019 crash plus inventory data were used for the model prediction validation. Chapter 5 provides details of the modeling efforts.

4.5 Descriptive Statistics

After integrating data from various sources, descriptive statistics of the variables used through the model estimation and evaluation process are presented. Note that for each crossing there is one observation for each year. Figure 4.8 shows a histogram plot demonstrating the distribution of the studied highway rail grade crossings by natural logarithmic values of AADT. It can be observed that the maximum and minimum values for AADT are around 50,000 vehicles per day and one vehicle per day, respectively. The average AADT for all considered crossings is approximately 672 vehicles per day.

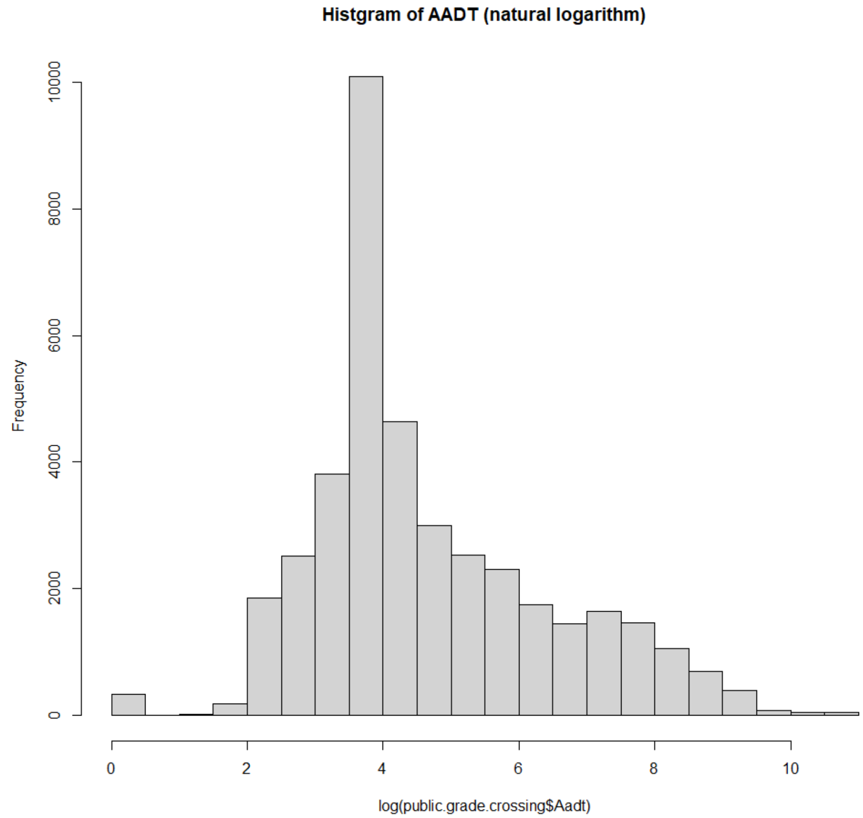


Figure 4.8 Histogram of Highway Rail Grade Crossings by AADT (natural logarithm)

Figure 4.9 shows a histogram plot demonstrating the distribution of the studied highway rail grade crossings by the number of through trains (including day and night). It can be observed that the maximum and minimum values for the number of through trains are 118 and zero trains per day, respectively. The average value for all considered crossings is approximately 16.47 trains per day.

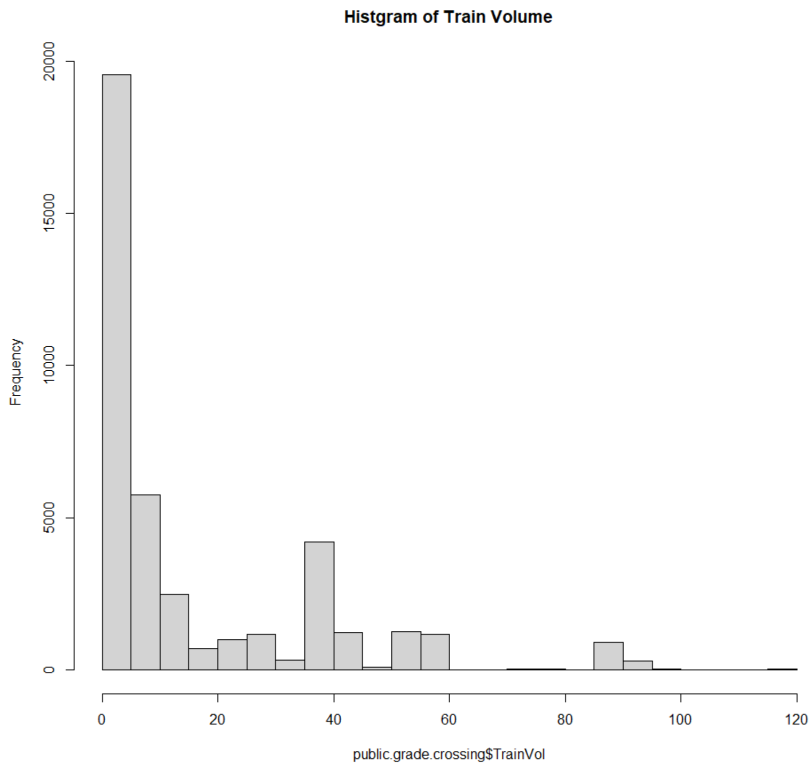


Figure 4.9 Histogram of Highway Rail Grade Crossings by Number of Through Trains

Figure 4.10 shows the distribution of the studied highway rail grade crossings by highway classification (urban or rural). It can be observed that 92.2% of the roadways (a total of 2,192 crossings, excluding missing values) at HRGCs were classified as rural.

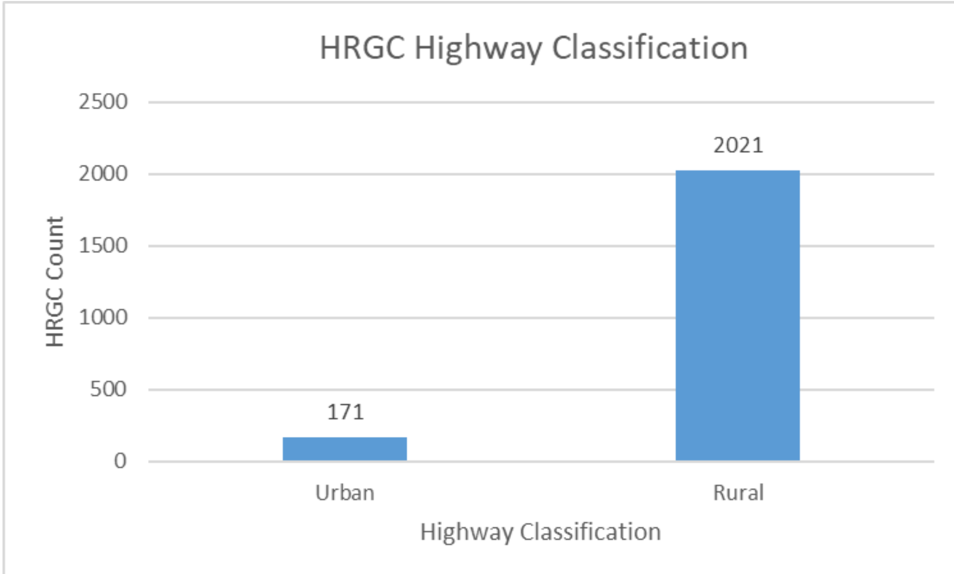


Figure 4.10 Distribution of Highway Rail Grade Crossings by Highway Classification

According to FRA's classification of highway functional classification, roadways can be categorized as six levels: (1) interstate; (2) other freeway and expressway; (3) other principal arterial; (4) minor arterial; (5) collector; and (6) local roadway. Figure 4.11 shows the distribution of the HRGCs by highway functional classification. It can be observed that 1,693 roadways were classified as local roads (77.5% of all the roadways). In addition, there were 136 minor collector roadways, 269 major collector roadways, 73 minor arterial roadways, 13 other principal arterial roadways and 1 other freeways and expressway.

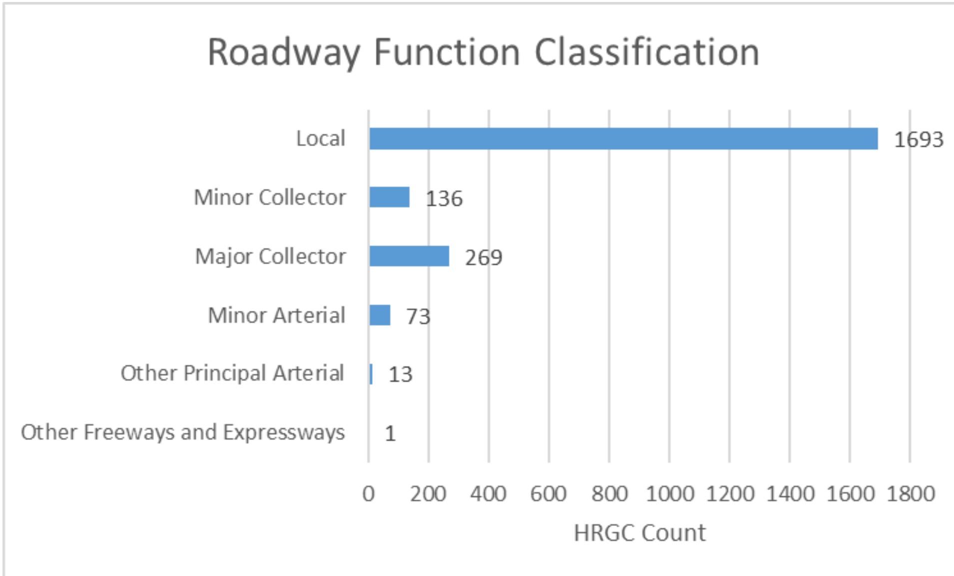


Figure 4.11 Distribution of Highway Rail Grade Crossings by Highway Function Classification

Figure 4.12 presents the distribution of the HRGCs by highway lanes. As shown in the figure, the minimum number of traffic lanes at the HRGCs was one lane, while the maximum value is eight lanes. The distribution indicates the majority of the roadways at HRGCs (85.2%) consisted of two traffic lanes.

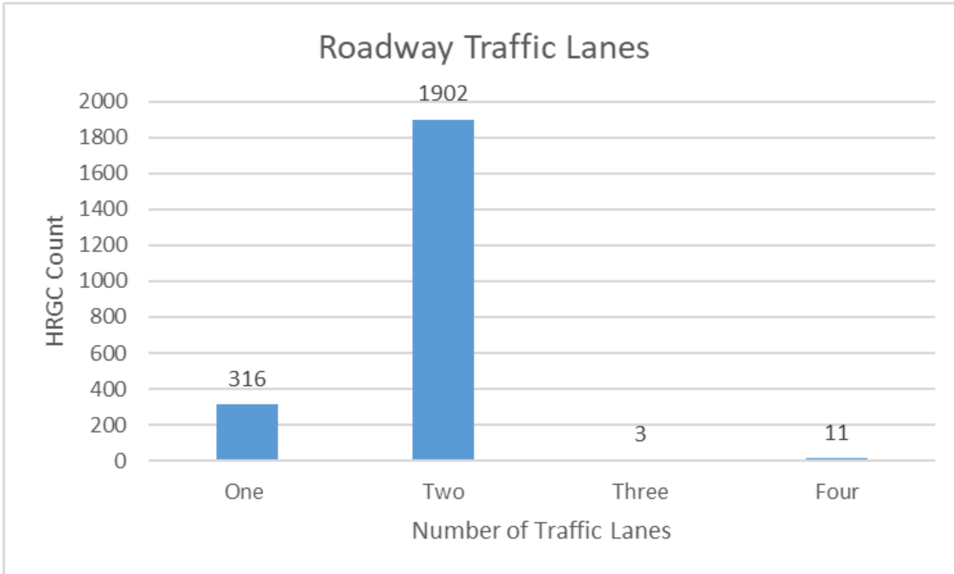


Figure 4.12 Distribution of Highway Rail Grade Crossings by Number of Traffic Lanes

In terms of the dependent variable, crash frequency at HRGCs, based on the crash history data on a yearly basis, only five HRGCs were associated with two crashes while 388 HRGCs had only one crash and the rest of the dataset had zero crashes. It can be observed that the majority of observations (99.0%) did not involve a crash. The disproportionate distribution of zero values warrants the investigation of a zero-inflated model as discussed in Chapter 3.

Chapter 5 HRGC Crash Prediction Model Estimation

This chapter covers the first research objective, which was to update NDOT's 1999 Nebraska Accident Prediction Model for rail crossings using the latest crash and rail crossing inventory data. It presents the estimation of the 2020 Nebraska HRGC Crash Prediction Model based on the dataset created for model estimation including the model estimation process and the different variants that were explored.

The 1999 Nebraska Accident Prediction Model for rail crossings (HNTB, 1999) was based on 5-year data. This research utilized 11-year (2008-2018) crash and HRGC inventory data for the 2020 Nebraska HRGC Crash Prediction Model estimation and 2019 crash and HRGC inventory data for validation of the model predictions. The 11-year dataset is also referred to as training data in this report. The model estimation process aimed to investigate statistical associations of various factors (e.g., crossing characteristics, exposure measures, land use, etc.) with crashes at HRGCs. In this chapter, various statistical modeling techniques (e.g., Poisson or Negative Binomial) are explored and evaluated based on characteristics of the data and statistical tests. The corresponding results present a set of models (equations) for the expected number of crashes per year at Nebraska public HRGCs. Note that the data utilized for model estimation included HRGC corrected inventory data resulting from field visits and use of the NDOT PathWeb system as described previously in this report.

The estimated model equations were validated by predicting crashes for 2019 and comparing those results with the actual crashes reported in 2019. Additionally, results of the model equations were compared to those obtained from the 1999 Nebraska Accident Prediction Model as well as the new FRA Accident Prediction Model (Brod and Gillen, 2020) when applied to Nebraska data. Consequently, the 2020 Nebraska HRGC Crash Prediction Model outperformed the 1999 Nebraska Accident Prediction Model and the new FRA Accident Prediction Model.

5.1 Analysis of Accident Prediction Models Based on Various Criteria

This section presents the estimation results of candidate crash prediction models with descriptions of the model selection procedure. The HRGC crash data for 2019 were used to assess goodness of fit of the candidate crash models based on various performance metrics. Specifically, the model selection criteria included mean of squared error (MSE) and root mean of squared error (RMSE), logarithm score, Akaike information criteria (AIC) and the percent difference in 2019 crash predictions.

There were a few models that could be appropriate for modeling the HRGC crash data: Conventional Poisson regression and Negative Binomial regression (to address over-dispersion), Zero-inflated Poisson/Negative Binomial models (to account for excess zero crashes) and Poisson/Negative Binomial models with mixed effects, assuming normality and homogeneity of variance of residuals. For each model framework, variable selection was performed based on the results of AIC, logarithm score and forward selection. In addition, a "small" model was also considered as an important benchmark, which was based on the variables used in the existing NDOT Accident Prediction Model (HNTB, 1999). To determine the best performing model, several procedures were conducted such as over-dispersion test, model selection, variable selection, etc. and the results are as follows.

5.1.1 Over-dispersion test

A standard Poisson regression models the conditional mean $E(Y) = \mu$, which is assumed equal to the variance of the dependent/response variable. The over-dispersion test assesses the hypothesis that this assumption holds against the alternative that the variance is of the form:

$$var(Y) = \mu(1 + \alpha\mu)$$

Where a quadratic function of the mean for $\alpha > 0$, equivalent to the Poisson variance if $\alpha = 0$. Over-dispersion corresponds to $\alpha > 0$ and underdispersion to $\alpha < 0$. The coefficient α can be tested with the corresponding z statistic which is asymptotically standard normal under the null hypothesis. By building a Poisson model on the model estimation dataset, the over-dispersion test yields a p-value of 0.24 which indicated a lack of significant evidence of over-dispersion or under-dispersion. It can also be validated by examining the mean and variance of the response variable. The yearly mean crash frequency of the training dataset was 0.0098 (crashes) while the variance was 0.0010 ($crashes^2$). Thus, estimating a Poisson model was viable for this dataset and there is was no need for estimating a Negative Binomial models.

5.1.2 Candidate model performance

The US DOT formula has an initial model and two variants (referred to as weighted and normalized). The initial model can be estimated using the following equation:

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

Where:

a is the initial crash prediction outcome;

K is the constant;

EI is the factor for exposure index based on the product of highway and train traffic;

MT indicates the factor for the number of main tracks;

DT indicates the factor for the number of through trains per day during daylight;

HP indicates the factor for highway pavement status;

MS indicates the factor for maximum timetable speed;

HT indicates the factor for highway type;

MS indicates the factor for the number of highway lanes;

This initial model has two variants, based on the values of the highway-rail grade crossing characteristic factors such as traffic control devices installed at a given highway-rail grade crossing: (a) passive; (b) flashing lights; and (c) gates.

For instance, the "weighted" model or the second crash prediction is formulated as follows:

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

Where:

a is the initial crash prediction outcome;

B is the second crash prediction outcome;

N is the number of crashes occurred in T years;

$T_0$is a weighting factor that equals $\frac{1}{0.05+a}$;

The "normalized" model can be formulated by normalizing the constant, which is the sum of the predicted crashes multiplied by the corresponding normalizing constant equal to the number of crashes, which occurred in a recent period. The normalizing procedure is different depending on the installed control devices at each highway-rail grade crossings separately. Similarly, the 1999 NDOT Accident Prediction model (HNTB, 1999) has "weighted" and "normalized" formulas as well.

Table 5.1 presents the performances of different candidate crash prediction models. Evaluation metrics such as AIC, MSE, logarithm score and prediction outcome are reported for comparison.

Table 5.1 Performance of Candidate Nebraska Crash Prediction Models

Candidate models		AIC	MSE	RMSE	Logarithm score	Predicted outcome	Percentage difference in prediction results for 2019
Poisson	All variables	2408.142	0.008132886	0.090182515	0.05391566	21.078	-15.69%
	Small	2452.267	0.008001879	0.089453222	0.05533638	21.262	-14.95%
	Selected variables based on AIC	2422.401	0.008027479	0.0895962	0.05217232	21.078	-15.68%
	Selected variables based on LR test	2418.829	0.008021923	0.089565189	0.05388908	21.079	-15.68%
	Selected variables based on stepwise selection	2434.824	0.008712828	0.09334253	0.04775778	**26.156**	**+4.62%**
	Mixed effects all variables	2407.311	0.008138168	0.090211795	0.05409535	21.077	-15.69%
	Mixed effects small	2442.536	0.007960817	0.089223411	0.05557907	21.262	-14.95%

	Mixed effects all variables based on AIC	2421.009	0.008036451	0.089646255	0.05273017	21.079	-15.68%
	Mixed effects all variables based on LR test	2411.967	0.008080538	0.089891813	0.05287434	21.078	-15.69%
Zero-inflated Poisson	All variables	2382.547	0.008289844	0.09104858	0.04591305	21.086	-15.65%
	Small	2437.525	0.00802454	0.089579797	0.09737134	21.266	-14.93%
	Selected variables based on AIC	2393.524	0.008146457	0.090257725	0.1037519	21.080	-15.68%
	Selected variables based on stepwise selection	3246.723	0.008712828	0.09334253	0.0759834	28.082	+12.33%
	Mixed effects small	2408.802	0.008115772	0.09008758	0.06357789	14.731	-41.07%
	Mixed effects all variables based on LR test	2399.02	0.008027471	0.089596155	0.07578753	14.439	-42.24%

Table 5.1 (cont.) Performance of Candidate Nebraska Crash Prediction Models

Candidate models		AIC	MSE	RMSE	Logarithm score	Predicted outcome	Percentage difference in prediction results
1999 NDOT model	Raw	-	0.008087182	0.089928761	-	0.1469295	-99.41%
	Weighted	-	0.008007666	0.089485563	-	6.05532	-75.78%
	Normalized	-	0.008019007	0.089548908	-	3.749482	-85.00%
US DOT model	Raw	-	0.008056705	0.08975915	-	4.871511	-80.51%
	Weighted	-	0.007994272	0.089410693	-	10.04063	-59.84%
	Normalized	-	0.008007084	0.089482311	-	6.206282	-75.17%
	New ZINB model variables	2398.908	0.008194754	0.090524881	0.05509472	21.09046	-15.64%

As there were 25 crashes reported in 2019, the most accurate model prediction outcome was 26.16 from the Poisson model with stepwise selection of independent variables, even though it did not have the smallest AIC value (a smaller AIC value usually indicates a better goodness of fit). The results indicated that most Poisson models achieved similar outcomes with approximately 21 crashes predicted for 2019, including the New FRA ZINB model (Brod and Gillen, 2020). All regression models were able to outperform variants of the existing US DOT crash prediction models. However, based on the prediction outcome, mixed effects models and zero-inflated models did not demonstrate significant improvement compared to the variants of the conventional Poisson models.

In terms of the variable selection procedure, AIC, likelihood-ratio test and stepwise selection were all utilized. For instance, since missing values were common in the dataset and could lead to potential convergence issues, the modeling started with inclusion of all available variables and then narrowed them down to a smaller set of variables in the stepwise regression process. The stepwise regression started from containing the constant only as the base and moved forward towards a set of variables containing train volume, maximum timetable speed, the number of main tracks, AADT, presence of gates, presence of flashing light, exposure (AADT multiplied by daily trains), etc. The Poisson model selected by the stepwise algorithm was the best performing model in terms of prediction closest to the 2019 reported HRGC crashes. Therefore, it was chosen as the 2020 Nebraska HRGC Crash Prediction Model and it updates the 1999 Nebraska Accident Prediction Model.

In addition, efforts were made to estimate candidate models based on different data, such as using 2008-2016 data as training data and 2017 data for validation, and using 2008-2017 data as training data and using 2018 data for validation. Appendix C presents the results of these scenarios. However, the crash prediction model for Nebraska was based on 2008-2018 data with 2019 data used for validation.

5.1.3 The 2020 Nebraska HRGC Crash Prediction Model estimated coefficients

Tables 5.2 and 5.3 present the output for the stepwise-based Poisson regression model for the 2020 Nebraska HRGC Crash Prediction Model. Note that all variables with numerical values were scaled before model estimation using the formula $\sqrt{\sum(x^2)/(n-1)}$. Indicator variables were created for categorical variables.

Table 5.2 Deviance Residuals

Min	1Q	Median	3Q	Max
-1.0358	-0.1731	-0.1020	-0.0761	3.9529

Table 5.3 Estimated Coefficients of the 2020 Nebraska Crash Prediction Model

Parameter	Estimate	Std. Error	Z-statistic	P-value
Intercept/Constant	-7.1427	0.26092	-27.375	< 2e-16
MaxTtSpd.scaled	1.57265	0.18292	8.597	< 2e-16
Expo.scaled	0.11558	0.02555	4.524	6.08e-06
MainTrk.scaled	0.58506	0.14191	4.123	3.74e-05
Aadt.scaled	0.16671	0.06706	2.486	0.0129

The column labeled “estimates” indicates the βs from the count model. The “Std. Error” corresponds to the standard error calculated for the variable to the left. The “Z-statistic” column is the coefficient divided by the standard error. The statistical significance of the variable was indicated by p-value in the last column, which all showed strong statistical significance.

5.2 Interpreting Regression Output

According to the model estimation results, all of the estimated coefficients for MaxTtSpd, MainTrk, Expo and Aadt were positive indicating positive relationships between the explanatory variables and the response variable (i.e., as the variables increase in values so does the expected HRGC yearly crash frequency).

Using the estimated coefficients, the crash prediction model can be formulated as:

$$log(y) = -7.1427 + 1.573 * \mathrm{MaxTtSpd.scaled} + 0.1156 * \mathrm{Expo.scaled}$$
$$+0.5851 * \mathrm{MainTrk.scaled} + 0.1667 * \mathrm{Aadt.scaled}$$

Where:

y indicates the expected HRGC yearly crash frequency;

MaxTtSpd is the scaled maximum timetable speed;

Expo is the scaled exposure;

MainTrk is the scaled number of main tracks; and

Aadt is the scaled average annual daily traffic.

These variables were found statistically significant at the 95% confidence level. Interpretation of the model output is as follows:

1) With 95% confidence, a one-unit increase in the scaled maximum timetable speed leads to a 381.9% increase in the expected HRGC yearly mean crash frequency;

2) The yearly mean crash frequency is estimated to increase by 12.3% for every one-unit increase in the scaled exposure;
3) With 95% confidence, a one-unit increase in the scaled number of main tracks leads to a 79.5% increase in the yearly mean crash frequency;
4) The yearly mean crash frequency increases by 18.1% for every one-unit increase in the scaled AADT;

The estimated model can predict the number of expected crashes for a HRGC. For instance, figure 5.1 shows predicted values based on the stepwise-based Poisson model, grouped by presence of crossbucks. In the figure, the vertical lines indicate the average exposure for each grouping and the horizontal lines indicate the predicted crash frequency for each grouping.

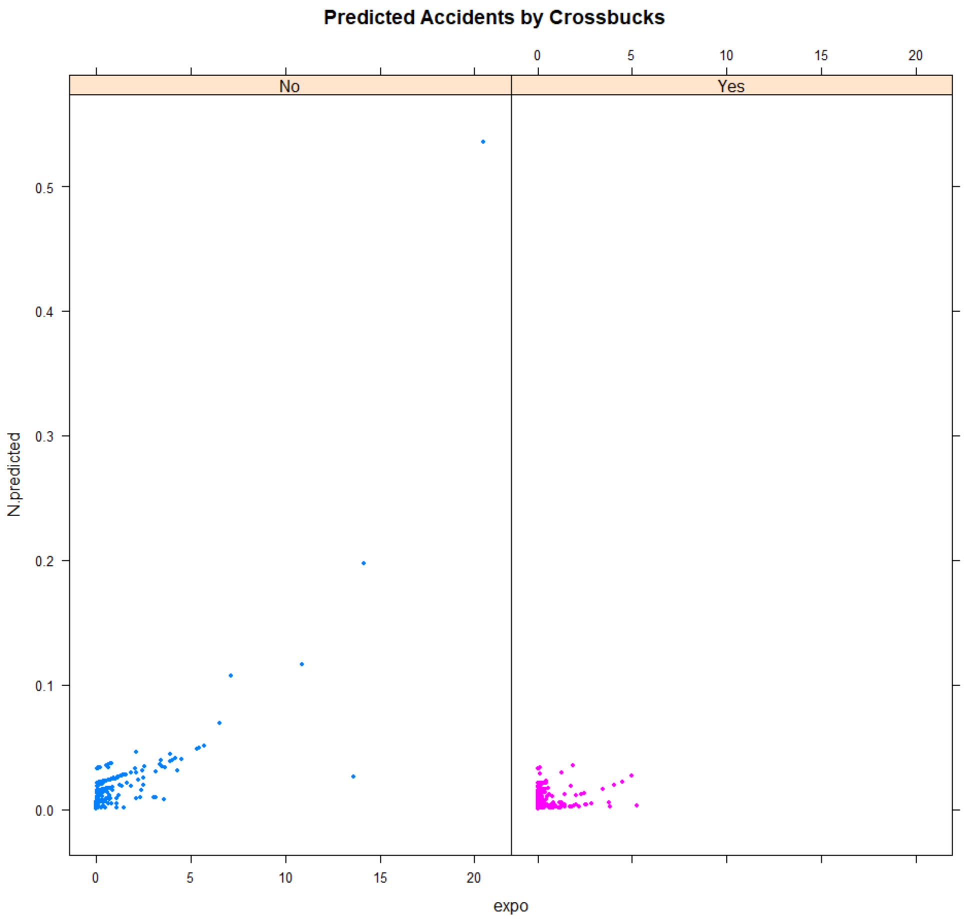


Figure 5.1 Predicted Crashes by Presence of Crossbucks

5.3 Empirical Bayes Prediction Adjustment

To address "regression to the mean" bias, the Empirical Bayes (EB) adjustment accounts for crash history. This technique is described in Hauer (2015) and applied in the FRA's New Accident Prediction Model (Brod and Gillen, 2020). The adjustment can be formulated as follows for each HRGC:

$$N_{expected} = \omega \cdot N_{predicted} + (1 - \omega) \cdot N_{observed}$$

$$\omega = 1/(1 + \frac{var(N_{predicted})}{N_{predicted}})$$

Where:

$N_{expected}$ indicates the adjusted number of predicted crashes;

$N_{predicted}$ indicates the prediction result from the estimated Poisson model;

$N_{observed}$ indicates the number of observed accidents.

Note that, if needed, this procedure is required to account for additional coefficients for zero-inflated models and Negative Binomial models. After applying the EB adjustment, the results of predicted crashes by presence of crossbucks are presented in figure 5.2. Compared to figure 5.1, the values reflected on y-axis are centered on either 1 or 0.

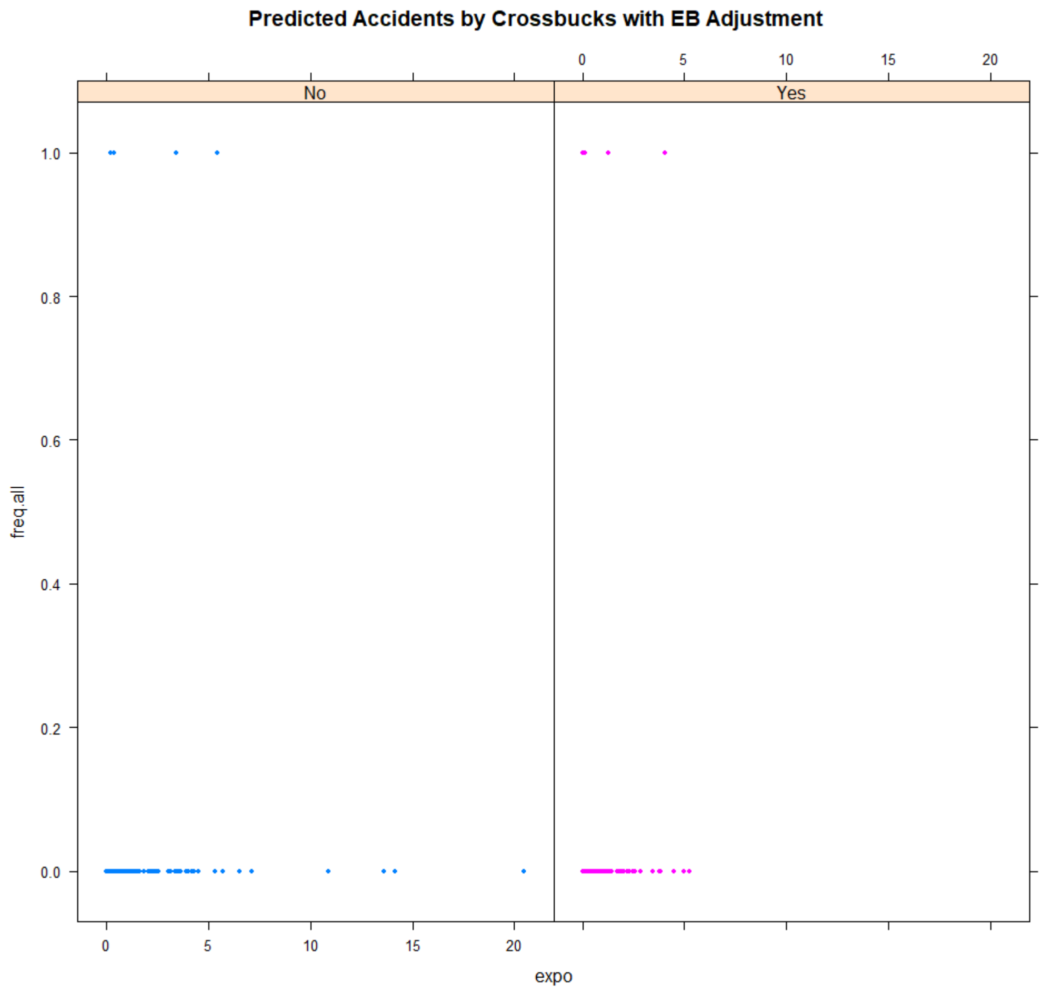


Figure 5.2 Predicted Crashes by Presence of Crossbucks with EB Adjustment

Chapter 6 Crossing Consistency Assessment

This chapter covers the second research objective, which was to develop guidelines for improving safety at urban HRGCs that are not designated quiet zones but are in vicinity of existing quiet zone crossings. The focus was on HRGCs located in Lancaster County, Nebraska. HRGCs in Lancaster County were visited and inventory data corrected with field conditions as described in Chapter 4.

6.1 Methodology for Consistency Assessment

The methodology consisted of a selection of HRGCs that were in the vicinity of established quiet zones by using buffer zones of varying sizes and then investigating within range HRGCs that may have histories of high crash frequencies and crash severities. A risk index assessment was made for the HRGCs using the FRA's Quiet Zone (QZ) Calculator. The quiet zone risk index represents the average severity weighted collision risk for all public HRGCs that are part of a quiet zone and includes added risk caused by the lack of a train horn and risk reductions caused by the implementation of FRA approved supplemental safety measures (SSMs). Based on crash histories and the QZ Calculator results, improvement suggestions were developed for specific HRGCs in Lancaster County.

6.2 HRGC Consistency Assessment

Chapter 4 described creation of the GIS database for Lancaster County. The consistency analysis was based on this GIS database with the first step of buffer creation around fifteen existing HRGCs comprising of quiet zones. Buffers of 0.25 mile, 0.5 mile,1 mile, 2 mile, 3 mile and 5 mile radii were created to identify gated non-quiet zone HRGCs for consideration in the consistency analysis. Figure 6.1 shows buffers of varying size around quiet zone HRGCs in

Lancaster County. The number of non-quiet zone gated HRGCs within a buffer increased as the buffer size around existing quiet zone HRGCs increased.

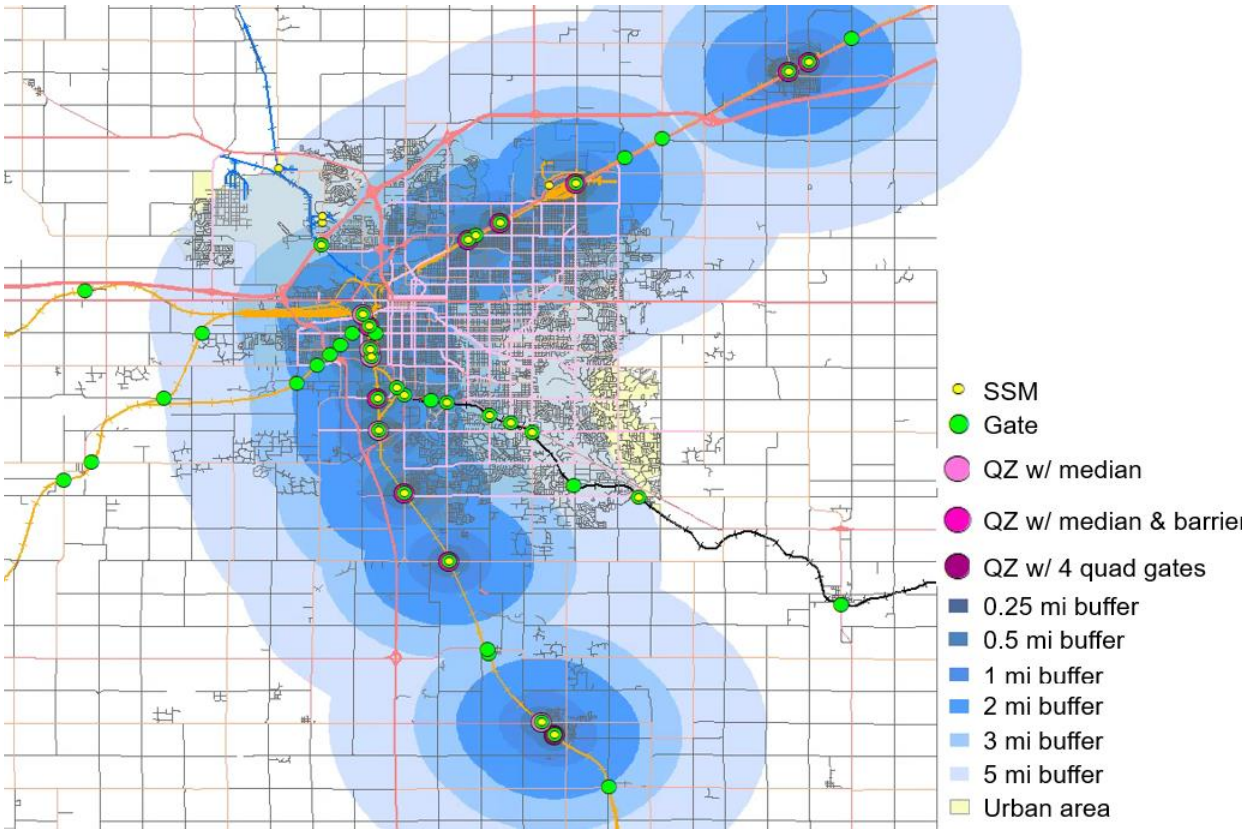
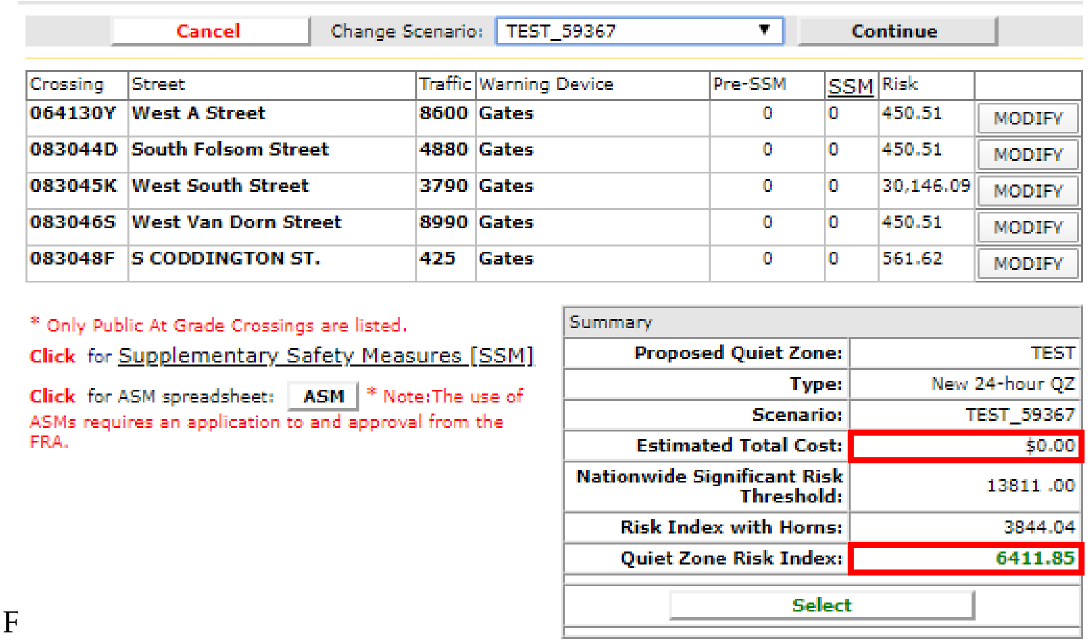


Figure 6.1 Buffers of Varying Size Around Existing Quiet Zone HRGCs in Lancaster County

Table 6.1 shows the results of different buffers and the corresponding HRGCs with their respective crash history. The FRA QZ Calculator was used to assess the risk index for gated crossings that were within the 5-mile buffer (Figure 6.2). These HRGCs were visited for consistency assessment. As quiet zone HRGCs are distinguished by the presence of FRA approved SSMs, the primary criteria for consistency was the practicality of installing SSMs at non-quiet zone HRGCs.

Table 6.1 Results of Buffers of Varying Size and Corresponding HRGC Crash Histories

Buffer Size →	0.5 mi		1 mi		2 mi		3 mi		5 mi		Crash	Year	Fatalities	Injured
	Count	Xing #	Count	Xing #	Count	Xing #	Count	Xing #	Count	Xing #				
# of Gated Non-QZ	3		7		13		19		24					
	1		3	064129E	4	064129E	6	064129E	9	064129E	1	2017	0	0
				072946C		072946C		072946C		072946C				
				083884		083884		083884		083884M				
						083886B		083886B		083886B				
With SSMs								083890R		083890R				
								815572E		815572E				
										083891X				
										083895A				
										083900U				
	2	064130Y	4	064130Y	9	064130Y	13	064130Y	15	064130Y	1	2009	1	0
		064359F		064359F		064359F		064359F		064359F				
				083044		083044		083044		083044D	1	2007	1	0
				083045K		083045K		083045K		083045K	1	2017	2	0
						083885U		083885U		083885U				
						083048F		083048F		083048F				
						083046S		083046S		083046S				
Without SSMs						098443J		098443J		098443J				
						074945C		074945C		074945C				
								083528S		083528S	1	2011	0	0
								083519T		083519T				
								083518L		083518L				
								074934P		074934P	1	2013	0	0
										077809M				
										070129T				



Cancel | Change Scenario: TEST_59367 | Continue

Crossing	Street	Traffic	Warning Device	Pre-SSM	SSM	Risk	
064130Y	West A Street	8600	Gates	0	0	450.51	MODIFY
083044D	South Folsom Street	4880	Gates	0	0	450.51	MODIFY
083045K	West South Street	3790	Gates	0	0	30,146.09	MODIFY
083046S	West Van Dorn Street	8990	Gates	0	0	450.51	MODIFY
083048F	S CODDINGTON ST.	425	Gates	0	0	561.62	MODIFY

* Only Public At Grade Crossings are listed.

Click for Supplementary Safety Measures [SSM]

Click for ASM spreadsheet: ASM * Note:The use of ASMs requires an application to and approval from the FRA.

Summary	
Proposed Quiet Zone:	TEST
Type:	New 24-hour QZ
Scenario:	TEST_59367
Estimated Total Cost:	$0.00
Nationwide Significant Risk Threshold:	13811 .00
Risk Index with Horns:	3844.04
Quiet Zone Risk Index:	6411.85

Select

F

- Temporary closure of a public highway-rail grade crossing,
- Permanent closure of a public highway-rail grade crossing,
- Grade separation of a public highway-rail grade crossing,
- 4-Quad gates upgrade from 2-quad gates, no vehicle presence detection,
- 4-Quad gates upgrade from 2-quad gates, with medians, no vehicle presence detection,
- 4-Quad gates upgrade from 2-quad gates, with vehicle presence detection,
- 4-Quad gates upgrade from 2-quad gates, with medians and vehicle presence detection,
- 4-Quad gates new installation, no vehicle presence detection,
- 4-Quad gates new installation with medians and no vehicle presence detection,
- 4-Quad gates new installation with vehicle presence detection,
- 4-Quad gates new installation with medians and vehicle presence detection,
- Mountable medians with reflective traffic channelization devices,
- Non-Traversable Curb Medians with or without Channelization Devices, and

- One-way Streets with gates.

Many of these SSMs are not practical to implement at the Lancaster County HRGCs under consideration for consistency assessment. The two realistic options for Lancaster County HRGCs are mountable medians with reflective traffic channelization devices and the non-traversable curb medians with or without channelization devices.

6.3 Recommendations for Lancaster County HRGCs

The general recommendation for gated HRGCs in Lancaster County that have a history of crashes or a high FRA QZ Calculator risk index is the installation of mountable medians with reflective traffic channelization devices (vertical panels or tubular delineators) or non-traversable curb medians with or without channelization devices. However, many HRGCs may have specific characteristics that prevent or limit adoption of this general recommendation. For example, a non-paved crossing surface may limit installation of a raised median. Therefore, site characteristics must be taken into account before consideration of the general recommendation. Based on the analysis conducted in this research study, Table 6.2 presents measures that may be considered for implementation at the listed HRGCs.

Table 6.2 Suggested Improvements for Consideration at Selected Lancaster County HRGCs

Crossing ID	Crossing Street	Risk Index	Non-traversable Curb Median, w/ or w/t channelization (vertical panels or tubular delineators)	Traversable/Mountable Channelization Device (vertical panels or tubular delineators)	Median and Mountable Channelization Device (vertical panels or tubular delineators)
074945C	N 162nd St	11626.13			X (one side is unpaved)
074934P	N 98th St	12839.29			X (one side is unpaved)
098443J	84th St	39187.31			X (one side is unpaved)
064129E	Adams St	138020	Planned for Grade Separation		
064359F	A St	47.43			X (one side may block roadway)
064130Y	West A St	450.51	X (one side may block driveway)		
083044D	South Folsom St	450.51	X		
083045K	West South St	30146.09	X		
083046S	West Van Dorn St	450.51			X
083048F	S Coddington St	155.64			X
072946C	Pioneers Blvd	245.61		X (limited train traffic)	
083884M	S 14th St	587.16		X (complex HRGC, needs further study)	
083885U	Southwood Dr	200.95			X (narrow road, may block driveway)
083886B	S 27th St	200.95		X	
083890R	S 40th St	266.71		X	
083891X	S 48th St	307.24	X	X	
083895A	S 56th St	9045.45		X (complex HRGC, may not work)	
077809M	S 70th St	4518.91			X
815387K	HWY 79	7037.14			X
815572E	Northwest 12th St	5544.63		X (very short distance on one side)	
070129T	W A St	427.32			X (one side is unpaved, may block maintenance road)
083518L	Main St (Roca)	30400.31	X (one side may block roadway for the factory)		
083519T	Roy St (Roca)	14799.53		X (one side is unpaved road)	
083528S	Panama Rd	31229.56	X		
083274E	3rd St (Firth)	24213.56	X (one side may block roadway for the factory)		

Note: an ‘X’ indicates a recommendation for consideration

Chapter 7 Summary and Recommendations

This chapter provides a summary of the research effort and the recommendations pertaining to the outcomes from this research as well as future efforts.

7.1 Summary

This research had two objectives: to update the 1999 Nebraska Accident Prediction Model for HRGCs and to develop guidelines for improving safety at urban rail crossings that are not designated quiet zones but are in vicinity of existing quiet zone crossings. HRGCs located in Lancaster County, Nebraska were candidates for the second objective. The 1999 Nebraska Accident Prediction Model is dated and needed an update in view of the availability of new modeling techniques and changing transportation patterns.

FRA crash and HRGC inventory data were utilized for estimation of the new model. Lancaster County HRGCs were visited to verify HRGC inventory data and this effort was subsequently extended to eight additional Nebraska counties. Inventory information for some HRGCs was validated using NDOT's PathWeb system. The combined effort of field visits and use of the PathWeb system resulted in inventory verification of 742 HRGCs in Nebraska. In total 2,911 values were corrected and 1,841 missing values were added to the HRGC inventory database while 33 HRGCs in the database were either abandoned, private crossings listed as public or altogether non-existent. Corrected Nebraska HRGC inventory and reported HRGC crashes (2008-2019) were combined to obtain a dataset; model estimation utilized 2008-2018 data and the 2019 data were used for validation of model predictions.

The FRA's New Model for HRGC Accident Prediction and Severity was used for guidance in the model estimation process. Several model formulations were explored for the 2020 Nebraska Crash Prediction Model. Based on the data characteristics, statistical test results, prediction performance and validation, a Poisson regression model with scaled parameters was chosen as the 2020 Nebraska Crash Prediction Model.

Lancaster County HRGCs that are not designated as quiet zones were assessed for safety improvements (using SSMs) to reduce the chances of violating drivers' expectations. Gated HRGCs were selected based on proximity to existing quiet zone crossings and their risk index as calculated by the FRA QZ Calculator. The selected HRGCs were visited in the field and recommendations developed for each HRGC. However, the general guidance revolves around application of two SSMs that are more practical compared to the other options available via FRA approved SSMs. The guidance is to consider installation of mountable medians with reflective traffic channelization devices (vertical panels or tubular delineators) or non-traversable curb medians with or without channelization devices at non-quiet zone gated HRGCs that are in proximity of established quiet zones.

7.2 Recommendations

Based on the results of this research study, the following recommendations are made to NDOT (and/or other relevant agencies).

- Adopt the 2020 Nebraska HRGC Crash Prediction Model in lieu of the 1999 Nebraska Accident Prediction Model as the newer 2020 model better predicts expected crashes compared to the 1999 model as well as compared to the new FRA's Accident Prediction Model.
- Consider installation of mountable medians with reflective traffic channelization devices or non-traversable curb median with or without channelization devices at gated HRGCs that are in proximity (5-mile radius) of established quiet zones.

- A complete update of the statewide HRGC inventory is recommended to remove errors and missing values from the existing data.
- Establish an update cycle for the Nebraska HRGC Crash Prediction Model to prevent it from becoming dated; a 5-year update cycle appears reasonable.
- Crash severity was not considered in this research but it is an important element of safety. Therefore, a crash severity prediction model is recommended for future research.

Acknowledgments

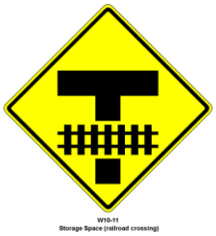
The authors thank members of the project Technical Advisory Committee for their guidance and input throughout this research study. The authors are also thankful to the NDOT Research Division (Mark Fischer, Lieska Halsey, Angela Andersen) for help with all aspects of this research project. Amirfarrokh Iranitalab is acknowledge for initial modeling of the crash data while Harrison Redepenning, Jonathan Camenzind and Joel Smith are acknowledged for help with HRGC inventory data validation.

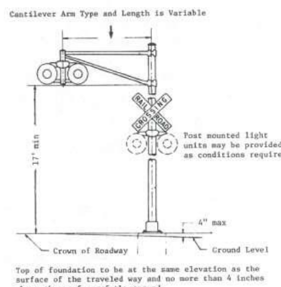
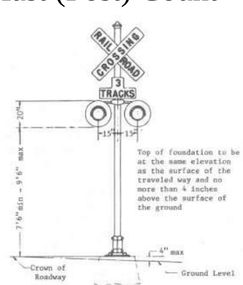
Appendix A

Variable descriptions for HRGC inventory.

Variable Descriptions		
FRA crossing inventory database field	**RIMS**	**Value**
CrossingID	Crossing Number	
ReportingAgencyTypeID		
ReasonID	Reason	
RevisionDate		
LastUpdated	Latest Inventory Approve Date/Time	
Railroad	Primary Operating Railroad	
Nearest	In or Near City	0 = In 1 = Near
CityName	City	
Street	Street or Road Name	
XPurpose	Crossing Purpose	1 = Highway 2 = Pathway, Pedestrian 3 = Station, Pedestrian
DevelTypID	Type of Land Use	11 = Open Space 12 = Residential 13 = Commercial 14 = Industrial 15 = Institutional 16 = Farm 17 = Recreational 18 = RR Yard
Whistban	Quiet Zone	0 = No 1 = 24 hr 2 = Partial 3 = Chicago Excused
Latitude	Latitude	
Longitude	Longitude	
MilePost	RR Milepost	
TotTrk		
MainTrk	Main Tracks	
SidingTrk	Siding Tracks	
YardTrk	Yard Tracks	
TransitTrk	Transit Tracks	
IndustryTrk	Industry Tracks	
OthrTrk		

Sgnleqp	Signaling for Train Operation: Is Track Equipped with Train Signals?	
NoSigns	Are There Signs or Signals?	1 = Yes 2 = No
XBuck	Number of Crossbuck Assemblies	
StopStd	Number of STOP Signs (R1-1)	
YieldStd	Number of YIELD Signs (R1-2)	
AdvWarn	Advance Warning Signs http://www.trafficsign.us/w10.html	
AdvW10_1	Advance Warning Signs: W10-1  W10-1  W10-1a	
AdvW10_2	Advance Warning Signs: W10-2  W10-2 Parallel Railroad Crossing (crossroad)	
AdvW10_3	Advance Warning Signs: W10-3  W10-3 Parallel Railroad Crossing (side road)	
AdvW10_4	Advance Warning Signs: W10-4  W10-4 Parallel Railroad Crossing (T)	

AdvW10_11	Advance Warning Signs: W10-11  	
AdvW10_12	Advance Warning Signs: W10-12  	
PaveMrkIDs	Pavement Markings	0 = None 1 = Stop Lines 2 = RR Xing Symbols 3 = Dynamic Envelope
Channel	Channelization Devices/Medians	1 = All Approaches 2 = One Approach 3 = Median – All Approaches 4 = Median – One Approach 5 = None
SSM		0 = None 1 = normal medians 2 = Non-Traversable Curb Medians with or without Channelization Devices 3 = Mountable medians with Reflective Traffic channelization Devices 4 = four quadrant gate systems, 5 = one-way streets with gates, 6 = temporary or permanent crossing closures
EnsSign	ENS Sign (I-13) Displayed?  	1 = Yes 2 = No
OthSgn	Other MUTCD Signs?	1 = Yes 2 = No

	https://mutcd.fhwa.dot.gov/ser-shs_millennium_eng.htm; http://www.trafficsign.us/	
OthSgn1	MUTCD Code (1)	
OthDes1	Other MUTCD Signs Count (1)	
OthSgn2	MUTCD Code (2)	
OthDes2	Other MUTCD Signs Count (2)	
OthSgn3	MUTCD Code (3)	
OthDes3	Other MUTCD Signs Count (3)	
Gates	Count of Roadway Gate Arms	
GateConf	Gate Configuration	1 = 2 Quad 2 = 3 Quad 3 = 4 Quad
GateConfType	Gate Configuration Type	4 = Full (Barrier) Resistance 6 = Median Gates
FlashOv	Number of Cantilevered (or Bridged) Flashing Light Structures Over Traffic Lane  	
FlashNov	Number of Cantilevered (or Bridged) Flashing Light Structures Not Over Traffic	
CFlashType	Cantilevered (or Bridged) Flashing Light Types	0 = None 1 = Incandescent 2 = LED
FlashPost	Mast-Mounted Flashing Lights: Mast (Post) Count  	
FlashPostType	Mast-Mounted Flashing Lights: Light Types	0 = None 1 = Incandescent 2 = LED

Bkl_FlashPost	Mast-Mounted Flashing Lights: Back Lights?	1 = Yes 2 = No
Sdl_FlashPost	Mast-Mounted Flashing Lights: Side Lights?	1 = Yes 2 = No
FlashPai	Total Count of Flashing Light Pairs	
Bells	Number of Bells	
HwynrSig	Does Nearby Hwy. Intersection Have Traffic Signals?	1 = Yes 2 = No
HwtrfPsigsdis	Highway Traffic Pre-Signals: Storage Distance	
HwtrfPsiglndis	Highway Traffic Pre-Signals: Stop Line Distance	
WdCode		
TraficLn	Number of Traffic Lanes Crossing Track	
TraflnType	Traffic Lane Type	1 = One-way Traffic 2 = Two-way Traffic 3 = Divided Traffic
HwyPved	Is Roadway/Pathway Paved?	1 = Yes 2 = No
Downst	Does Track Run Down a Street (Y/N)?	1 = Yes 2 = No
Illumina	Is Crossing Illuminated?	1 = Yes 2 = No
XSurfWidth	Crossing Surface: Width (Feet)	
XSurfLength	Crossing Surface: Length (Feet)	
XSurfaceIDs	Crossing Surface (Main Track)	11 = 1. Timber 12 = 2. Asphalt 13 = 3. Asphalt and Timber 14 = 4. Concrete 15 = 5. Concrete and Rubber 16 = 6. Rubber 17 = 7. Metal 18 = 8. Unconsolidated 19 = 9. Composite 20 = 10. Other (specify)
HwyNear	Intersecting Roadway Within 500 Feet?	1 = Yes 2 = No
HwynDist	Approximate Roadway Distance (Feet) (Feet)	
XAngle	Smallest Crossing Angle	1 = 0° – 29° 2 = 30° – 59° 3 = 60° - 90°

ComPower	Commercial Power Available Within 500 Feet (Y/N)?	
HwyClassCD	Functional Classification: Development	0 = (0) Rural 1 = (1) Urban
HwyClassrdtpID	Functional Classification: Road Function	11 = (1) Interstate 12 = (2) Other Freeways and Expressways 13 = (3) Other Principal Arterial 16 = (4) Minor Arterial 17 = (5) Major Collector 18 = (6) Minor Collector 19 = (7) Local
HwySpeed	Posted Highway Speed (mph)	
open		
flash		
gate		
	NDOT Crossing Number:	
	Structure Number:	
	Highway District:	
	NDOT County Map Reference:	
	NDOT Reference Post:	
	NDOT-Specified Location:	
	NDOT Control Number:	
	NDOT Highway Number:	
	Passenger Trains Per Day:	
	Passenger Train Speed:	
	Track Category Code:	
	Storage Distance:	
	Approach Surface Type:	
	Approach Surface Width:	
	Grade First:	
	Grade Second:	
	First Location of Change:	
	Second Location of Change:	
	Distance Between Tracks (RRX Width):	
	Type of Service:	
	NDOT Narrative:	
	NDOT Crossing Rank:	

Appendix B

Highway Rail Accident/Incident Variables (FRA Form 6180.57)

FIELD NAME	DEFINITION	Values
amtrak	amtrak involvement	
iyr	year of incident	
imo	Month of incident	
railroad	Railroad code (reporting RR)	
incdtno	Railroad assigned number	
iyr2	Year of incident	
imo2	Month of incident	
rr2	Railroad code (other RR involved)	
incdtno2	Other railroad assigned number	
iyr3	Year of incident	
imo3	Month of incident	
rr3	Railroad code (RR responsible for track maintenance)	
incdtno3	RR assigned number	
dummy1	Blank data expansion field	
casinjrr	# of injured for reporting railroad calculated from F6180.55a's submitted	
gxid	Grade crossing id number	
year	Year of incident	
month	Month of incident	
day	Day of incident	
timehr	Hour of incident	
timemin	Minute of incident	
ampm	Am or pm	
station	Nearest timetable station	
county	County name (see FIPS codes for associated code)	
state	FIPS state code	
region	FRA designated region	
dummy2	Blank data expansion field	
city	City name (see FIPS codes for associated code)	
highway	Highway name	
vehspd	Vehicle estimated speed (blank = unknown)	

typveh	Highway User	A = auto G = school bus B = truck H = motorcycle C = truck-trailer J = other motor vehicle. D = pick-up truck K = pedestrian E = van M = other F = bus
vehdir	Highway user direction	1 = north 3 = east 2 = south 4 = west
position	Position of highway user	1 = stalled or stuck on crossing* 2 = stopped on crossing 3 = moving over crossing 4 = trapped on crossing by traffic* 5 = blocked on crossing by gates**
rrequip	RR equipment involved	1 = train (units pulling) A = train pulling (RCL) 2 = train (units pushing) B = train pushing (RCL) 3 = train (standing) C = train standing (RCL) 4 = car(s) (moving) D = EMU Locomotive(s)* 5 = car(s) (standing) E = DMU Locomotive(s)* 6 = light loco(s) (moving) 7 = light loco(s) (standing) 8 = other
rrcar	Position of car unit in train	
typacc	Circumstance of accident	1 = rail equipment struck highway user 2 = rail equipment struck by highway user
hazard	Entity transporting hazmat	1 = highway user 3 = both 2 = rail equipment 4 = neither
temp	temperature in degrees Fahrenheit	
visiblty	Visibility	1 = dawn 3 = dusk 2 = day 4 = dark
weather	Weather conditions	1 = clear 4 = fog 2 = cloudy 5 = sleet 3 = rain 6 = snow
typeq	Type of consist	1 = freight train 2 = passenger train(pulling)* 3 = commuter train(pulling)* 4 = work train 5 = single car 6 = cut of cars 7 = yard/switching 8 = light loco(s) 9 = maint/inspec car A = special MoW equipment B=passenger train (pushing)** C=commuter train (pushing)** D=EMU** E=DMU**
typtrk	Type of track	1 = main 3 = siding 2 = yard 4 = industry
trkname	track identification	
trkclas	FRA track class: 1-9, X	
nbrlocos	Number of locomotive units	

nbrcars	Number of cars	
trnspd	Speed of train in miles per hour (if field is blank = unknown)	
typspd	Train speed type	E = estimated R = recorded Blank = unknown
trndir	Time table direction	1=north 2=south 3=east 4=west
signal	Type of signaled crossing warning	
locwarn	Location of warning	1 = both sides 2 = side of vehicle approach 3 = opposite side of vehicle approach
warnsig	Crossing warning interconnected with highway signal	1 = yes 2 = no 3 = unknown
lights	Crossing Illuminated by Street Lights or Special Lights	
standveh	Driver passed highway standing vehicle	
train2	Highway user went behind or in front of train and struck or was struck by second train	
motorist	Action of highway user	1 = went around the gates* 2 = stopped and then proceeded 3 = did not stop 4 = stopped on crossing 5 = other 6 = went around/thru temporary barricade (if yes, see instructions)*** 7 = went thru the gate*** 8 = suicide/attempted suicide***
view	Primary obstruction of track view	1 = permanent structure 2 = standing RR equipment 3 = passing train 4 = topography 5 = vegetation 6 = highway vehicles 7 = other 8 = not obstructed
vehdmg	Highway vehicle property damage in $	
driver	Driver was	1 = killed 2 = injured 3 = uninjured
inveh	Driver in vehicle	1 = yes 2 = no
totkld	Total killed for railroad as reported on F6180.57	
totinj	Total injured for railroad as reported on F6180.57	
totocc	Total # of vehicle occupants (including driver)*	
incdrpt	F6180.54 filed:	1 = yes 2 = no
jointcd	Indicates railroad reporting	
typrr	Type railroad – ICC categories	1st position indicates class 1, 2, or 3 railroad
dummy3	Blank data expansion field	

caskldrr	# killed for reporting RR – calculated from F6180.55a's submitted	
dummy4	Blank data expansion field	
crossing	Type of warning device at crossing (series of 2 digit codes)	01 = gates 07 = cross bucks 02 = cantilever FLS 08 = stop signs 03 = standard FLS 09 = watchman 04 = wig wags 10 = flagged by crew 05 = highway traffic 11 = other (specify) signals 12 = none 06 = audible
narrlen	Length of narrative	
dummy5	Blank data expansion field	
year4	4 digit year of incident	
division	Railroad division	
public	Public crossing	1 = public 2 = private
cntycd	FIPS county code	
stcnty	FIPS state and county code	
hzmrlsed	Hazmat released by	1=highway user 3=both 2=rail equipment 4=neither Blank=unknown
hzmname	Name of hazmat released	
hzmqnty	Quantity of hazmat released	
hzmmeas	Measure used in hazmat quantity field	
sigwarnx	Further definition of signal field	
whisban	Whistle ban in effect	1=yes 2=no 3=not provided blank=unknown
drivage	Highway user's age	
drivgen	Highway user's gender	1 = male 2 = female blank = unknown
pleontrn	Total # of people on train (includes passengers and crew)	
ssb1	Special study block 1	
ssb2	Special study block 2	
userkld	# of highway-rail crossing users killed as reported by railroad on F6180.57	
userinj	# of highway-rail crossing users injured as reported by railroad on F6180.57	
rrempkld	# of railroad employees killed as reported by railroad on F6180.57	
rrempinj	# of railroad employees injured as reported by railroad on F6180.57	
passkld	# of train passengers killed as reported by railroad on F6180.57	
passinj	# of train passengers injured as reported by railroad on F6180.57	
subdiv	Railroad Subdivision	

roadcond	Roadway Conditions	A = dry B = wet C = snow/slush D = Ice E = sand, mud, dirt, oil, gravel F = water (standing, moving)

Appendix C

Table C.1 Performances of all candidate models. Trained on 2008-2016 data and validated on 2017 data (observed crash frequency was 18).

Candidate models		AIC	MSE	RMSE	Logarithm score	Predicted outcome	Percentage difference in prediction results
Poisson	All variables	2718.188	0.006198456	0.078730274	0.03429069	28.4339	+57.97%
	Small	2718.7	0.00614256	0.078374486	0.03580901	28.43493	+57.97%
	Selected variables based on AIC	2713.919	0.006179902	0.078612353	0.03616989	28.4335	+57.96%
	Selected variables based on LR test	2710.83	0.006169171	0.07854407	0.03430084	28.43382	+57.97%
	Selected variables based on stepwise selection	2078.909	0.005816272	0.076264487	0.0356827	27.32421	+51.80%
	Mixed effects all variables	2719.8	0.006094613	0.078068002	0.03575563	28.43491	+57.97%
	Mixed effects small	2717.6	0.006188576	0.078667503	0.03657963	28.43394	+57.97%

	Mixed effects all variables based on AIC	2717.007	0.006155857	0.07845927	0.03626977	28.43361	+57.96%
	Mixed effects all variables based on LR test	2726.093	0.006112371	0.078181654	0.03621087	28.43511	+57.97%
Zero-inflated Poisson	All variables	2674.205	0.08721682	0.29532494	0.03616989	43.53799	+141.88%
	Small	2700.951	0.006124286	0.078257818	0.0525844	28.42062	+57.89%
	Selected variables based on AIC	2698.154	0.006209432	0.078799949	0.05655367	28.44069	+58.00%
	Selected variables based on stepwise selection	2699.15	0.005937432	0.077054734	0.05277615	28.43731	+57.99%
	Mixed effects small	2664.1	0.006117588	0.078215011	0.03761002	17.73504	-1.47%

Table C.1 (cont.)

Candidate models		AIC	MSE	RMSE	Logarithm score	Predicted outcome	Percentage difference in prediction results
NDOT model	Raw	-	0.0064	0.08	-	0.2	-98.89%
	Weighted	-	0.006409604	0.08006	-	5.74	-68.11%
	Normalized	-	0.006390404	0.07994	-	3.48	-80.67%
USDOT model	Raw	-	0.006366444	0.07979	-	6.39	-64.50%
	Weighted	-	0.006393602	0.07996	-	10.86	-39.67%
	Normalized	-	0.006371232	0.07982	-	6.62	-63.22%
	New ZINB model variables	2716.415	0.006104412	0.07813	0.03590733	28.43499	+57.97%

Table C.2 Performances of all candidate models. Trained on 2008-2017 data and validated on 2018 data (observed crash frequency was 35).

Candidate models		AIC	MSE	RMSE	Logarithm score	Predicted outcome	Percentage difference in prediction results
Poisson	All variables	2183.977	0.01000041	0.10000205	0.06935734	20.98835	-40.03%
	Small	2224.541	0.009691069	0.098443227	0.06901735	21.19004	-39.46%
	Selected variables based on AIC	2192.455	0.009842123	0.099207475	0.06874995	20.98958	-40.03%
	Selected variables based on LR test	2195.8	0.00985581	0.099276432	0.06818701	20.98931	-40.03%
	Selected variables based on stepwise selection	2205.661	0.01274362	0.112887643	0.06421917	26.03446	-25.62%
	Mixed effects all variables	2223.318	0.009684737	0.098411061	0.06961588	21.18985	-39.46%
	Mixed effects small	2184.44	0.01000043	0.10000215	0.06929384	20.98839	-40.03%

	Mixed effects all variables based on AIC	2190.078	0.009836715	0.099180215	0.06793833	20.98891	-40.03%
	Mixed effects all variables based on LR test	2196.825	0.009848963	0.099241942	0.06820578	20.98941	-40.03%
Zero-inflated Poisson	All variables	2178.809	0.01000117	0.10000585	0.07162996	20.98894	-40.03%
	Small	2206.44	0.009657594	0.098273058	0.1193884	21.1899	-39.46%
	Selected variables based on AIC	2186.503	0.009828321	0.099137889	0.08618688	14.16269	-59.54%
	Selected variables based on stepwise selection	2893.301	0.01274362	0.112887643	0.1071404	27.39226	-21.74%
	Mixed effects small	2170.543	0.009791449	0.098951751	0.08247168	14.0481	-59.86%

Table C.2 (cont.)

Candidate models		AIC	MSE	RMSE	Logarithm score	Predicted outcome	Percentage difference in prediction results
NDOT model	Raw	-	0.009883361	0.099415094	-	0.1469295	-99.58%
	Weighted	-	0.009854559	0.099270131	-	5.172267	-85.22%
	Normalized	-	0.009835792	0.099175562	-	3.189546	-90.89%
USDOT model	Raw	-	0.009861255	0.099303852	-	4.871511	-86.08%
	Weighted	-	0.009853375	0.099264168	-	9.12246	-73.94%
	Normalized	-	0.009840335	0.099198463	-	5.635839	-83.90%
	New ZINB model variables	2164.743	0.009969616	0.099847964	0.07271544	20.99934	-40.00%

References

Austin, R. D., and J. L. Carson. An alternative accident prediction model for highway-rail interfaces. Accident Analysis & Prevention, No. 34.1, 2002, pp. 31-42.

Brod, D. and D. Gillen. A New Model for Highway-Rail Grade Crossing Accident Prediction and Severity. Report DOT/FRA/ORD-20/40, October 2020.

Connecticut Department of Transportation. Connecticut Railway-Highway Crossing Program 2014 Annual Report, 2015.

Farr, Edwin H. Rail-highway crossing resource allocation procedure: User's guide. No. DOT-TSC-FRA-87-1. 1987.

Federal Railroad Administration. Federal Railroad Administration Highway-Rail Crossing Database Files and Reports, 2020. Accessed on March 2020, from: https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/downloads/downloads.aspx

HNTB Corp. Development of an Accident Prediction Model for Highway-Rail Grade Crossings in Nebraska, June 1999.

Hauer, E. The Art of Regression Modeling in Road Safety, Springer Publications, 2015.

Hitz, J. S. Rail-Highway Crossing Resource Allocation Procedure. User's Guide (No. DOT-TSC-FHWA-85-2; FHWA-IP-86-11). United States. Federal Highway Administration. 1986.

Hu, S., C. Li, and C. Lee. Investigation of Key Factors for Accident Severity at Railroad Grade Crossings by using a Logit Model. Safety Science, Vol. 48, 2010, pp. 186-194.

Lord, D., & Mannering, F. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation research part A: policy and practice, 44(5), 291-305. 2010.

Malyshkina, N. V., & Mannering, F. L. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents. Accident Analysis & Prevention, 42(1), 131-139. 2010.

McCollister, G. M. and C. C. Pflaum. A Model to Predict the Probability of Highway Rail Crossing Accidents. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2007, pp. 321-329.

Medina, J.C. and R.F. Benekohal. Macroscopic Models for Accident Prediction at Railroad Grade Crossings: Comparisons with U.S. Department of Transportation Accident Prediction Formula. Transportation Research Record 2476, Journal of the Transportation Research Board, Washington, D.C., 2015, pp. 85-93.

Miaou, S. P., & Lum, H. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis & Prevention, 25(6), 689-709. 1993.

Milton, J., & Mannering, F. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation, 25(4), 395-413. 1998.

Miranda-Moreno, L.F. and L. Fu. A Comparative Study of Alternative Model Structures and Criteria for Ranking Locations for Safety Improvements. Networks and Spatial Economics, Vol. 6, 2006, pp. 97-110.

NCHRP Report 50. Factors Influencing Safety at Highway-Rail Grade Crossings. Highway Research Board, National Research Council, Washington D.C., 1968.

Niu, X., Q. Chen, and H.M. Dowell. 2014. Statistical Modeling and Prediction of Highway-Railroad Grade Crossing Safety in Florida. Florida State University, Tallahassee.

North Carolina Department of Transportation. North Carolina Railway-Highway Crossing Program 2014 Annual Report, 2015.

Ogden B. D. Railroad-highway grade crossing handbook. U.S. Department of Transportation, Federal Highway Administration, 2007.

Ogden, B. D. and C. Cooper. Highway-Rail Crossing Handbook, 3rd Edition. U.S. Department of Transportation, Federal Highway Administration Report No. FHWA-SA-18-040/FRA-RRS-18-001, July 2020.

Peabody, L. E., and T. B. Dimmick. Accident Hazards at Grade Crossings. Public Roads 22.6, 1941, pp. 123-130.

Qureshi, M., M. Virkler, K. Bernhardt, G. Spring, S. Avalokita, N. Yathapu, V. Chilukuri, T. King and K. Gibbons. 2004. Highway/Rail Crossing Project Selection. Report Number UTC R60, Center for Infrastructure Engineering Studies, University of Missouri-Rolla.

Sperry, Benjamin R., Bhaven Naik, and Jeffery E. Warner. Evaluation of Grade Crossing Hazard Ranking Models. 2017 Joint Rail Conference. American Society of Mechanical Engineers, 2017.

U.S. Department of Transportation. GradeDec Crossing Evaluation Tool, 2018. Accessed from: https://gradedec.fra.dot.gov/

U.S. Department of Transportation. Federal-Aid Policy Guide: Part 924 - Highway Safety Improvement Program. 1991. Accessed Oct. 2018 from: https://www.fhwa.dot.gov/legsregs/directives/fapg/cfr0924.htm.

Weissmann, A.J., J. Weissmann, J.L. Kunisetty, J.E. Warner, E.S. Park, S. Sunkari, A. Protopapas, and S. Venglar. 2013. Integrated Prioritization Method for Active and Passive Highway-Rail Crossings. Texas Department of Transportation, Report No. FHWA/TX-12/0-6642-1.